Development of CRISPR/Cas9-based gene drive biotechnology in *S. cerevisiae*

by

Yao Yan

B.S., Dalian Medical University, 2014
M.S., Emporia State University, 2017

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Biochemistry and Molecular Biophysics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2022

Abstract

The budding yeast *Saccharomyces cerevisiae* is a well-studied eukaryotic model organism for investigating the fundamental aspects of molecular and cellular biology. It also can be used to evaluate the technical feasibility of new technologies in eukaryotic cells. In our lab, we aimed to use *Saccharomyces cerevisiae* as a model to develop CRISPR/Cas9-based gene drive biotechnology.

The CRISPR/Cas9 biotechnology is a powerful gene editing tool used to modify target genomic sequences in all forms of living creatures including animals, plants, fungi, and bacteria. One potential application of this molecular method is within a "gene drive" system. This unique arrangement of CRISPR within a genome may one day allow for global control of biological populations and be used to eliminate pests, parasites, and invasive species. However, there are many concerns regarding the utilization of this technology, including gene drive design, control, and development of resistance, etc. In our study, we developed an artificial gene drive system in budding yeast, which could edit multiple loci at the same time. We demonstrated that this triple gene drive system could successfully edit three DNA targets independently with only a single copy of *S. pyoge*nes Cas9. We also found that the occurrence of NHEJ could be repressed by modifying DNA Ligase IV. However, successful gene drives still allowed for the occurrence of a small number of "resistant" clones. We investigated potential causes of this imperfect drive activity. Our work illustrated that imperfect activation of the inducible promoter driving expression of the Cas9 nuclease or issues with multiplexing to artificial sequences may have resulted in a small percentage of resistant/inactive clones. The CRISPR/Cas9 system can also be used to regulate gene transcription. This involves a mutated Cas9 variant that has lost its nuclease activity (dCas9). We developed a CRISPR/dCas9 system by tagging dCas9 with transcriptional regulators. Our

experiments demonstrated that CRISPR/dCas9 could activate target gene transcription when tagged with the transcriptional activator VPR and repress gene transcription when tagged with the transcriptional repressor Mxi1.

Development of CRISPR/Cas9-based gene drive biotechnology in *S. cerevisiae*

by

Yao Yan

B.S., Dalian Medical University, 2014
M.S., Emporia State University, 2017

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Biochemistry and Molecular Biophysics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2022

Approved by:

Major Professor
Dr. Gregory Finnigan

Copyright

Abstract

The budding yeast *Saccharomyces cerevisiae* is a well-studied eukaryotic model organism for investigating the fundamental aspects of molecular and cellular biology. It also can be used to evaluate the technical feasibility of new technologies in eukaryotic cells. In our lab, we aimed to use *Saccharomyces cerevisiae* as a model to develop CRISPR/Cas9-based gene drive biotechnology.

The CRISPR/Cas9 biotechnology is a powerful gene editing tool used to modify target genomic sequences in all forms of living creatures including animals, plants, fungi, and bacteria. One potential application of this molecular method is within a "gene drive" system. This unique arrangement of CRISPR within a genome may one day allow for global control of biological populations and be used to eliminate pests, parasites, and invasive species. However, there are many concerns regarding the utilization of this technology, including gene drive design, control, and development of resistance, etc. In our study, we developed an artificial gene drive system in budding yeast, which could edit multiple loci at the same time. We demonstrated that this triple gene drive system could successfully edit three DNA targets independently with only a single copy of *S. pyoge*nes Cas9. We also found that the occurrence of NHEJ could be repressed by modifying DNA Ligase IV. However, successful gene drives still allowed for the occurrence of a small number of "resistant" clones. We investigated potential causes of this imperfect drive activity. Our work illustrated that imperfect activation of the inducible promoter driving expression of the Cas9 nuclease or issues with multiplexing to artificial sequences may have resulted in a small percentage of resistant/inactive clones. The CRISPR/Cas9 system can also be used to regulate gene transcription. This involves a mutated Cas9 variant that has lost its nuclease activity (dCas9). We developed a CRISPR/dCas9 system by tagging dCas9 with transcriptional regulators. Our

experiments demonstrated that CRISPR/dCas9 could activate target gene transcription when tagged with the transcriptional activator VPR and repress gene transcription when tagged with the transcriptional repressor Mxi1.

Table of Contents

List of Figures

List of Tables

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my major professor Dr. Gregory Finnigan for his invaluable advice, continuous support, and patience at every stage of my graduate education. His immense knowledge and contribution in this field inspired me to growth as a research scientist. I am extremely grateful to be his graduate student over the years.

I would like to thank my committee professors Dr. Anna Zolkiewska, Dr. Erika Geisbrecht, and Dr. Zhilong Yang for their insightful comments and suggestions during my PhD study.

I also want to extend my sincere thanks to the Department of Biochemistry and Molecular Biophysics at Kansas State University for supporting me to finish my graduate study.

Finally, I would like to offer my special thanks to my family and friends for their unwavering support and belief in me.

Dedication

I dedicate my dissertation work to my loving parents, Shixiong Yan and Meifeng Wang, who have encouraged me attentively with their fullest and truest attention.

Also I would like to dedicate this project to my fiancé, Guanpeng Wang, who has been a source of strength, support, patience, and motivation for me throughout this entire experience.

I would not have been able to finish my study without their unconditional love.

Chapter 1 - Introduction and application of CRISPR/Cas9 technology

History of genome editing

Genome editing is a critical biological process in maintaining genome stability in cell development. The basis of gene editing is built on the fact that cells can repair DNA damage by stimulating cellular repair mechanisms[1]. One type of DNA damage is a DNA double strand break (DSB), which can be induced by ionizing radiation, chemical treatment, or oxidative stress[2]. DNA DSBs can be repaired by homology-directed repair (HDR) or non-homologous end joining (NHEJ) (Fig. 1.1)[2]. HDR repairs DSBs by homologous recombination. This process occurs in S/G2-phase of the cell cycle when the duplicated chromosome is available to serve as a homologous template[3,4]. Based on this property, HDR can be used to insert target genes in mammalian cells. For example, Capecchi *et al.* generated transgenic mice by inserting genes into mouse embryos through HDR-mediated gene editing[5]. Non-homologous end joining is an efficient repair pathway, which is activated throughout the cell cycle. However, NHEJ pathways are error-prone, and they fix DSBs by connecting two exposed chromosomal ends and creating insertions/deletions (indels) at the break site[6].

Previous studies demonstrated that HDR has been widely applied in developing transgenic mouse models. However, the efficiency of HDR was low since it required the selection of exogenous survival markers (Fig. 1.1). To enhance the working efficiency of HDR-mediated gene editing, Jasin *et al.* developed a site-specific editing system by making site-specific DSBs with the yeast endonuclease I-SceI[7,8]. Other studies showed that Zinc Finger Nucleases (ZFNs) and Transcription Activator-Like Effector Nucleases (TALENs) also could be used to create DSBs at restriction sites[9,10]. ZFNs are synthetic proteins that contain both the zinc finger DNA-binding

domain at the N-terminal end and the FokI type II restriction endonuclease (Fn domain) at the C-terminal end[11]. A previous study demonstrated that the Fn domain must dimerize to cut DNA, and double Zinc fingers enhance the specificity of target site locations[12]. Similar to ZFNs, TALENs also use the Fn domain as the DNA cleavage module. The DNA binding domain of TALENs consists a series of tandem repeats and each repeat can recognize a single nucleotide[13]. Application of these enzymes significantly increased the HDR working efficiency at target loci. However, these enzymes were limited because of their poor targeting capability. Thus, it would be better to develop a more efficient gene editing system.

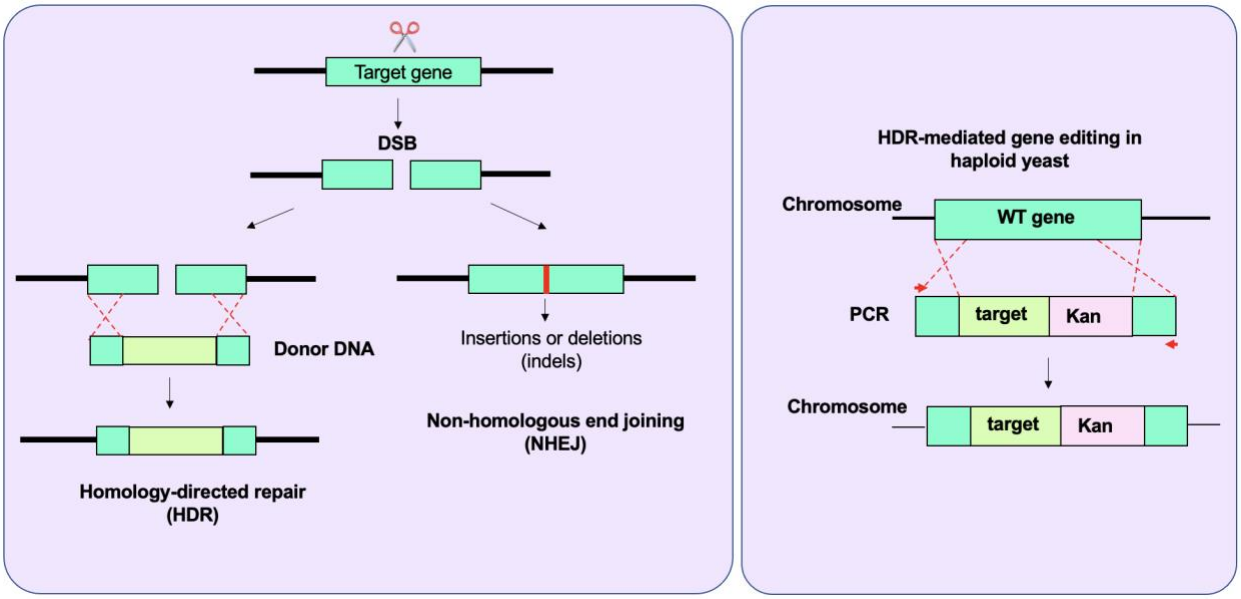


Figure 1.1. DNA double strand breaks can be repaired by homology-directed repair or non-homologous end joining repair.

Left, schematic of two pathways for DNA DSBs. Homology-directed repair pathway requires a homologous template, and the occurrence rate for HDR is higher in yeast than mammalian cells. The nonhomologous end joining pathway fixes DNA DSBs without a DNA template. *Right*, schematic of working mechanism for traditional HDR-meditated gene editing technology in yeast. The target DNA fragment (PCR) contains sufficient flanking homologous sequences at both ends, and it can integrate into yeast genome by homologous recombination without a targeted break. Figure adapted from Esvelt *et al.* 2014[14],Yan and Finnigan 2018[15].

Development of CRISPR/Cas9 system

The clustered regularly interspaced short palindromic repeat (CRISPR)–CRISPR-associated protein 9 (Cas9) system has been demonstrated to be an important genome editing technology in many organisms and cell types[16-20]. CRISPR was first documented as repeat sequences present in the *Escherichia coli* genome in 1987[21]. Scientists found that in bacteria and archaea, CRISPR systems could work as an adaptive immune system to defend against bacteriophage infection and plasmid invasion when incorporated with Cas nucleases[22,23]. There are three steps in the CRISPR/Cas defense mechanism. The first step is called "CRISPR adaptation". Phages or plasmids are recognized by the Cas1–Cas2 adaptation module proteins and excised into short fragments; these short fragments are named "spacers"[24]. The spacers are then incorporated into the CRISPR repeat array. The second step is CRISPR RNA (crRNA) biogenesis. During this process, the CRISPR array is transcribed into the pre-crRNA, then this pre-crRNA is modified into a mature crRNA to form a tracrRNA (trans-activating crRNA): crRNA: Cas complex. Finally, the CRISPR/Cas complex binds to a target DNA sequence and creates a double strand break[24,25].

The CRISPR/Cas systems are split into two classes (class I and class II) based on the structure and function of the Cas protein, which are further separated into six types (type I–VI)[26]. The type I, III, and IV belong to the class I family and type II, V, and VI are members of the class II family[27]. In all these CRISPR/Cas systems, the type I, II, and V systems are responsible for DNA cleavage, while the type VI recognize and cleave RNA, and the type III system can edit both DNA and RNA[27]. The Cas9 nuclease belongs to the type II CRISPR/Cas system, and it contains both a recognition (REC) lobe, and a nuclease lobe[26]. This nuclease can be found in many different bacterial species, including *Streptococcus pyogenes* (SpCas9), *Neisseria meningitidis* (NmCas9)

and *Staphylococcus aureus* (SaCas9)[26]. In the CRISPR/Cas9 system, the activated Cas9 nuclease cleaves the DNA target strand using its HNH nuclease domain and the non-target strand using its RuvC domain[26].

In 2012, a simplified CRISPR/Cas9 system was developed by Doudna and Charpentier [27], and their discovery resulted in the 2020 Nobel Prize in Chemistry. They discovered that the tracrRNA could fuse with the crRNA to form a single guide RNA (sgRNA), and only 20 base pairs (bp) of crRNA were required to recognize and bind the target DNA[27]. In this CRISPR/Cas9 system, three components are essential for genome editing: a Cas9 nuclease, a sgRNA, and a DNA target that also contains a protospacer adjacent motif (PAM) sequence 5'-NGG-3' at the 3' end. Cas9 binds to the sgRNA to form a Cas9/sgRNA complex, which can recognize specific DNA targets and make a double strand break at three bp upstream of the PAM site[28,40]. In eukaryotic cells, DNA DSBs can be repaired by either HDR (presence of donor DNA) or NHEJ (absence of donor DNA) repair mechanisms[29-32]. Typically, NHEJ systems repair DSBs by connecting DNA ends and creating indels at cleavage sites. As a result, these mutated sites can no longer be targeted by Cas9/sgRNA complex a second time. Thus, the NHEJ mechanism can be used to introduce frameshift mutations to knockout genes after CRISPR editing[33]. On the other hand, HDR systems edit genomes precisely by providing sufficient flanking homology around DSBs[28]. Therefore, HDR systems can be used to introduce sequences or mutations into target DNA.

Another CRISPR associated nuclease named Cas12a (Cpf1) was identified in 2015[34]. Cas12a belongs to the type V CRISPR/Cas system[35-37], and has different properties compared to Cas9. First, in the CRISPR/Cas12a system, the guide RNA does not contain a 3' tracrRNA structure. Second, Cas12a requires a T-rich (5'-TTN-3') PAM site at the 5' end of the target DNA. Third, instead of making a blunt ended DNA break, Cas12a creates a staggered double strand

break[34,38,39]. All together, these properties make it possible for Cas12a to edit AT-rich genomic regions.

Applications of CRISPR technology

Based on the role of CRISPR technology in each system, applications of this methodology can be classified into a number of separate categories. First, CRISPR/Cas9 technology is used to knock in or knock out target DNA in living creatures[41-43]. In mammalian cells, CRISPR technology can be used to generate different disease models. Previous studies demonstrated that CRISPR/Cas9 systems could assist in study of human cancers, neurological disease, infectious disease, and immunodeficient models by deleting or mutating different genes[33]. For instance, immunodeficient mice were generated when multiple genes (B2m, Il2rg, Prf1, Prkdc and Rag1) were edited by CRISPR/Cas9 technology[44]. Additionally, CRISPR technology is a potential gene therapy strategy for human genetic diseases. Yin *et al.* identified that CRISPR/Cas9 gene editing systems could correct genetic mutations and phenotypes of human hereditary tyrosinemia[45]. This is a fatal genetic disease, which can be induced when mutations occur within the fumarylacetoacetate hydrolase (FAH) gene. Their results demonstrated that the CRISPR/Cas9 system could be used to correct the FAH mutation by replacing the mutated nucleotides with wild-type nucleotides through homologous recombination. Previous studies demonstrated that CRISPR gene therapy also can be used for patients suffering from β-thalassemia or sickle cell disease, which can be caused by mutations in the hemoglobin β-subunit gene (*HBB*). Patients with these mutations lose the ability to produce fetal hemoglobin (HbF; hemoglobin F) in red blood cells[46,47]. In 2018, an FDA approved *ex vivo* CRISPR gene-edited therapy (named CTX001) was developed, and this treatment could enhance the level of HbF. Moreover, CRISPR/Cas9 can edit genomes in insects.

Previous studies demonstrated that this technology could induce gene knock-outs and mutations in different species of insects, including *Drosophila*[48,49], *Anopheles*[50], and *Bombyx mori*[51]. Taken together, CRISPR/Cas9 technology enables scientists to edit genomes in a precise and highly efficient way, facilitating the investigation of gene functions in many organisms.

Second, studies demonstrated that CRISPR technology plays roles in genome wide screening when the nuclease (Cas9) combines with pooled-sgRNA libraries[52-56]. These genetic screens are mediated by NHEJ, HDR, or CRISPR based transcriptional regulation[57]. The third application involves a "gene drive" system, which propagates genetic elements through populations by *Super*-Mendelian inheritance. This process requires participation of site-specific endonucleases (Cas9), and it has the potential to modify entire populations within a few years[15,58,59]. Finally, CRISPR based gene interference (CRISPRi) and CRISPR-mediated gene activation (CRISPRa) are two transcriptional regulation systems, and both include a mutated Cas9 (D10A/H840A) nuclease, a sgRNA and a fused regulator domain[60]. The mutated Cas9 (dCas9) loses its nuclease activity but retains the ability to recognize target DNA[61]. Thus, this system can precisely regulate gene transcription when dCas9 is tagged with different transcriptional regulators, like VPR or Mxi1[62,63]. Additionally, CRISPR/dCas9 can also work as a "base editor" to create point mutations when tagged with either cytosine base editors (CBEs) or adenine base editors (ABEs)[64]. For instance, cytosine base editors catalyze the conversion of C•G to T•A point mutations at target sites when fused to dCas9. Moreover, the CRISPR/dCas9 system is critical for the development of genetic detection tools, such as CRISPR Chip[65]. This technique detects unamplified target genes within 15 mins, and it contains two main components: the CRISPR/dCas9 complex and the sensitive detector graphene-based field-effect transistor (gFET). In this system,

the dCas9/sgRNA complex interacts with its target sequences, and gFET is functionally sensitive to the adsorption/interaction of charged molecules at its surface.

CRISPR based gene drive technology

The CRISPR based genome editing technology has been considered as a revolutionary technology across many different organisms, including mammalian cells[42,66,67], plants[43,68], insects[69-71], different species of bacteria[72,73] and fungi[15,58,59,74,75]. This powerful technology has the potential to modify an entire population. CRISPR/Cas technology can be utilized to solve a specific set of problems involving the control of biological populations. For example, human health is threatened by many insect-borne diseases, like Zika, malaria or Lyme disease[76,77]. Infected insects work as vectors to transmit pathogenetic organisms into humans and cause severe symptoms or death. Malaria is a mosquito-borne disease caused by infection of *Plasmodium*. Symptoms of malaria include fever, tiredness, vomiting, and headaches[78]. The *Plasmodium* parasite is a unicellular eukaryote, and it can be transmitted to humans and animals through the bite of infected mosquitoes[79]. This disease has been spread widely across the globe, including Africa, Asia, and Latin America. In 2020, more than 627,000 people died because of malaria, and 95% of the deaths occurred in Sub-Saharan Africa[80]. To inhibit the spread of these insect-borne diseases, one of the current most effective strategies is to eliminate the population of these infected insects.

Additionally, control of populations is critical in maintaining human food supplies, natural habitats, and ecological stability. A report showed that control of weed and pest populations benefits the development of agriculture. Their results displayed that the potential loss of global crop production varied from 50% to 80%[81], of which the reduction from weeds and pests were around 34% and 20%, respectively. Moreover, the emergence of invasive species in different

ecosystems has raised concerns about the future of their ecological stability. These invasive species can be animals, plants, and fungi. Since most invasive species arrive without their native predators, they can increase their populations and reproduce rapidly in the new environment. The existence of invasive species seriously threatens the survival of native species. In Australia, invasive cats and red foxes serve as predators and contribute to extinction of many native animals. A report demonstrated that around 400 vertebrate species were preyed on by feral cats[82]. Invasive plants can also kill native plants by preventing access to sunlight. For instance, the *Kudzu* plant could inhibit the growth of mature trees and change the structure of their ecosystem[83]. In marine ecosystems, some invasive fish take resources away from native species and eliminate their populations by preying on their eggs, such as Asian carp in America. Hence, it is critical to control or eradicate populations of these organisms.

First, populations of invasive species can be eliminated through biological control. The natural predators of invasive species can be introduced to target invasive species and reduce their populations. This strategy has been considered as a safe and cost-effective tool for pest management. However, a study demonstrated that these natural predators also affect populations of non-target species[84], indicating that this type of strategy may be not specific. Second, pests, weeds, other invasive insects, and fungi can be killed by different kinds of chemicals, including pesticides, herbicides, fungicides, and insecticides. However, other species may be exposed to these same chemicals. For instance, inhalation of pesticides can cause short-term effects on human health. Many pesticides and insecticides also can induce long-term effects when ingested or absorbed through the skin. These chemicals poison the human body by disrupting the immune system, nervous systems, and cell division[85]. Furthermore, the extensive use of pesticides and herbicides causes the occurrence of pesticide and/or herbicides resistance. Chemical control has

the potential to develop drug-resistant species and is toxic to human and other non-target species. Finally, we can restrict the access of the invasive species by physical barriers. This strategy is limited because of its high costs and low efficiency[86]. Taken together, a more specific, highly efficient but inexpensive strategy is required to control these biological populations. CRISPR-based gene drive technology has the potential to solve many of these problems.

In gene drive systems, the DNA encoding both the Cas9 nuclease gene and sgRNA expression construct are integrated at a specific genomic location (Fig. 1.2). The target is the corresponding gene on the homologous chromosome in a polyploid genome. Then, the DSB will be repaired by HDR using the gene drive allele as a donor DNA; the wild-type allele will be replaced with the gene drive allele[14]. The heterozygote will be converted to a homozygous state. As a result, two gene drive alleles will be inherited, and WT alleles are destroyed. This is referred to as "*Super*-Mendelian" genetics.

CRISPR/Cas9 gene drive technology biases inheritance of desired genetic elements and propagates these elements rapidly through the entire population (Fig. 1.2). For instance, gene drive modified mosquitoes would mate with wild-type mosquitoes, resulting in a 100% inheritance rate of gene drive alleles in the offspring. Therefore, CRISPR/Cas9 gene drive systems could specifically modify entire populations within a few generations. In theory, gene drive systems can be utilized to eliminate invasive species or disease vectors by causing the population to crash. This could be done by targeting genes essential for sex determination; causing all male (or female) offspring will result in the prevention of reproduction. In theory, this process could be extended to the entire earth to eliminate the species. CRISPR/Cas9 dependent gene drive technology also can be used in agricultural weed management by either inserting population-suppression drives or population-sensitizing drives[87]. Scientists developed CRISPR gene drive systems in different

species and have demonstrated use in fungi, mammals, and insects.

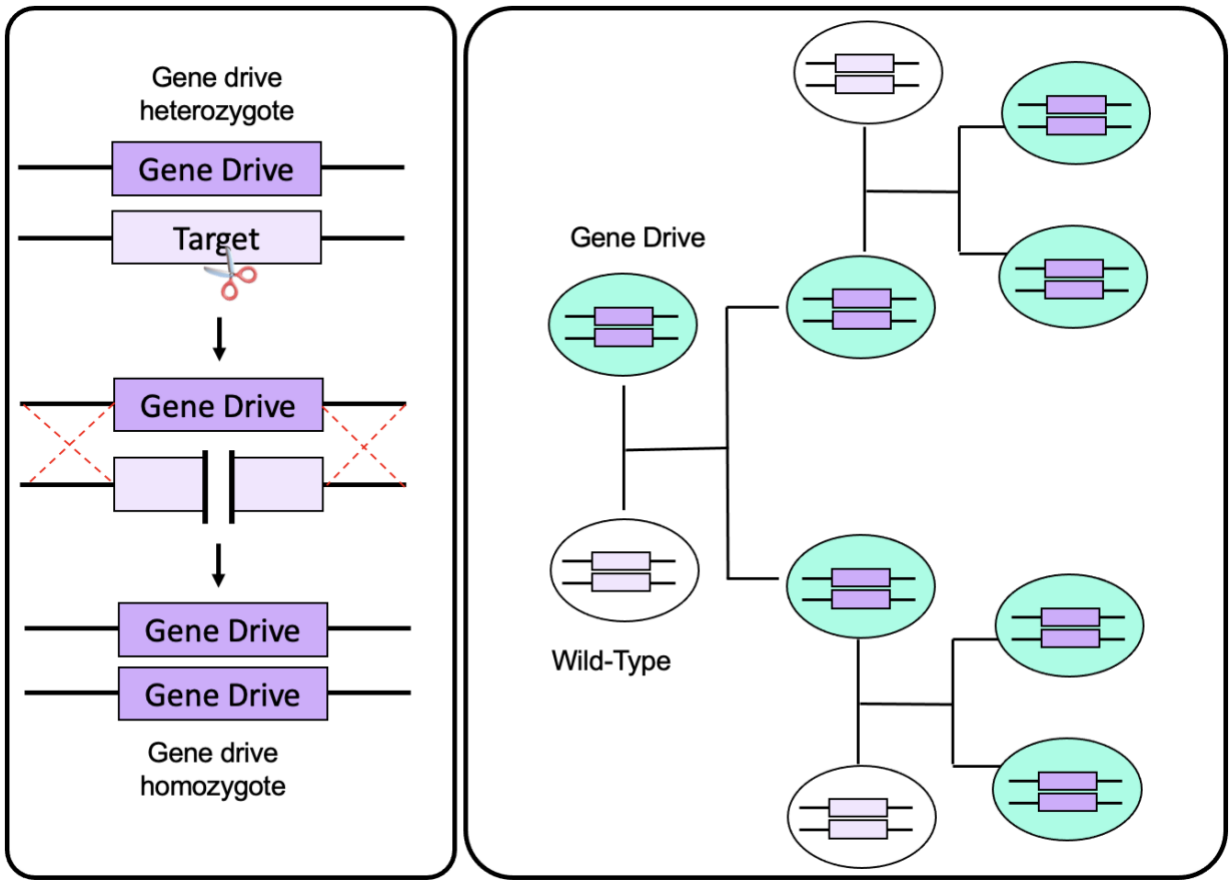


Figure 1.2. Mechanism of CRISPR/Cas9 gene drive system.

Left, CRISPR-based gene drive cuts the wild-type copy of a target chromosome. Next, the gene drive DNA serve as the donor to fix the broken chromosome using HDR. The heterozygous gene drive/WT cell converts to the homozygous gene drive/gene drive state. *Right,* expected outcome when gene drive technology is applied within a population. Figure adapted from Yan and Finnigan 2018[15], and DiCarlo *et al.* 2015[88].

Recently, a study developed a male-biased sex-distorter gene drive (SDGD) system to eliminate a lab population of female mosquitos, since malaria is transmitted by only female insects[89]. They fused a CRISPR-based gene drive with the endonuclease I-PpoI at the *doublesex* (*dsx*) gene locus in *Anopheles gambiae*. The CRISPR gene drive was responsible for targeting the haplosufficient female fertility genes (*AGAP011377, AGAP007280* and *AGAP005958*), and the I-PpoI endonuclease could cleave a conserved sequence on the X chromosome. Consequently, their

study predicted that the SDGD system could inhibit the spread of malaria by impairing female fertility and producing a large population of male mosquitos.

However, there are potential safety concerns, including i) issues of gene drive control, ii) gene drive design, and iii) the occurrence of gene drive resistance and off-target effects. First, gene drive systems are very powerful at spreading genetic elements through entire populations, including cases of malicious or accidental release. The challenge is that safety mechanisms to control or shut off the gene drive system do not currently exist. To solve this problem, scientists have developed the gene drive reversal systems[90] and gene drive inhibition systems[91] to safeguard against active drives. Second, many types of gene drive designs have been demonstrated, such as "daisy-chain drives" or "underdominance gene drives"[92,93], but these designs will be limited when there are multiple genetic elements required for editing. Third, a major concern for CRISPR-based gene drive systems is the evolution of gene drive resistance[94-96], which can be induced by NHEJ based repair rather than HDR. The introduction of indels at target sites prevents Cas9 editing and prevents gene drive action (forming resistant alleles). Previous studies demonstrated several strategies to reduce the occurrence of resistance alleles, including utilization of conserved target sequences, multi-sgRNAs, and paired nickases[97].

In this dissertation, I developed an improved artificial gene drive system in budding yeast. This system only requires a single copy of the Cas9 nuclease and three sgRNAs to propagate three separate loci precisely. Our results revealed that this triple drive system was highly efficient and could edit multiple loci in budding yeast. We found that gene drives in NHEJ deficient cells still functioned similar to gene drives in NHEJ competent cells (Chapter 2). Next, we analyzed formation of drive resistant cells, and tested the editing activity with two independent sgRNAs (Chapter 3). Finally, we examined transcriptional regulation using the CRISPR/dCas9 system. We

showed that dCas9 could activate the transcription of target genes when it was tagged with the transcriptional activator (VPR) and inhibit transcription when tagged with the transcriptional repressor (Mxi1) (Chapter 4). Together, these findings will assist in the development of future gene drive constructs that are safe, reversable, and highly effective to address the many challenges of controlling biological populations.

References

1 Takata, M. *et al.* Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. *The EMBO journal* **17**, 5497-5508 (1998).

2 Salsman, J. & Dellaire, G. Precision genome editing in the CRISPR era. *Biochemistry and cell biology* **95**, 187-201 (2017).

3 Mladenov, E., Magin, S., Soni, A. & Iliakis, G. in *Seminars in cancer biology.* 51-64 (Elsevier).

4 Branzei, D. & Foiani, M. Regulation of DNA repair throughout the cell cycle. *Nature reviews Molecular cell biology* **9**, 297-308 (2008).

5 Capecchi, M. R. Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. *Nature reviews genetics* **6**, 507-512 (2005).

6 Chiruvella, K. K., Liang, Z. & Wilson, T. E. Repair of double-strand breaks by end joining. *Cold Spring Harbor perspectives in biology* **5**, a012757 (2013).

7 Smih, F., Rouet, P., Romanienko, P. J. & Jasin, M. Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells. *Nucleic acids research* **23**, 5012-5019 (1995).

8 Rouet, P., Smih, F. & Jasin, M. Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proceedings of the National Academy of Sciences* **91**, 6064-6068 (1994).

9 Miller, J. C. *et al.* A TALE nuclease architecture for efficient genome editing. *Nature biotechnology* **29**, 143-148 (2011).

10 Kim, Y.-G., Cha, J. & Chandrasegaran, S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences* **93**, 1156-1160 (1996).

11 Porteus, M. H. & Carroll, D. Gene targeting using zinc finger nucleases. *Nature biotechnology* **23**, 967-973 (2005).

12 Carroll, D. Genome engineering with zinc-finger nucleases. *Genetics* **188**, 773-782 (2011).

13 Sun, N. & Zhao, H. Transcription activator-like effector nucleases (TALENs): a highly efficient and versatile tool for genome editing. *Biotechnology and bioengineering* **110**, 1811-1821 (2013).

14 Esvelt, K. M., Smidler, A. L., Catteruccia, F. & Church, G. M. Concerning RNA-guided gene drives for the alteration of wild populations. *Elife* **3**, doi:10.7554/eLife.03401 (2014).

15 Yan, Y. & Finnigan, G. C. Development of a multi-locus CRISPR gene drive system in budding yeast. *Sci Rep* **8**, 17277, doi:10.1038/s41598-018-34909-3 (2018).

16 Yan, M. Y. *et al.* CRISPR-Cas12a-Assisted Recombineering in Bacteria. *Appl Environ Microbiol* **83**, doi:10.1128/aem.00947-17 (2017).

17 Wu, Y. *et al.* Correction of a genetic disease in mouse via use of CRISPR-Cas9. *Cell Stem Cell* **13**, 659-662, doi:10.1016/j.stem.2013.10.016 (2013).

18 Vermersch, E., Jouve, C. & Hulot, J. S. CRISPR/Cas9 gene-editing strategies in cardiovascular cells. *Cardiovasc Res* **116**, 894-907, doi:10.1093/cvr/cvz250 (2020).

19 Kistler, K. E., Vosshall, L. B. & Matthews, B. J. Genome engineering with CRISPR-Cas9 in the mosquito Aedes aegypti. *Cell Rep* **11**, 51-60, doi:10.1016/j.celrep.2015.03.009 (2015).

20 Espino-Saldaña, A. E., Rodríguez-Ortiz, R., Pereida-Jaramillo, E. & Martínez-Torres, A. Modeling Neuronal Diseases in Zebrafish in the Era of CRISPR. *Curr Neuropharmacol* **18**, 136-152, doi:10.2174/1570159x17666191001145550 (2020).

21 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. *Journal of bacteriology* **169**, 5429-5433 (1987).

22 Mojica, F. J., Díez-Villaseñor, C., García-Martínez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* **60**, 174-182, doi:10.1007/s00239-004-0046-3 (2005).

23 Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709-1712, doi:10.1126/science.1138140 (2007).

24 Semenova, E. & Severinov, K. in *Evolutionary Biology: Self/Nonself Evolution, Species and Complex Traits Evolution, Methods and Concepts* 51-62 (Springer, 2017).

25 Patterson, A. G., Yevstigneyeva, M. S. & Fineran, P. C. Regulation of CRISPR–Cas adaptive immune systems. *Current opinion in microbiology* **37**, 1-7 (2017).

26 Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR–Cas systems. *Nature Reviews Microbiology* **13**, 722-736 (2015).

27 Liu, Z., Dong, H., Cui, Y., Cong, L. & Zhang, D. Application of different types of CRISPR/Cas-based systems in bacteria. *Microbial cell factories* **19**, 1-14 (2020).

28 Tsai, S. Q. & Joung, J. K. Defining and improving the genome-wide specificities of CRISPR–Cas9 nucleases. *Nature Reviews Genetics* **17**, 300-312 (2016).

29 Jinek, M. *et al.* A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *science* **337**, 816-821 (2012).

30 Sternberg, S. H. & Doudna, J. A. Expanding the biologist's toolkit with CRISPR-Cas9. *Molecular cell* **58**, 568-574 (2015).

31 Hryhorowicz, M., Lipiński, D., Zeyland, J. & Słomski, R. CRISPR/Cas9 Immune System as a Tool for Genome Engineering. *Arch Immunol Ther Exp (Warsz)* **65**, 233-240, doi:10.1007/s00005-016-0427-5 (2017).

32 Harrison, M. M., Jenkins, B. V., O'Connor-Giles, K. M. & Wildonger, J. A CRISPR view of development. *Genes & development* **28**, 1859-1872 (2014).

33 Yao, X. *et al.* Homology-mediated end joining-based targeted integration using CRISPR/Cas9. *Cell research* **27**, 801-814 (2017).

34 Lieber, M. R. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annual review of biochemistry* **79**, 181-211 (2010).

35 Martinez-Lage, M., Torres-Ruiz, R. & Rodriguez-Perales, S. CRISPR/Cas9 technology: applications and human disease modeling. *Progress in molecular biology and translational science* **152**, 23-48 (2017).

36 Zetsche, B. *et al.* Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759-771 (2015).

37 Paul, B. & Montoya, G. CRISPR-Cas12a: Functional overview and applications. *Biomed J* **43**, 8-17, doi:10.1016/j.bj.2019.10.005 (2020).

38 Dong, L. *et al.* An anti-CRISPR protein disables type V Cas12a by acetylation. *Nat Struct Mol Biol* **26**, 308-314, doi:10.1038/s41594-019-0206-1 (2019).

39 Chen, J. S. *et al.* CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* **360**, 436-439, doi:10.1126/science.aar6245 (2018).

40 Yao, R. *et al.* CRISPR-Cas9/Cas12a biotechnology and application in bacteria. *Synth Syst Biotechnol* **3**, 135-149, doi:10.1016/j.synbio.2018.09.004 (2018).

41 Swarts, D. C. & Jinek, M. Cas9 versus Cas12a/Cpf1: Structure-function comparisons and implications for genome editing. *Wiley Interdiscip Rev RNA* **9**, e1481, doi:10.1002/wrna.1481 (2018).

42 Redman, M., King, A., Watson, C. & King, D. What is CRISPR/Cas9? *Archives of Disease in Childhood-Education and Practice* **101**, 213-215 (2016).

43 Garrood, W. T. *et al.* Analysis of off-target effects in CRISPR-based gene drives in the human malaria mosquito. *Proc Natl Acad Sci U S A* **118**, doi:10.1073/pnas.2004838117 (2021).

44 Ghaemi, A. *et al.* CRISPR-cas9 genome editing delivery systems for targeted cancer therapy. *Life Sci* **267**, 118969, doi:10.1016/j.lfs.2020.118969 (2021).

45 Zhang, Y., Ma, X., Xie, X. & Liu, Y. G. CRISPR/Cas9-Based Genome Editing in Plants. *Prog Mol Biol Transl Sci* **149**, 133-150, doi:10.1016/bs.pmbts.2017.03.008 (2017).

46 Zhou, J. *et al.* One-step generation of different immunodeficient mice with multiple gene modifications by CRISPR/Cas9 mediated genome engineering. *The international journal of biochemistry & cell biology* **46**, 49-55 (2014).

47 Yin, H. *et al.* Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. *Nature biotechnology* **32**, 551-553 (2014).

48 Frangoul, H. *et al.* CRISPR-Cas9 gene editing for sickle cell disease and β-thalassemia. *New England Journal of Medicine* **384**, 252-260 (2021).

49 Demirci, S., Leonard, A., Haro-Mora, J. J., Uchida, N. & Tisdale, J. F. CRISPR/Cas9 for sickle cell disease: applications, future possibilities, and challenges. *Cell Biology and Translational Medicine, Volume 5*, 37-52 (2019).

50 Yu, Z. *et al.* Highly efficient genome modifications mediated by CRISPR/Cas9 in Drosophila. *Genetics* **195**, 289-291 (2013).

51 Bassett, A. R., Tibbit, C., Ponting, C. P. & Liu, J.-L. Highly efficient targeted mutagenesis of Drosophila with the CRISPR/Cas9 system. *Cell reports* **4**, 220-228 (2013).

52 Hammond, A. *et al.* A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae. *Nature biotechnology* **34**, 78-83 (2016).

53 Daimon, T., Kiuchi, T. & Takasu, Y. Recent progress in genome engineering techniques in the silkworm, B ombyx mori. *Development, growth & differentiation* **56**, 14-25 (2014).

54 Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods* **11**, 783-784, doi:10.1038/nmeth.3047 (2014).

55 Ringel, T. *et al.* Genome-Scale CRISPR Screening in Human Intestinal Organoids Identifies Drivers of TGF-β Resistance. *Cell Stem Cell* **26**, 431-440.e438, doi:10.1016/j.stem.2020.02.007 (2020).

56 Koike-Yusa, H., Li, Y., Tan, E. P., Velasco-Herrera Mdel, C. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat Biotechnol* **32**, 267-273, doi:10.1038/nbt.2800 (2014).

57 Zhang, C., Quan, R. & Wang, J. Development and application of CRISPR/Cas9 technologies in genomic editing. *Human molecular genetics* **27**, R79-R88 (2018).

58 Schumann, K. *et al.* Generation of knock-in primary human T cells using Cas9 ribonucleoproteins. *Proceedings of the National Academy of Sciences* **112**, 10437-10442 (2015).

59 Klann, T. S. *et al.* CRISPR–Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. *Nature biotechnology* **35**, 561-568 (2017).

60 Yan, Y. & Finnigan, G. C. Analysis of CRISPR gene drive design in budding yeast. *Access Microbiol* **1**, e000059, doi:10.1099/acmi.0.000059 (2019).

61 Goeckel, M. E. *et al.* Modulating CRISPR gene drive activity through nucleocytoplasmic localization of Cas9 in S. cerevisiae. *Fungal Biol Biotechnol* **6**, 2, doi:10.1186/s40694-019-0065-x (2019).

62 Gilbert, L. A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442-451, doi:10.1016/j.cell.2013.06.044 (2013).

63 Karlson, C. K. S., Mohd-Noor, S. N., Nolte, N. & Tan, B. C. CRISPR/dCas9-based systems: Mechanisms and applications in plant sciences. *Plants* **10**, 2055 (2021).

64 Perez-Pinera, P. *et al.* RNA-guided gene activation by CRISPR-Cas9–based transcription factors. *Nature methods* **10**, 973-976 (2013).

65 Chavez, A. *et al.* Highly efficient Cas9-mediated transcriptional programming. *Nature methods* **12**, 326-328 (2015).

66 Anzalone, A. V., Koblan, L. W. & Liu, D. R. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nature biotechnology* **38**, 824-844 (2020).

67 Hajian, R. *et al.* Detection of unamplified target genes via CRISPR–Cas9 immobilized on a graphene field-effect transistor. *Nature biomedical engineering* **3**, 427-437 (2019).

68 Khan, F. A. *et al.* CRISPR/Cas9 therapeutics: a cure for cancer and other genetic diseases. *Oncotarget* **7**, 52541-52552, doi:10.18632/oncotarget.9646 (2016).

69 Chen, M. *et al.* CRISPR-Cas9 for cancer therapy: Opportunities and challenges. *Cancer Lett* **447**, 48-55, doi:10.1016/j.canlet.2019.01.017 (2019).

70 Xing, H. L. *et al.* A CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC Plant Biol* **14**, 327, doi:10.1186/s12870-014-0327-y (2014).

71 Tong, X. L. *et al.* [Applications of the CRISPR/Cas9 system in insects]. *Yi Chuan* **40**, 266-278, doi:10.16288/j.yczz.17-263 (2018).

72 Taning, C. N. T., Van Eynde, B., Yu, N., Ma, S. & Smagghe, G. CRISPR/Cas9 in insects: Applications, best practices and biosafety concerns. *J Insect Physiol* **98**, 245-257, doi:10.1016/j.jinsphys.2017.01.007 (2017).

73 Chen, L., Wang, G., Zhu, Y. N., Xiang, H. & Wang, W. Advances and perspectives in the application of CRISPR/Cas9 in insects. *Dongwuxue Yanjiu* **37**, 220-228, doi:10.13918/j.issn.2095-8137.2016.4.220 (2016).

74 Vigouroux, A. & Bikard, D. CRISPR Tools To Control Gene Expression in Bacteria. *Microbiol Mol Biol Rev* **84**, doi:10.1128/mmbr.00077-19 (2020).

75 Cho, S., Shin, J. & Cho, B. K. Applications of CRISPR/Cas System to Bacterial Metabolic Engineering. *Int J Mol Sci* **19**, doi:10.3390/ijms19041089 (2018).

76 Roggenkamp, E. *et al.* CRISPR-UnLOCK: Multipurpose Cas9-Based Strategies for Conversion of Yeast Libraries and Strains. *Front Microbiol* **8**, 1773, doi:10.3389/fmicb.2017.01773 (2017).

77 Roggenkamp, E. *et al.* Tuning CRISPR-Cas9 Gene Drives in Saccharomyces cerevisiae. *G3 (Bethesda)* **8**, 999-1018, doi:10.1534/g3.117.300557 (2018).

78 Nelder, M. P. *et al.* Human pathogens associated with the blacklegged tick Ixodes scapularis: a systematic review. *Parasites & vectors* **9**, 1-14 (2016).

79 Berthet, N. *et al.* Molecular characterization of three Zika flaviviruses obtained from sylvatic mosquitoes in the Central African Republic. *Vector-Borne and Zoonotic Diseases* **14**, 862-865 (2014).

80 Caraballo, H. & King, K. Emergency department management of mosquito-borne illness: malaria, dengue, and West Nile virus. *Emergency medicine practice* **16**, 1-23; quiz 23 (2014).

81 Courtier-Orgogozo, V., Morizot, B. & Boëte, C. Agricultural pest control with CRISPR-based gene drive: time for public debate: Should we use gene drive for pest control? *EMBO Rep* **18**, 878-880, doi:10.15252/embr.201744205 (2017).

82 Prevention, C. f. D. C. a. Malaria's Impact Worldwide. (2021).

83 Oerke, E.-C. Crop losses to pests. *The Journal of Agricultural Science* **144**, 31-43 (2006).

84 Doherty, T. S. *et al.* A continental-scale analysis of feral cat diet in Australia. *Journal of Biogeography* **42**, 964-975 (2015).

85 Montagnani, C., Casazza, G., Gentili, R., Caronni, S. & Citterio, S. Kudzu in Europe: niche conservatism for a highly invasive plant. *Biological Invasions*, 1-16 (2022).

86 Messing, R. H. & Wright, M. G. Biological control of invasive species: solution or pollution? *Frontiers in Ecology and the Environment* **4**, 132-140 (2006).

87 Nicolopoulou-Stamati, P., Maipas, S., Kotampasi, C., Stamatis, P. & Hens, L. Chemical pesticides and human health: the urgent need for a new concept in agriculture. *Frontiers in public health* **4**, 148 (2016).

88 Hanley, N. & Roberts, M. The economic benefits of invasive species management. *People and Nature* **1**, 124-137 (2019).

89 Neve, P. Gene drive systems: do they have a place in agricultural weed management? *Pest Manag Sci* **74**, 2671-2679, doi:10.1002/ps.5137 (2018).

90 DiCarlo, J. E., Chavez, A., Dietz, S. L., Esvelt, K. M. & Church, G. M. Safeguarding CRISPR-Cas9 gene drives in yeast. *Nature biotechnology* **33**, 1250-1255 (2015).

91 Simoni, A. *et al.* A male-biased sex-distorter gene drive for the human malaria vector Anopheles gambiae. *Nature biotechnology* **38**, 1054-1060 (2020).

92 Vella, M. R., Gunning, C. E., Lloyd, A. L. & Gould, F. Evaluating strategies for reversing CRISPR-Cas9 gene drives. *Scientific reports* **7**, 1-8 (2017).

93 Basgall, E. M. *et al.* Gene drive inhibition by the anti-CRISPR proteins AcrIIA2 and AcrIIA4 in Saccharomyces cerevisiae. *Microbiology* **164**, 464 (2018).

94 Dhole, S., Vella, M. R., Lloyd, A. L. & Gould, F. Invasion and migration of spatially self-limiting gene drives: A comparative analysis. *Evolutionary applications* **11**, 794-808 (2018).

95 Burt, A. & Crisanti, A. Gene drive: evolved and synthetic. *ACS chemical biology* **13**, 343-346 (2018).

96 Unckless, R. L., Clark, A. G. & Messer, P. W. Evolution of Resistance Against CRISPR/Cas9 Gene Drive. *Genetics* **205**, 827-841, doi:10.1534/genetics.116.197285 (2017).

97 Hammond, A. M. *et al.* The creation and selection of mutations resistant to a gene drive over multiple generations in the malaria mosquito. *PLoS genetics* **13**, e1007039 (2017).

98 Champer, J. *et al.* Reducing resistance allele formation in CRISPR gene drive. *Proceedings of the National Academy of Sciences* **115**, 5522-5527 (2018).

99 Koo, T., Lee, J. & Kim, J.-S. Measuring and reducing off-target activities of programmable nucleases including CRISPR-Cas9. *Molecules and cells* **38**, 475 (2015).

Chapter 2 - Development of a multi-locus CRISPR gene drive system in budding yeast

This chapter has been published as a journal article in *Scientific reports.*

Yao Yan[1] and Gregory C. Finnigan[1*]

[1]Department of Biochemistry and Molecular Biophysics, Kansas State University, 141 Chalmers Hall, Manhattan, KS 66506 USA

*Correspondence to: Gregory C. Finnigan, Phone: (785) 532-6939; FAX; (785) 532-7278; E-mail: gfinnigan@ksu.edu

Running title: Multi-locus CRISPR gene drive

Keywords: CRISPR, Cas9, gene drive, biotechnology, yeast, multi-locus

Abstract

The discovery of CRISPR/Cas gene editing has allowed for major advances in many biomedical disciplines and basic research. One arrangement of this biotechnology, a nuclease-based gene drive, can rapidly deliver a genetic element through a given population and studies in fungi and metazoans have demonstrated the success of such a system. This methodology has the potential to control biological populations and contribute to eradication of insect-borne diseases, agricultural pests, and invasive species. However, there remain challenges in the design, optimization, and implementation of gene drives including concerns regarding biosafety, containment, and control/inhibition. Given the numerous gene drive arrangements possible, there is a growing need for more advanced designs. In this study, we use budding yeast to develop an artificial multi-locus gene drive system. Our minimal setup requires only a single copy of *S. pyogenes* Cas9 and three guide RNAs to propagate three gene drives. We demonstrate how this system could be used for targeted allele replacement of native genes and to suppress NHEJ repair systems by modifying DNA Ligase IV. A multi-locus gene drive configuration provides an expanded suite of options for complex attributes including pathway redundancy, combatting evolved resistance, and safeguards for control, inhibition, or reversal of drive action.

Introduction

The discovery and implementation of the clustered regularly interspaced short palindromic repeat (CRISPR) gene editing system has revolutionized countless fields and sub-specialties across molecular biology and biotechnology to improve human health, agriculture, ecological control, and beyond. Briefly, alteration of the genetic code is accomplished using (i) a bacterial derived nuclease (typically Cas9 or Cas12a), (ii) a single-stranded fragment of "guide" RNA (sgRNA), and (iii) an optional exogenous repair fragment of DNA[1-4]. Priming of the nuclease with a pre-programmed guide RNA fragment targets a specific genomic sequence for a double stranded break (DSB). Following DNA cleavage, eukaryotic cells activate repair systems to either fuse broken chromosomal ends together via non-homologous end joining (NHEJ) or, in the presence of donor DNA, introduce exogenous sequence via homologous recombination (HR). Moreover, the CRISPR methodology is not restricted to DSB-induced alteration of the genome—recent efforts have demonstrated that nuclease-dead variants (e.g. dCas9) can serve as delivery systems to modulate transcriptional activity[5], alter epigenetic landscapes[6], or introduce mutational substitutions *sans* any DNA cleavage event[7].

One powerful biotechnological application of the CRISPR methodology is within a "gene drive" (GD) system. The basic design of a homing drive includes the expression constructs for the CRISPR nuclease, and the corresponding guide RNA positioned at a desired locus of choice—the mechanism of propagation involves targeting of the homologous chromosome (within a diploid or polyploid organism) at the same genetic position (typically cleaving the wild-type gene). Creation of a DSB followed by HR-based repair (using the gene drive-containing DNA as a donor) causes the entire artificial construct (Cas9, the sgRNA, and any desired "cargo") to be copied; in this way, a heterozygous cell is automatically converted to the homozygous state. This *super*-Mendelian

genetic arrangement allows for the forced propagation of a genetic element within a population and has the potential to modify entire species on a global scale[8,9]. Some of the possible benefits of this technology include eradication of invasive species[10,11], agricultural pest management[12], and elimination of insect-borne diseases such as malaria[9,13,14]. A number of recent studies have demonstrated the potency and success of CRISPR-based gene drives in fungi[15-18], and metazoans[19-22]. While ongoing technical challenges remain in the design, optimization, and field testing of gene drive-harboring organisms, there are also serious biosafety and ethical concerns regarding use of this biotechnology as even current drive systems are expected to be highly invasive within native populations[23]. There is an immediate need for further study (*in silico* and *in vivo*) of gene drive systems that focus on issues of safety[15,24,25], control and reversal[26,27], and optimal design[11].

There are many types of gene drive designs including "daisy-chain drives," "underdominance drives," and "anti-drives," each with a distinct arrangement of the basic CRISPR components that is predicted to sweep through native populations at varying levels/rates[9,28,29]. Moreover, the need for additional drive components (more than one guide RNA construct), genetic safeguards, and built-in redundancy, calls for a new level of complexity within drive architecture. Here, we demonstrate use of *multiple gene drives* across three chromosomal loci within an artificial budding yeast system. Our "minimal" multi-locus gene drive arrangement requires only a single copy of the *S. pyogenes* Cas9 gene (installed at one position), along with three distinct guide RNAs to multiplex the nuclease throughout the genome. We demonstrate that this technique could be used to perform targeted replacement of a native gene (under its endogenous promoter) *in trans* from the Cas9-harboring locus. Finally, reducing or modulating NHEJ by targeting the highly conserved DNA Ligase IV may provide a means to further bias HR-dependent repair and action

of gene drives across eukaryotic systems. Our method includes multiple layers of genetic safeguards as well as recommendations for future designs of multi-locus drive systems.

Materials and Methods

Yeast Strains and Plasmids

Standard molecular biology protocols were used to engineer all *S. cerevisiae* strains (Supplementary Table B.1) used in this study[65]. The overall methodology for construction of the triple gene drive strain utilized both standard HR-based chromosomal integrations (*sans* any DSB) and Cas9-based editing (Supplementary Fig. B.1). Briefly, DNA constructs were first assembled onto *CEN*-based plasmids (typically pRS315) using *in vivo* assembly in yeast[66]. If necessary, point mutations were introduced using PCR mutagenesis[67]. Next, the engineered cassette was amplified with a high-fidelity polymerase (KOD Hot Start, EMD Millipore), transformed into yeast using a modified lithium acetate method[68], and integrated at the desired genomic locus. PCR was used to diagnose proper chromosomal position for each integration event followed by DNA sequencing. The DNA maps for manipulated yeast loci are included in Supplementary Fig. B.2. DNA plasmids used in this study can be found in Supplementary Table B.2. Expression cassettes for sgRNA were based on a previous study[69], purchased as synthetic genes (Genscript), and sub-cloned to high-copy plasmids using unique flanking restriction sites. All vectors were confirmed by Sanger sequencing.

Culture Conditions

Budding yeast were cultured in liquid or solid medium. YPD-based medium included 2% peptone, 1% yeast extract, and 2% dextrose. Synthetic (drop-out) medium included yeast nitrogen

base, ammonium sulfate, and amino acid supplements. The supplement mixture included adenine, arginine, tyrosine, isoleucine, phenylalanine, glutamic acid, aspartic acid, threonine, serine, valine, lysine, and methionine. For specific drop-out combinations, one or more of the following were removed: leucine, uracil, and/or histidine. Tryptophan (filter sterilized solution) was also added to media before final plating. A raffinose/sucrose mixture (2%/0.2%) was used to pre-induce cultures prior to treatment with galactose (2%). Yeast cultures were all grown in a 30°C incubator with shaking. All media was autoclaved or filter sterilized (sugars). For agar plates containing G418 sulfate, the final concentration was 240 µg/mL.

Cas9-based editing in vivo

Editing of haploid *S. cerevisiae* strains was performed as previously described[17]. Briefly, an integrated copy of *S. pyogenes* Cas9 was designed with two flanking "unique" (u2) sites—23 base pairs artificially introduced into the genome. This sequence contains a maximum mismatch to the native yeast genome and is used in order to (i) multiplex at two separate sites using a single guide RNA construct, (ii) minimize (or likely eliminate) potential off-target effects, and (iii) allow for increased biosecurity in testing of active CRISPR gene drive systems[38]. Haploid yeast were pre-induced overnight in a raffinose/sucrose mixture to saturation, back-diluted to an OD_{600} of approximately 0.35 in rich medium containing galactose, and cultured for 4.5 hr at 30°C. Equimolar amounts (1,000 ng) of high-copy plasmid (sgRNA) were transformed into yeast followed by recovery overnight in galactose and a final plating onto SD-LEU medium. Colonies were imaged and quantified after 3-4 days of growth. Haploid yeast editing experiments included three replicates in triplicate—all as separate transformation events—for each strain (n=9).

Gene drive activation and containment

Haploid yeast strains harboring the gene drive (Cas9) system were first transformed with the sgRNA-containing plasmid (*LEU2*-marked). Next, drive strains were mated to target strains of the opposite mating type on rich medium for 24 hr. Third, yeast were velvet-transferred to synthetic drop-out medium to select diploids (e.g. SD-URA-LEU or SD-URA-LEU-HIS); each haploid genome contained at least one unique selectable marker. Diploids were selected three consecutive rounds with 1-2 days incubation at each step. Fourth, yeast were cultured in pre-induction medium (raffinose/sucrose) lacking leucine overnight, back-diluted into rich medium containing galactose, and grown for 5 hr (or appropriate time intervals). Strains were diluted to approximately 100-500 cells per mL and plated onto SD-LEU for 2 days. Finally, colonies were transferred to the appropriate selection plates (e.g. SD-HIS, G418, SD-URA, and a fresh SD-LEU plate) for 1 additional day of growth before being imaged. The number of surviving colonies on each media type was quantified; experiments were performed in at least triplicate.

A number of safeguards were included in the design of all gene drive systems. First, the genomic targets for all guide RNAs included only non-yeast sequences (u1, GFP, and Kan^R)[18,38]. Second, the primary guide RNA cassette (u1) for targeting of the *HIS3* locus which included Cas9 was maintained on an unstable high-copy (2μ) plasmid; previous work has demonstrated loss of this vector type in the absence of active selection[15,17]. Third, the *S. cerevisiae* BY4741/BY4742 genetic background does not readily undergo sporulation, even under optimal conditions. Fourth, Cas9 expression was repressed by growth on dextrose until gene drives were activated. And finally, all diploid strains, plates, and consumables were autoclaved and inactivated.

Images, Graphics, Data and Evolutionary Analysis

Images (DNA gels, agar plates) were processed using ImageJ (National Institute of Health). For PCR reactions demonstrating the absence of a gene target (following gene drive activation), the original, unedited raw images were also included in Supplementary Figs. B.S7, B.S8, and B.S9 for comparison.

Data analysis (Fig. 2.4C) included error illustrated as the standard deviation of multiple independent trials and statistical comparisons were performed using an unpaired t-test.

Molecular graphics were generated using the Chimera software package from the Univ. of California, San Francisco[57]. Homologous sequences to the yeast DNA Ligase IV (Dnl4) protein were obtained using multiple BLAST (NCBI) searches within either the fungal or metazoan clade (Supplementary Table B.3). The phylogenetic tree of DNA Ligase IV was created using the Phylogeny.fr software[52,53]. Multiple sequence alignments were performed using Clustal Omega[54]. The predicted structures of the human, yeast, and mosquito Ligase IV enzyme were generated using I-TASSER[56]. The template structures included the human Lig4 N-terminus (PDB:3W1B)[70] and the yeast Dnl4 C-terminus (PDB:1Z56)[55]. Predicted models were individually aligned against the crystal structures using MatchMaker in Chimera. Metrics for the predicted structures are included in Supplementary Table B.4.

Results

Rationale and design of a multi-locus CRISPR gene drive

To date, a number of studies in fungi, insects, and now vertebrates, have demonstrated that CRISPR-based gene drive systems are effective in both single-celled and multicellular eukaryotes[15-17,19-22]. One of the benefits of homing systems is the ability to install additional genetic "cargo" proximal to the gene drive (consisting of a nuclease gene and an expression cassette for

the guide RNA). Current strategies use the gene drive cassette itself to delete and replace an endogenous gene, and/or include exogenous material as a desired cargo. However, there are a number of limitations to the use of a single locus harboring the entirety of the gene drive. First, addition of entire genetic pathways or large numbers of gene expression systems may be less efficient at HR-based copying of the drive. Second, introduction of additional endogenous gene(s) or modified alleles may require the native promoter system and/or epigenetic landscape to provide accurate and timely expression—this would not be possible at a single generic drive-containing locus. Third, given the observation of both natural (e.g. single nucleotide polymorphisms) and evolved resistance to gene drives through insertions or deletions (indel) resulting from NHEJ within insect populations[20,30-34], mechanisms for fortifying drive systems are still being elucidated. The proposal to increase the number of targeted double stranded breaks (and corresponding sgRNAs) to the single nuclease of choice (e.g. *S. pyogenes* Cas9) would greatly aid in combatting resistance[35-37]. However, an independent means to both minimize or escape resistance *and* ensure the intended biological outcome (deletion of the intended gene or introduction of the exogenous cargo) would involve a redundant delivery system. In this way, multiple gene drives (with multiple guide RNAs) within the same organism could target independent genetic loci either from the same, distinct, or parallel genetic pathways to achieve the desired outcome(s).

We envisioned two general strategies for the development of a gene drive system across distinct chromosomal positions: (i) each multi-locus "Complete" Gene Drive (CGD) would contain both a nuclease and corresponding guide RNA or (ii) a multi-locus "Minimal" Gene Drive (MGD) would include a nuclease and sgRNA, and all other genetic loci would *only* contain additional guide RNA cassettes (Fig. 2.1). We chose to focus on the latter strategy for a number of reasons, but we recognize that both would have distinct challenges and advantages. For one, a

possible technical hurdle to development of a modified organism with multiple CGDs would be the generation of distinct "large" expression system consisting of the entire nuclease gene, flanking untranslated region (UTR), the guide expression cassette(s), and any optional cargo compared to the MGD which removes the bulk of the drive system (nuclease expression) at additional loci.

Along these lines, the issue of appropriate expression of each of the nuclease gene(s) (whether identical or distinct) would need to be addressed using identical or modified promoter elements; this issue does not exist for a MGD with only a single copy of Cas9. Second, the issue of biosecurity and safeguarding against accidental or malicious release was taken into consideration. Given that a MGD would only harbor one copy of the nuclease, it would provide far less hurdles to counter and inactivate—either through use of an anti-drive system[15,26], by induced self-excision[17], or by removal of the Cas9-containing drive guide RNA[17]. Therefore, we have chosen to focus our study on design and testing of a three-locus MGD in budding yeast using the *S. pyogenes* Cas9 nuclease.

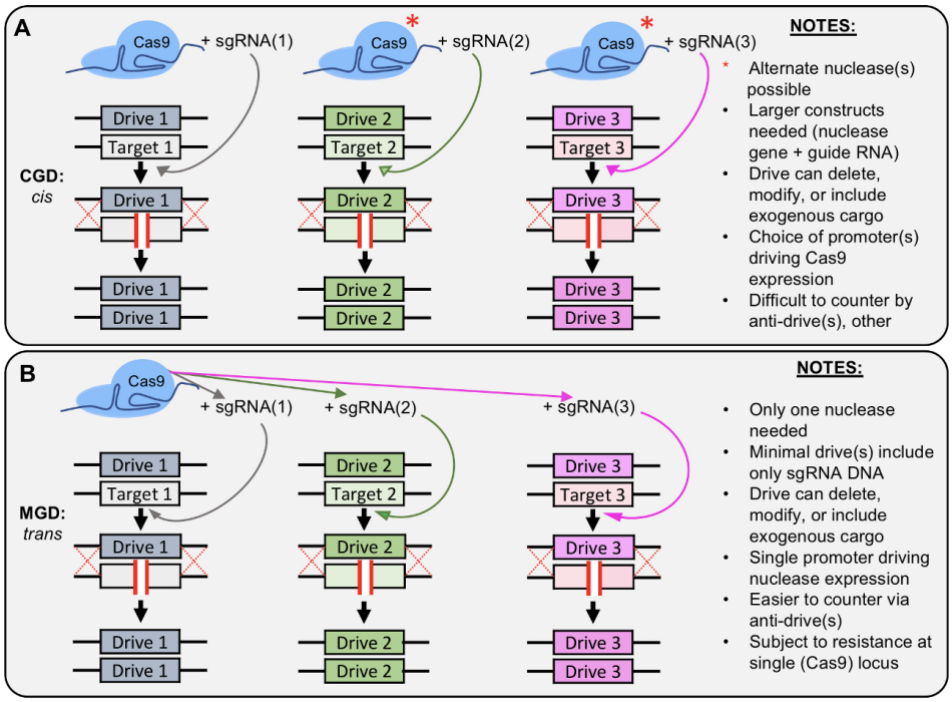


Figure 2.1. Models for multi-locus CRISPR gene drive systems.

(A) A proposed gene drive arrangement *in cis*. Each locus to be modified contains a "complete" system (drive and guide RNA cassette). These may be identical nuclease genes, altered variants, or sourced from separate species (e.g. Cas9 versus Cas12a). The action of each drive is fully independent from other drive-containing loci. (B) A single nuclease functions *in trans* across multiple loci with separate guide RNAs. This "minimal" design allows for greater safety and security (easily countered by a single anti-drive system or other means) but may be more susceptible to resistance at the primary (Cas9-harboring) locus.

An efficient triple gene drive system functions independently at each locus

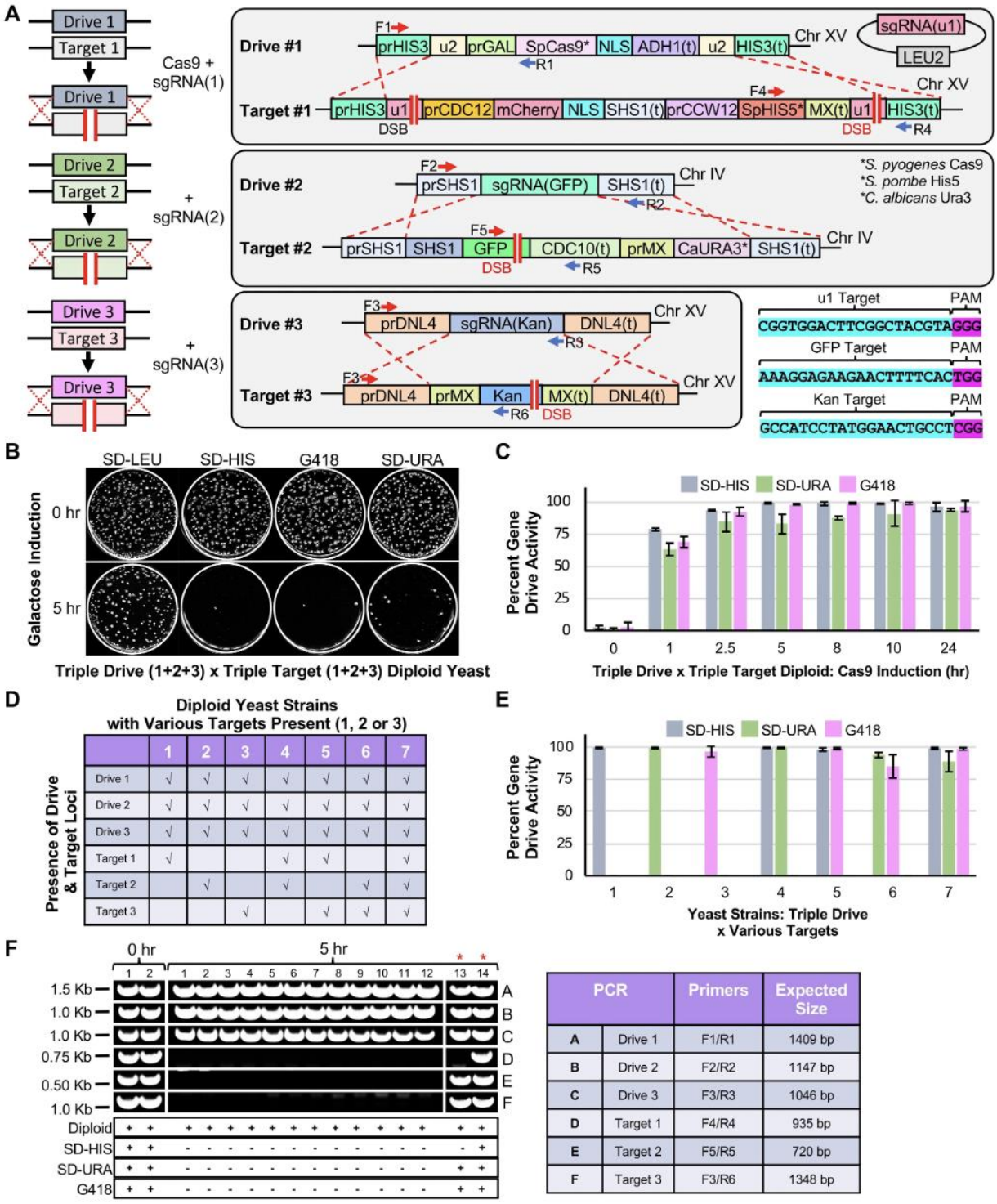
Our novel system includes the most potent genetic safeguard known to date used within a gene drive: artificial and non-native sequences used as targets. In this way, we have not only generated a haploid yeast strain harboring the MGD system at three genetic loci (*HIS3, SHS1,* and *DNL4*), but have also created a corresponding haploid strain with three distinct artificial *targets* at the same three loci (Fig. 2.2A). The "primary" drive at the *HIS3* locus includes (i) Cas9 under an inducible promoter (*GAL1/10*) commonly used for overexpression, (ii) flanking (u2) artificial sequences to be used for self-excision as a safeguard, and (iii) the absence of any selectable marker. The corresponding guide RNA cassette was installed on a high-copy plasmid for security reasons, but could have also been integrated proximal to the drive itself. The "secondary" and "tertiary" drive systems (*SHS1* and *DNL4*) are both non-essential genes and contain the minimum required components in the MGD design; in both cases, the native gene was deleted and fully replaced by the guide expression cassette (455 bp, although this could be reduced further) with no selectable marker. Construction of this complex haploid yeast strain used a combination of traditional HR-based integrations (with selectable markers), universal Cas9-targeting systems (CRISPR-UnLOCK)[18], and novel "self-editing" integration events (Supplementary Fig. B.1). To test the efficacy of the MGD, a three-locus "target" strain was generated: the *HIS3* locus was flanked by two (u1) sequences and included the *S.p.HIS5* selectable marker, the *SHS1* gene was fused with

GFP and contained the *C.a.URA3* marker, and finally, *DNL4* locus was deleted and replaced with the Kan^R drug cassette (Fig. 2.2A).

The triple MGD strain was first mated with the triple target strain to form a diploid, and Cas9 was activated by culturing in medium containing galactose (Fig. 2.2B). In the absence of nuclease expression (Fig. 2.2B, *top*), nearly all yeast colonies contained the (u1) guide plasmid (*LEU2*), and three selectable markers (*HIS5*, *URA3*, and Kan^R—providing resistance to G418). However, following a 5 hr incubation in galactose, >95% of all colonies were sensitive to all three growth conditions indicating a loss of all three selectable markers and replacement via the MGD (Fig. 2.2B, *bottom*). A time course of galactose induction illustrated highly efficient drive activity for all three loci by five hours; we noticed a slight lag in efficiency for the loss of the *URA3* marker (*SHS1*) locus until the 24 hr mark (Fig. 2.2C). This observation may be due to the *HIS3* and *DNL4* loci both being present on chromosome XV whereas *SHS1* was located on chromosome IV. Alternatively, differences in available guide RNAs (plasmid-borne versus integrated) or local epigenetic effects could cause this slight reduction in editing. Next, to ensure that action of the MGD at each locus was not dependent on the presence or absence of one or more of the intended targets (simulating "resistance" at one or more loci), we retested the triple drive strain against six additional strains, each lacking one or two of the proper targets and instead, contained the native yeast sequence: *his3Δ1*, *SHS1*, or *DNL4* (Fig 2.2D). We obtained similar results for each combination as the triple MGD strain (#7) indicating that each gene drive functioned independent of the presence of additional target(s) (Fig. 2.2E). We also observed that drive success at the *SHS1* locus slightly increased when fewer targets were presented. Finally, to ensure that the loss of the selectable marker was coupled to replacement of the target locus by the drive locus, we isolated clonal yeast from the MGD triple cross (Fig. 2.2B) and confirmed both the growth profile and

ploidy status of random samples (Fig. 2.2F, *bottom*). Diagnostic PCRs were performed on all six distinct loci to assay for the presence or absence of each engineered drive and target (Fig. 2.2F, Supplementary Fig. B.3). Oligonucleotides (Supplementary Table B.5) unique to specific drive/target elements were chosen; prior to Cas9 activation (0 hr), diploids contain all six distinct loci (two isolates). However, following activation of the nuclease, diploids maintained all three drive loci (PCRs A,B, and C), but lost all three target loci (PCRs D, E, and F) (twelve independent isolates).

We recognized that following the 5 hr drive activation, a small number of yeast colonies (<5% in most cases) still contained one or more selectable marker(s). We reasoned that these rare colonies likely arose from either complete or partial failure of the gene drive system for various possible reasons (poor expression, loss of guide RNA plasmid, NHEJ, alterations in ploidy, etc.). Therefore, we isolated and tested additional clones that displayed incomplete growth profiles across the three selection plates (isolates 13,14) (Fig. 2.2F). One isolate (13) had lost the (u1)-flanked target at the *HIS3* locus yet still contained the *SHS1* and *DNL4* markers. The second isolate (14) appeared to have lost the *LEU2*-based plasmid and all three target loci were still present (Fig. 2.2F). Following transformation with the (u1) guide vector, we examined a second round of drive activation from these two isolates and obtained a similar growth profile with a loss of the remaining loci indicating that at least some of the "failed" drive occurrences resulted from improper activation and/or targeting (Supplementary Fig. B.4). Of note, our gene drive system was activated in the absence of any selection—diploids were grown in rich medium containing galactose, and grown on SD-LEU plates prior to testing of the drive status on various medium. In this way, the action of the gene drive was performed in the absence of any selection or challenge.



PCR		Primers	Expected Size
A	Drive 1	F1/R1	1409 bp
B	Drive 2	F2/R2	1147 bp
C	Drive 3	F3/R3	1046 bp
D	Target 1	F4/R4	935 bp
E	Target 2	F5/R5	720 bp
F	Target 3	F3/R6	1348 bp

Figure 2.2. Design of a CRISPR/Cas9-based gene drive system in *S. cerevisiae* across three loci.

(A) *Left*, An artificial gene drive was installed at three loci in haploid yeast. Each drive system (Drive 1-3) contained a guide RNA cassette targeting an artificial target (Target 1-3) at the same

locus. Only Drive 1 contained the cassette for *S. pyogenes* Cas9. *Right,* Artificial (u1) and (u2) sites[38] were used flanking the gene drive at the *HIS3* locus (Chromosome XV) and the *S.p.HIS5* selectable marker. The *SHS1* locus (Chromosome IV) included a C-terminal GFP and *C.a.URA3*. *DNL4* (Chromosome XV) was deleted with the Kan^R cassette. All sgRNAs were targeted to non-native sequences. The sgRNA(u1) cassette was on a high-copy plasmid (*LEU2* marker). *S.p.*Cas9 was under control of the inducible *GAL1/10* promoter. (B) Haploid yeast harboring the triple drive (GFY-3675) were mated to the triple target strain (GFY-3596) to form diploids. Cas9 expression was induced by galactose (0 or 5 hr). Cultures were diluted to 100-500 cells per plate, grown for 2 days, and transferred to SD-LEU, SD-HIS, SD-URA, and G418 plates. (C) A time course of galactose activation using the [GFY-3675 x GFY-3596] diploid in triplicate. Error, SD. (D) Seven haploid strains (GFY-3206, 3593, 3264b, 3578, 3594, 3623, and 3596) were tested as in (B) against the triple drive strain (GFY-3675). (E) Each of the diploids from (D) were cultured for 5 hr and quantified for drive success. Error, SD. (F) Clonal isolates were obtained from diploids generated in (B) at either 0 hr (2 isolates) or 5 hr activation of Cas9 (14 isolates). All yeast were confirmed as diploids and assayed on each media type (*below*). Diagnostic PCRs were performed on genomic DNA to detect the presence (or absence) of each locus; oligonucleotide (Supplementary Table B.5) positions can be found in (A) and the expected sizes are illustrated (*right*). Two isolates (13,14) were chosen for their incomplete growth profile (red asterisks). Images were cropped from separate portions of larger gels or from independent DNA gels and are separated by white lines. The unedited images can be found in Supplementary Fig. B.S8.

DNA Ligase IV as a target for gene drives

Our choice of the yeast *DNL4* gene as one of the MGD targets was intended to highlight the ability of a drive itself to modify or eliminate non-homologous end joining (NHEJ)—the DNA repair process that directly counteracts the action of gene drives. Following DSB formation by Cas9, the function of the homing drive requires repair of the broken chromosome via homology directed repair using the homologous chromosome (and drive itself) as the source of the donor DNA. However, should NHEJ repair systems ligate the broken chromosome ends prior to HR-based copying, the drive will fail to copy; in fact, imprecise repair by NHEJ may even generate alleles of the target that would be resistant to further rounds of editing. Therefore, this competing DNA repair system remains one major technical hurdle to optimal gene drive design in higher eukaryotes. Of note, interest in modulating, tuning, or inhibiting NHEJ-based repair pathways is

not unique to CRISPR gene drives as this mode of repair still competes with the introduction of exogenous DNA via HR[39-44].

The NHEJ pathway is highly conserved from yeast to humans and functions to directly fuse exposed DNA ends[45,46]. DNA Ligase IV (Dnl4 in yeast, Lig4 in humans) is required for the final step of DNA ligation along with other conserved binding partners[47]. We examined the genomes of other fungi and metazoans using the yeast Dnl4 protein sequence as a query and a phylogenetic history of this enzyme illustrated the evolution of this enzyme through deep time (Fig. 2.3A). Note, the branching of *Z. nevadensis* (termite) was poorly supported and has been previously shown to be included within the *Insecta* class[48]. The DNA Ligase IV enzyme is divided into multiple subdomains including DNA binding, adenylation, oligonucleotide binding, and a C-terminal BRCA1 C-terminal domain (BRCT) that interacts with binding partner Lif1 (XRCC4 in human). A previous study identified a number of mutational substitutions within the C-terminus of yeast Dnl4 that resulted in a *partial* loss of function of NHEJ[49]. Examination of protein sequence alignments between yeast, mosquito, and human DNA Ligase IV C-terminal domains revealed only a minor conservation of sequence identity (Fig. 2.3B). However, several of the identified yeast residues were conserved by either insects and/or humans (yeast T744, D800, G868, and G869). Using the crystal structure of the C-terminus of yeast Dnl4 as a template, we generated models (I-TASSER) for the corresponding domains of mosquito and human Lig4—both displayed a much higher conservation of structure as opposed to primary sequence (Fig. 2.3C, Supplementary Table B.4). The N-terminal region also displayed strong structural homology using the human Lig4 crystal structure as a template (Supplementary Fig. B.5).

While total loss of NHEJ (e.g. *dnl4*Δ) is tolerated in yeast, it is unclear whether a DNA Ligase IV null allele would be viable in higher eukaryotes. Along these lines, truncations or

mutations of Lig4 in humans can lead to the rare DNA Ligase IV syndrome[50,51]. However, given that reduction in transcript or replacement by a partially functioning allele could reduce, but not eliminate NHEJ repair, it could be utilized in other systems to maximize gene drive efficiency, even at the (potential) expense of overall fitness. Therefore, we utilized a "self-editing" methodology to integrate six *dnl4* alleles—five partial loss of function substitutions, one truncation, and a WT control (Fig. 2.3D). In a strain harboring integrated Cas9 at the *HIS3* locus[17], we introduced two DSBs within the C-terminus of native *DNL4* and integrated two different constructs: (i) a modified *dnl4* allele with a sgRNA(Kan) cassette and (ii) a modified *dnl4* locus using the native terminator sequence. Both Cas9 target sites were also mutated within the repair (donor) DNA to prevent subsequent rounds of unintended editing.

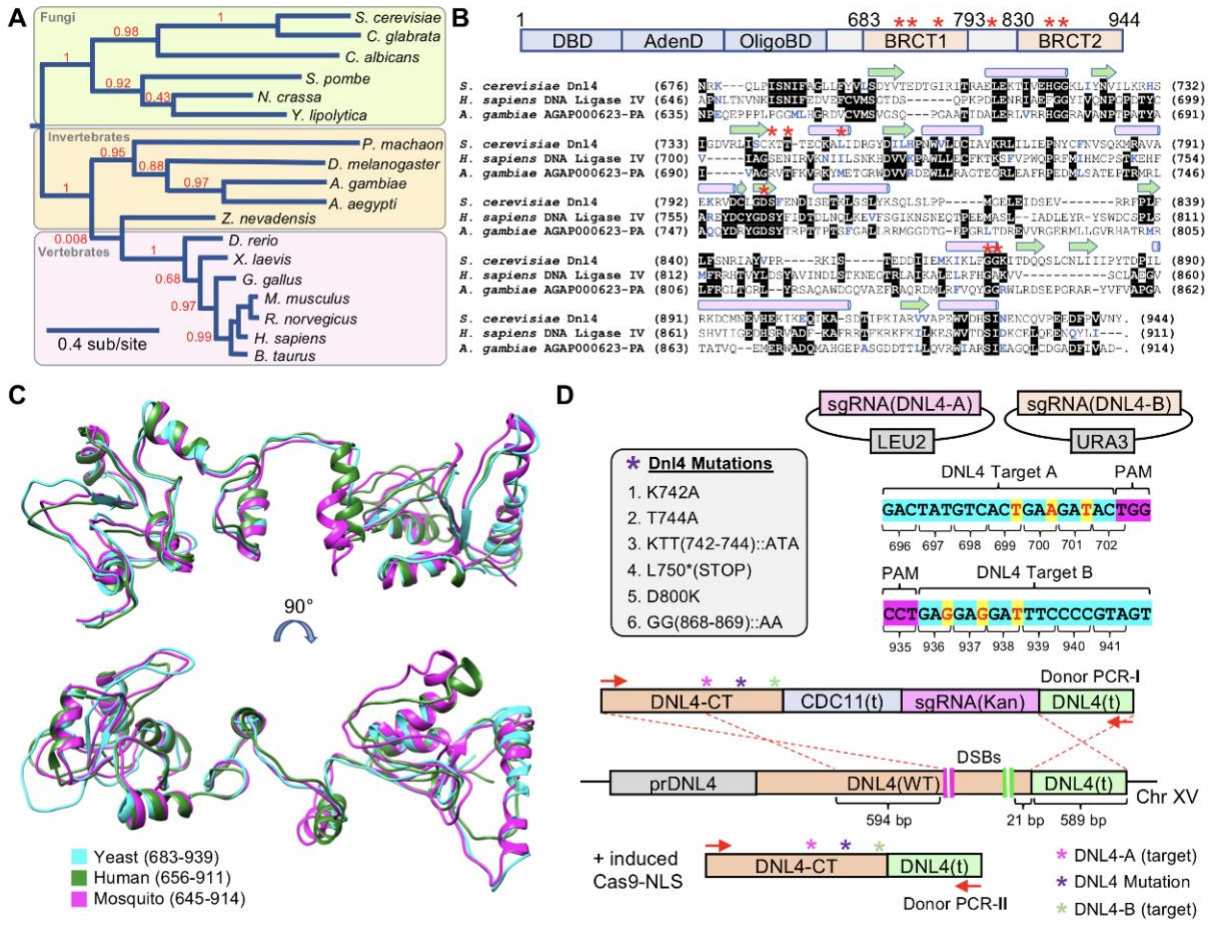


Figure 2.3. DNA Ligase IV, critical for NHEJ and conserved across eukaryotes, provides a unique candidate for gene drives.

(A) Phylogenic analysis of Ligase IV candidates (Supplementary Table B.3) across fungi and metazoans by Phylogeny.fr[52,53]. Branch lengths correspond to the number of substitutions per site and the confidence of most branches is illustrated as a decimal (red text). (B) *Top,* Illustration of the domain structure of yeast Dnl4. The catalytic N-terminal portion includes a DNA binding domain, adenylation domain, and oligonucleotide domain (blue). The C-terminal portion includes tandem BRCA1 C-Terminal domains (BRCT). *Bottom,* A multiple sequence alignment was performed using Clustal Omega[54] of the yeast, mosquito, and human Ligase IV protein C-termini. Identical residues are shown against a black background and similar residues are colored in blue. Secondary structures (pink cylinder, α-helix; green arrow, β-strand) for the yeast Dnl4 C-terminal as determined by the crystal structure are illustrated[55]. The position of six alleles (K742, T744, L750, D800, G868, and G869) are also illustrated (red asterisk) that were identified from a previous study[49]. (C) The protein sequences of the *A. gambiae* (645-914) and *H. sapiens* (656-911) Ligase IV were modeled against the crystal structure of the *S. cerevisiae* (683-939) Dnl4 (PDB:1Z56) using I-TASSER[56] (Supplementary Table B.4) and illustrated using Chimera[57]. (D) Cas9-based genomic integration methodology for introduction of mutational substitutions to the native *DNL4* locus in yeast. Two sgRNA-expressing cassettes were cloned onto high copy plasmids (marked with *LEU2* and *URA3*) to induce two DSBs within the C-terminus of *DNL4*. Silent substitutions were generated within the intended repair DNA to prevent re-targeting of Cas9 (silent alterations in yellow). Two repair strategies were used to include either a non-native terminator coupled with a sgRNA(Kan) cassette, or the native *DNL4* terminator; the included amount of homology (bp) is illustrated.

We utilized these eight haploid strains to quantify the level of NHEJ repair (Fig. 2.4). Our system of DSB formation followed by DNA repair utilized the dual programmed (u2) sites flanking the Cas9 expression cassette (Fig. 2.4A). With only a single guide construct, Cas9 would be multiplexed to both identical sites, causing complete excision of the nuclease gene and Kan^R marker. Following transformation of the sgRNA(u2) plasmid, yeast were analyzed for the number of surviving colonies on SD-LEU medium (Fig. 2.4B). Editing by Cas9 at both (u2) sites followed by precise DNA ligation of the broken ends would generate a "new" (u2) site, and would be subject to a second round of Cas9-dependent cleavage—continual DSB formation followed by exacting repair causes inviability in yeast[17].

However, introduction of an insertion, deletion, or substitution within the target sequence would render the site immune from subsequent rounds of editing. Furthermore, loss of the Kan^R

marker provided a growth phenotype associated with targeting of the (u2) sites and excision of the entire cassette at the *HIS3* locus. Both the total number of surviving colonies as well as the percentage of isolates with an excised marker were quantified in triplicate (Fig. 2.4C). In our assay, the presence of WT *DNL4* allowed for approximately 7 colonies/experimental trial, whereas *dnl4*Δ yeast resulted in 0-1 colonies on average. Importantly, of the WT *DNL4* isolates, 73% had properly excised the entire cassette whereas this was found to be 0% for *dnl4*Δ yeast across numerous independent trials (Fig. 2.4C). The partial loss of function *dnl4* alleles provided a range of NHEJ efficiencies: the K742A mutant averaged 4 colonies/trial with an excision rate of nearly 50% and other substitutions displayed excision rates of between 0-25%. As expected, the C-terminal *dnl4* truncation at L750 phenocopied the null allele. Diagnostic PCRs confirmed the presence or absence of the Cas9-Kan^R expression cassette for clonal isolates from each of the aforementioned haploid strains tested (Fig. 2.4D). For strains that had undergone editing and marker excision, the *HIS3* locus was amplified and sequenced; NHEJ followed by imprecise ligation introduced either insertions or deletions at the site of Cas9 cleavage 3 bp upstream of the 5' end of the PAM sequence (Fig. 2.4E). Finally, each of the *dnl4* alleles was tested within our MGD system as a native cargo-based delivery system (Fig. 2.4F, *top*). Given that our artificial *DNL4* target was the *dnl4*Δ*Kan^R* null allele, we recognize that in the context of a [gene drive x WT] diploid genome, further modifications would be required to bias the HR-based repair of the intended *dnl4* allele. This could include recoding (silent substitutions) of the *DNL4* C-terminal domain sequence to prevent promiscuous cross-over downstream of the intended mutation(s). Following expression of Cas9 and activation of the MGD, the growth profiles of 7 diploid strains were assessed in triplicate and demonstrated efficient drive activity at all three loci (Fig. 2.4F, *bottom*). Moreover, PCRs from clonal isolates confirmed the presence or absence of each drive and target locus (Supplementary

Fig. B.6). These data demonstrate that the MGD strategy can be used as a knock-out or allele replacement strategy (at a native locus) with only minimal added sequence (782 bp).

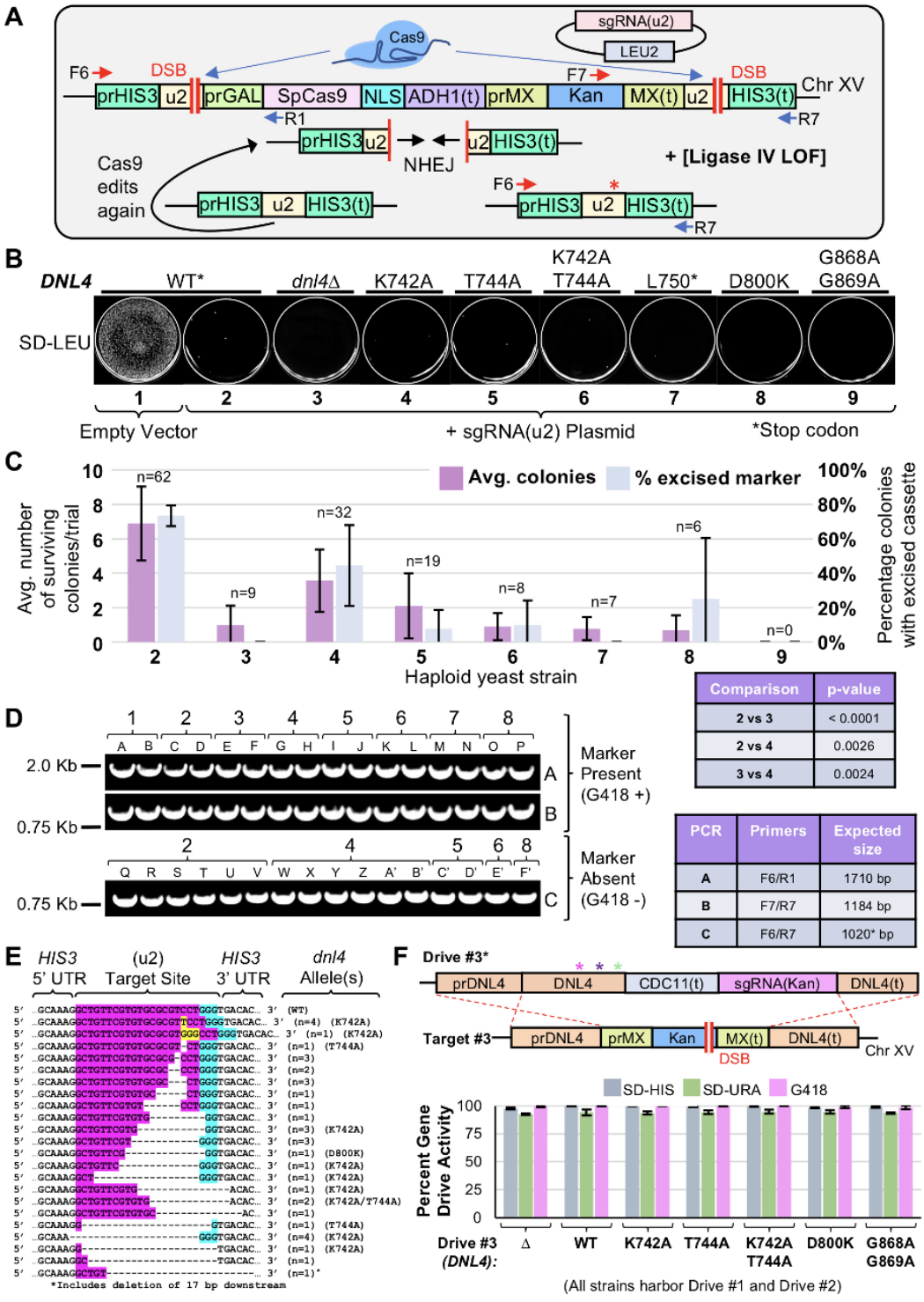


Figure 2.4. Partial loss of function alleles of yeast DNA Ligase IV reduce NHEJ.

(A) Design of a self-excising Cas9-based assay for NHEJ. Strain GFY-2383 included an inducible Cas9 cassette paired with the Kan^R marker. Transformation of the sgRNA(u2) plasmid would result in multiplexing to two flanking (u2) sites. Repair via NHEJ would result in the formation of the original (u2) site, and would be subject to further rounds of editing; introduction of an indel (red asterisk) would cause destruction of the target. (B) Strains GFY-3850 through GFY-3856 and GFY-3864 (Supplementary Table B.1, *Conditions* 1-9) were transformed with the sgRNA(u2) plasmid (pGF-V809) or empty vector control (pRS425) and plated onto SD-LEU for three days. The *DNL4* (WT) gene contained six silent substitutions (asterisk). (C) The average number of surviving colonies was quantified for all trials—labeled as in (B); the number of colonies (n) obtained across all experiments is displayed. Error, SD. For the pRS425 vector, 2897 +/- 357 colonies were obtained. The percentage of isolates that excised the cassette at the *HIS3* locus (by sensitivity to G418) is displayed. Error, SD. Statistical analyses of strain comparisons (colonies per trial) were performed using an unpaired t-test. (D) Diagnostic PCRs were performed on chromosomal DNA from isolates from (B) to illustrate presence (2 isolates each) or loss (between 2-6 shown) of the Cas9 cassette. *Conditions* 2-8 correspond to (B). Oligonucleotides (Supplementary Table B.5) are in (A) and the expected sizes are illustrated (*right*). Images of independent DNA gels are separated by white lines; unedited gel images can be found in Supplemental Fig. B.S8. (E) DNA sequencing of the *HIS3* locus following NHEJ (on isolates sensitive to G418). For each insertion or deletion, the number of identical clones is displayed. All sequences were obtained from WT yeast unless otherwise noted. Target, pink. PAM, blue. Insertions, yellow. (F) Triple-drive containing strains were constructed with a modified *DNL4* and sgRNA(Kan) cassette. Haploid strains (GFY-3675, 3865-3867, 3871, 3872, and 3875) containing the sgRNA(u1) plasmid were mated with GFY-3596, diploids selected, and drives activated. Percentage of colonies sensitive to each condition, gene drive activity. Error, SD.

Discussion

In this study, we have developed a multi-locus CRISPR gene drive with a minimal design (MGD) that allows for multiplexing of Cas9 *in trans* across three distinct chromosomal locations (Fig. 2.1). An alternative strategy could also be employed to create more than one gene drive system within a single genome—a CGD where each locus of interest contains the full complement of genetic information (nuclease, UTR, sgRNA, and optional cargo). In this way, each drive would be completely independent from all other drive(s). While this design clearly provides a maximum level of potential redundancy, there are other technical and safety issues inherent to this multi-nuclease arrangement. For one, countering or inhibiting a CGD with more than one active nuclease would require more sophisticated anti-drive systems, the discovery of additional anti-CRISPR

proteins, or complex regulatory systems to ensure inactivation or destruction of each drive. In contrast, our minimal GD design can be inhibited by the AcrIIA2/A4 proteins[27], self-excised by our flanking (u2) sites[17], or targeted by an anti-drive system no different than a traditional single-locus gene drive. We argue that this type of design provides a higher level of biosecurity and can still accomplish the same task as n-number of "full" gene drives. Moreover, the issue of tightly regulated control of the nuclease transcript may pose additional challenges if the same promoter elements are positioned across multiple chromosomes and epigenetic landscapes in the CGD design.

One potential issue facing our MGD design (or *any* gene drive design, for that matter) is that of natural or evolved resistance to the action of the drive. Since our multi-locus arrangement includes only a single nuclease gene powering all drives, any resistance or escaped action to the Cas9-containing locus would render all three gene drives inactivated in subsequent generations. However, evidence now exists both *in silico*[11,35,37] and *in vivo*[58] that the addition of multiple guide RNAs (to the same genetic target) can reduce (or potentially *eliminate*) resistance to the gene drive. Using current estimations for a given target, five separate guide RNAs may provide a sufficiently rare or improbable event requiring mismatch or mutation to occur at all five target DNA sites[35,37]. This would provide greater than 99% confidence in eliminating an *A. gambiae* population on a continent-wide scale[35]. Therefore, our recommendation would be to greatly bias multiplexing (via multiple guide RNAs) to the gene drive locus harboring the sole copy of Cas9 in the MGD design (in our system, Cas9 creates two DSBs flanking the entire locus). To ensure even higher fidelity of the nuclease and to combat resistance, one could combine the two strategies (CGD and MGD) to have a secondary copy of the nuclease (of the same variant or a different species) positioned at a second locus—additional multiplexing across numerous other loci could include the minimal

design (guide RNA cassette only). Finally, while our MGD methodology includes a gene drive consisting of only 455 nucleotides (sgRNA expression cassette), this could be reduced even further to be only a few bases or the absence of any base pairs. Additional sgRNA cassettes could be installed at one or more loci to allow for targeting of chromosomal positions where the "drive" is nothing more than a single base substitution or deletion. Provided few bases separate the DSB site and the intended mutation(s), HR-based repair would allow for propagation of the few bases no different than a "full" gene drive (consisting of many thousands or tens of thousands of bases) into the homozygous condition. The only requirement would involve the sgRNA expression cassette(s) to also be installed within a drive-containing locus *in trans*.

Finally, we chose one of the chromosomal targets within our minimal gene drive system to include both deletions, truncations, and substitution alleles of *DNL4*—one of the essential components of the NHEJ repair pathway. While loss of DNA Ligase IV is non-lethal in yeast[59] and flies[60], it is embryonic lethal in mouse[61] and not tolerated in mosquito[62]. However, suppression or inhibition of this enzyme or the NHEJ repair pathway has been shown to increase rates of recombination and genomic integration of exogenous DNA using CRISPR systems *in vivo*[39,41-44,62,63]. We demonstrate that this critical factor could be an additional target for a multi-locus gene drive system—suppression of NHEJ, whether by mutated alleles, regulation of transcript, or direct inhibition of the enzyme—would aid in successful HR-based copying of the drive and further reduce the possibility for drive resistance, especially when coupled with multiple guide RNAs. Numerous strategies might be employed to accomplish targeted suppression of NHEJ including testing of additional Ligase IV loss of function alleles that may be widely conserved across eukaryotes; our study focused on the C-terminal BRCT-domain containing portion of Dnl4, but other substitutions have been also been characterized within the N-terminal catalytic domain[64].

A multi-locus CRISPR gene drive system should help advance current designs and provide additional options for (i) biosecurity, (ii) drive redundancy, (iii) combatting of evolved resistance, (iv) native gene replacement, (v) multiple gene cargo/genetic pathway delivery, (vi) suppression of NHEJ or activation of HR-promoting repair pathways, and (vii) multiple phenotypic outcomes. Advanced drive arrangements[28] could accomplish multiple outcomes within a single-genome system—the additional of exogenous cargo could also be paired with (native) allele introduction *and* modulating of organism fitness by perturbing numerous other genetic pathways in a single step. As the design and application of CRISPR gene drives continues to advance, we continue to stress the need for multiple levels of control, tunability, inhibition, and drive reversal.

Acknowledgements & Funding

We thank Emily Wedeman (Kansas State University) and Muriel Eaton (Purdue University) for useful comments and Megan Halloran (Kansas State University) for laboratory assistance. This project was supported by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under grant number P20 GM103418 to G.C.F. This work was also supported by the USDA National Institute of Food and Agriculture, Hatch Project 1013520 to G.C.F. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institute of Health.

Author Contributions

Y.Y. and G.C.F built all reagents (plasmids and yeast), performed all experiments, and performed data analyses and figure preparation. G.C.F. wrote the manuscript.

Competing Interest Statement

G.C.F. (Kansas State University) has filed for a provisional patent entitled "Multi-Locus Gene Drive System," U.S. Serial No. 62/697,855 on July 13, 2018 for the intellectual property and technology described within this work. G.C.F declares no potential non-financial conflict of interest. Y.Y. declares no potential conflict of interest of any kind.

Data Availability

The datasets generated during and analyzed during the current study are available from the corresponding author on reasonable request.

Animal and Human Subject Statement

This study does not use any animals or human subjects.

References

1 Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.)* **337**, 816-821, doi:10.1126/science.1225829 (2012).

2 Jinek, M. *et al.* RNA-programmed genome editing in human cells. *eLife* **2**, e00471, doi:10.7554/eLife.00471 (2013).

3 Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science (New York, N.Y.)* **346**, 1258096, doi:10.1126/science.1258096 (2014).

4 Zetsche, B. *et al.* Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759-771, doi:10.1016/j.cell.2015.09.038 (2015).

5 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).

6 Thakore, P. I. *et al.* Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nature methods* **12**, 1143-1149, doi:10.1038/nmeth.3630 (2015).

7 Gaudelli, N. M. *et al.* Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature* **551**, 464-471, doi:10.1038/nature24644 (2017).

8 Bull, J. J. & Barrick, J. E. Arresting Evolution. *Trends in genetics : TIG* **33**, 910-920, doi:10.1016/j.tig.2017.09.008 (2017).

9 Godfray, H. C. J., North, A. & Burt, A. How driving endonuclease genes can be used to combat pests and disease vectors. *BMC biology* **15**, 81, doi:10.1186/s12915-017-0420-4 (2017).

10 Esvelt, K. M. & Gemmell, N. J. Conservation demands safe gene drive. *PLoS biology* **15**, e2003850, doi:10.1371/journal.pbio.2003850 (2017).

11 Prowse, T. A. A. *et al.* Dodging silver bullets: good CRISPR gene-drive design is critical for eradicating exotic vertebrates. *Proceedings. Biological sciences* **284**, doi:10.1098/rspb.2017.0799 (2017).

12 Courtier-Orgogozo, V., Morizot, B. & Boete, C. Using CRISPR-based gene drive for agriculture pest control. *EMBO reports*, doi:10.15252/embr.201744822 (2017).

13 Lambert, B., North, A., Burt, A. & Godfray, H. C. J. The use of driving endonuclease genes to suppress mosquito vectors of malaria in temporally variable environments. *Malaria journal* **17**, 154, doi:10.1186/s12936-018-2259-8 (2018).

14 Hammond, A. M. & Galizi, R. Gene drives to fight malaria: current state and future directions. *Pathogens and global health* **111**, 412-423, doi:10.1080/20477724.2018.1438880 (2017).

15 DiCarlo, J. E., Chavez, A., Dietz, S. L., Esvelt, K. M. & Church, G. M. Safeguarding CRISPR-Cas9 gene drives in yeast. *Nature biotechnology* **33**, 1250-1255, doi:10.1038/nbt.3412 (2015).

16 Shapiro, R. S. *et al.* A CRISPR-Cas9-based gene drive platform for genetic interaction analysis in Candida albicans. *Nature microbiology* **3**, 73-82, doi:10.1038/s41564-017-0043-0 (2018).

17 Roggenkamp, E. *et al.* Tuning CRISPR-Cas9 Gene Drives in Saccharomyces cerevisiae. *G3 (Bethesda, Md.)* **8**, 999-1018, doi:10.1534/g3.117.300557 (2018).

18 Roggenkamp, E. *et al.* CRISPR-UnLOCK: multipurpose Cas9-based strategies for Conversion of yeast libraries and strains. *Frontiers in microbiology* **8**, 1773, doi:10.3389/fmicb.2017.01773 (2017).

19 Gantz, V. M. *et al.* Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E6736-6743, doi:10.1073/pnas.1521077112 (2015).

20 Champer, J. *et al.* Novel CRISPR/Cas9 gene drive constructs reveal insights into mechanisms of resistance allele formation and drive efficiency in genetically diverse populations. *PLoS genetics* **13**, e1006796, doi:10.1371/journal.pgen.1006796 (2017).

21 Hammond, A. *et al.* A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae. *Nature biotechnology* **34**, 78-83, doi:10.1038/nbt.3439 (2016).

22 Grunwald, H. A. *et al.* Super-Mendelian inheritance mediated by CRISPR/Cas9 in the female mouse germline. *bioRxiv*, doi:10.1101/362558 (2018).

23 Noble, C., Adlam, B., Church, G. M., Esvelt, K. M. & Nowak, M. A. Current CRISPR gene drive systems are likely to be highly invasive in wild populations. *eLife* **7**, doi:10.7554/eLife.33423 (2018).

24 James, S. *et al.* Pathway to Deployment of Gene Drive Mosquitoes as a Potential Biocontrol Tool for Elimination of Malaria in Sub-Saharan Africa: Recommendations of a Scientific Working Group(dagger). *The American journal of tropical medicine and hygiene* **98**, 1-49, doi:10.4269/ajtmh.18-0083 (2018).

25 Najjar, D. A., Normandin, A. M., Strait, E. A. & Esvelt, K. M. Driving towards ecotechnologies. *Pathogens and global health* **111**, 448-458, doi:10.1080/20477724.2018.1452844 (2017).

26 Vella, M. R., Gunning, C. E., Lloyd, A. L. & Gould, F. Evaluating strategies for reversing CRISPR-Cas9 gene drives. *Scientific reports* **7**, 11038, doi:10.1038/s41598-017-10633-2 (2017).

27 Basgall, E. M. *et al.* Gene drive inhibition by the anti-CRISPR proteins AcrIIA2 and AcrIIA4 in Saccharomyces cerevisiae. *Microbiology (Reading, England)* **164**, 464-474, doi:10.1099/mic.0.000635 (2018).

28 Dhole, S., Vella, M. R., Lloyd, A. L. & Gould, F. Invasion and migration of spatially self-limiting gene drives: A comparative analysis. *Evolutionary applications* **11**, 794-808, doi:10.1111/eva.12583 (2018).

29 Burt, A. & Crisanti, A. Gene Drive: Evolved and Synthetic. *ACS chemical biology* **13**, 343-346, doi:10.1021/acschembio.7b01031 (2018).

30 Drury, D. W., Dapper, A. L., Siniard, D. J., Zentner, G. E. & Wade, M. J. CRISPR/Cas9 gene drives in genetically variable and nonrandomly mating wild populations. *Science advances* **3**, e1601910, doi:10.1126/sciadv.1601910 (2017).

31 Hammond, A. M. *et al.* The creation and selection of mutations resistant to a gene drive over multiple generations in the malaria mosquito. *PLoS genetics* **13**, e1007039, doi:10.1371/journal.pgen.1007039 (2017).

32 Buchman, A., Marshall, J. M., Ostrovski, D., Yang, T. & Akbari, O. S. Synthetically engineered Medea gene drive system in the worldwide crop pest Drosophila suzukii. *Proceedings of the National Academy of Sciences of the United States of America* **115**, 4725-4730, doi:10.1073/pnas.1713139115 (2018).

33 Unckless, R. L., Clark, A. G. & Messer, P. W. Evolution of Resistance Against CRISPR/Cas9 Gene Drive. *Genetics* **205**, 827-841, doi:10.1534/genetics.116.197285 (2017).

34 KaramiNejadRanjbar, M. *et al.* Consequences of resistance evolution in a Cas9-based sex conversion-suppression gene drive for insect pest management. *Proceedings of the National Academy of Sciences of the United States of America* **115**, 6189-6194, doi:10.1073/pnas.1713825115 (2018).

35 Marshall, J. M., Buchman, A., Sanchez, C. H. & Akbari, O. S. Overcoming evolved resistance to population-suppressing homing-based gene drives. *Scientific reports* **7**, 3776, doi:10.1038/s41598-017-02744-7 (2017).

36 Bull, J. J. & Malik, H. S. The gene drive bubble: New realities. *PLoS genetics* **13**, e1006850, doi:10.1371/journal.pgen.1006850 (2017).

37 Noble, C., Olejarz, J., Esvelt, K. M., Church, G. M. & Nowak, M. A. Evolutionary dynamics of CRISPR gene drives. *Science advances* **3**, e1601964, doi:10.1126/sciadv.1601964 (2017).

38 Finnigan, G. C. & Thorner, J. mCAL: a new approach for versatile multiplex action of Cas9 using one sgRNA and loci flanked by a programmed target sequence. *G3 (Bethesda, Md.)* **6**, 2147-2156, doi:10.1534/g3.116.029801 (2016).

39 Vartak, S. V. & Raghavan, S. C. Inhibition of nonhomologous end joining to increase the specificity of CRISPR/Cas9 genome editing. *The FEBS journal* **282**, 4289-4294, doi:10.1111/febs.13416 (2015).

40 Schwartz, C., Frogue, K., Ramesh, A., Misa, J. & Wheeldon, I. CRISPRi repression of nonhomologous end-joining for enhanced genome engineering via homologous recombination in Yarrowia lipolytica. *Biotechnology and bioengineering* **114**, 2896-2906, doi:10.1002/bit.26404 (2017).

41 Robert, F., Barbeau, M., Ethier, S., Dostie, J. & Pelletier, J. Pharmacological inhibition of DNA-PK stimulates Cas9-mediated genome editing. *Genome medicine* **7**, 93, doi:10.1186/s13073-015-0215-6 (2015).

42 Canny, M. D. *et al.* Inhibition of 53BP1 favors homology-dependent DNA repair and increases CRISPR-Cas9 genome-editing efficiency. *Nature biotechnology* **36**, 95-102, doi:10.1038/nbt.4021 (2018).

43 Maruyama, T. *et al.* Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nature biotechnology* **33**, 538-542, doi:10.1038/nbt.3190 (2015).

44 Chu, V. T. *et al.* Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nature biotechnology* **33**, 543-548, doi:10.1038/nbt.3198 (2015).

45 Lieber, M. R. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annual review of biochemistry* **79**, 181-211, doi:10.1146/annurev.biochem.052308.093131 (2010).

46 Chiruvella, K. K., Liang, Z. & Wilson, T. E. Repair of double-strand breaks by end joining. *Cold Spring Harbor perspectives in biology* **5**, a012757, doi:10.1101/cshperspect.a012757 (2013).

47 Ellenberger, T. & Tomkinson, A. E. Eukaryotic DNA ligases: structural and functional insights. *Annual review of biochemistry* **77**, 313-338, doi:10.1146/annurev.biochem.77.061306.123941 (2008).

48 Misof, B. *et al.* Phylogenomics resolves the timing and pattern of insect evolution. *Science (New York, N.Y.)* **346**, 763-767, doi:10.1126/science.1257570 (2014).

49 Chiruvella, K. K. *et al.* Yeast DNA ligase IV mutations reveal a nonhomologous end joining function of BRCT1 distinct from XRCC4/Lif1 binding. *DNA repair* **24**, 37-45, doi:10.1016/j.dnarep.2014.10.003 (2014).

50 Altmann, T. & Gennery, A. R. DNA ligase IV syndrome; a review. *Orphanet journal of rare diseases* **11**, 137, doi:10.1186/s13023-016-0520-1 (2016).

51 Chistiakov, D. A. Ligase IV syndrome. *Advances in experimental medicine and biology* **685**, 175-185 (2010).

52 Dereeper, A., Audic, S., Claverie, J. M. & Blanc, G. BLAST-EXPLORER helps you building datasets for phylogenetic analysis. *BMC evolutionary biology* **10**, 8, doi:10.1186/1471-2148-10-8 (2010).

53 Dereeper, A. *et al.* Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic acids research* **36**, W465-469, doi:10.1093/nar/gkn180 (2008).

54 Li, W. *et al.* The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic acids research* **43**, W580-584, doi:10.1093/nar/gkv279 (2015).

55 Dore, A. S. *et al.* Structure of an Xrcc4-DNA ligase IV yeast ortholog complex reveals a novel BRCT interaction mode. *DNA repair* **5**, 362-368, doi:10.1016/j.dnarep.2005.11.004 (2006).

56 Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols* **5**, 725-738, doi:10.1038/nprot.2010.5 (2010).

57 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).

58 Champer, J. *et al.* Reducing resistance allele formation in CRISPR gene drive. *Proceedings of the National Academy of Sciences of the United States of America*, doi:10.1073/pnas.1720354115 (2018).

59 Wilson, T. E., Grawunder, U. & Lieber, M. R. Yeast DNA ligase IV mediates non-homologous DNA end joining. *Nature* **388**, 495-498, doi:10.1038/41365 (1997).

60 Gorski, M. M. *et al.* The Drosophila melanogaster DNA Ligase IV gene plays a crucial role in the repair of radiation-induced DNA double-strand breaks and acts synergistically with Rad54. *Genetics* **165**, 1929-1941 (2003).

61 Barnes, D. E., Stamp, G., Rosewell, I., Denzel, A. & Lindahl, T. Targeted disruption of the gene encoding DNA ligase IV leads to lethality in embryonic mice. *Current biology : CB* **8**, 1395-1398 (1998).

62 Basu, S. *et al.* Silencing of end-joining repair for efficient site-specific gene insertion after TALEN/CRISPR mutagenesis in Aedes aegypti. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 4038-4043, doi:10.1073/pnas.1502370112 (2015).

63 Cen, Y., Timmermans, B., Souffriau, B., Thevelein, J. M. & Van Dijck, P. Comparison of genome engineering using the CRISPR-Cas9 system in C. glabrata wild-type and lig4 strains. *Fungal genetics and biology : FG & B* **107**, 44-50, doi:10.1016/j.fgb.2017.08.004 (2017).

64 Chiruvella, K. K., Liang, Z., Birkeland, S. R., Basrur, V. & Wilson, T. E. Saccharomyces cerevisiae DNA ligase IV supports imprecise end joining independently of its catalytic activity. *PLoS genetics* **9**, e1003599, doi:10.1371/journal.pgen.1003599 (2013).

65 Sambrook, J. & Russell, D. W. *Molecular Cloning: A Laboratory Manual*. 3rd edn, (Cold Spring Harbor Laboratory Press, 2001).

66 Finnigan, G. C. & Thorner, J. Complex *in vivo* ligation using homologous recombination and high-efficiency plasmid rescue from *Saccharomyces cerevisiae*. *Bio-protocol* **5**, e1521. http://www.bio-protocol.org/e1521 (2015).

67 Zheng, L., Baumann, U. & Reymond, J. L. An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic acids research* **32**, e115, doi:10.1093/nar/gnh110 (2004).

68 Eckert-Boulet, N., Pedersen, M. L., Krogh, B. O. & Lisby, M. Optimization of ordered plasmid assembly by gap repair in Saccharomyces cerevisiae. *Yeast (Chichester, England)* **29**, 323-334, doi:10.1002/yea.2912 (2012).

69 DiCarlo, J. E. *et al.* Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic acids research* **41**, 4336-4343, doi:10.1093/nar/gkt135 (2013).

70 Ochi, T., Gu, X. & Blundell, T. L. Structure of the catalytic region of DNA ligase IV in complex with an Artemis fragment sheds light on double-strand break repair. *Structure (London, England : 1993)* **21**, 672-679, doi:10.1016/j.str.2013.02.014 (2013).

Chapter 3 - Analysis of CRISPR gene drive design in budding yeast

This chapter has been published as a journal article in *Access Microbiology*.

Yao Yan[1] and Gregory C. Finnigan[1*]

[1]Department of Biochemistry and Molecular Biophysics, Kansas State University, 141 Chalmers Hall, Manhattan, KS 66506 USA

*Correspondence to: Gregory C. Finnigan, Phone: (785) 532-6939; FAX; (785) 532-7278;

E-mail: gfinnigan@ksu.edu

Abstract

Control of biological populations remains a critical goal for challenges facing ecosystems, agriculture, and human disease including pests, parasites, pathogens, and invasive species. A particular architecture of the CRISPR/Cas biotechnology—a gene drive—has the potential to modify or eliminate populations on a massive scale. Super-Mendelian inheritance has now been demonstrated in both fungi and metazoans including disease vectors such as mosquitos. Studies in yeast and fly model systems have developed a number of molecular safeguards to increase biosafety and control over drive systems *in vivo* including titration of nuclease activity, anti-CRISPR dependent inhibition, and use of non-native DNA target sites. We have developed a CRISPR/Cas9 gene drive in *S. cerevisiae* that allows for safe and rapid examination of alternative drive designs and control mechanisms. In this study, we tested whether non homologous end joining had occurred within diploid cells displaying a loss of the target allele following drive activation; we did not detect any instances of NHEJ within multiple sampled populations. We also demonstrated successful multiplexing using two additional non-native yeast target sequences. Furthermore, we extended our analysis of "resistant" clones that still harbored both the drive and target selection markers following expression of *S. pyogenes* Cas9; *de novo* mutation or NHEJ-based repair could not explain the majority of these heterozygous clones. Finally, we developed a second-generation gene drive in yeast with a guide RNA cassette integrated within the drive locus with a near 100% success rate; resistant clones in this system could also be reactivated during a second round of Cas9 induction.

Introduction

The use of CRISPR/Cas genomic editing has allowed for recent advances across many fields including agriculture, biotechnology, and basic laboratory research[1-3]. Introduction of targeted chromosomal breaks within a genome of interest coupled with DNA repair allows for the generation of nearly any conceivable genetic modification[4,5]. One powerful arrangement that utilizes CRISPR/Cas has the ability to rapidly "force" a genetic element of choice through a native population—a gene drive system[6-8]. In theory, this biotechnology could be used to either deliver a desired trait to a population or eliminate native populations (for example, through extreme bias of sex determination). Control of specific biological populations is a critical challenge facing numerous industries and global health epidemics including animal and plant pests and parasites, spread of pathogens (via insects vectors), and alteration of native environments by invasive organsims[9,10]. Given the widespread future applicability of CRISPR gene drives, additional laboratory study into biosafety[6,11], control[12-15], and reversal systems[16] is critical.

Current CRISPR-based gene drives (GD) have been developed and tested under laboratory conditions in fungi[12,13,17-19], insects[7,8,20,21], and even vertebrates[22] with varying success rates. Our previous work in budding yeast has developed a highly tractable drive system that can be used as a platform for testing of novel drive arrangements, CRISPR components, DNA repair, and modes of inhibition or control[12,13,17,23,24]. The basic mechanism of a GD includes integration of the basic CRISPR system (nuclease and corresponding guide RNA expression cassettes) at a particular genomic locus. Expression of Cas9/sgRNA within a diploid genome allows for introduction of a double strand break (DSB) in the target allele within the same locus the GD was positioned on the homologous chromosome. The DSB will be repaired using homology directed repair (HDR) and the drive-containing chromosome as the source of donor DNA; the gene drive cassette is then

copied to replace the entire target locus. The heterozygous pairing of a drive-containing individual with a WT individual (heterozygote) yields all *homozygous* diploid progeny for the gene drive; this allows for *super*-Mendelian inheritance of the drive locus through a population.

To date, some of our findings in *S. cerevisiae* have also been tested in higher eukaryotes including targeting of non-native DNA sequences within target loci (such as eGFP or other programmed artificial sequences)[17,25,26], titration of Cas9 activity within a gene drive[13,15], and split drives that separate guide RNAs from nucleases across multiple loci[14,24]. A number of important ongoing questions remain for gene drive research including the occurrence of drive resistance, mechanisms to slow, inhibit, or reverse active drives, and how population control may impact larger ecosystems during possible field deployment of drive systems. In this study, we examine our original first-generation CRISPR system in budding yeast and test (i) the possible occurrence of DNA repair by the non-homologous end joining system (NHEJ), (ii) additional controls for our examination of the drive and target loci following drive action, (iii) multiplexing to independent target sequences, (iv) examination of "resistant" clone formation, and (v) a modified second-generation drive harboring an integrated guide RNA cassette.

Materials and Methods

Yeast strains and plasmids

Budding yeast (*S. cerevisiae*) strains used are in Table S1 (also see Fig. C.S1). Molecular techniques were used for generation of all artificial constructs[27]. Prior to genomic integration, engineered DNA assemblies were first created on *CEN*-based plasmids using *in vivo* ligation and homologous recombination.[28] Plasmids were confirmed with diagnostic PCRs and DNA sequencing. Amplified PCRs of assembled expression cassettes (from isolated chromosomal DNA,

generated plasmids, or synthetic genes as templates) were transformed into cells using a lithium acetate-based protocol and integrated at the *HIS3* locus, often in multiple overlapping fragments using standard selection markers (*SpHIS5* or Kan^R). Chromosomal insertions were confirmed by both PCRs and DNA sequencing. Plasmids used in this study are found in Table S2 (see Fig. C.S1 for sequences). The expression cassettes for *S. pyogenes* guide RNAs were modeled on previous work.[29] Briefly, these included the yeast *SNR52* promoter, 20 bp variable crRNA sequence, a 79 bp tracrRNA sequence, and the *SUP4* terminator (Fig. C.S1). Custom genes were synthetized (Genscript) into a pUC57-Kan^R vector and sub-cloned to pRS425 or pRS426 using two unique flanking restriction sites.

Culture conditions

Yeast were propagated on solid agar plates or in liquid media cultures. Rich medium included 2% peptone, 1% yeast extract, and 2% dextrose. Synthetic drop-out media included yeast nitrogen base, ammonium sulfate, amino acids, and a carbon source. For pre-induction, a mixture of raffinose (2%) and sucrose (0.2%) was used for overnight cultures. For galactose activation (2%), rich medium (YP) was used. Sugars were not autoclaved and, instead, were filter sterilized. Synthetic drop-out medium also contained added tryptophan (not autoclaved). A concentration of 240 μg/mL was used for G418-containing plates. Yeast strains were all maintained at 30°C on plates or in liquid culture for the indicated times.

CRISPR editing and gene drives

CRISPR editing of haploid yeast included culturing GFY-2383 yeast in a pre-induction medium overnight, followed by back-dilution into rich medium containing galactose for 5 h. Cells

were harvested, transformed with the sgRNA(u2) (pGF-V809) high-copy plasmid, recovered overnight in rich medium containing galactose, and plated onto SD-LEU medium for 3-4 days. Clonal isolates were tested by growth (loss of G418 resistance), chromosomal DNA extraction, PCR, and DNA sequencing. For creation and activation of CRISPR gene drives, haploid strains (harboring Cas9) were first transformed with the sgRNA-containing plasmid(s), if necessary. Second, haploids were mated to the haploid target strain of the opposite mating type for 20 - 24 h on rich medium. Third, diploids were selected on SD-LEU-HIS, SD-LEU-URA-HIS, or SD-URA-HIS for three consecutive rounds (24 - 48 h incubation times). Fourth, cells were cultured in pre-induction medium overnight: S+Raff/Suc-LEU, S+Raff/Suc-LEU-URA, or S+Raff/Suc-URA-HIS and S+Raff/Suc-URA-LEU-HIS. Fifth, strains were transferred into rich medium containing galactose (typically for 5 h). Sixth, cells were harvested, diluted in sterile water to approximately 200-500 cells/mL and plated onto SD-LEU, SD-LEU-URA, or SD-URA for 48 h. Finally, yeast colonies were transferred (sterile velvet) to additional plates such as SD-HIS or SD-URA-HIS for 20 - 24 h before imaging. Clonal isolates were sampled at random from the final growth plates in at least triplicate. Safety mechanisms in place included a number of features including targeting of *S. pyogenes* Cas9 to sequences not found within the native yeast genome (*SpHIS5*, mCherry, u1', u2)[25], control of Cas9 expression under a galactose-responsive promoter (and use of dextrose in all plate types to repress nuclease expression), use of unstable high-copy plasmids[6] to harbor the sgRNA cassettes, and programmed self-excision (u2) or (u2') sites flanking all drive modules[13]. Generated diploid yeast strains prior to and after drive activation (or mock activation) were all destroyed following experimentation (only original haploids were preserved). Of note, our modified gene drive system (GD2, strains GFY-4325/4226) contained an integrated sgRNA

cassette proximal to the Cas9 gene. However, all other safety features still applied including the choice of (u1) as the crRNA sequence included within the drive.

Results

Additional controls for a first-generation CRISPR gene drive system

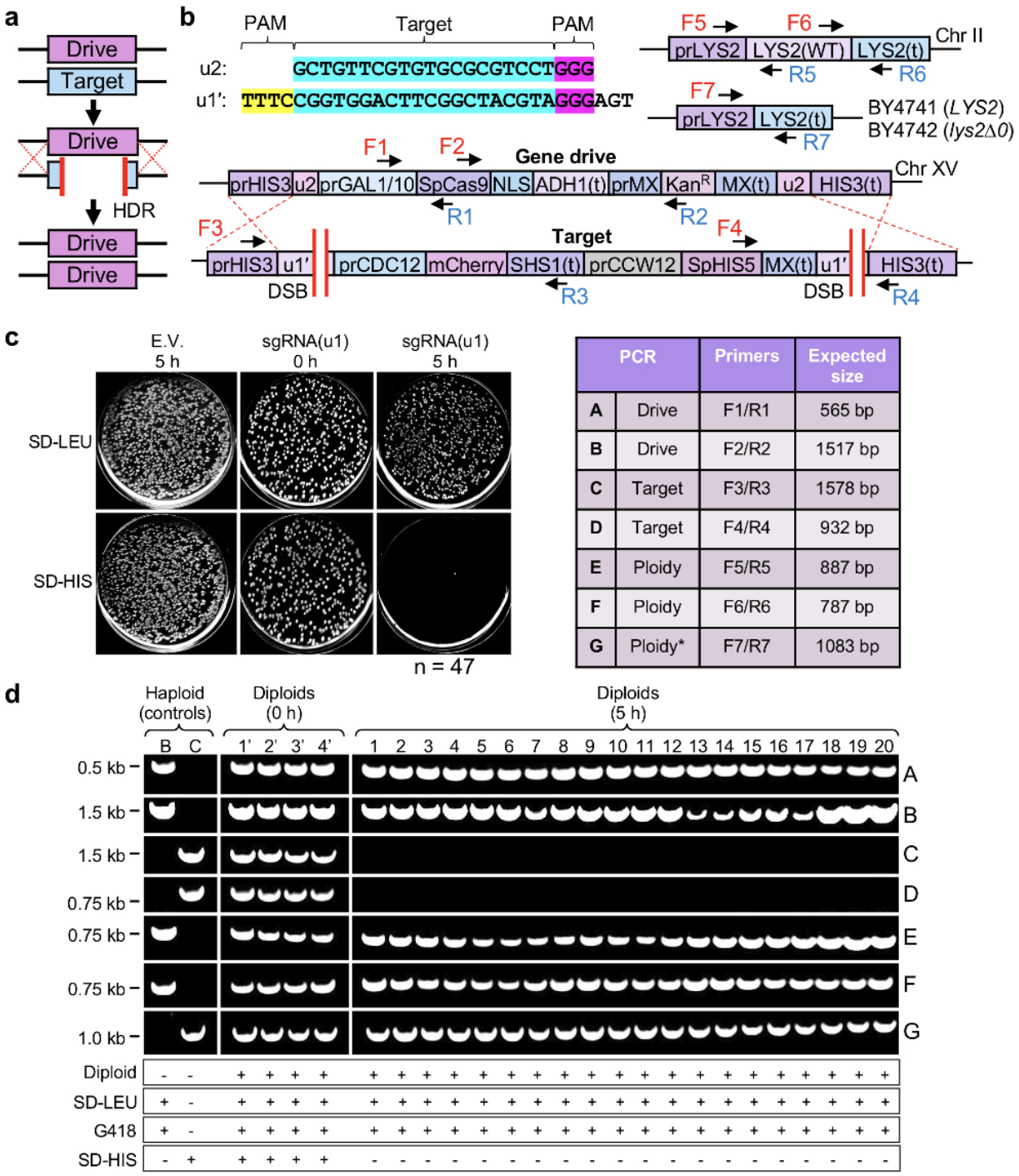
The intended mechanism of an artificial gene drive includes DSB formation within the target (often multiplexing to more than one cut site), excision of the target DNA, and replacement of the entire locus using the drive-containing chromosome as the source of donor material and HDR (Fig. 3.1A). This system requires a polyploid (diploid) genome in order to convert heterozygous cells to the homozygous state for the gene drive locus. Therefore, we re-examined our first-generation GD system (GD1)[13] with several modifications and additional experimental controls (Fig. 3.1B). First, we developed a new target haploid strain that included identical flanking (u1) sequences[25] distinct from the yeast genome. The new version (u1') included the same target sequence and PAM for use with *S. pyogenes* Cas9, and also a 5' PAM sequence (5'-TTTV-3') for the *F. novicida* Cas12a nuclease[30] as well as one possible PAM sequence for *S. aureus* Cas9 (5'-NNGRRT-3')[31] for future studies that could all utilize an overlapping core target sequence motif (Fig. 3.1B).

Activation of our GD1 system required treatment of diploid cells (pairing between the two yeast mating types, one harboring the gene drive, the other harboring the target locus) with galactose for induction of Cas9 expression. The sgRNA(u1)-expression cassette was included on a high-copy plasmid. Success of the gene drive system included testing colonies on synthetic medium lacking histidine. The target locus included constitutive expression of *S. pombe HIS5* (the functional equivalent of *S. cerevisiae HIS3*) that allowed for growth on SD-HIS plates; copying of

the drive allele to replace the target removed the selection marker and caused sensitivity to this condition. Inclusion of an empty vector control (no guide RNA) resulted in no loss of growth on SD-HIS (Fig. 3.1C, *left*). Following pre-induction (0 h, raffinose/sucrose mixture), only a small number of colonies were sensitive (Fig. 3.1C, *middle*). However, after 5 h galactose treatment, nearly every yeast colony was inviable on SD-HIS plates across numerous independent trials (Fig. 3.1C, *right*, Fig. C.S2). Subsequent analyses of individual diploid clones were performed to test whether the target locus was removed as expected (Fig. 3.1D). Using diagnostic PCRs (oligonucleotides found in Table C.S3) to unique DNA elements within the drive or target loci, we tested four randomly chosen isolates from the 0 h condition versus twenty isolates for the 5 h condition that all maintained growth on plates containing G418 due to the presence of the Kan^R selection marker present within the drive locus (PCRs A-D). Importantly, we included the two original haploid strains (drive and target) as further controls for this analysis. In all clones sampled after expression of Cas9, the drive locus was present (PCRs A,B) whereas the target locus was unable to be amplified (PCRs C,D) compared to the 0 h control and the original haploid strain.

We chose clonal isolates that were diploids using two independent assays for subsequent analyses. Examination of gene drive success requires conversion of the heterozygous diploid genome to the homozygous state (copying of the drive and loss of the target allele). Selected clones (that also maintained G418 resistance) were mated against known haploid strains of the two mating types and tested for the presence of two selectable markers (one from each haploid). Only strains that were able to mate to form diploids and contain both markers would survive under dual selection; diploid strains would be unable to mate with haploid controls and would be sensitive to the final selection challenge (Fig. C.S3). Clones that failed to mate in these initial assays (diploids) were chosen for further analysis by PCR. Second, we included a set of diagnostic PCRs to analyze

the diploid genome. The BY4741 and BY4742 laboratory strains include distinct alleles of the *LYS2* and *MET15* markers (Table S1, Fig. C.S1). Two regions of the *LYS2* coding sequence (PCRs E,F) were amplified to test for the presence of the *LYS2* gene (from BY4741, haploid drive genome) and one PCR was performed (PCR G) using primers flanking the *LYS2* coding sequence to test for the *lys2*Δ*0* allele (from BY4742, haploid target genome). While haploid controls demonstrated inclusion of only one of the two possible *LYS2* alleles, diploids from our gene drive experiments (first confirmed through the mating test) included amplification of both *LYS2* and *lys2*Δ*0* alleles (Fig. 3.1D). Finally, as an independent test of a different chromosome not containing *LYS2*, we examined ten isolates by PCR at the *MET15* locus (Fig. S4). In this case, the BY4741 genome (drive) included the *met15*Δ*0* allele whereas the BY4742 genome (target) included *MET15*. In these samples, the gene drive clones contained both alleles also supporting that these were diploid yeast.



PCR		Primers	Expected size
A	Drive	F1/R1	565 bp
B	Drive	F2/R2	1517 bp
C	Target	F3/R3	1578 bp
D	Target	F4/R4	932 bp
E	Ploidy	F5/R5	887 bp
F	Ploidy	F6/R6	787 bp
G	Ploidy*	F7/R7	1083 bp

Figure 3.1. A first-generation CRISPR gene drive in budding yeast.

(A) Mode of action for an artificial gene drive. Pairing between drive and target (typically WT) individuals and activation of the drive system results in cleavage of the target allele, copying of the drive cassette through homology directed repair, and conversion to the homozygous condition. (B) Design of an artificial drive/target system in *S. cerevisiae*. Unique CRISPR target sites (u1' and u2) were introduced flanking both the drive and target cassettes at the yeast *HIS3* locus. The

GAL1/10 promoter is repressed when cells are exposed to dextrose and activated in medium containing galactose. The *CCW12* promoter (cell wall component) provides constitutive expression of the target selection marker—*S. pombe HIS5* (functional equivalent of *S. cerevisiae HIS3* and allows for growth on medium lacking histidine). A codon-optimized *S. pyogenes* Cas9 contains a C-terminal SV40 NLS. Haploid yeast strains contain unique alleles for *LYS2* (BY4741, gene drive strain, *LYS2*; BY4742, target strain, *lys2*Δ0). The sgRNA(u1) expression cassette is present on a high-copy *LEU2*-based plasmid (not illustrated). (C) *Left*, Following treatment with galactose (0 or 5 h), strains were plated onto permissive medium (SD-LEU) followed by a transfer to both SD-LEU and SD-HIS plates (examples shown) before imaging. E.V., empty pRS425 vector. The gene drive condition (*far right*) was tested in 47 independent diploid strains (see Fig. C.S2). Separate plate images (entire plate) were edited for contrast and clarity. (D) Diagnostic PCRs of clonal yeast isolates from (C) following gene drive activation and resistance to G418. *Top right*, table of oligonucleotides used and expected DNA fragment sizes (bp). Haploid controls included strains GFY-2383 (drive, Control-B) and GFY-3733 (target, Control-C). Asterisk, the expected PCR size for PCR-G was 5,573 bp for Control-B and 1,083 bp for Control-C. Horizontal white lines designate separate DNA gels; vertical white lines have been included for clarity. Determination of diploid status also utilized mating assays to control strains (Fig. C.S3) and amplification of the *LYS2* (PCRs E-G) and *MET15* loci (Fig. C.S4). Unmodified DNA gels are presented in Fig. C.S7.

Gene drive action utilizes HDR over NHEJ

Studies with CRISPR gene drive systems in insects have found that a competing DNA repair pathway, non-homologous end joining, can provide a source of drive resistant alleles[20,32,33]. This occurs due to the following mechanism of repair: the nuclease induces a double strand break within the target, yet rather than the drive coping and replacing the target locus through HDR, the cleaved chromosome repairs via NHEJ including destruction of the original CRISPR site(s) (Fig. 3.2A). Give that our yeast GD1 includes two identical (u1') sites flanking the entire target locus, cleavage followed by NHEJ-based repair would result in complete excision of the target DNA. This mode of action was demonstrated within haploids harboring GD1 at the *HIS3* locus and a guide RNA that targeted the flanking (u2) sequences[13] (Fig. 3.2B). In haploid cells, Cas9 was induced through culturing haploid yeast with galactose and the high-copy sgRNA(u2)-containing plasmid was transformed followed by selection and recovery on dextrose-containing plates. Surviving clones were then tested for loss of G418 resistance (indicating removal of the entire GD

locus and Kan^R selection marker) and analyzed using PCR and DNA sequencing. Independent clones had indels at the site of cleavage +3 bp upstream of the 5' end of the PAM within the remaining (u2) CRISPR site (Fig. 3.2B). These surviving haploid strains had excised the GD locus; one such isolate served as a control for analysis of possible NHEJ-based repair within our gene drive system.

From our first-generation CRISPR drive experiments, we randomly selected 100 independent clones across three trials involving 47 separately generated diploid strains between the drive and target parental strains (Fig. 3.1). These were tested as clonal diploids (using the mating test) that were also G418-resistant and sensitive to the SD-HIS condition; diagnostic PCRs were also performed demonstrating that these isolates had maintained the drive and lost the target allele (Fig. C.S5). We next performed a PCR using oligonucleotides to the *HIS3* promoter and terminator for all 100 strains compared to controls (Fig. 3.2C). For the haploid strain containing a self-cleaved *HIS3* locus and a modified (u2) site from NHEJ-based repair (Fig. 3.2B), a small PCR fragment was generated (369 bp). PCR conditions were optimized for the generation of this fragment size. However, for all 100 separate drives that had successfully lost the target allele, none produced a band of similar size, indicating the locus had been replaced by the drive, rather than repaired by NHEJ. A similar near-100% effectiveness of a yeast gene drive was also observed in a previous CRISPR system using sporulation of diploids and subsequent testing of generated haploids[6]. While it remains possible that NHEJ may still occur within this yeast-based system, it may be significantly below that of HDR given the effectiveness of homologous recombination in budding yeast.

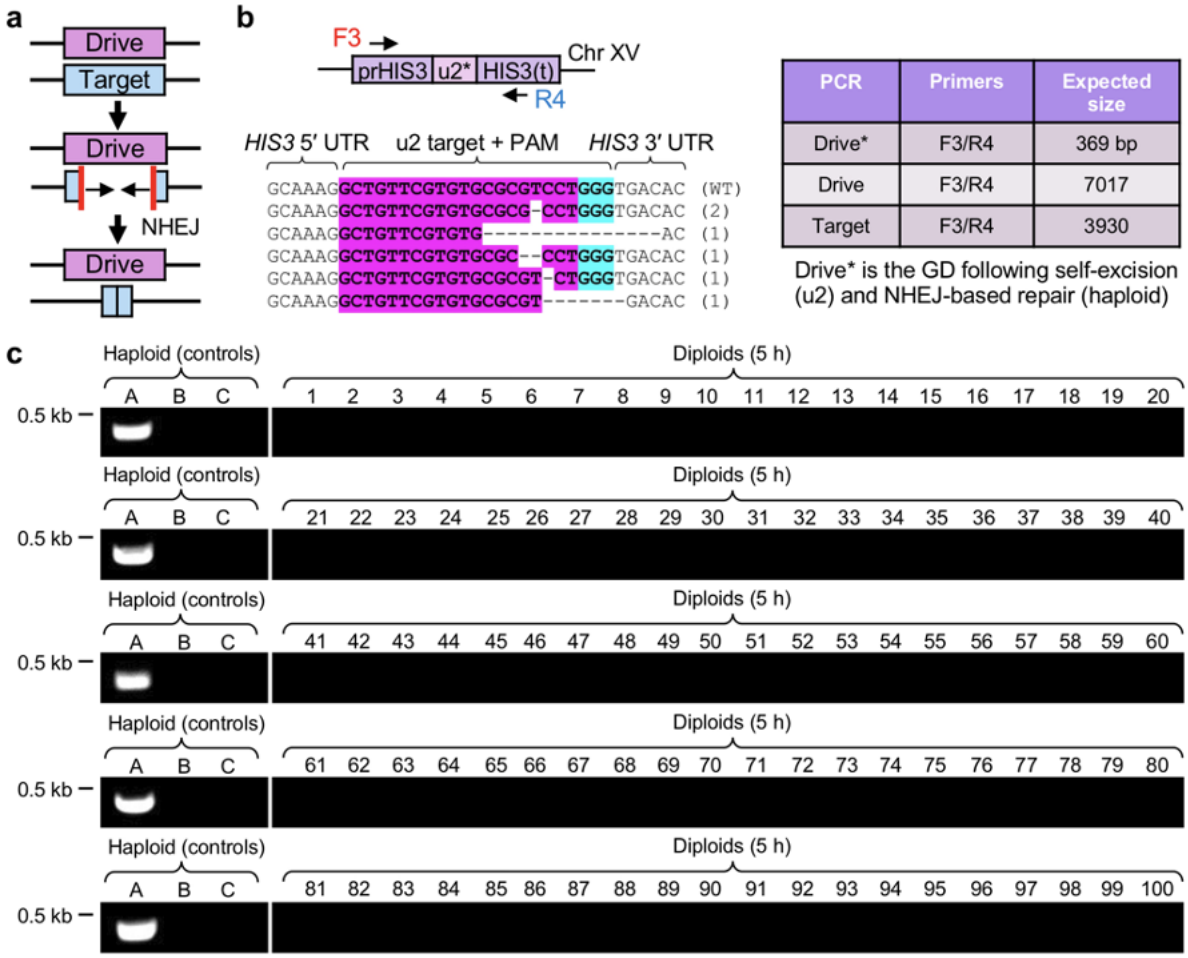


Figure 3.2. Homology directed repair is the primary mechanism for gene drive action compared to non-homologous end joining.

(A) Schematic of the potential for NHEJ-based repair of the target allele that would prevent HDR-based repair and successful propagation of a gene drive. (B) Generation of a control allele in a haploid strain using self-cleavage of Cas9 and artificial u2 sites. The gene drive haploid strain from Fig. 3.1 (GFY-2383) was activated and a plasmid expressing sgRNA(u2) (pGF-V809) was transformed with no additional donor DNA. Surviving clonal isolates were obtained on SD-LEU medium and the *HIS3* locus was analyzed. Sample sequences obtained are illustrated for six independent isolates. The expected PCR product size for NHEJ-based repair within this target allele (designed as Control-A) would equal 369 bp (for an isolate lacking a single base at the cleavage site). Use of the same primers to amplify the entire drive or target allele would result in a product of 7,017 bp and 3,930 bp, respectively. (C) From 47 gene drive diploids tested (Fig. 3.1), one to three isolates (100 total) were obtained at random that were resistant to G418, had lost the target allele (sensitivity to SD-HIS plates), and were diploid (via the mating test, Fig. C.S3). For all strains, the *HIS3* locus was also examined by diagnostic PCR (Fig. C.S5). Vertical white lines were added for clarity. Control-B and Control-C (haploid) strains were used from Fig. 3.1.

Multiplexing within the target locus

Our GD1 design utilized flanking (u2) or (u1') sites surrounding the engineered drive or target alleles, respectively. While this increases biosafety and containment, it also allows dual cleavage using only a single guide RNA (both sites are identical sequences). Previous work has demonstrated that multiplexing to identical sites across the genome can be accomplished for both native and artificial sequences for genomic editing or recruitment of enzymatically-dead Cas9 fusions[25,34]. Moreover, multiplexing the Cas9 nuclease to the intended target has become an important strategy in insect systems to promote drive success and reduce drive resistance[20]. However, our (u1'/u2) site arrangements present a unique scenario for NHEJ-based repair. Cleavage of both sites followed by precise repair (of the 5' end of the upstream site and the 3' end of the downstream site) would result in recreation of a new (u1') site that could be theoretically be subjected to additional rounds of cleavage and repair. In order to test whether our gene drive system would still be effective when the flanking (u1') sites were not used, we designed new guide RNAs to positions within the coding sequences of mCherry and *S. pombe HIS5* (both non-native yeast genes). We chose sequences (Fig. 3.3A) that still provided a maximum mismatch from the budding yeast genome for biosafety reasons and to prevent (or minimize) off-target effects.

First-generation drive strains were transformed with empty vector controls, a high-copy plasmid expressing sgRNA(mcherry), a high-copy plasmid expressing sgRNA(*SpHIS5*), or both guide RNAs; these were tested against the target strain for drive efficiency similar to Fig. 3.1. Inclusion of only one guide RNA still allowed for strong gene drive activity with nearly all yeast colonies sensitive to the SD-HIS condition (Fig. 3.3B). We noticed that for drives that multiplexed to both target sites, the number of surviving colonies was either 0 or very close to 0 (near 100% drive activity), a slight improvement over the same drives that utilized only a single guide; this

had not been previously tested given our dual (u1)-based system. Experiments were repeated after switching the plasmid backbone (pRS426 and pRS425) for the sgRNA(mCherry) and sgRNA(*SpHIS5*) constructs and similar results were obtained (data not shown).

Subsequent analyses of selected diploid clones from each condition demonstrated the presence of the drive allele, loss of the target (for active conditions), and diploid status based on the *LYS2* locus (Fig. 3.3C). Of note, because the Cas9 cleavage sites in this experiment were designed to internal sequences within the target allele, it remained possible that recombination may have occurred between the drive and target MX(t) DNA segment (rather than within the flanking *HIS3* terminator). However, in both cases, the entire target locus would still be replaced by the drive allele for an active GD; the only difference would be preservation of the downstream (u1') site (29 bp) rather than replacement by the (u2) sequence (23 bp). Diagnostic PCRs to the target locus did not discriminate between these two possibilities because the *CCW12* promoter sequence was excised in both cases (Fig. 3.3C, PCR-D).

For drives harboring only one guide RNA to *SpHIS5* (isolates 5-12), it was possible that NHEJ-based repair may have allowed for formation of indels that could have disrupted the reading frame and/or introduced a premature stop codon (also providing sensitivity on SD-HIS plates and phenocopying loss of the entire allele through HDR). However, NHEJ-based repair should have still allowed for one or two amplified PCRs (Fig. 3.3C, PCRs C and/or D) within the target locus given that positioning of the DNA primers was sufficiently distant from the cleavage site within *SpHIS5*. Our sampled isolates did not allow for amplification of either PCR, demonstrating that HDR had converted the entire target locus to GD1/GD1. Finally, for isolates harboring both guide RNA constructs (21-28), it remained possible that the intervening sequence between the cleavage sites in mCherry and *SpHIS5* (Fig. 3.3A) were excised and the chromosome was repaired by NHEJ

(and eliminated *SpHIS5* expression and/or function). Our amplified fragments of the target locus (PCRs C,D) both relied on primers to sequences that may have been excised in this scenario. Therefore, we also tested for amplification of the target locus (isolates 21-28) for the presence of pr*CDC12* positioned at the *HIS3* locus; we could not detect any evidence supporting NHEJ-based repair for these sampled isolates (Fig. C.S6).

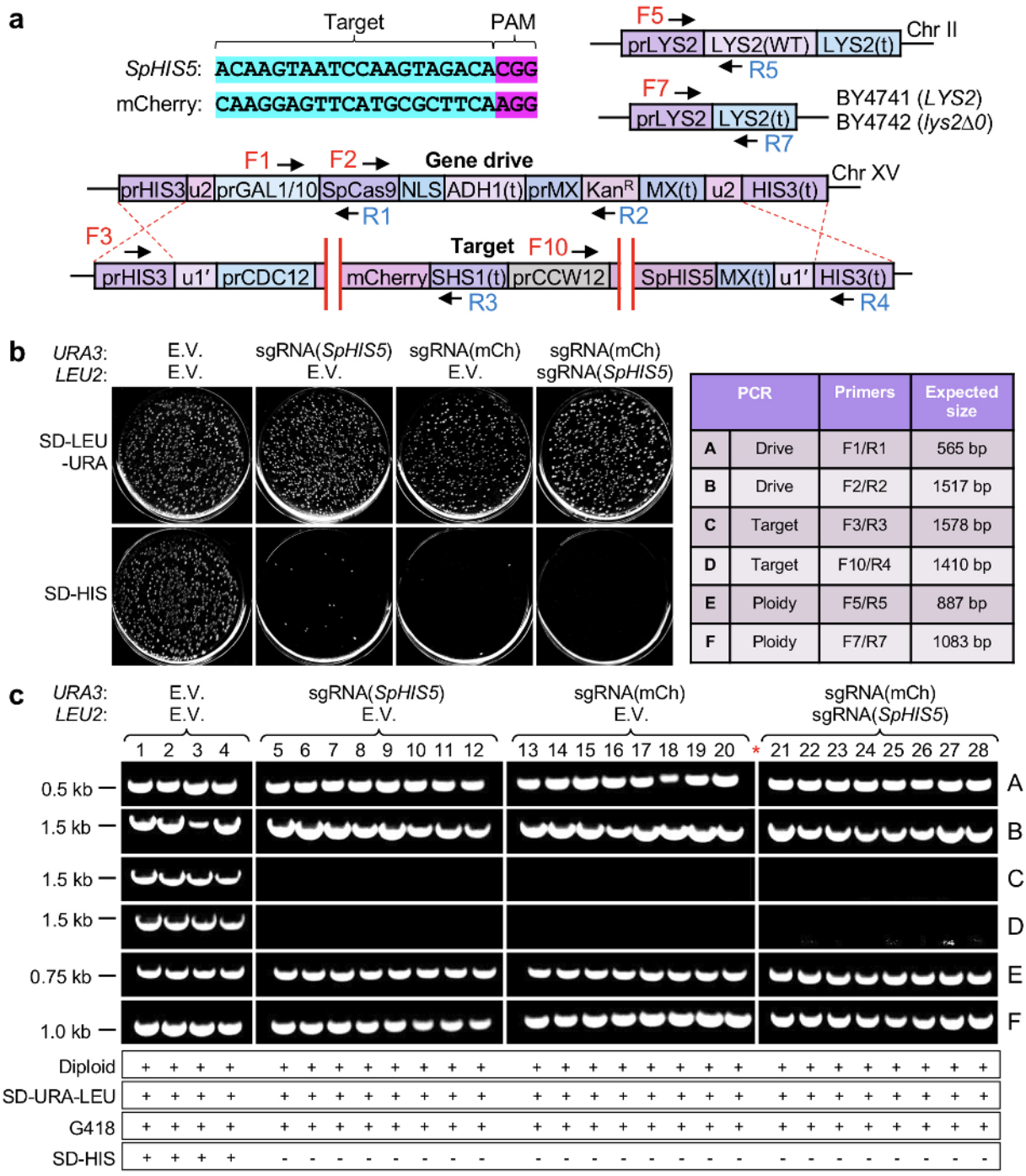


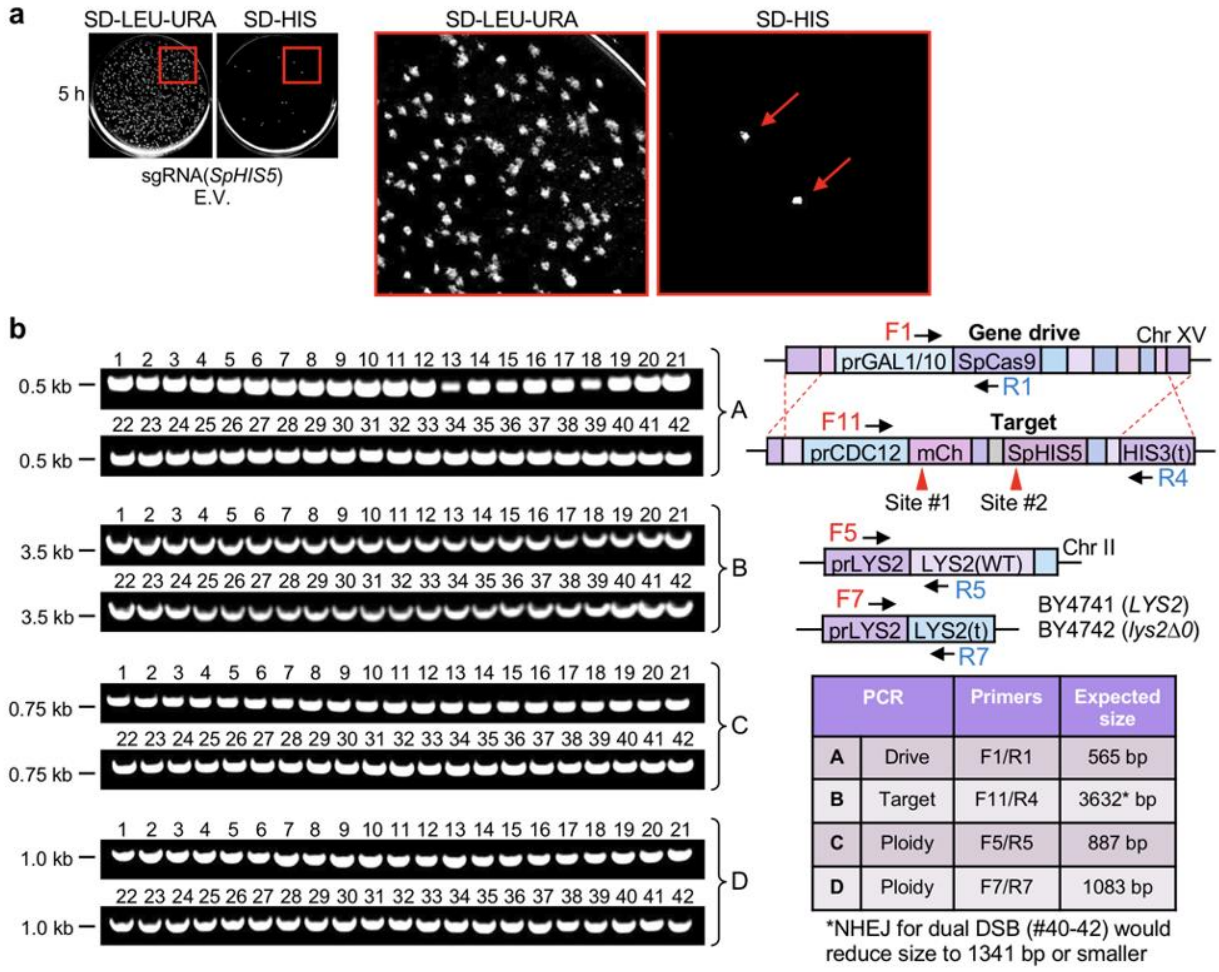
Figure 3.3. Use of alternative sgRNAs within the target allele allows for successful CRISPR gene drive action.

(A) Schematic of a modified drive/target system that included target sites within the mCherry and *SpHIS5* coding sequences. Cross-over between the MX(t) DNA is also possible upstream of the (u1') and (u2) sequences, but is not illustrated for clarity. (B) Two high-copy plasmids (*URA3*-based, pGF-V2153, pGF-V2159; *LEU2*-based, pGF-V2152) were included within the drive strain

GFY-2383. E.V., empty pRS425 or pRS426 vectors. Following activation in galactose, diploids were plated as in Fig. 3.1 using SD-LEU-URA and SD-HIS medium. Individual plate images (full plates) were edited for contrast. (C) Diagnostic PCRs of clonal diploid isolates from (B). *Top right*, table of primers chosen for analysis and the expected product sizes. Horizontal white lines designate separate DNA gels. Vertical white lines are included for clarity; red asterisk, a separate DNA gel was used for samples 21-28.

From all our work involving CRISPR drives in yeast[12,13,17,23,24], we have noticed that there are typically a very small number of surviving colonies on the SD-HIS condition, regardless of the choice of guide RNA used (Figs. 3.1, 3.3). It is important to note that our developed gene drive system does not impose selection or challenge concurrent with drive action. There is no (or little) selective advantage or disadvantage to maintaining or losing the target locus (cells were allowed to recover on SD-LEU or SD-URA-LEU plates to select for any included plasmids prior to examination on SD-HIS). This system is distinct from other types of gene drives that might directly select for loss of the target allele (for example, loss of *URA3* for resistance to 5-fluoroorotic acid) or include selective pressure within the population directly dependent on drive action and the subsequent outcome of the progeny (for example, sex-determination bias or alteration of traits required for optimal fitness). Our current system does not provide additional challenge to "resist" at the point of drive action and repair. Therefore, we expected that our drives would allow for near 100% activity under standard conditions. However, to better understand the occurrence of this small percentage of "resistant" clones that still maintained growth on SD-HIS plates, we analyzed 42 independent diploid isolates from the drives that utilized mCherry and *S. pombe HIS5* targets and were still G418 resistant (drive allele) and SD-HIS resistant (target allele) (Fig. 3.4). Following growth on SD-HIS plates (Fig. 3.4A), yeast were selected as clonal isolates (a second round of selection on SD-HIS) and chromosomal DNA was isolated and subjected to diagnostic PCRs of the *HIS3* and *LYS2* loci; all 42 strains were diploids and still contained both the drive and target

cassettes (Fig. 3.4B). The target loci were amplified and sequenced surrounding both the mCherry and *S. pombe HIS5* sites for all clones. Interestingly, we confirmed that all 42 strains had unmodified sequences at both CRISPR sites including several hundred bases upstream and downstream of the intended cleavage site identical to the original haploid parental strain. Moreover, for fifteen isolates from the sgRNA(*SpHIS5*) condition, the *GAL1/10* promoter within the drive allele was also sequenced and did not contain any alterations (Fig. 3.4C). Of note, for the three resistant isolates that originally included both guide RNAs, the intervening sequence between target sites was maintained (Fig. 3.4C, 40-42). Finally, we tested for the occurrence of the *URA3*-based and *LEU2*-based high-copy plasmids harboring guide RNA cassettes. For some isolates, one or both plasmids were lost in the absence of continual selection (Fig. 3.4C). However, loss of the guide RNA plasmids has been previously demonstrated in our system and other yeast gene drives[6,13]. From these data, we conclude that it is unlikely that NHEJ (through formation of indels) or *de novo* mutation of the target site(s) could account for the accumulation of some "resistant" yeast colonies following action of our gene drive.



PCR		Primers	Expected size
A	Drive	F1/R1	565 bp
B	Target	F11/R4	3632* bp
C	Ploidy	F5/R5	887 bp
D	Ploidy	F7/R7	1083 bp

*NHEJ for dual DSB (#40-42) would reduce size to 1341 bp or smaller

c

Guide(s) included (*URA3*/*LEU2*)	SD-HIS-resistant colonies tested*	Growth on SD-URA	Growth on SD-LEU	Growth on both SD-URA & SD-LEU	Diploid status	CRISPR sites with wild-type seq	prGAL1/10 with wild-type seq**
sgRNA(*SpHIS5*) E.V.	21 (#1-21)	9/21	4/21	2/21	21/21	42/42	15/15
sgRNA(mCherry) E.V.	18 (#22-39)	5/18	2/18	1/18	18/18	36/36	ND
sgRNA(*SpHIS5*) sgRNA(mCherry)	3 (#40-42)	2/3	2/3	2/3	3/3	6/6	ND
sgRNA(mCherry) sgRNA(*SpHIS5*)	0	-	-	-	-	-	-

*Colonies growing on SD-HIS plates that were also G418 resistant
**At least 500 bp of 5′ UTR confirmed via DNA sequencing

Figure 3.4. Analysis of yeast isolates still displaying growth on SD-HIS medium following gene drive activation.

(A) Example plates from Fig. 3.3 are illustrated with an area highlighted, enlarged for clarity (*right*), and edited for increased contrast (each plate separately). Red arrows illustrate colonies resistant to the SD-HIS condition. (B) 42 separate clonal isolates were obtained from independent

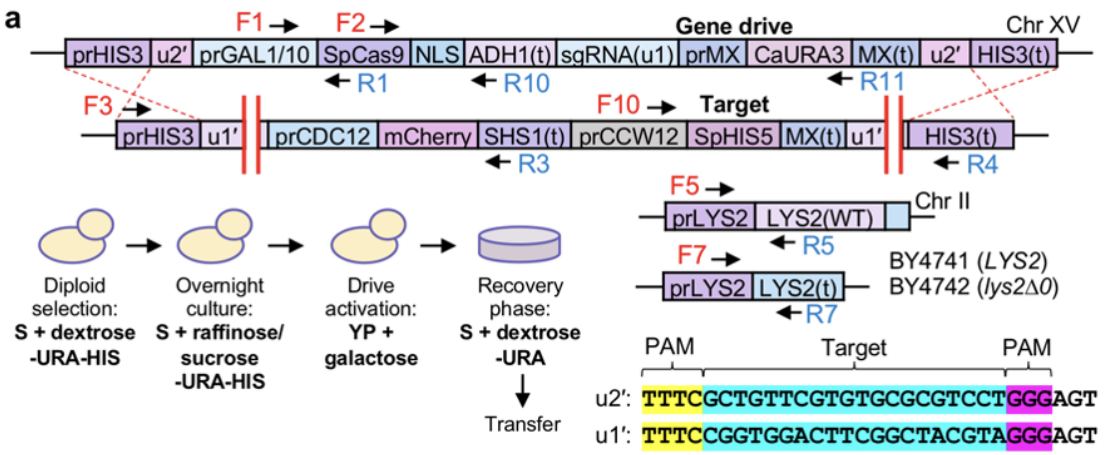
SD-HIS plates used for gene drives in Fig. 3.3 that were also resistant to G418. Diagnostic PCRs were performed on purified chromosomal DNA for both the *HIS3* and *LYS2* loci. Asterisk, for isolates 40-42, if NHEJ had repaired the *HIS3* locus following dual cleavage, the amplified product size (PCR-B) would be approximately 1,341 bp (depending on included indels). (C) Yeast clones still displaying growth on SD-HIS from gene drive experiments from Fig. 3.3 (that were also resistant to G418) were subjected to additional growth and ploidy tests, diagnostic PCRs, and DNA sequencing. ND, not determined. Sequencing of the mCherry and *SpHIS5* target sites included confirmation of several hundred bases upstream and downstream of the site at minimum. 15 of the 21 isolates from the sgRNA(*SpHIS5*) condition were chosen for sequencing of the *GAL1/10* promoter.

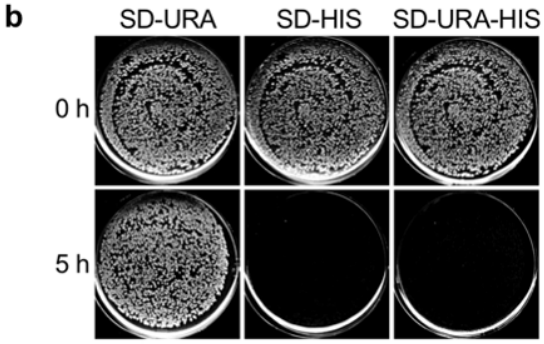
A second-generation CRISPR drive and imperfect nuclease activation

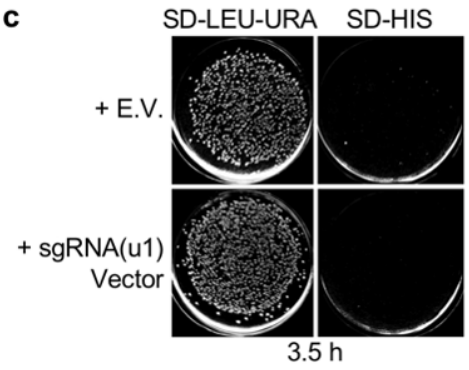
The design of eukaryotic drives includes placement of both the nuclease and corresponding guide RNA expression cassette(s) at the intended locus (or loci) of interest. In our yeast system, we could rely on plasmid-borne expression for guide RNAs. While this is convenient for testing of multiple guide sequences, altered variants, multiplexing, and biosafety (rapid loss without selection), it does not examine action of a gene drive harboring both critical components within the genome. Therefore, we developed a second-generation CRISPR drive (GD2) in yeast to test the effectiveness of this altered arrangement. Modifications to our original drive design (Fig. 3.1) included the following: (i) new (u2') sites flanking the gene drive that included additional PAM sequences for *F. novicida* Cas12a and *S. aureus* Cas9, (ii) an integrated guide RNA cassette downstream of the *S. pyogenes* Cas9 terminator sequence, and (iii) switching of the Kan^R selection marker for the *C. albicans URA3* marker. Moreover, we adjusted our methodology for selection and activation of the gene drive to include continual selection for the drive and target alleles (using media lacking both uracil and histidine) within our pre-induction culture and continual selection for the drive allele within the recovery plate following galactose treatment (Fig. 3.5A). Using this system, we tested yeast colonies on both SD-URA-HIS and SD-HIS plates following a 5 h galactose induction of Cas9. We found no difference between these two conditions; both resulted

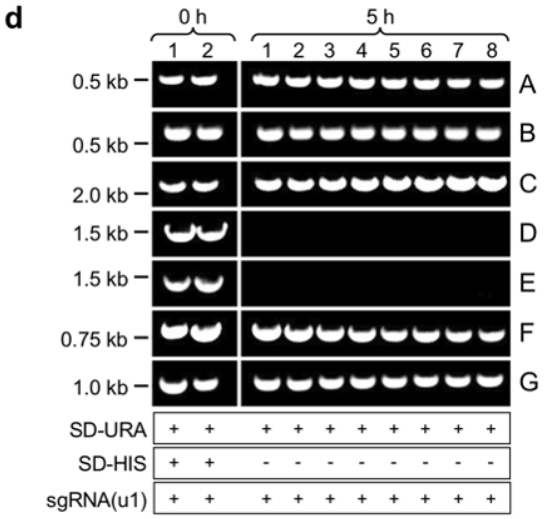
in nearly 100% active drives (Fig. 3.5B). We also tested whether integration of the guide RNA expression cassette might result in a less efficient drive system (due to limiting sgRNA expression). We included either an empty vector control or second copy of the sgRNA(u1)-expressing high-copy plasmid within GD2 and utilized a 3.5 h galactose induction. We observed only a minimal difference between one integrated guide cassette and a second cassette on a high-copy plasmid, although this may be specific to the *SNR52* promoter, the *S. pyogenes* Cas9 nuclease, and/or use of the *HIS3* locus for our drive system (Fig. 3.5C). Analysis of clonal isolates following drive activation demonstrated the presence of the drive allele (PCRs A-C), loss of the target allele (PCRs D,E), diploid status using *LYS2* (PCRs F,G), and also the presence of the guide RNA cassette itself (DNA sequencing) proximal to the Cas9 gene (Fig. 3.5D).

Finally, given that the guide RNA expression cassette was integrated within the drive, we examined 50 separate "resistant" isolates across multiple plates that maintained survival on the SD-URA-HIS condition. It is important to note that these sampled 50 clones represent only a very small percentage of all colonies examined following drive activation (Fig. 3.5B). Yeast were selected as clonal isolates on the same media type once more—selecting for one marker within the drive (*C. albicans URA3*) and one marker within the target (*S. pombe HIS5*). Next, strains were pre-induced overnight, cultured in galactose a second time for 5 h, recovered on SD-URA plates, and tested once more on SD-HIS (Fig. 3.5E). This experiment examined whether the remaining "resistant" clones were still competent to express Cas9 and function as an active drive. Interestingly, 88% of the 50 clones (44/50) were able to reactivate the drive to near 100% activity, demonstrating that imperfect expression (or function) of the nuclease could explain the occurrence of at least some remaining heterozygous clones (Fig. 3.5E).









PCR		Primers	Expected size
A	Drive	F1/R1	565 bp
B	Drive	F2/R10	671 bp
C	Drive	F2/R11	2380 bp
D	Target	F3/R3	1578 bp
E	Target	F10/R4	1410 bp
F	Ploidy	F5/R5	887 bp
G	Ploidy	F7/R7	1083 bp

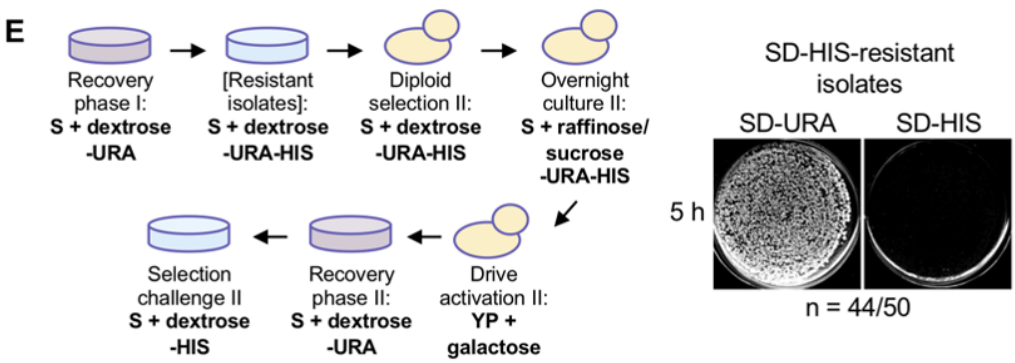


Figure 3.5. A second generation CRISPR gene drive includes an integrated sgRNA cassette. (A) An updated gene drive system includes a number of alterations: (i) an integrated sgRNA(u1) expression cassette downstream of the *ADH1* terminator, (ii) *C. albicans URA3* as a selection marker for the drive locus (as opposed to Kan^R), (iii) and a modified (u2') site similar to (u1') (includes a 5'-TTTV-3' sequence at the 5' end that is compatible with Cas12a systems). Changes to the activation protocol include dual selection for both drive/target loci (SD-URA-HIS) during diploid selection as well as pre-induction medium. The recovery phase maintains selection for the drive allele on SD-URA plates. (B) Strains GFY-4325 and 4326 were mated with GFY-3733, diploids selected, and Cas9 expression was activated by culturing for 5 h in galactose. Following recovery on dextrose, colonies were transferred to both SD-HIS and SD-URA-HIS plates. (C) Drive strains were transformed with either an empty pRS425 vector or sgRNA(u1)-expressing cassette (pGF-V1220), diploids selected on SD-LEU-URA-HIS, and gene drives were activated for 3.5 h in galactose before plating. (D) *Left*, clonal isolates following drive activation (B) were tested by diagnostic PCRs. *Right*, table of examined primer combinations and the expected amplified product sizes. The presence of the integrated sgRNA(u1) cassette was determined by PCR and DNA sequencing for all 10 isolates. Horizontal white lines indicate separate DNA gels used; vertical white line was included for clarity. (E) Across 12 independent GD2 diploid strains activated (B), a total of 50 clonal isolates that maintained growth on SD-URA-HIS were collected. *Left*, strains were grown on SD-URA-HIS plates a second time and prepared for a second round of gene drive activation. *Right*, representative plates of one isolate that was originally resistant to the SD-URA-HIS condition (44/50). All agar plates were edited for contrast and clarity (each plate separately).

Discussion

A model system for study of gene drive action

Our yeast gene drive model includes a number of important benefits: simplicity, genetic tractability, tunability, and biosafety. While previous work with our initial GD1 system highlighted a number of variations to guide RNA sequences[13], Cas9 subcellular localization[13], and anti-CRISPR-based inhibition[12], numerous other alterations to drive action (Cas9 enzymatic activity), drive control/regulation (anti-CRISPRs or nuclease degradation), and DNA repair could be tested in future iterations. Our findings from this study included controls to allow for efficient examination of diploid status (amplification of the *LYS2* and/or *MET15* loci), comparisons to parental haploid genomes, and the lack of detectible NHEJ-based repair within our gene drive regardless of the choice of target sites. We recognize that NHEJ-based repair may still occur within

our yeast model, but this will require development of a distinct drive/target system for simultaneous detection of (i) loss of the target allele and (ii) maintenance of only a *single* copy of the drive allele within a diploid genome. HDR and/or NHEJ repair systems themselves could also be examined or modulated within our drive system; this might allow for future modulation of DNA repair to aid in drive optimization or control[24].

Optimizing nuclease expression in vivo

Our study also examined a large number of isolated "resistant" colonies that continued to be present within our yeast system on the final medium (SD-HIS plates) albeit at a small percentage within sampled populations of asynchronous cultures. Various sources can provide alterations within the target DNA that will prevent action of the drive (and develop true resistance) including *de novo* mutation[32], sequence variation within a mixed population[35], and NHEJ-based repair followed by indel formation or DNA excision[20]. In these cases, the nuclease would be unable to successfully cleave the altered target site(s) in subsequent generations and this would prevent *super*-Mendelian inheritance. In our model, we examined the target loci of 42 clonal diploid isolates still harboring both the drive and target alleles following drive activation; we did not detect any incidence of mutation within the 23 bp target site and PAM as well as a large amount of flanking DNA. We suspected that one explanation for the formation of these "resistant" colonies was failed or poor activation of Cas9 (rather than inappropriate mutation or repair of the target site). Our data suggest that this is the case for many *SpHIS5* positive clones as a second round of galactose culturing provided successful gene drive action in 88% of isolates from our GD2 system. Also, the presence of the MX terminator (Fig. 3.5) within both the drive and target locus upstream of the (u1') cut site may provide a source of inappropriate crossover if there is also lack of cleavage

on the upstream (u1') position. The goal of obtaining a 100% effective gene drive may be useful in some scenarios; however, even gene drives that allow for resistance may still be effective at spreading within natural populations[36]. Future iterations of our drive/target system may rely on targeting of the mCherry and *SpHIS5* coding sequences (or others) given the success of multiplexing to these sites. It may also be possible to further optimize our yeast drives (and activation of the nuclease) by including a selective advantage or requirement for Cas9 expression coupled to survival of an environmental challenge through translational fusions or *GAL1/10*-dependent transcription of other markers.

Our gene drive model provides a highly tractable system to explore alternative drive arrangements (split drives[14,24], daisy chain drives[37]), nuclease inhibitors (small molecules[15] or naturally evolved anti-CRISPRs[12]), and DNA repair systems[24]. Future work may aid in informing complex drive designs, arrangements to reduce or counter resistance, or safe practices for study or possible application in higher eukaryotes.

Author Contributions

Y.Y. and G.C.F. performed all experiments, analyzed data, and generated all figures and tables. G.C.F. wrote the manuscript. All authors have read and approved this manuscript.

Conflict of Interest

G.C.F. (Kansas State University) has filed for a provisional patent on September 29, 2017 (U.S. Provisional Patent Application Serial No. 62/565,651, "Programmed modulation of CRISPR/Cas9 activity") followed by patent filing on January 31, 2018 regarding some of the data presented in this work. A patent was filed (G.C.F.) on April 20, 2017 by the Univ. of California,

Berkeley, “Methods and Compositions for Genomic Editing” International Application No. PCT/US2017/028676 and published as No. WO 2017/189336 A1 on November 2, 2017, for the artificial target sites (u1’/u2/u2’) used in this study. GCF declares no non-financial conflict of interest. Y.Y. declares no conflict of interest of any kind.

Funding Information

This project was supported by the USDA National Institute of Food and Agriculture, Hatch Project 1013520 (G.C.F.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the United States Department of Agriculture.

Ethical Statement

This study did not involve humans or animals of any kind.

Abbreviations

CRISPR, clustered regularly interspaced short palindromic repeats; PAM, protospacer adjacent motif; GD1, first-generation CRISPR gene drive; GD2, second generation CRISPR gene drive; NHEJ, non-homologous end joining; HDR, homology-directed repair; DSB, double strand break; (u1’)/(u2)/(u2’), unique artificial CRISPR target sites programmed within the yeast genome; sgRNA, single guide RNA; pr, promoter; t, terminator; SD, synthetic drop-out with dextrose; E.V., empty vector; indel, insertion or deletion.

Reference

1 Sternberg, S. H. & Doudna, J. A. Expanding the Biologist's Toolkit with CRISPR-Cas9. *Molecular cell* **58**, 568-574, doi:10.1016/j.molcel.2015.02.032 (2015).

2 Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science (New York, N.Y.)* **346**, 1258096, doi:10.1126/science.1258096 (2014).

3 Wright, A. V., Nunez, J. K. & Doudna, J. A. Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell* **164**, 29-44, doi:10.1016/j.cell.2015.12.035 (2016).

4 Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.)* **337**, 816-821, doi:10.1126/science.1225829 (2012).

5 Jinek, M. *et al.* RNA-programmed genome editing in human cells. *eLife* **2**, e00471, doi:10.7554/eLife.00471 (2013).

6 DiCarlo, J. E., Chavez, A., Dietz, S. L., Esvelt, K. M. & Church, G. M. Safeguarding CRISPR-Cas9 gene drives in yeast. *Nature biotechnology* **33**, 1250-1255, doi:10.1038/nbt.3412 (2015).

7 Gantz, V. M. *et al.* Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E6736-6743, doi:10.1073/pnas.1521077112 (2015).

8 Hammond, A. *et al.* A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae. *Nature biotechnology* **34**, 78-83, doi:10.1038/nbt.3439 (2016).

9 Esvelt, K. M. & Gemmell, N. J. Conservation demands safe gene drive. *PLoS biology* **15**, e2003850, doi:10.1371/journal.pbio.2003850 (2017).

10 Esvelt, K. M., Smidler, A. L., Catteruccia, F. & Church, G. M. Concerning RNA-guided gene drives for the alteration of wild populations. *eLife* **3**, doi:10.7554/eLife.03401 (2014).

11 Akbari, O. S. *et al.* BIOSAFETY. Safeguarding gene drive experiments in the laboratory. *Science (New York, N.Y.)* **349**, 927-929, doi:10.1126/science.aac7932 (2015).

12 Basgall, E. M. *et al.* Gene drive inhibition by the anti-CRISPR proteins AcrIIA2 and AcrIIA4 in Saccharomyces cerevisiae. *Microbiology (Reading, England)* **164**, 464-474, doi:10.1099/mic.0.000635 (2018).

13 Roggenkamp, E. *et al.* Tuning CRISPR-Cas9 Gene Drives in Saccharomyces cerevisiae. *G3 (Bethesda, Md.)* **8**, 999-1018, doi:10.1534/g3.117.300557 (2018).

14 Del Amo, V. L. *et al.* Split-gene drive system provides flexible application for safe laboratory investigation and potential field deployment. *bioRxiv*, 684597, doi:10.1101/684597 (2019).

15 López Del Amo, V. *et al.* Small-molecule control of super-Mendelian inheritance in gene drives. *bioRxiv*, 665620, doi:10.1101/665620 (2019).

16 Vella, M. R., Gunning, C. E., Lloyd, A. L. & Gould, F. Evaluating strategies for reversing CRISPR-Cas9 gene drives. *Scientific reports* **7**, 11038, doi:10.1038/s41598-017-10633-2 (2017).

17 Roggenkamp, E. *et al.* CRISPR-UnLOCK: multipurpose Cas9-based strategies for Conversion of yeast libraries and strains. *Frontiers in microbiology* **8**, 1773, doi:10.3389/fmicb.2017.01773 (2017).

18 Halder, V., Porter, C. B. M., Chavez, A. & Shapiro, R. S. Design, execution, and analysis of CRISPR-Cas9-based deletions and genetic interaction networks in the fungal pathogen Candida albicans. *Nature protocols*, doi:10.1038/s41596-018-0122-6 (2019).

19 Shapiro, R. S. *et al.* A CRISPR-Cas9-based gene drive platform for genetic interaction analysis in Candida albicans. *Nature microbiology* **3**, 73-82, doi:10.1038/s41564-017-0043-0 (2018).

20 Champer, J. *et al.* Reducing resistance allele formation in CRISPR gene drive. *Proceedings of the National Academy of Sciences of the United States of America*, doi:10.1073/pnas.1720354115 (2018).

21 Kyrou, K. *et al.* A CRISPR-Cas9 gene drive targeting doublesex causes complete population suppression in caged Anopheles gambiae mosquitoes. *Nature biotechnology* **36**, 1062-1066, doi:10.1038/nbt.4245 (2018).

22 Grunwald, H. A. *et al.* Super-Mendelian inheritance mediated by CRISPR-Cas9 in the female mouse germline. *Nature*, doi:10.1038/s41586-019-0875-2 (2019).

23 Goeckel, M. E. *et al.* Modulating CRISPR gene drive activity through nucleocytoplasmic localization of Cas9 in S. cerevisiae. *Fungal biology and biotechnology* **6**, 2, doi:10.1186/s40694-019-0065-x (2019).

24 Yan, Y. & Finnigan, G. C. Development of a multi-locus CRISPR gene drive system in budding yeast. *Scientific reports* **8**, 17277, doi:10.1038/s41598-018-34909-3 (2018).

25 Finnigan, G. C. & Thorner, J. mCAL: a new approach for versatile multiplex action of Cas9 using one sgRNA and loci flanked by a programmed target sequence. *G3 (Bethesda, Md.)* **6**, 2147-2156, doi:10.1534/g3.116.029801 (2016).

26 Champer, J. *et al.* Molecular safeguarding of CRISPR gene drive experiments. *eLife* **8**, doi:10.7554/eLife.41439 (2019).

27 Sambrook, J. & Russell, D. W. *Molecular Cloning: A Laboratory Manual*. 3rd edn, (Cold Spring Harbor Laboratory Press, 2001).

28 Finnigan, G. C. & Thorner, J. Complex *in vivo* ligation using homologous recombination and high-efficiency plasmid rescue from *Saccharomyces cerevisiae*. *Bio-protocol* **5**, e1521. http://www.bio-protocol.org/e1521 (2015).

29 DiCarlo, J. E. *et al.* Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic acids research* **41**, 4336-4343, doi:10.1093/nar/gkt135 (2013).

30 Swiat, M. A. *et al.* FnCpf1: a novel and efficient genome editing tool for Saccharomyces cerevisiae. *Nucleic acids research* **45**, 12585-12598, doi:10.1093/nar/gkx1007 (2017).

31 Lian, J., HamediRad, M., Hu, S. & Zhao, H. Combinatorial metabolic engineering using an orthogonal tri-functional CRISPR system. *Nature communications* **8**, 1688, doi:10.1038/s41467-017-01695-x (2017).

32 Unckless, R. L., Clark, A. G. & Messer, P. W. Evolution of Resistance Against CRISPR/Cas9 Gene Drive. *Genetics* **205**, 827-841, doi:10.1534/genetics.116.197285 (2017).

33 Hammond, A. M. *et al.* The creation and selection of mutations resistant to a gene drive over multiple generations in the malaria mosquito. *PLoS genetics* **13**, e1007039, doi:10.1371/journal.pgen.1007039 (2017).

34 Chen, B. *et al.* Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479-1491, doi:10.1016/j.cell.2013.12.001 (2013).

35 Buchman, A., Marshall, J. M., Ostrovski, D., Yang, T. & Akbari, O. S. Synthetically engineered Medea gene drive system in the worldwide crop pest Drosophila suzukii. *Proceedings of the National Academy of Sciences of the United States of America* **115**, 4725-4730, doi:10.1073/pnas.1713139115 (2018).

36 Noble, C., Adlam, B., Church, G. M., Esvelt, K. M. & Nowak, M. A. Current CRISPR gene drive systems are likely to be highly invasive in wild populations. *eLife* **7**, doi:10.7554/eLife.33423 (2018).

37 Noble, C. *et al.* Daisy-chain gene drives for the alteration of local populations. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 8275-8282, doi:10.1073/pnas.1716358116 (2019).

Chapter 4 - Development of a CRISPR/dCas9-based assay to study CRISPR components using budding yeast growth as a convenient genetic readout for CRISPR function.

Abstract

The CRISPR/Cas9 system's programmable DNA cleavage allows for efficient, site-specific genome modification in different organisms. This technology has been employed in research for a variety of purposes, including gene modification, imaging genomic loci, managing genome-wide screening, and regulating transcription. The CRISPR based transcription system can regulate gene expression by either CRISPR based gene interference (CRISPRi) or CRISPR-mediated gene activation (CRISPRa). Both systems contain a catalytically inactive Cas9 (dCas9), a single-guide RNA (sgRNA) and a transcriptional regulation domain. In this study, we developed two CRISPR/dCas9 transcriptional regulation systems by fusing dCas9 with either the transcriptional activator VPR, or the transcriptional repressor Mxi1. Our experiments demonstrated that dCas9-VPR could activate the transcription of an artificial target gene, which was inhibited by the repressive promoter *HED1* or *GAL7*. We also demonstrated that dCas9-Mxi1 could repress the expression of yeast septin genes (*CDC3* or *CDC11*) when the sgRNA was targeted to the native septin promoter.

Introduction

CRISPR/Cas9 has been utilized as the most efficient genomic tool to edit genomes in a variety of organisms[1-9]. It was discovered in prokaryotic organisms, such as, bacteria and archaea[10] as an adaptive immune system. In this system, single guide RNA (sgRNA) binds the endonuclease Cas9 and together they target a complementary sequence of DNA. This creates a double strand break (DSB) at 3 bps upstream of the protospacer-adjacent motif (PAM) site[2,7,8]. The editing process allows scientists to either make an insertion, deletion, or modifications to DNA break sites. However, Cas9 variants lose their nuclease activity when mutations are introduced. Scientists discovered that a nuclease-null Cas9 (dCas9; D10A and H840A) retains its ability to bind target DNA[11,12] but is unable to introduce a double strand break. The dCas9 has been repurposed as a target-specific platform that can be translationally fused with different cellular programming molecules to regulate gene functions. For example, dCas9 was reported to work as a base editing tool when fused with cytidine deaminase enzyme[13,14]. This strategy causes base substitution by mediating the conversion from cytidine (C) to uridine (U). The dCas9 can also be used for other genetic manipulations when fused with epigenetic modifiers[15]. When dCas9 is tagged with fluorescent proteins, such as GFP, it can be used to visualize chromosomes in real time[16]. The dCas9 can even be fused with different transcriptional regulators, such as the transcriptional activation domains, VPR, or repressive domain, Mxi1[17] to regulate gene expression.

The transcriptional activator VPR is the combination of three constituent domains (VP64, p65 and Rta)[18]. Scientists demonstrated that these three activating domains displayed high transcriptional activity; any one of them replaced by mCherry led to low transcriptional activity, suggesting the essentiality of all three domains[18]. Previous work with dCas9 in yeast demonstrated that the dCas9-VPR could activate gene transcription[19], such as, the promoters of *HED1* and

GAL7[18]. The *HED1* promoter is a meiotic specific promoter, which is only activated during meiosis[20-22], and the *GAL7* promoter is an inducible promoter, which is activated by galactose and repressed by dextrose[23]. Mxi1 is a transcriptional repressor domain that was discovered in mammalian cells. It is reported to interact with the histone deacetylase Sin3 homolog in yeast[24]. It is a potential tumor suppressor by inhibiting the transcriptional activity of MYC. Previous studies demonstrated that dCas9-Mxi1 can suppress the transcription of endogenous eukaryotic genes when targeted to specific regions[11]. Unfortunately, it is still unknown if dCas9-Mxi1 repression is position-dependent in eukaryotic cells. It remains unclear if i) dCas9-VPR could reactivate gene expression that is repressed by various mechanisms, ii) dCas9-Mxi1 can be applied to other genes in budding yeast, and iii) whether patterns exist relating to dCas9-Mxi1 based repression of gene expression.

In this study, we developed dCas9-VPR/Mxi1 systems within budding yeast. Each system contained a dCas9-based transcriptional fusion, a corresponding guide RNA, and a target allele. Using the dCas9-VPR to activate transcription, we examined if this construct could restore expression of the yeast selection marker *S.p.HIS5* that was under control of the repressive promoters of *HED1* or *GAL7*. Our results revealed that dCas9-VPR could significantly activate the expression of *S.p.HIS5*. Using the dCas9-Mxi1 repression system, we tested repression of native septin genes (*CDC3* and *CDC11*) in yeast. We designed 12 independent sgRNAs against different positions of septin *prCDC3* and *prCDC11*. This approach provided evidence that when the sgRNA targeted to a certain region (50 ~180 bps upstream of the start codon), dCas9-Mxi1 inhibited septin gene expression. Together, these findings could be used in the development of genome wide screening tools.

Materials and Methods

Yeast strains and plasmids

Saccharomyces cerevisiae strains and plasmids used in this study can be found in Table 4.1. Standard molecular biology protocols were applied to manipulate all DNA and yeast strains[25]. Yeast strains were created using the following steps. First, modified septin alleles were assembled onto *CEN*-based plasmids (pRS315) using *in vivo* ligation[26]. Then, the constructed plasmid cassettes were amplified with a high-fidelity polymerase into two fragments: the upstream PCR product included a promoter, septin gene, mCherry and an *ADH1* terminator, whereas the downstream fragment included the same *ADH1* terminator, drug resistance cassette and the septin gene terminator. The PCR products contained several hundred base pairs of homology and were transformed into yeast using a modified lithium acetate method[27]; and integrated at the septin loci using homologous recombination (HR). Diagnostic PCR was applied to examine if each cassette integrated successfully, followed by DNA sequencing to confirm the integration results. The study contained three types of plasmids: *LEU2*-labeled dCas9-Mxi1, *URA3*-labeled septin covering vectors and *HIS3*-labeled sgRNA. First, we built the covering vector using *in vivo* ligation; the WT septin gene and GFP were under control of the *SHS1* promoter (this plasmid was labeled with *C.a.URA3*). Then, the dCas9 plasmids were created by *in vivo* ligation of a promoter, *S. pyogenes* dCas9, NLS (nuclear localization signal), Mxi1 and drug resistance cassette together into *CEN*-based vectors (pRS315). Finally, the sgRNA expression constructs were custom synthesized and subcloned to yeast vectors with two flanking unique restriction sites. All vectors were confirmed by Sanger DNA sequencing.

Culture conditions

Budding yeast were cultured in liquid or solid medium. YPD-based medium included 2% peptone, 1% yeast extract, and 2% dextrose. Synthetic (drop-out) medium included yeast nitrogen base, ammonium sulfate, and amino acid supplements. All media was autoclaved, and the sugar solution was filter sterilized. For drug-containing agar plates, the final concentration of G418 sulfate was 240 µg/mL, and concentration for 5-fluoro-orotic acid (5-FOA) plates was 500 µg/mL. To eliminate potential contaminants, this 5-FOA solution was heated to 70°C for 30 min before being cooled and filtered.

Yeast growth assay

Yeast strains for VPR assays grew overnight at 30°C on synthetic medium SD-URA-LEU as a square patch. Then, yeast cells were examined on SD-URA-LEU and triple drop-out agar plates (SD-URA-LEU-HIS) individually by streaking for single colonies. All strains were grown at identical temperatures. Images were taken after 3 days of growth.

All plasmid-carrying strains for Mxi1 assays were grown at 30°C on triple drop-out agar plates (SD-URA-LEU-HIS) as a square patch for 24 h. All tested strains were inoculated on triple drop-out medium (SD-URA-LEU-HIS). Strains that failed to survive on this medium after transformation experiments were not tested on 5-FOA medium. Yeast cells were tested on SD-LEU-HIS+5-FOA agar plates by streaking for single colonies. All strains were handled in a consistent way. Images were taken after 3 days of growth.

Fluorescence microscopy

Yeast cultures were grown in triple drop-out (SD-URA-LEU-HIS) liquid medium at 30°C overnight, then back diluted into rich medium YPD for 4.5 h at 30°C. Cells were then collected by

centrifugation, washed in water, and observed under a fluorescence microscope (Leica, model DMI6000; Leica Microsystems, Buffalo Grove, IL, USA) within 1 h. Cells were imaged by a 100x lens with GFP and mCherry fluorescence filters (Semrock, GFP- 4050B-LDKM-ZERO, mCherry-C-LDMK-ZERO). Images were captured by a Leica DFC340 FX camera. All images were analyzed by using ImageJ (National Institute of Health). Samples for each figure were all treated in the same day under identical conditions. Representative cells were chosen for each image. White light (DIC) was used to bring cells into the plane of focus.

Results

Development of a genetic assay using dCas9-VPR

A previous study demonstrated that in the presence of sgRNA, the Cas9 nuclease bound to target DNA precisely even without its nuclease activity[28]. The dCas9 construct was tagged with different types of transcriptional regulators to develop a CRISPRi and CRISPRa system in eukaryotes[11,28,29]. We chose the gene *S.p.HIS5* (non-native gene in *S. cerevisiae*) and yeast septin genes *CDC3*/*CDC11* (essential) as transcriptional targets for dCas9.

In our transcriptional activation system, the *S.p.HIS5* allele was integrated at the *HIS3* locus under the control of the repressive promoters *HED1* or *GAL7*. A previous study developed and tested sgRNAs targeting the promoters of *HED1* and *GAL7* at their native loci[18]. These same sgRNA expression cassettes and the dCas9-VPR construct were included on *LEU2*- and *URA3*-labeled plasmids (Fig. 4.1A). In our transcriptional repression system, septin allele(s) were modified at the native *CDC3*/*CDC11* loci with their native promoters. The sgRNAs, dCas9-Mxi1 and covering vectors were expressed on *HIS3*-, *LEU2*- and *URA3*-labeled plasmids, respectively (Fig. 4.1B).

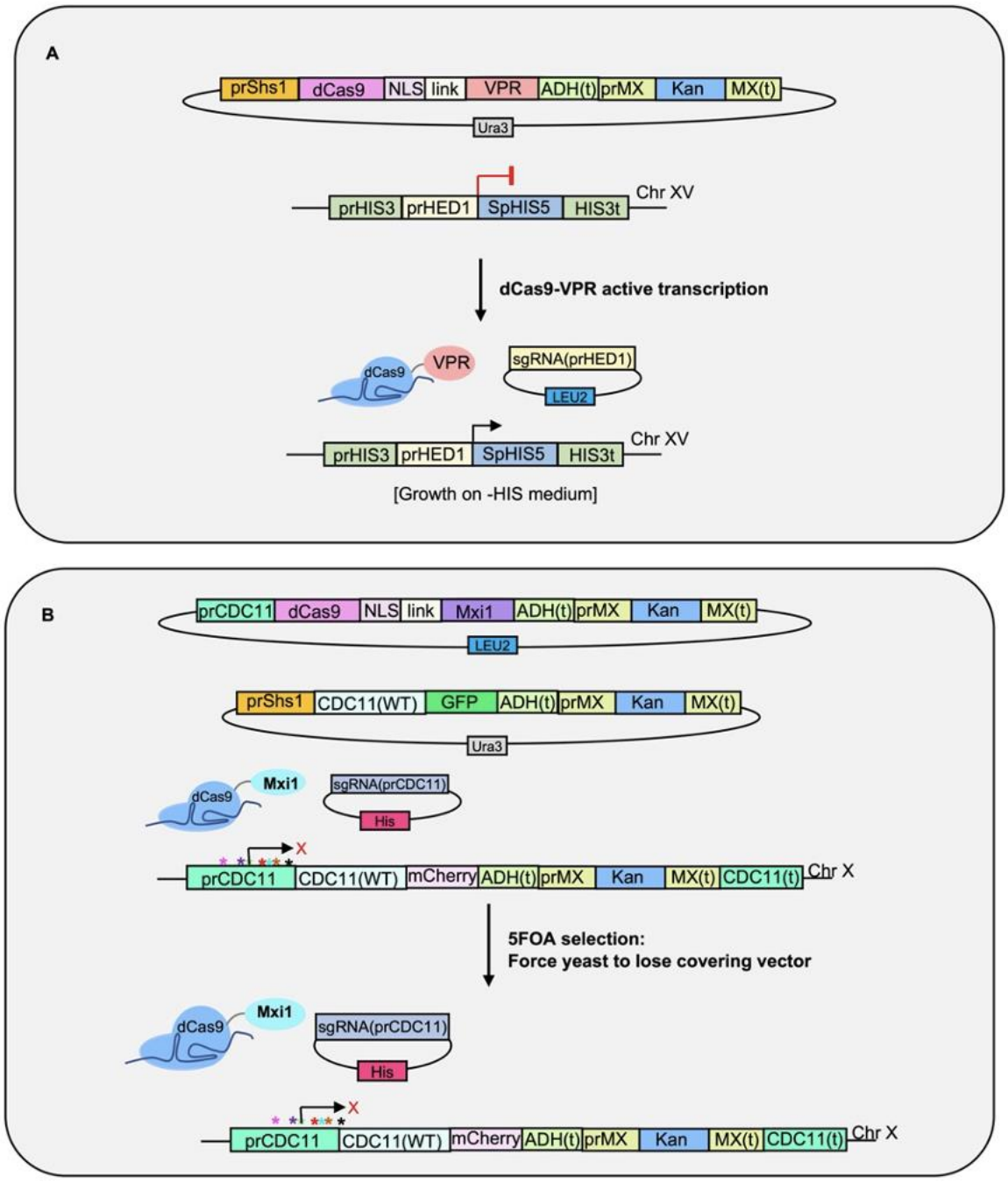


Figure 4.1. Design of a CRISPR/dCas9-based transcriptional regulation system in *S. cerevisiae.*

(A) The dCas9-VPR transcriptional regulator was included on a *URA3* labeled plasmid. The *prHED1-S.p.HIS5* was integrated at *HIS3* locus. The sgRNA(*prHED1*) cassette was included on a *LEU2* labeled plasmid. (B) Our CRISPR/dCas9 transcriptional repression system contained i) a native copy of *CDC3/CDC11*-mCherry within the genome, ii) the dCas9-Mxi1 transcriptional regulator, a covering vector expressing wild-type *S. cerevisiae CDC3/CDC11* and a sgRNA on *LEU2, URA3* and *HIS3* plasmids, respectively.

In the CRISPR/dCas9 system, the target DNA contained the selection marker *S. pombe HIS5* that allowed cells to grow on medium lacking histidine. The transcription of *S.p.HIS5* was inhibited by promoter *HED1* (Fig. 4.2A). In order to activate the transcription of *S.p.HIS5*, we introduced plasmids containing either sgRNA(*prHED1*), or dCas9-VPR, or both into the *prHED1-S.p.HIS5* strain. This would test if the dCas9-VPR system could activate the inhibited promoter. These cells were transferred onto SD-URA-LEU and/or SD-URA-LEU-HIS plates to obtain single colonies. Only the strain (GFY-4817) that contained both dCas9-VPR and sgRNA(*prHED1*) plasmids could grow on SD-URA-LEU-HIS medium, suggesting the combination of both could activate the transcription of *S.p.HIS5* and allowed for growth on medium lacking histidine. All other strains, GFY-4814 (dCas9-VPR plasmid only), GFY-4820 (empty vectors) and GFY-4823 (sgRNA(*prHED1*) plasmid only) failed to grow on SD-URA-LEU-HIS medium (Fig. 4.2A). To confirm these results, we repeated the assay with a separate promoter; we substituted *prHED1* with *prGAL7*[23](Fig. 4.2B). We observed a similar phenomenon: only the strain containing both dCas9-VPR and sgRNA(*prGAL7*) plasmids could survive on medium lacking histidine (Fig. 4.2B). These results indicated that *prHED1* and *prGAL7* repressed the expression of *S.p.HIS5*, but the combination of dCas9-VPR with a sgRNA targeting these promoters could successfully activate a synthetic target gene allowing for cell surviving.

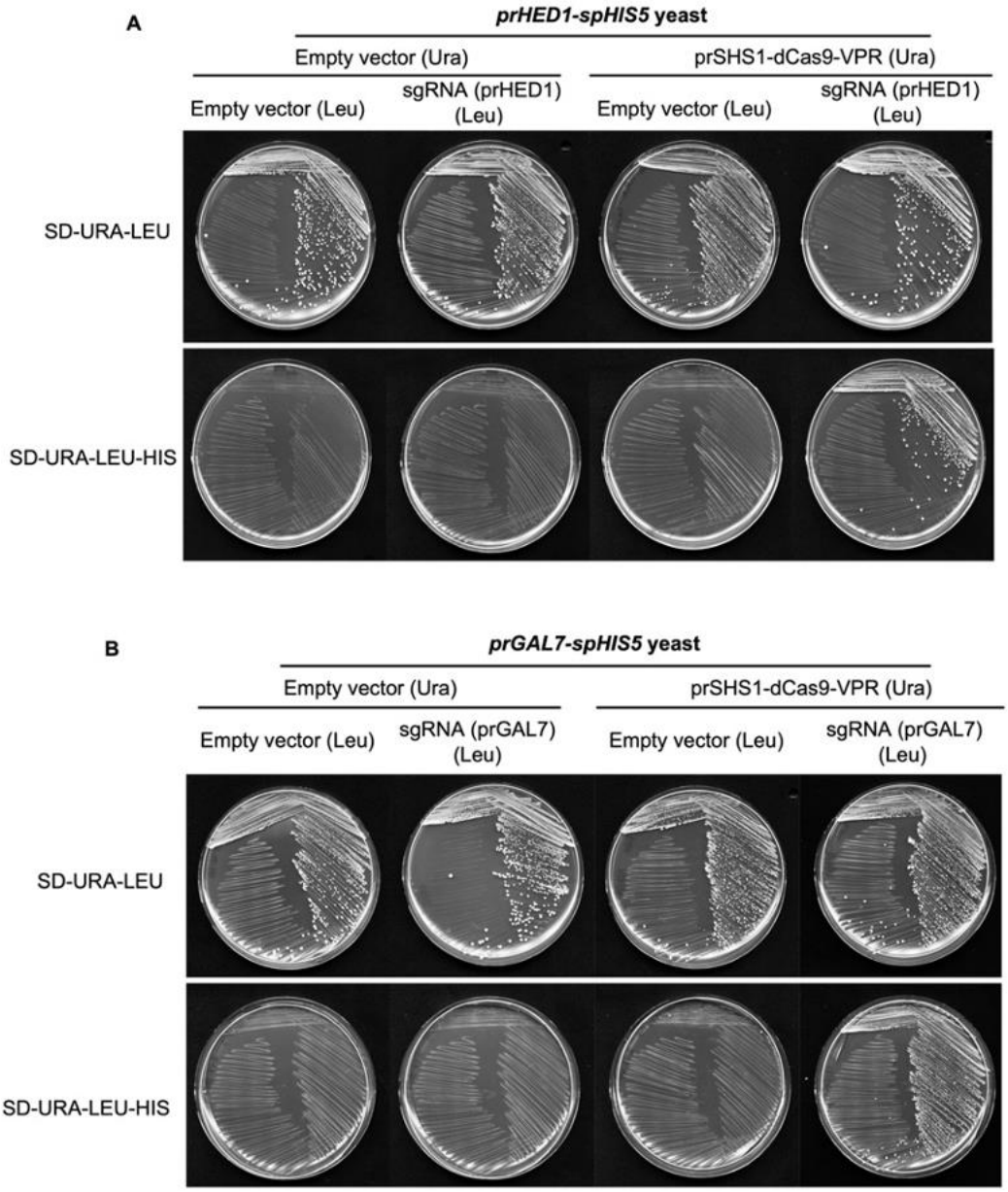


Figure 4.2. The dCas9-VPR construct activates transcription of *S.p.HIS5* in yeast.

(A) The dCas9-VPR activated transcription of *S.p.HIS5*, whose expression was repressed by the *prHED1* in budding yeast. Yeast strains GFY-4814, GFY-4817, GFY-4820 and GFY-4823 were grown at 30°C on SD-URA-LEU medium as a square patch for 24 h. Then cells were tested on both SD-URA-LEU and SD-URA-LEU-HIS medium by streaking for single colonies. Images were taken after 3 days growth. All strains were handled in a consistent manner. (B) The dCas9-VPR could activate the expression of *S.p.HIS5*, when it was repressed by *prGAL7* in budding yeast. Yeast strains GFY-4826, GFY-4829, GFY-4832 and GFY-4835 were treated as described above. Images were captured after 3 days incubation.

CRISPR/dCas9-Mxi1 transcriptional downregulation of essential yeast septin genes

A previous study demonstrated that dCas9-Mxi1 could repress gene transcription in eukaryotic cells[11]. To test this transcriptional inhibition strategy, we transformed a plasmid that contains dCas9-Mxi1 into yeast. Our yeast strains included a native copy of *CDC3/CDC11*-mCherry within the genome (Fig. 4.1B). The septin genes (*CDC3* and *CDC11*) are essential for completion of the cell cycle. The Cdc3/Cdc11-deficient strains needed to contain a *URA3*-marked plasmid expressing wild-type *S. cerevisiae CDC3/CDC11* to maintain cell viability. However, we could not rely on a WT copy of *CDC11* on a plasmid under control of its native promoter. The dCas9-Mxi1 and guide RNA constructs were constitutively expressed, and both the *CDC11* genomic copy and covering vector copy would be repressed upon transformation of these constructs into yeast. Therefore, we designed a covering vector to include a separate promoter (promoter *SHS1*) driving expression of *CDC11* to escape dCas9-Mxi1 based transcriptional inhibition. Finally, the genotype (native *CDC11* gene, guide RNA, and dCas9-Mxi1) of interest would be exposed by selecting for loss of the *URA3*-marked plasmid on 5-FOA containing medium[31].

We first tested *cdc11*Δ yeast with two empty plasmids (*LEU2/HIS3*) as our positive control, demonstrating lack of the *CDC11* gene leads to cell death (Fig. 4.3A, plate 1). To ensure that all components in our transcriptional system (dCas9-Mxi1, sgRNA, and empty plasmids) individually did not result in cell death, we tested these constructs in yeast containing a native WT copy of *CDC11* tagged with mCherry (Fig. 4.3A, plates 2-5). We observed that the conditions that included two empty vectors (*LEU2/HIS3*), sgRNA/empty vector, dCas9-Mxi1/empty vector, and a separate dCas9-Mxi1/empty vector (dCas9 expressed from a different promoter) displayed robust growth, suggesting these components were not toxic.

Previous studies found that the upstream promoter region could impact expression when targeted by transcriptional effectors[29,30]. We designed 7 sgRNAs (212, 184, 163, 82, 65, 52, 23) to target the *CDC11* promoter and 6 sgRNAs (223, 170, 122, 75, 45, 16) to target the *CDC3* promoter (Table 4.1). Each guide RNA number indicated the distance from the last base pair of 5'-NGG-3' PAM site to first base pair of the "ATG" start codon. Cells were tested on SD-URA-LEU-HIS and SD-LEU-HIS+5-FOA medium (Fig. 4.3B). In cells expressing Cdc11-mCherry, we noticed robust cell growth when sgRNAs targeted positions 212, 184, and 23 of promoter *CDC11* (Fig. 4.3B, left). Interestingly, there were still a small percentage of cells that survived when sgRNAs targeted at position 163 and 82. We do not have a strong explanation for this phenotype, and it is possible that some mutations allowed for bypass of our repression system. However, when sgRNAs targeted at position 65 or position 52, we observed that nearly 100% of the colonies were sensitive to the SD-LEU-HIS+5-FOA condition, indicating a strong repression occurred. We also found that there was no growth difference in the *prCDC11*-dCas9-Mxi1 mediated group or the *prSHS1*-dCas9-Mxi1 mediated group (Fig. 4.3B, right). Altogether, these results indicated that dCas9-Mxi1 transcriptional regulation was position dependent within the *CDC11* promoter. Additionally, similar results were obtained in cells expressing Cdc3-mCherry (Fig. 4.3C). The dCas9-Mxi1 failed to repress *CDC3* expression at positions 223, 45, and 16, but it showed strong inhibition when the sgRNA targeted position 75. However, at positions 170 and 122, colonies were almost completely invisible when they were repressed by *prSHS1*-dCas9-Mxi1, whereas a small percentage of colonies were still visible when regulated by *prCDC11*-dCas9-Mxi1 (Fig. 4.3C). Together, these results demonstrated that dCas9-Mxi1 based repression of essential yeast gene promoters was highly dependent upon the positioning of guide RNAs.

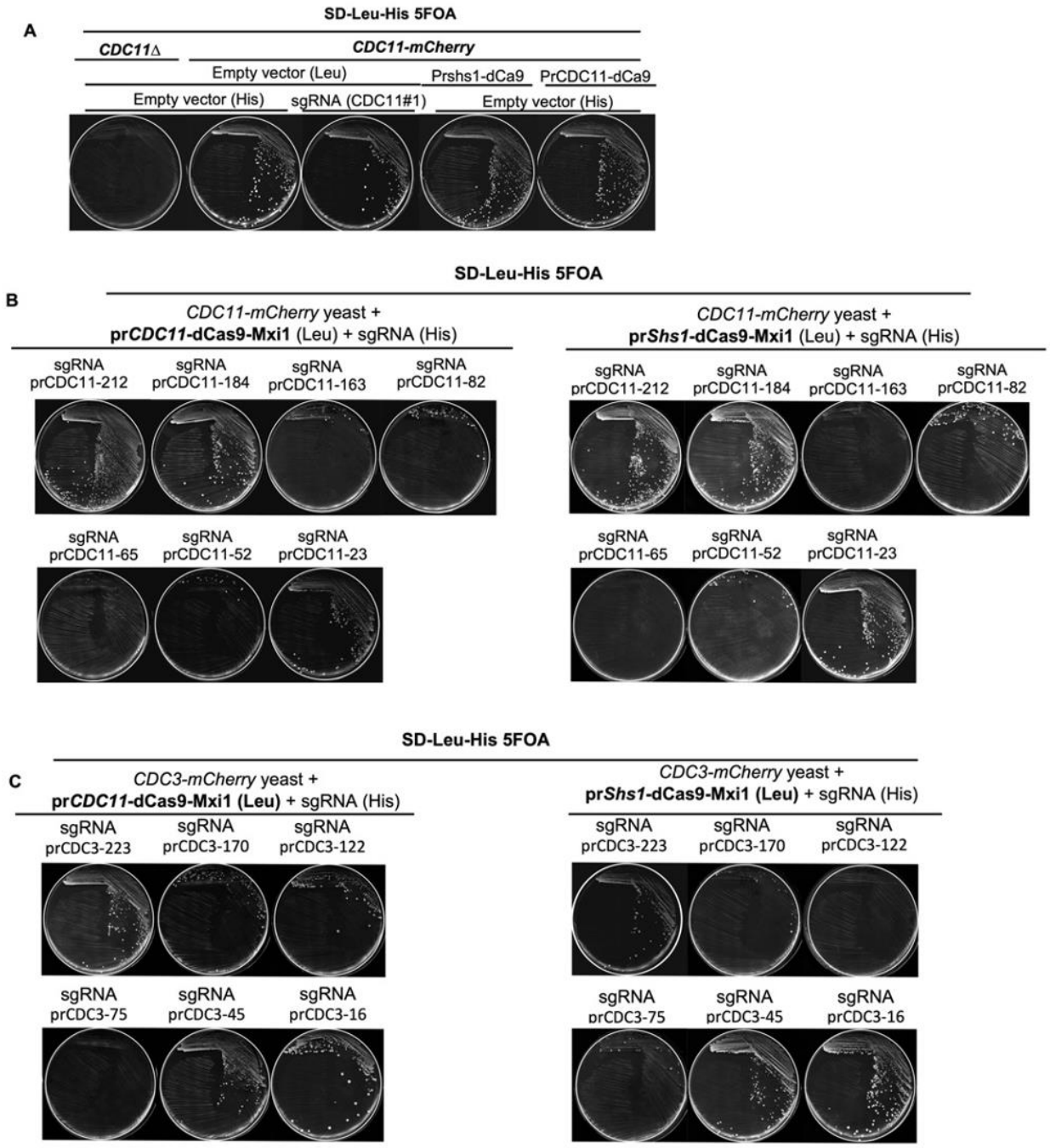


Figure 4.3. The dCas9-Mxi1 system inhibits transcription of septin genes by targeting the promoter region.

(A) Plasmids controls do not affect yeast growth. Yeast strains GFY-4334, GFY-4336, GFY-4337, GFY-4338 and GFY-4339 (all initially harboring a *URA3*-marked covering plasmid expressing WT *CDC11*) were cultured on SD-URA-LEU-HIS media overnight and tested on SD-LEU-HIS+5 FOA media, as indicated. Images were captured after 3 days incubation. (B) (C) left, the dCas9-Mxi1 was under the control of *prCDC11*; right, the dCas9-Mxi1 was under control of *prSHS1*. All other settings were identical between left and right in each figure. (B) The dCas9-Mxi1

inhibited transcription of yeast septin gene *CDC11*. Yeast strains GFY-4209 to GFY-4215 and GFY-4202 to GFY-4208 (all initially harboring a covering vector expressing WT *CDC11*) were cultured on SD-URA-LEU-HIS media overnight at 30ºC, then streaked for single on SD-LEU-HIS+5 FOA media. Images were captured after 3 days incubation. (C) The dCas9-Mxi1strongly reduced the transcription of *CDC3* gene. Yeast strains GFY-4222 to GFY-4227 and GFY-4216 to GFY-4221 were assayed as described in (B). SD, synthetic drop-out medium with dextrose.

To independently verify that dCas9-Mxi1 inhibited septin expression by transcriptional repression, we examined the localization of septins C-terminal tagged with mCherry using fluorescence microscopy (Fig. 4.4). These cells also expressed *URA3*-marked *CDC3/CDC11*-GFP plasmids to clearly show the location of septin-based structures (Fig. 4.4A). In Cdc11 expressing strains, we observed that mCherry signal almost completely disappeared when sgRNAs were targeted to positions 65 or 82 of the promoter, suggesting that expression of Cdc11-mCherry was strongly repressed by *prCDC11*-dCas9-Mxi1 or *prSHS1*-dCas9-Mxi1 at this position. However, when the sgRNA targeted to position 212, we could observe robust mCherry signal, revealing that dCas9-Mxi1 did not inhibit transcription at this position (Fig. 4.4B). Similar results were obtained in yeast strains expressing Cdc3-mCherry with an extra copy of Cdc3-GFP on a plasmid (Fig. 4.4C). We observed that mCherry tagged Cdc3 localized to the yeast bud neck when the sgRNA was targeted to position 223 of the *CDC3* promoter in either *prCDC11*-dCas9-Mxi1 or *prSHS1*-dCas9-Mxi1 expression groups. However, signals started to show differences in these two groups when sgRNAs bound to positions 122 and 75. In the *prCDC11*-dCas9-Mxi1 (Fig. 4.4C, left) group, we could observe strong Cdc3-mCherry signal when targeting position 122 and very weak signal at position 75. However, in the *prSHS1*-dCas9-Mxi1 (Fig. 4.4C, right) group, there was no mCherry signal detected for guides targeting position 122 or position 75. These data suggested that dCas9-Mxi1 under different promoters may have different transcriptional efficiencies when regulating the *CDC3* septin gene.

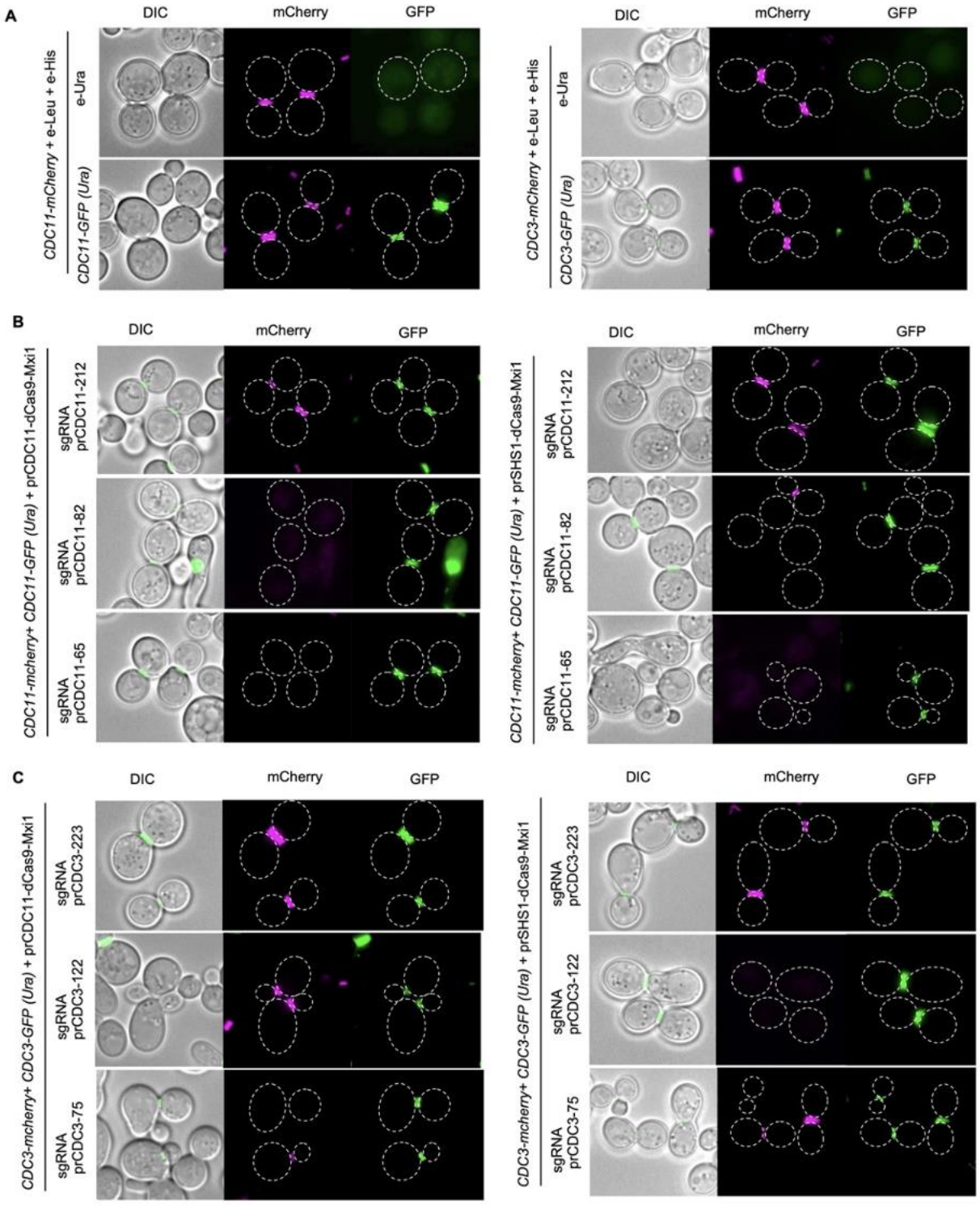


Figure 4.4. The dCas9-Mxi1 system repressed the transcription of septin *CDC3* and *CDC11*.

(A) Yeast strains GFY-4335, GFY-4336, GFY-4340 and GFY-4341 were cultured overnight in SD-LEU-URA-HIS media, back diluted into YPD media, and grown for 4.5 h at 30ºC. Images

were captured by fluorescence microscopy. Representative images are displayed. (B) The dCas9-Mxi1 repressed the transcription of *CDC11*-mCherry at *CDC11* locus. Yeast strains GFY-4209, GFY-4212, GFY-4213, GFY-4202, GFY-4205 and GFY-4206 were treated as indicated in Figure 4.4 (A) and imaged by fluorescence microscopy. Representative images are displayed. (C) The dCas9-Mxi1 inhibited *CDC3*-mCherry transcription at *CDC3* locus. Yeast strains were cultured as described in Figure 4.4(A) and imaged by fluorescence microscopy. Representative images are displayed.

Discussion

In this study, we developed a CRISPR/dCas9 transcriptional regulation system, which could control gene transcription efficiently. We fused dCas9 with different transcriptional regulators, like VPR or Mxi1, and targeted these proteins to different positions of gene promoters. Our study demonstrated that in budding yeast, dCas9-VPR and dCas9-Mxi1 could dramatically promote or repress the transcription of target genes. The dCas9-VPR complex could activate the transcription of target *S.p.HIS5* even though it was repressed by the *HED1* promoter. We also noticed that dCas9-Mxi1 repression regulation might be position-dependent. It only functioned when dCas9-Mxi1 targeted the septin promoters within a certain range. A previous study also demonstrated that CRISPR/dCas9 transcriptional systems could activate/repress gene transcription by targeting promoters[29]. They designed 88 sgRNAs to target the promoters of 12 native yeast genes and examined the effect of the dCas-Mxi1 and dCas9-VPR systems. Similar to our findings, Jensen *et al.* found that 29 out of 88 sgRNAs displayed significant regulation (activation/repression).

Our results suggested that the dCas9-Mxi1 transcriptional regulation was position dependent for the septin promoters we examined. Another study claimed that this position-dependent modification also occurred within the dCas9-VPR transcriptional system[19]. Deaner *et al.* demonstrated that the dCas9-VPR could work as a dual-mode activator/repressor when it targeted different positions of genes. This system activated transcription when it was targeted to

the promoter of genes, but, surprisingly, repressed transcription when dCas9-VPR targeted within the open reading frame. Therefore, it is extremely important to investigate how sgRNA targeting position affects dCas9-mediated transcriptional regulation.

However, some limitations remained in our system. First, the dCas9-Mxi1 constructs contained the *CDC11* promoter, which might also be targeted by sgRNA(*prCDC11*) and led to self-targeting inhibition in a feedback loop. Surprisingly, our data presented no difference between the *prCDC11*-dCas9-Mxi1 group and *prSHS1*-dCas9-Mxi1 group. Second, this study used a small number of sgRNAs to each septin promoter. Third, we only included the 5'-NGG-3' PAM sites on the coding strand of septin genes; it is unclear whether use of 5'-CCN-3' PAM sites and the corresponding sgRNAs would affect the efficiency of transcriptional regulation.

A previous study illustrated that the combination of multiple adjacent sgRNAs that target a single gene could dramatically increase editing capability[32]. Thus, this strategy was utilized to increase the CRISPR/Cas9 editing efficiency in mammalian cells[33]. Similar to this approach, we could examinate the CRISPR/dCas9 transcriptional regulation of target genes by using multiple sgRNAs.

Together, these findings provide insight into how genome wide screens could be developed to i) assay transcriptional regulation of native genes, ii) investigate development of synthetic genetic pathways, and/or iii) study CRISPR/Cas9 function *in vivo*.

Funding Information

This project was supported by an Institutional Development Award (IDeA) from the National Institute of General Medical Science of the National Institute of Health under grant number P20 GM103418 to G.C.F. This work was also supported by the USDA National Institute

of Food and Agriculture, Hatch Project 1013520 (to G.C.F.). Kansas State Univ. College of Art & Sciences and the Johnson Cancer Research Center also supported this work.

Author Contributions

E.B. (Erianna Basgall) and M.G. (Megan Goeckel) built all VPR related constructs and Y.Y. built all Mxi1 related constructs. Y.Y. streaked all VPR and Mxi1 plates and captured all images presented in this study. Y.Y. and G.C.F. analyzed data and generated all figures and tables. Y.Y. and G.C.F. wrote the manuscript. All authors have read and approved this manuscript.

Conflict of Interest

G.C.F. has applied for CRISPR-related patents: i) April 20, 2017 by the Univ. of California, Berkeley, "Methods and Compositions for Genomic Editing" International Application No. PCT/US2017/028676 and published as No. WO 2017/189336 A1 on November 2, 2017, for the artificial target sites (u1'/u2/u2') and non-native sites used in this study, ii) September 29, 2017 (U.S. Provisional Patent Application Serial No. 62/565,651, "Programmed modulation of CRISPR/Cas9 activity") followed by patent filing on January 31, 2018, iii) a PTC application "Multi-Locus Gene Drive System" No. PCT/US2019/041538 on July 12, 2019. Authors declare no other conflicts of interest.

Ethical Statement

This work did not involve any human or animal subjects.

Table 4.1. Yeast strains and plasmids used in this study

Strain	**Genotype**	**Reference**
GFY-4334	*MAT***a** *cdc11*Δ*::Kan*^{R} + pRS315 (empty) + pRS313 (empty) + pSB1/JT1520: (pRS316 URA3 CDC11*)*	Versele *et al.* (2004)
GFY-4335	*MAT***a** *cdc11*Δ*::S.c.CDC11::mCherry::ADH1(t)::Kan*^{R} + pRS316 (empty) + pRS315 (empty) + pRS423 (empty)	This study
GFY-4336	*MAT***a** *cdc11*Δ*::S.c.CDC11::mCherry::ADH1(t)::Kan*^{R} + pRS315 (empty) + pRS423 (empty) + pYY-55: (pRS316; *prSHS1::S.c.CDC11::GFP::ADH1(t)::Kan*^{R})	This study
GFY-4337[a]	*MAT***a** *cdc11*Δ*::S.c.CDC11::mCherry::ADH1(t)::Kan*^{R} + pRS423 (empty) + pYY-55: (pRS316; *prSHS1::S.c.CDC11::GFP::ADH1(t)::Kan*^{R}) + pGF-IVL-1594: (pRS315; *prSHS1::dCas9::NLS::Mxi1::ADH1(t)::Kan*^{R})	This study
GFY-4338	*MAT***a** *cdc11*Δ*::S.c.CDC11::mCherry::ADH1(t)::Kan*^{R} + pRS423 (empty) + pYY-55: (pRS316; *prSHS1::S.c.CDC11::GFP::ADH1(t)::Kan*^{R}) + pGF-IVL-1595: (pRS315; *prCDC11::dCas9::NLS::Mxi1::ADH1(t)::Kan*^{R})	This study
GFY-4339	*MAT***a** *cdc11*Δ*::S.c.CDC11::mCherry::ADH1(t)::Kan*^{R} + pRS315 (empty) + pYY-55: (pRS316; *prSHS1::S.c.CDC11::GFP::ADH1(t)::Kan*^{R}) + pGF-V2112: (pRS423; *prSNR52::Sp-sgRNA(prCDC11-212)::SUP4(t)*)	This study
GFY-4340	*MAT***a** *cdc3*Δ*::S.c.CDC3::mCherry::ADH1(t)::Kan*^{R} + pRS316 (empty) + pRS315 (empty) + pRS423 (empty)	This study
GFY-4341	*MAT***a** *cdc3*Δ*::S.c.CDC3::mCherry::ADH1(t)::Kan*^{R} +	This study

	pRS315 (empty) + pRS423 (empty) + pYY-57: (pRS316; *prSHS1::S.c.CDC3::GFP::ADH1(t)::Kan^R*)	
GFY-4202[b]	*MAT***a** *cdc11*Δ*::S.c.CDC11::mCherry::ADH1(t)::Kan^R* + pYY-55: (pRS316; *prSHS1::S.c.CDC11::GFP::ADH1(t)::Kan^R*) + pGF-IVL-1594: (pRS315; *prSHS1::dCas9::NLS::Mxi1::ADH1(t)::Kan^R*) + pGF-V2112: (pRS423; *prSNR52::sgRNA(prCDC11-212)::SUP4(t)*)	This study
GFY-4203[c]	*MAT***a** *cdc11*Δ*::S.c.CDC11::mCherry::ADH1(t)::Kan^R* + pYY-55: (pRS316; *prSHS1::S.c.CDC11::GFP::ADH1(t)::Kan^R*) + pGF-IVL-1594: (pRS315; *prSHS1::dCas9::NLS::Mxi1::ADH1(t)::Kan^R*) + pGF-V2113: (pRS423; *prSNR52::sgRNA(prCDC11-184)::SUP4(t)*)	This study
GFY-4204[d]	*MAT***a** *cdc11*Δ*::S.c.CDC11::mCherry::ADH1(t)::Kan^R* + pYY-55: (pRS316; *prSHS1::S.c.CDC11::GFP::ADH1(t)::Kan^R*) + pGF-IVL-1594: (pRS315; *prSHS1::dCas9::NLS::Mxi1::ADH1(t)::Kan^R*) + pGF-V2114: (pRS423; *prSNR52::sgRNA(prCDC11-163)::SUP4(t)*)	This study
GFY-4205[e]	*MAT***a** *cdc11*Δ*::S.c.CDC11::mCherry::ADH1(t)::Kan^R* + pYY-55: (pRS316; *prSHS1::S.c.CDC11::GFP::ADH1(t)::Kan^R*) + pGF-IVL-1594: (pRS315; *prSHS1::dCas9::NLS::Mxi1::ADH1(t)::Kan^R*) + pGF-V2115: (pRS423; *prSNR52::sgRNA(prCDC11-82)::SUP4(t)*)	This study
GFY-4206[f]	*MAT***a** *cdc11*Δ*::S.c.CDC11::mCherry::ADH1(t)::Kan^R* + pYY-55: (pRS316; *prSHS1::S.c.CDC11::GFP::ADH1(t)::Kan^R*) + pGF-IVL-1594: (pRS315; *prSHS1::dCas9::NLS::Mxi1::ADH1(t)::Kan^R*) + pGF-V2116: (pRS423; *prSNR52::sgRNA(prCDC11-65)::SUP4(t)*)	This study
GFY-4207[g]	*MAT***a** *cdc11*Δ*::S.c.CDC11::mCherry::ADH1(t)::Kan^R* +	This study

	pYY-55: (pRS316; *prSHS1::S.c.CDC11::GFP::ADH1(t)::Kan^R*) + pGF-IVL-1594: (pRS315; *prSHS1::dCas9::NLS::Mxi1::ADH1(t)::Kan^R*) + pGF-V2117: (pRS423; *prSNR52::sgRNA(prCDC11-52)::SUP4(t)*)	
GFY-4208[h]	***MAT*a** *cdc11Δ::S.c.CDC11::mCherry::ADH1(t)::Kan^R* + pYY-55: (pRS316; *prSHS1::S.c.CDC11::GFP::ADH1(t)::Kan^R*) + pGF-IVL-1594: (pRS315; *prSHS1::dCas9::NLS::Mxi1::ADH1(t)::Kan^R*) + pGF-V2118: (pRS423; *prSNR52::sgRNA(prCDC11-23)::SUP4(t)*)	This study
GFY-4209	***MAT*a** *cdc11Δ::S.c.CDC11::mCherry::ADH1(t)::Kan^R* + pYY-55: (pRS316; *prSHS1::S.c.CDC11::GFP::ADH1(t)::Kan^R*) + pGF-IVL-1595: (pRS315; *prCDC11::dCas9::NLS::Mxi1::ADH1(t)::Kan^R*)+ pGF-V2112: (pRS423; *prSNR52::sgRNA(prCDC11-212)::SUP4(t)*)	This study
GFY-4210	***MAT*a** *cdc11Δ::S.c.CDC11::mCherry::ADH1(t)::Kan^R* + pYY-55: (pRS316; *prSHS1::S.c.CDC11::GFP::ADH1(t)::Kan^R*) + pGF-IVL-1595: (pRS315; *prCDC11::dCas9::NLS::Mxi1::ADH1(t)::Kan^R*) + pGF-V2113: (pRS423; *prSNR52::sgRNA(prCDC11-184)::SUP4(t)*)	This study
GFY-4211	***MAT*a** *cdc11Δ::S.c.CDC11::mCherry::ADH1(t)::Kan^R* + pYY-55: (pRS316; *prSHS1::S.c.CDC11::GFP::ADH1(t)::Kan^R*) + pGF-IVL-1595: (pRS315; *prCDC11::dCas9::NLS::Mxi1::ADH1(t)::Kan^R*) + pGF-V2114: (pRS423; *prSNR52::sgRNA(prCDC11-163)::SUP4(t)*)	This study
GFY-4212	***MAT*a** *cdc11Δ::S.c.CDC11::mCherry::ADH1(t)::Kan^R* +	This study

	pYY-55: (pRS316; *prSHS1::S.c.CDC11::GFP::ADH1(t)::Kan^R*) + pGF-IVL-1595: (pRS315; *prCDC11::dCas9::NLS::Mxi1::ADH1(t)::Kan^R*) + pGF-V2115: (pRS423; *prSNR52::sgRNA(prCDC11-82)::SUP4(t)*)	
GFY-4213	***MAT*a** *cdc11*Δ*::S.c.CDC11::mCherry::ADH1(t)::Kan^R* + pYY-55: (pRS316; *prSHS1::S.c.CDC11::GFP::ADH1(t)::Kan^R*) + pGF-IVL-1595: (pRS315; *prCDC11::dCas9::NLS::Mxi1::ADH1(t)::Kan^R*) + pGF-V2116: (pRS423; *prSNR52::sgRNA(prCDC11-65)::SUP4(t)*)	This study
GFY-4214	***MAT*a** *cdc11*Δ*::S.c.CDC11::mCherry::ADH1(t)::Kan^R* + pYY-55: (pRS316; *prSHS1::S.c.CDC11::GFP::ADH1(t)::Kan^R*) + pGF-IVL-1595: (pRS315; *prCDC11::dCas9::NLS::Mxi1::ADH1(t)::Kan^R*) + pGF-V2117: (pRS423; *prSNR52::sgRNA(prCDC11-52)::SUP4(t)*)	This study
GFY-4215	***MAT*a** *cdc11*Δ*::S.c.CDC11::mCherry::ADH1(t)::Kan^R* + pYY-55: (pRS316; *prSHS1::S.c.CDC11::GFP::ADH1(t)::Kan^R*) + pGF-IVL-1595: (pRS315; *prCDC11::dCas9::NLS::Mxi1::ADH1(t)::Kan^R*) + pGF-V2118: (pRS423; *prSNR52::sgRNA(prCDC11-23)::SUP4(t)*)	This study
GFY-4216[i]	***MAT*a** *cdc3*Δ*::S.c.CDC3::mCherry::ADH1(t)::Kan^R* + pYY-57: (pRS316; *prSHS1::S.c.CDC3::GFP::ADH1(t):: Kan^R*) + pGF-IVL-1594: (pRS315; *prSHS1::dCas9::NLS::Mxi1::ADH1(t)::Kan^R*) + pGF-V2100: (pRS423; *prSNR52::sgRNA(prCDC3-223)::SUP4(t)*)	This study

GFY-4217[j]	*MAT***a** *cdc3*Δ*::S.c.CDC3::mCherry::ADH1(t)::Kan^R* + pYY-57: (pRS316; *prSHS1::S.c.CDC3::GFP::ADH1(t):: Kan^R*) + pGF-IVL-1594: (pRS315; *prSHS1::dCas9::NLS::Mxi1::ADH1(t)::Kan^R*) + pGF-V2101: (pRS423; *prSNR52::sgRNA(prCDC3-170)::SUP4(t)*)	This study
GFY-4218[k]	*MAT***a** *cdc3*Δ*::S.c.CDC3::mCherry::ADH1(t)::Kan^R* + pYY-57: (pRS316; *prSHS1::S.c.CDC3::GFP::ADH1(t):: Kan^R*) + pGF-IVL-1594: (pRS315; *prSHS1::dCas9::NLS::Mxi1::ADH1(t)::Kan^R*) + pGF-V2102: (pRS423; *prSNR52::sgRNA(prCDC3-122)::SUP4(t)*)	This study
GFY-4219[l]	*MAT***a** *cdc3*Δ*::S.c.CDC3::mCherry::ADH1(t)::Kan^R* + pYY-57: (pRS316; *prSHS1::S.c.CDC3::GFP::ADH1(t):: Kan^R*) + pGF-IVL-1594: (pRS315; *prSHS1::dCas9::NLS::Mxi1::ADH1(t)::Kan^R*) + pGF-V2103: (pRS423; *prSNR52::sgRNA(prCDC3-75)::SUP4(t)*)	This study
GFY-4220[m]	*MAT***a** *cdc3*Δ*::S.c.CDC3::mCherry::ADH1(t)::Kan^R* + pYY-57: (pRS316; *prSHS1::S.c.CDC3::GFP::ADH1(t):: Kan^R*) + pGF-IVL-1594: (pRS315; *prSHS1::dCas9::NLS::Mxi1::ADH1(t)::Kan^R*) + pGF-V2104: (pRS423; *prSNR52::sgRNA(prCDC3-45)::SUP4(t)*)	This study
GFY-4221[n]	*MAT***a** *cdc3*Δ*::S.c.CDC3::mCherry::ADH1(t)::Kan^R* + pYY-57: (pRS316; *prSHS1::S.c.CDC3::GFP::ADH1(t):: Kan^R*) + pGF-IVL-1594: (pRS315; *prSHS1::dCas9::NLS::Mxi1::ADH1(t)::Kan^R*) + pGF-V2105: (pRS423; *prSNR52::sgRNA(prCDC3-16)::SUP4(t)*)	This study
GFY-4222	*MAT***a** *cdc3*Δ*::S.c.CDC3::mCherry::ADH1(t)::Kan^R* + pYY-57: (pRS316; *prSHS1::S.c.CDC3::GFP::ADH1(t)::Kan^R*) + pGF-IVL-1595:	This study

	(pRS315; *prCDC11::dCas9::NLS::Mxi1::ADH1(t)::Kan^R*) + pGF-V2100: (pRS423; *prSNR52::sgRNA(prCDC3-223)::SUP4(t)*)	
GFY-4223	*MAT***a** *cdc3*Δ*::S.c.CDC3::mCherry::ADH1(t)::Kan^R* + pYY-57: (pRS316; *prSHS1::S.c.CDC3::GFP::ADH1(t)::Kan^R*) + pGF-IVL-1595: (pRS315; *prCDC11::dCas9::NLS::Mxi1::ADH1(t)::Kan^R*) + pGF-V2101: (pRS423; *prSNR52::sgRNA(prCDC3-170)::SUP4(t)*)	This study
GFY-4224	*MAT***a** *cdc3*Δ*::S.c.CDC3::mCherry::ADH1(t)::Kan^R* + pYY-57: (pRS316; *prSHS1::S.c.CDC3::GFP::ADH1(t)::Kan^R*) + pGF-IVL-1595: (pRS315; *prCDC11::dCas9::NLS::Mxi1::ADH1(t)::Kan^R*) + pGF-V2102: (pRS423; *prSNR52::sgRNA(prCDC3-122)::SUP4(t)*)	This study
GFY-4225	*MAT***a** *cdc3*Δ*::S.c.CDC3::mCherry::ADH1(t)::Kan^R* + pYY-57: (pRS316; *prSHS1::S.c.CDC3::GFP::ADH1(t)::Kan^R*) + pGF-IVL-1595: (pRS315; *prCDC11::dCas9::NLS::Mxi1::ADH1(t)::Kan^R*) + pGF-V2103: (pRS423; *prSNR52::sgRNA(prCDC3-75)::SUP4(t)*)	This study
GFY-4226	*MAT***a** *cdc3*Δ*::S.c.CDC3::mCherry::ADH1(t)::Kan^R* + pYY-57: (pRS316; *prSHS1::S.c.CDC3::GFP::ADH1(t)::Kan^R*) + pGF-IVL-1595: (pRS315; *prCDC11::dCas9::NLS::Mxi1::ADH1(t)::Kan^R*) + pGF-V2104: (pRS423; *prSNR52::sgRNA(prCDC3-45)::SUP4(t)*)	This study
GFY-4227	*MAT***a** *cdc3*Δ*::S.c.CDC3::mCherry::ADH1(t)::Kan^R* +	This study

	pYY-57: (pRS316; *prSHS1::S.c.CDC3::GFP::ADH1(t)::Kan^R*) + pGF-IVL-1595: (pRS315; *prCDC11::dCas9::NLS::Mxi1::ADH1(t)::Kan^R*) + pGF-V2105: (pRS423; *prSNR52::sgRNA(prCDC3-16)::SUP4(t)*)	
GFY-4814[o]	*MAT***a** *his3*Δ*::prHIS3:prHED1::SpHIS5::HIS3(t)* + pRS425 (empty) pGF-IVL-1596: (pRS316; *prSHS1::dCas9::NLS::VPR::ADH1(t)::Kan^R*)	
GFY-4817[p]	*MAT***a** *his3*Δ*::prHIS3:prHED1::SpHIS5::HIS3(t)* + pGF-IVL-1596: (pRS316; *prSHS1::dCas9::NLS::VPR::ADH1(t)::Kan^R*) + pGF-V2042: (pRS425; *prSNR52::sgRNA(prHED1)::SUP4(t)*)	
GFY-4820	*MAT***a** *his3*Δ*::prHIS3:prHED1::SpHIS5::HIS3(t)* + pRS425 (empty) + pRS316 (empty)	
GFY-4823	*MAT***a** *his3*Δ*::prHIS3:prHED1::SpHIS5::HIS3(t)* + pRS316 (empty) + pGF-V2042: (pRS425; *prSNR52::sgRNA(prHED1)::SUP4(t)*)	
GFY-4826[o]	*MAT***a** *his3*Δ*::prHIS3:prGAL7::SpHIS5::HIS3(t)* + pRS316 (empty) + pRS425 (empty)	
GFY-4829	*MAT***a** *his3*Δ*::prHIS3:prGAL7::SpHIS5::HIS3(t)* + pRS425 (empty) pGF-IVL-1596: (pRS316; *prSHS1::dCas9::NLS::VPR::ADH1(t)::Kan^R*)	
GFY-4832[q]	*MAT***a** *his3*Δ*::prHIS3:prGAL7::SpHIS5::HIS3(t)* + pRS316 (empty) + pGF-V2043: (pRS425; *prSNR52::sgRNA(prGAL7)::SUP4(t)*)	
GFY-4835	*MAT***a** *his3*Δ*::prHIS3:prGAL7::SpHIS5::HIS3(t)* + pGF-IVL-1596: (pRS316; *prSHS1::dCas9::NLS::VPR::ADH1(t)::Kan^R*) + pGF-V2043: (pRS425; *prSNR52::sgRNA(prGAL7)::SUP4(t)*)	
GFY-1192	*his3*Δ*::prHIS3:(u2)::prGAL1/10::SpCas9::NLS::ADH1(t):prMX::*	

	Kan^R::MX(t)::(u2)::HIS3(t)	

[a]The dCas9 was created from *S. pyogenes*. It contains two mutations D10A/H840A, which makes Cas9 lose its nuclease activity. dCas9 fused to a nuclear localization sequence (SV40 NLS). Mxi1 domain sequence is: MERVRMINVQRLLEAAEFLERRERECEHGYASSFPSMPSPRG[37].

[b]The sgRNA(*prCDC11*-212) sequence is 5' AGATAAATATAGCTTAAATA 3'. All sgRNA constructs include 269 bp of the SNR52 promoter, the 79 bp tracrRNA, and the 20 bp SUP4 terminator. PAM sequence 5'-NGG-3' for each sgRNA are present on the coding strand of all septin loci[3].
[c]The sgRNA(*prCDC11*-184) sequence is 5' TTTACACAACACAATGCTCG 3'.
[d]The sgRNA(*prCDC11*-163) sequence is 5' GGGTCGTTACCCGACCATGC 3'.
[e]The sgRNA(*prCDC11*-82) sequence is 5' ATTTGATCGAAAAGTAAAAT 3'.
[f]The sgRNA(*prCDC11*-65) sequence is 5' AATAGGTAGACACCACGTAT 3'.
[g]The sgRNA(*prCDC11*-52) sequence is 5' CACGTATTGGCGACCCGATC 3'.
[h]The sgRNA(*prCDC11*-23) sequence is 5' TTAGAAAGTCAATCATCACA 3'.
[i]The sgRNA(*prCDC3*-223) sequence is 5' GTAAAAAGGAACCACCTGGG 3'.
[j]The sgRNA(*prCDC3*-170) sequence is 5' ACGTGTGTAATGATATGGTG 3'.
[k]The sgRNA(*prCDC3*-122) sequence is 5' CCATTTTTTGTCTCGGTTTT 3'.
[l]The sgRNA(*prCDC3*-75) sequence is 5' TATTAAGTTAAAATCTCTCG 3'.
[m]The sgRNA(*prCDC3*-45) sequence is 5' ATCAAGAAAATATCAAATCC 3'.
[n]The sgRNA(*prCDC3*-16) sequence is 5' CGACAACTGAACGATTACAT 3'.
[o]Strain GFY-4814 and GFY-4826 were derived from GFY-1192. First, *S. pyogenes* Cas9 was activated by adding galactose and then followed by transforming sgRNA(u2) plasmid and the amplified target PCR fragment: *prHIS3::prHED1::SpHIS5::HIS3(t)* (GFY-4814), and *prHIS3:prGAL7::SpHIS5::HIS3(t)* (GFY-4826), respectively. During this activation process, Cas9 made DSB on these two u2 artificial sites and the entire Cas9-Kan cassette was replaced by the target PCR fragment. Then, Yeast were plated on SD-LEU medium to select for guide RNA containing cells.
[p]The sgRNA(*prHED1*) sequence is 5' ACGGCTTTAATTAGCGTACG 3'[18].
VPR domain sequence is:
DALDDFDLDMLGSDALDDFDLDMLGSDALDDFDLDMLGSDALDDFDLDMLINSRSSGS
PKKKRKVGSQYLPDTDDRHRIEEKRKRTYETFKSIMKKSPFSGPTDPRPPPRRIAVPSRSS
ASVPKPAPQPYPFTSSLSTINYDEFPTMVFPSGQISQASALAPAPPQVLPQAPAPAPAPAM
VSALAQAPAPVPVLAPGPPQAVAPPAPKPTQAGEGTLSEALLQLQFDDEDLGALLGNST
DPAVFTDLASVDNSEFQQLLNQGIPVAPHTTEPMLMEYPEAITRLVTGAQRPPDPAPAPL
GAPGLPNGLLSGDEDFSSIADMDFSALLGSGSGSRDSREGMFLPKPEAGSAISDVFEGRE
VCQPKRIRPFHPPGSPWANRPLPASLAPTPTGPVHEPVGSLTPAPVPQPLDPAPAVTPEAS
HLLEDPDEETSQAVKALREMADTVIPQKEEAAICGQMDLSHPPPRGHLDELTTTLESMT
EDLNLDSPLTPELNEILDTFLNDECLLHAMHISTGLSIFDTSLF (Maselko *et al*., 2015)
[q]The sgRNA(prGAL7) sequence is 5' AACTGTTGACCGTGATCCGA 3'[18].

References

1 Cho, S. W., Kim, S., Kim, J. M. & Kim, J.-S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nature biotechnology* **31**, 230-232 (2013).

2 Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823 (2013).

3 DiCarlo, J. E. *et al.* Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic acids research* **41**, 4336-4343 (2013).

4 Gratz, S. J. *et al.* Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. *Genetics* **194**, 1029-1035 (2013).

5 Hwang, W. Y. *et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature biotechnology* **31**, 227-229 (2013).

6 Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature biotechnology* **31**, 233-239 (2013).

7 Jinek, M. *et al.* A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *science* **337**, 816-821 (2012).

8 Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823-826 (2013).

9 Wang, H. *et al.* One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *cell* **153**, 910-918 (2013).

10 Barrangou, R. The roles of CRISPR–Cas systems in adaptive immunity and beyond. *Current opinion in immunology* **32**, 36-41 (2015).

11 Gilbert, L. A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442-451, doi:10.1016/j.cell.2013.06.044 (2013).

12 Park, J.-J., Dempewolf, E., Zhang, W. & Wang, Z.-Y. RNA-guided transcriptional activation via CRISPR/dCas9 mimics overexpression phenotypes in Arabidopsis. *PLoS One* **12**, e0179410 (2017).

13 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).

14 Kim, Y. B. *et al.* Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nature biotechnology* **35**, 371-376 (2017).

15 Haldeman, J. M. *et al.* Creation of versatile cloning platforms for transgene expression and dCas9-based epigenome editing. *Nucleic acids research* **47**, e23-e23 (2019).

16 Nelles, D. A. *et al.* Programmable RNA tracking in live cells with CRISPR/Cas9. *Cell* **165**, 488-496 (2016).

17 Perez-Pinera, P. *et al.* RNA-guided gene activation by CRISPR-Cas9–based transcription factors. *Nature methods* **10**, 973-976 (2013).

18 Chavez, A. *et al.* Highly efficient Cas9-mediated transcriptional programming. *Nature methods* **12**, 326-328 (2015).

19 Deaner, M., Mejia, J. & Alper, H. S. Enabling graded and large-scale multiplex of desired genes using a dual-mode dCas9 activator in Saccharomyces cerevisiae. *ACS synthetic biology* **6**, 1931-1943 (2017).

20 Kong, Y.-J., Joo, J.-H., Kim, K. P. & Hong, S. Hed1 promotes meiotic crossover formation in Saccharomyces cerevisiae. *Journal of Microbiology and Biotechnology* **27**, 405-411 (2017).

21 Busygina, V. *et al.* Hed1 regulates Rad51-mediated recombination via a novel mechanism. *Genes & development* **22**, 786-795 (2008).

22 Tsubouchi, H. & Roeder, G. S. Budding yeast Hed1 down-regulates the mitotic recombination machinery when meiotic recombination is impaired. *Genes & development* **20**, 1766-1775 (2006).

23 Greger, I. H., Aranda, A. & Proudfoot, N. Balancing transcriptional interference and initiation on the GAL7 promoter of Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences* **97**, 8415-8420 (2000).

24 Schreiber-Agus, N. & DePinho, R. A. Repression by the Mad (Mxi1)-Sin3 complex. *Bioessays* **20**, 808-818 (1998).

25 Russell, D. W. & Sambrook, J. *Molecular cloning: a laboratory manual*. Vol. 1 (Cold Spring Harbor Laboratory Cold Spring Harbor, NY, 2001).

26 Finnigan, G. C. & Thorner, J. Complex in vivo ligation using homologous recombination and high-efficiency plasmid rescue from Saccharomyces cerevisiae. *Bio-protocol* **5** (2015).

27 Eckert-Boulet, N., Pedersen, M. L., Krogh, B. O. & Lisby, M. Optimization of ordered plasmid assembly by gap repair in Saccharomyces cerevisiae. *Yeast* **29**, 323-334 (2012).

28 Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173-1183 (2013).

29 Jensen, E. D. *et al.* Transcriptional reprogramming in yeast using dCas9 and combinatorial gRNA strategies. *Microbial cell factories* **16**, 1-16 (2017).

30 Smith, J. D. *et al.* Quantitative CRISPR interference screens in yeast identify chemical-genetic interactions and new rules for guide RNA design. *Genome biology* **17**, 1-16 (2016).

31 Boeke, J. D., LaCroute, F. & Fink, G. R. A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. *Molecular & general genetics : MGG* **197**, 345-346 (1984).

32 Cao, J. *et al.* An easy and efficient inducible CRISPR/Cas9 platform with improved specificity for multiple gene targeting. *Nucleic acids research* **44**, e149-e149 (2016).

33 Zuo, E. *et al.* One-step generation of complete gene knockout mice and monkeys by CRISPR/Cas9-mediated gene editing with multiple sgRNAs. *Cell research* **27**, 933-945 (2017).

34 Costanzo, M. *et al.* Global genetic networks and the genotype-to-phenotype relationship. *Cell* **177**, 85-100 (2019).

Chapter 5 - Conclusion of CRISPR/Cas9 technology

The work described in this dissertation aimed to investigate the following questions about CRISPR based gene drive systems: i) how to develop a multi-locus gene drive system (Chapter 2), ii) how to enhance gene drive efficiency and reduce gene drive resistance (Chapters 2&3), and iii) how CRISPR/dCas9 systems regulate gene transcription (Chapter 4). I will discuss three fields that extend beyond my current thesis, including the utilization of anti-CRISPR proteins, CRISPR titration technologies and mathematical models of gene drives.

Control of CRISPR gene drive using anti-CRISPR

Anti-CRISPR indicates a group of proteins that can inhibit the function of active CRISPR/Cas systems. The first anti-CRISPR protein was originally discovered in closely related phages that infect *Pseudomonas aeruginosa* in 2013[1]. The results demonstrated that when a phage carried an anti-CRISPR gene (acr), it could live and reproduce when it would otherwise be killed by the host's CRISPR/Cas system. To date, 45 anti-CRISPR proteins (Acr) have been identified against Class 1 (24), and Class 2 CRISPR/Cas systems (21), each having its own structure and no substantial sequence similarities[2-4]. Previous studies demonstrated that these Acr proteins inhibited CRISPR/Cas function by different mechanisms, including blocking DNA binding, DNA cleavage, or preventing sgRNA loading (Table 5.1). For instance, the type I-F CRISPR/Cas system can be inhibited by AcrF1, AcrF2 and AcrF3 proteins[5]. AcrF1 and AcrF2 function by directly interacting with the type I-F Cascade complex, resulting in the blocking of DNA binding. By contrast, the AcrF3 was demonstrated to block the target DNA cleavage by preventing the recruitment of Cas3 helicase–nuclease effector to the Cascade complex[6]. Also, both AcrIIA4 and AcrIIC3 could inhibit the function of type II CRISPR/Cas9 systems by interfering with the DNA binding process.

AcrIIA4 functions by mimicking a double-stranded DNA target, whereas AcrIIC3 functions by inducing the dimerization of Cas9 nucleases[6]. Moreover, another anti-CRISPR protein AcrIIC1 could block the target DNA cleavage by binding to the nuclease domain (HNH) of Cas9[7]. Although functions of some Acr proteins remain unidentified, it is still clear that CRISPR/Cas systems can be inhibited by Acr proteins with different strategies.

Table 5.1. Summary of mechanisms for anti-CRISPR (Acr) proteins.

Type	I			II				V
Mechanism	DNA binding	DNA Cleavage	Unclear	DNA binding	DNA Cleavage	Guide loading	Unclear	DNA binding
Acr Name	AcrIF1 AcrIF2 AcrIF4 AcrIF10 AcrID1	AcrIF3	AcrIE1 AcrIC1 AcrIE2–7 AcrIF5–9 AcrIF11–14	AcrIIA2 AcrIIA4 AcrIIA6 AcrIIC3 AcrIIC4 AcrIIC5	AcrIIC1 AcrIIA11	AcrIIC2	AcrIIA1 AcrIIA3 AcrIIA5 AcrIIA7–10	AcrVA1 AcrVA4 AcrVA5

Table adapted from Marino *et al.* 2020[8].

Expect for blocking the function of CRISPR/Cas systems, previous studies demonstrated that Acr proteins could also serve as critical molecules in a variety of applications. First, Acr proteins can be used to reduce toxic effects generated by expression of the Cas9 nuclease. Previous studies demonstrated that Cas9 toxicity had been observed in certain organisms as the use of Cas9 has increased, like mycobacteria[9], *Escherichia coli*[10], and *Synechococcus elongatus*[11]. Acr proteins could help reduce the toxicity by inhibiting the occurrence of undesired genome editing events. Second, Acr proteins can be used to develop a CRISPRi system by blocking the endonuclease activity of Cas9. For instance, AcrIIC1 disabled Cas9 activity by binding to the HNH domain, but it still allowed Cas9 to bind DNA[7]. Third, Acr proteins can also prevent the transcriptional regulation mediated by CRISPR/dCas9 through blocking its DNA binding process[12]. Taken

together, anti-CRISPR proteins were widely applied in different organisms for a variety of reasons related to genome engineering.

In 2018, our lab demonstrated that Acr proteins could serve as inhibitors of gene drive activity in budding yeast[13], and this was the first study to illustrate the effects of Acr proteins in a CRISPR gene drive system. We found that AcrIIA2 and AcrIIA4 proteins could reduce gene drive efficiency at different levels. To investigate which residues were critical for this inhibitory function, we deleted and mutated groups of residues within AcrIIA2 and AcrIIA4 proteins. Thus, our study illustrated that we could artificially titrate the gene drive efficiency by mutating different residues of Acr proteins.

However, no one has investigated if Acr proteins could collaborate with anti-drives to destroy the activity of gene drive systems. To solve this problem, one could develop an artificial "anti-drive" system by including Acr proteins within the CRISPR/Cas12a construct (Figure 5.1). Like we mentioned previously, in a traditional gene drive system (Figure 5.1, left), the gene drive (containing Cas9) allele will cleave the wild-type allele and produce a gene drive/gene drive homozygote. Another drive system (anti-drive) could attack the original Cas9 containing gene drive cassette and prohibit the spread of gene drive elements (Figure 5.1, right). First, the anti-drive system would contain i) a sgRNA (not shown in figure), ii) a Cas12a nuclease to serve as the anti-drive to edit the Cas9 containing drive elements, iii) a Acr protein (e.g. AcrIIA2/A4) that functions to inhibit the Cas9 nuclease, and iv) a DNA target (Cas9) that contains a T-rich (5'-TTN-3') PAM site. This anti-drive fragment would be constructed at the same locus as the original drive cassette. Next, the anti-drive (Cas12a) would cleave the original Cas9 gene drive and serve as a donor DNA to repair the DSB by using HDR. Thus, this would result in an anti-drive/anti-drive homozygote and destroy the gene drive propagation. We will include an *Acr* gene next to Cas12a,

in order to inhibit the activity of the Cas9 nuclease. Taken together, this design is much safer since i) it only attacks the Cas9 nuclease DNA, and ii) it contains redundant inhibitory mechanisms to destroy the original gene drive activity.

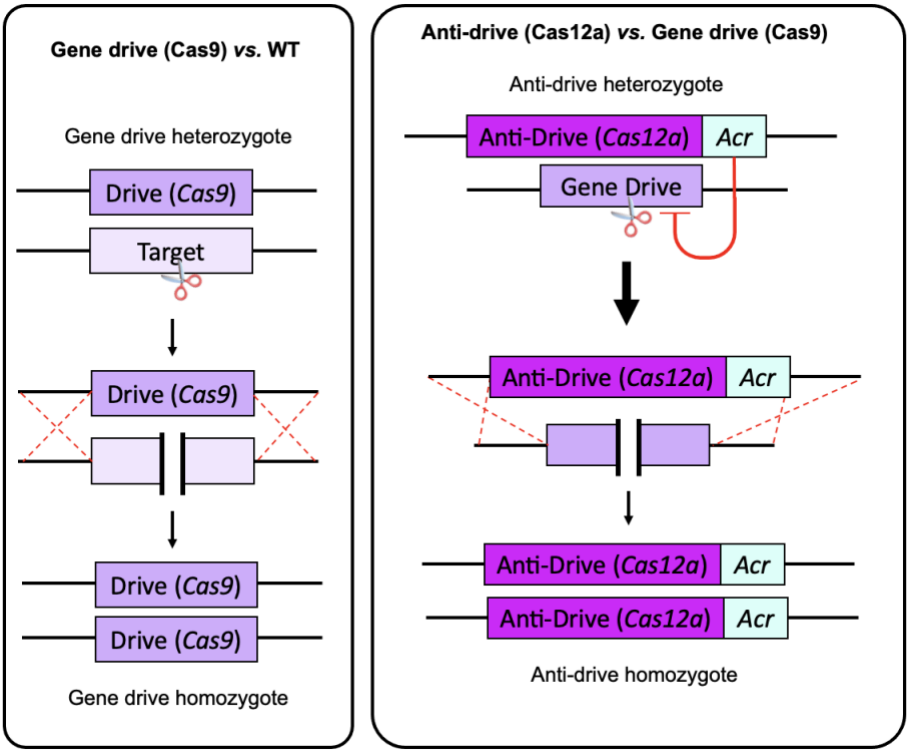


Figure 5.1. Design of an anti-CRISPR containing anti-drive system.

Left, the traditional gene drive design. *Right*, design of an anti-drive system. The Cas9 containing gene drive allele will be cleaved by CRISPR/Cas12a based anti-drive system. Then, the entire anti-drive (including *Acr*) will function as a donor DNA to fix the DSB by HDR. The Acr protein will inhibit the activity of the Cas9 nuclease. Consequently, the original gene drive will be destroyed.

How to titrate CRISPR gene drive activity

As mentioned previously, our lab could titrate the activity of CRISPR/Cas systems by controlling the Cas9 nuclease with Acr proteins (wt/mutants). Other studies have demonstrated that the CRISPR/Cas9 activity also can be titrated by other strategies, including the utilization of a CRISPR guide RNA pool[14], or introduction of mutations to target DNA sequences. Hsu *et al.* constructed a sgRNA pool that contained all possible single base-pair substitutions across the

crRNA. Their results demonstrated that the activity of Cas9 nuclease changed when combined with sgRNA containing different mutations[15]. Additionally, they mutated all possible target DNA sequences within 4 different target sites, and they got the similar results that Cas9 activity could be titrated by introducing mutations to target DNA sequence[15]. Scientists have developed systems to titrate gene expression by using libraries of CRISPR sgRNAs[16,17]. They developed an approach to precisely titrate expression of human genes by using CRISPR/dCas9 transcriptional regulation system with series of sgRNAs[16]. Their design enabled them to control the expression of ~2,400 genes in the human genome, which could potentially be helpful in identifying critical genes for diseases. However, all these studies focused on how the alteration of sgRNA affected the Cas9 activity, and more investigations are required to examine the properties of the Cas9 nuclease.

Our lab developed another strategy to titrate the efficiency of CRISPR gene drive systems in budding yeast[14]. We tested whether the nuclear localization of Cas9 nuclease affected the Cas9 editing efficiency. We constructed fragments that contained (multiple) NLS and/or NES at either C-terminal or N-terminal positions of Cas9. Our results showed that application of multiple NLS, direct fusion of an NES signal, or utilization of both NLS and NES around Cas9 greatly reduced gene drive activity, suggesting that direct competition between nuclear export and import can titrate the level of CRISPR gene drive editing[14].

In the future, we could design a "double nuclease" gene drive system that could be capable of both propagating the drive cassette into a population and being able to titrate the drive activity when not necessary or appropriate. Use of this system for genomic manipulation and nuclease titration requires: i) a Cas12a nuclease to serve as an active drive, ii) a dCas9 nuclease tagged with the Mxi1 repressive domain to function as a transcriptional repressor to titrate the activity of the drive nuclease when needed, iii) a sgRNA for Cas12a to edit DNA target, and another sgRNA for

dCas9 to target the promoter for Cas12a, and iv) a DNA target that contains 5'-TTN-3' PAM site at the 5' end for Cas12a's activity, and a 5'-NGG-3' PAM site at the 3' end for dCas9 transcriptional manipulation. In this system, the CRISPR/Cas12a based drive cassette will be under the control of an inducible promoter, followed by the expression of a sgRNAs (Cas12a). The dCas9-Mxi1 construct and its corresponding sgRNA will be driven by another inducible promoter at a separate locus, and it will be inhibited without addition of the external stimulus. When needed, the dCas9-Mxi1 fusion would be targeted to inhibit or down regulate the expression of Cas12a. This would allow for an externally controlled gene drive system.

Utilization of mathematical models in gene drive analysis

CRISPR based gene drive technology has already been developed successfully in yeast[18], mosquitoes, and fruit flies[19,20]. Application of gene drive mosquitoes could potentially save lives by eliminating the spread of malaria. To date, multiple gene drive approaches have been developed successfully under laboratory conditions. However, it is challenging to understand how gene drive populations will behave in real world settings outside of the laboratory. Thus, one method used to study population dynamics is by mathematical modeling[21]. In this way, various parameters and variables can be tested in simulations to understand their effects on gene drive spread and success.

Previous studies developed an advanced gene drive system, which could decrease the occurrence of gene drive resistance induced by NHEJ repair (failed DNA cleavage)[22,23]. Their approach could increase the drive cassette inheritance by cleaving the target multiple times with multiple sgRNAs. In that system, they constructed a Cas9 gene and 5 sgRNAs on one chromosome. The mathematical simulations of this new approach illustrated that i) in traditional gene drive systems (with 1 sgRNA), the occurrence frequency of HDR is above 95% for the first

200 generations but decreased dramatically when NHEJ occurred, and ii) in the proposed drive system (with 5 sgRNAs), the HDR probabilities keep above 95% for nearly 600 generations. Their mathematical prediction suggested this new approach could significantly improve the evolutionary stability of gene drives.

Studies listed above could potentially increase the drive efficiency. But there are still challenged that are not resolved by these simulations. For instance, Noble *et al.* demonstrated that utilization of multiple sgRNA could delay the occurrence of gene drive resistance[23], but it remains unclear how to eliminate the drive resistance that was induced by the single sgRNA/DNA target mismatch. Most importantly, it may be necessary to create drive systems that have the ability to either self-titrate or self-inhibit. In 2019, our lab developed a mathematical model testing whether induction of expression of a self-cleaving guide RNA could suppress drive propagation[24].

The target DNA/sgRNA mismatch is always a concern for the emergence of drive resistance. A gene drive system could be constructed with a pool of sgRNAs that contains all possible single nucleotide substitutions, where all sgRNAs are derived from an identical crRNA. Then, a mathematical model could be used to simulate if these mutated sgRNAs would help delay or prevent the occurrence of drive resistance alleles. In this design, each gene drive cassette would contain a Cas9 nuclease and multiple sgRNAs, including a *wt* sgRNA and many mutated sgRNA variants. In the model, one could simulate whether physical positioning or temporal release of this libraries of gene drives would assist and maximize the prevention of resistant allele formation. Other types of theoretical gene drive architectures could be developed to study challenges that may exist in real environments.

To date, a safe, yet highly effective gene drive system should be developed in order to satisfy all potential concerns for scientists, governments, and the general public. We envision

future gene drive architectures to include numerous safeguard mechanisms. These might include self-destruction drives, competing drives, or use of anti-CRISPRs for total inhibition. Also, choice of target DNA sequences that are specific to the species of choice would aid in both drive efficiency and prevention of drive cross over to neighboring species or sub-species. Finally, future studies on the occurrence of drive resistant alleles (and how to control them) will be required.

In conclusion, the research progress on CRISPR/Cas9 based gene drives in laboratories around the world has provided a promising, exciting, and powerful new genetic tool that could one day be used to combat human disease, invasive species, and protect agriculture.

References

1 Bondy-Denomy, J., Pawluk, A., Maxwell, K. L. & Davidson, A. R. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**, 429-432 (2013).

2 Trasanidou, D. *et al.* Keeping crispr in check: diverse mechanisms of phage-encoded anti-crisprs. *FEMS microbiology letters* **366**, fnz098 (2019).

3 Knott, G. J. *et al.* Broad-spectrum enzymatic inhibition of CRISPR-Cas12a. *Nature structural & molecular biology* **26**, 315-321 (2019).

4 Dong, L. *et al.* An anti-CRISPR protein disables type V Cas12a by acetylation. *Nature Structural & Molecular Biology* **26**, 308-314 (2019).

5 Bondy-Denomy, J. *et al.* Multiple mechanisms for CRISPR–Cas inhibition by anti-CRISPR proteins. *Nature* **526**, 136-139 (2015).

6 Liu, Z., Dong, H., Cui, Y., Cong, L. & Zhang, D. Application of different types of CRISPR/Cas-based systems in bacteria. *Microbial cell factories* **19**, 1-14 (2020).

7 Harrington, L. B. *et al.* A broad-spectrum inhibitor of CRISPR-Cas9. *Cell* **170**, 1224-1233. e1215 (2017).

8 Marino, N. D., Pinilla-Redondo, R., Csörgő, B. & Bondy-Denomy, J. Anti-CRISPR protein applications: natural brakes for CRISPR-Cas technologies. *Nature methods* **17**, 471-479 (2020).

9 Rock, J. M. *et al.* Programmable transcriptional repression in mycobacteria using an orthogonal CRISPR interference platform. *Nature microbiology* **2**, 1-9 (2017).

10 Cho, S. *et al.* High-level dCas9 expression induces abnormal cell morphology in Escherichia coli. *ACS synthetic biology* **7**, 1085-1094 (2018).

11 Wendt, K. E., Ungerer, J., Cobb, R. E., Zhao, H. & Pakrasi, H. B. CRISPR/Cas9 mediated targeted mutagenesis of the fast growing cyanobacterium Synechococcus elongatus UTEX 2973. *Microbial cell factories* **15**, 1-8 (2016).

12 Lee, J. *et al.* Tissue-restricted genome editing in vivo specified by microRNA-repressible anti-CRISPR proteins. *Rna* **25**, 1421-1431 (2019).

13 Basgall, E. M. *et al.* Gene drive inhibition by the anti-CRISPR proteins AcrIIA2 and AcrIIA4 in Saccharomyces cerevisiae. *Microbiology* **164**, 464 (2018).

14 Roggenkamp, E. *et al.* Tuning CRISPR-Cas9 gene drives in Saccharomyces cerevisiae. *G3: Genes, Genomes, Genetics* **8**, 999-1018 (2018).

15 Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology* **31**, 827-832 (2013).

16 Jost, M. *et al.* Titrating gene expression using libraries of systematically attenuated CRISPR guide RNAs. *Nature biotechnology* **38**, 355-364 (2020).

17 Bowman, E. K. *et al.* Bidirectional titration of yeast gene expression using a pooled CRISPR guide RNA approach. *Proceedings of the National Academy of Sciences* **117**, 18424-18430 (2020).

18 Lewis, I. C., Yan, Y. & Finnigan, G. C. Analysis of a Cas12a-based gene-drive system in budding yeast. *Access microbiology* **3** (2021).

19 Tong, X. L. *et al.* [Applications of the CRISPR/Cas9 system in insects]. *Yi Chuan* **40**, 266-278, doi:10.16288/j.yczz.17-263 (2018).

20 Chen, L., Wang, G., Zhu, Y. N., Xiang, H. & Wang, W. Advances and perspectives in the application of CRISPR/Cas9 in insects. *Dongwuxue Yanjiu* **37**, 220-228, doi:10.13918/j.issn.2095-8137.2016.4.220 (2016).

21 Eckhoff, P. A., Wenger, E. A., Godfray, H. C. J. & Burt, A. Impact of mosquito gene drive on malaria elimination in a computational model with explicit spatial and temporal dynamics. *Proceedings of the National Academy of Sciences* **114**, E255-E264 (2017).

22 Esvelt, K. M., Smidler, A. L., Catteruccia, F. & Church, G. M. Emerging technology: concerning RNA-guided gene drives for the alteration of wild populations. *elife* **3**, e03401 (2014).

23 Noble, C., Olejarz, J., Esvelt, K. M., Church, G. M. & Nowak, M. A. Evolutionary dynamics of CRISPR gene drives. *Science advances* **3**, e1601964 (2017).

24 Heffel, M. G. & Finnigan, G. C. Mathematical modeling of self-contained CRISPR gene drive reversal systems. *Scientific reports* **9**, 1-10 (2019).

Appendix A - Copyright Permissions

We hold the copyright for our publications.

We retain copyright and do not need to obtain permission from the publisher to use the material in my ETDR.

Appendix B - Supplementary Information for Development of a multi-locus CRISPR gene drive system in budding

Yao Yan[1] and Gregory C. Finnigan[1*]

[1]Department of Biochemistry and Molecular Biophysics, Kansas State University, 141 Chalmers Hall, Manhattan, KS 66506 USA

*Correspondence to: Gregory C. Finnigan, Phone: (785) 532-6939; FAX; (785) 532-7278; E-mail: gfinnigan@ksu.edu

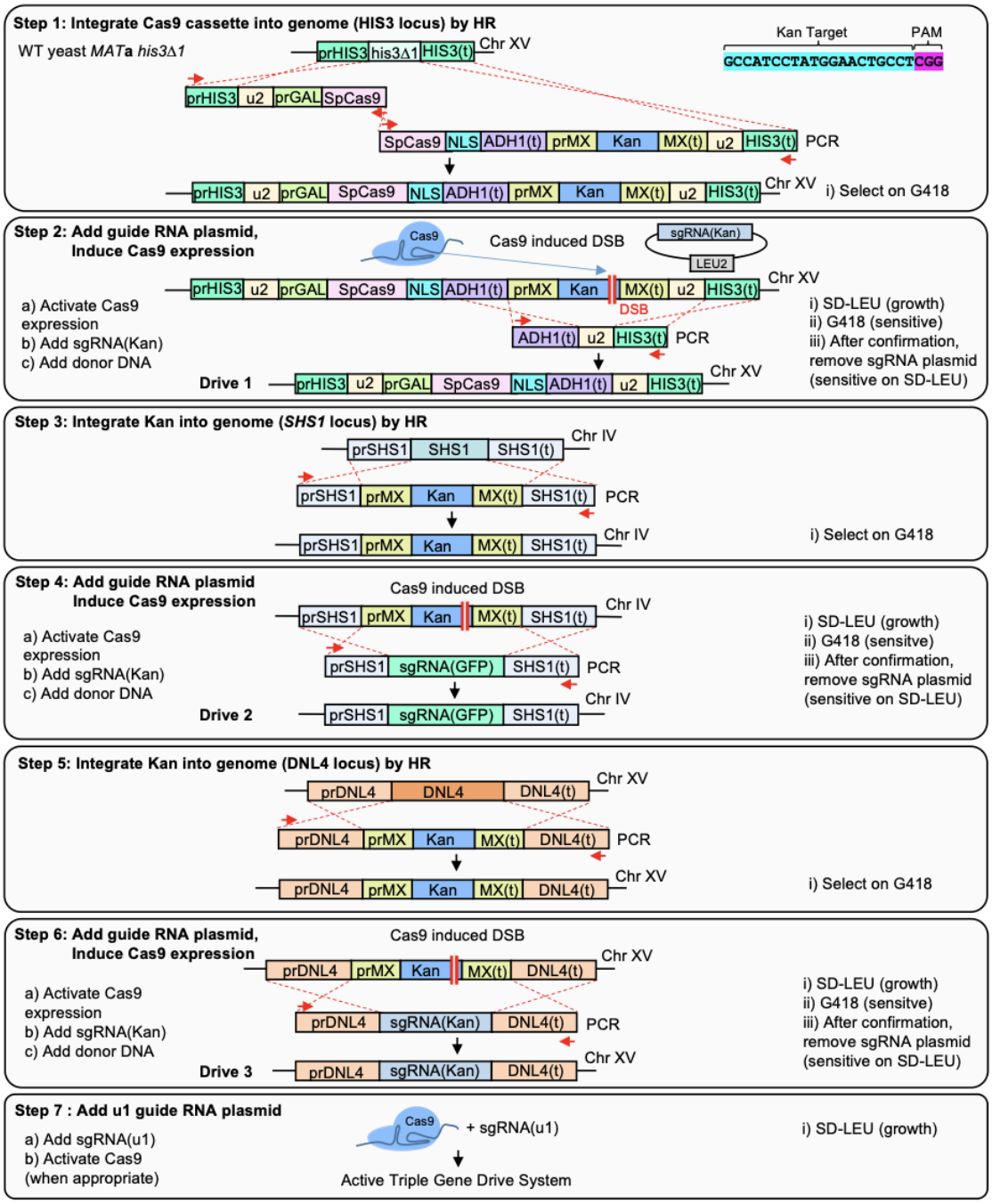


Figure B.S1. Methodology for construction of a CRISPR gene drive system across three loci.

This protocol is performed in haploid yeast. The final haploid strain harbors (inactive) Cas9 when grown on dextrose, and all three guide RNA cassettes. This method allows for a universal targeting strategy (Kan^R marker) and does not require gene-specific guides to be cloned or purchased. Moreover, integration of the guide cassette in place of the Kan^R marker allows for rapid screening of viable isolates following Cas9 editing (sensitivity to G418 indicates loss of the Kan^R cassette).

Figure B.S2. DNA sequences of engineered gene drives, guide RNA cassettes, and modified *DNL4* expression alleles used in this study.

GFY-3675

*prHIS3::**(u2)**::prGAL1/10::**SpCas9**::**NLS**::ADH1(t)::**(u2)**::HIS3(t)*

(u2) 20 bp target and 3 bp PAM sequence

(992 bp 5′ UTR to *HIS3* shown)

GGGTCAGTTATTTCATCCAGATATAACCCGAGAGGAAACTTCTTAGCGTCTGTTTTCGTACCATAAGGCAGTTCATG
AGGTATATTTTCGTTATTGAAGCCCAGCTCGTGAATGCTTAATGCTGCTGAACTGGTGTCCATGTCGCCTAGGTACG
CAATCTCCACAGGCTGCAAAGGTTTTGTCTCAAGAGCAATGTTATTGTGCACCCCGTAATTGGTCAACAAGTTTAAT
CTGTGCTTGTCCACCAGCTCTGTCGTAACCTTCAGTTCATCGACTATCTGAAGAAATTTACTAGGAATAGTGCCATG
GTACAGCAACCGAGAATGGCAATTTCTACTCGGGTTCAGCAACGCTGCATAAACGCTGTTGGTGCCGTAGACATATT
CGAAGATAGGATTATCATTCATAAGTTTCAGAGCAATGTCCTTATTCTGGAACTTGGATTTATGGCTCTTTTGGTTT
AATTTCGCCTGATTCTTGATCTCCTTTAGCTTCTCGACGTGGGCCTTTTTCTTGCCATATGGATCCGCTGCACGGTC
CTGTTCCCTAGCATGTACGTGAGCGTATTTCCTTTTAAACCACGACGCTTTGTCTTCATTCAACGTTTCCCATTGTT
TTTTTCTACTATTGCTTTGCTGTGGGAAAAACTTATCGAAAGATGACGACTTTTTCTTAATTCTCGTTTTAAGAGCT
TGGTGAGCGCTAGGAGTCACTGCCAGGTATCGTTTGAACACGGCATTAGTCAGGGAAGTCATAACACAGTCCTTTCC
CGCAATTTTCTTTTTCTATTACTCTTGGCCTCCTCTAGTACACTCTATATTTTTTTATGCCTCGGTAATGATTTTCA
TTTTTTTTTTTCCACCTAGCGGATGACTCTTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTA
TATAAAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG**GCTGTTCGT**
GTGCGCGTCCTGGGGACAGGTTATCAGCAACAACACAGTCATATCCATTCTCAATTAGCTCTACCACAGTGTGTGAA
CCAATGTATCCAGCACCACCTGTAACCAAAACAATTTTAGAAGTACTTTCACTTTGTAACTGAGCTGTCATTTATAT
TGAATTTTCAAAAATTCTTACTTTTTTTTTGGATGGACGCAAAGAAGTTTAATAATCATATTACATGGCATTACCAC
CATATACATATCCATATACATATCCATATCTAATCTTACTTATATGTTGTGGAAATGTAAAGAGCCCCATTATCTTA
GCCTAAAAAAACCTTCTCTTTGGAACTTTCAGTAATACGCTTAACTGCTCATTGCTATATTGAAGTACGGATTAGAA
GCCGCCGAGCGGGTGACAGCCCTCCGAAGGAAGACTCTCCTCCGTGCGTCCTCGTCTTCACCGGTCGCGTTCCTGAA
ACGCAGATGTGCCTCGCGCCGCACTGCTCCGAACAATAAAGATTCTACAATACTAGCTTTTATGGTTATGAAGAGGA
AAAATTGGCAGTAACCTGGCCCCACAAACCTTCAAATGAACGAATCAAATTAACAACCATAGGATGATAATGCGATT
AGTTTTTTAGCCTTATTTCTGGGGTAATTAATCAGCGAAGCGATGATTTTTGATCTATTAACAGATATATAAATGCA
AAAACTGCATAACCACTTTAACTAATACTTTCAACATTTTCGGTTTGTATTACTTCTTATTCAAATGTAATAAAAGT
ATCAACAAAAAATTGTTAATATACCTCTATACTTTAACGTCAAGGAGAAAAAACTATA**ATG**GATAAGAAATACTCTA
TCGGTTTGGATATTGGTACAAATTCAGTTGGTTGGGCAGTTATTACTGATGAATACAAGGTTCCATCTAAAAAGTTT
AAAGTTTTGGGTAACACTGATAGACATTCTATTAAGAAAAATTTGATTGGTGCTTTGTTATTTGATTCTGGTGAAAC
TGCTGAAGCAACAAGATTGAAAAGAACTGCAAGAAGAAGATACACAAGAAGAAAGAATAGAATCTGTTATTTGCAAG
AAATTTTCTCTAACGAAATGGCTAAGGTTGATGATTCTTTCTTTCATAGATTGGAAGAATCATTTTTAGTTGAAGAA
GATAAGAAACATGAAAGACATCCAATCTTCGGTAACATCGTTGATGAAGTTGCTTACCATGAAAAGTACCCAACAAT
CTATCATTTGAGAAAGAAATTGGTTGATTCAACTGATAAGGCAGATTTGAGATTGATATATTTGGCTTTAGCACATA
TGATCAAGTTTAGAGGTCATTTCTTGATCGAGGGTGACTTGAATCCAGATAATTCTGATGTTGATAAGTTGTTTATT
CAATTAGTTCAAACATATAATCAATTGTTTGAAGAAAATCCAATTAATGCTTCTGGTGTTGATGCTAAGGCAATCTT
GTCAGCAAGATTGTCTAAGTCAAGAAGATTGGAAAATTTGATCGCTCAATTACCAGGTGAAAAGAAAAATGGTTTGT
TCGGTAATTTGATCGCATTGTCTTTGGGTTTGACACCAAACTTCAAGTCAAACTTCGATTTGGCTGAAGATGCAAAG
TTGCAATTGTCTAAGGATACTTACGATGATGATTTGGATAATTTGTTGGCTCAAATTGGTGACCAATATGCAGATTT
GTTTTTGGCTGCTAAAAATTTGTCTGATGCTATCTTGTTGTCAGATATCTTGAGAGTTAACACTGAAATCACAAAGG
CTCCATTGTCTGCATCAATGATCAAGAGATACGATGAACATCATCAAGATTTGACTTTGTTGAAGGCATTGGTTAGA
CAACAATTACCAGAAAAGTACAAGGAAATTTTCTTTGATCAATCTAAAAATGGTTATGCTGGTTACATTGATGGTGG
TGCATCTCAAGAAGAATTCTACAAGTTTATTAAGCCAATCTTGGAAAAGATGGATGGTACAGAAGAATTGTTAGTTA
AATTGAACAGAGAAGATTTGTTAAGAAAACAAAGAACTTTCGATAACGGTTCTATCCCACATCAAATCCATTTGGGT
GAATTACATGCTATCTTGAGAAGACAAGAAGATTTCTACCCATTTTTAAAGGATAACAGAGAAAAGATTGAAAAGAT
TTTGACTTTTAGAATTCCATATTACGTTGGTCCATTAGCTCGTGGTAATTCTAGATTTGCATGGATGACTAGAAAGT
CAGAAGAAACTATCACACCATGGAATTTTGAAGAAGTTGTTGATAAAGGTGCTTCTGCACAATCTTTTATTGAAAGA
ATGACAAACTTCGATAAAAATTTGCCAAACGAAAAGGTTTTGCCAAAGCATTCATTGTTATATGAATACTTTACTGT
TTACAATGAATTGACAAAAGTTAAATATGTTACTGAGGGTATGAGAAAACCAGCATTTTTGTCTGGTGAACAAAAGA

AAGCAATCGTTGATTTGTTGTTTAAAACTAACAGAAAGGTTACAGTTAAACAATTGAAAGAAGATTACTTTAAGAAA
ATTGAATGTTTTGATTCTGTTGAAATTTCAGGTGTTGAAGATAGATTCAATGCTTCATTAGGTACTTACCATGATTT
GTTGAAGATTATTAAGGATAAAGATTTCTTGGATAATGAAGAAAATGAAGATATTTTAGAAGATATTGTTTTAACTT
TGACATTATTTGAAGATAGAGAAATGATCGAAGAAAGATTGAAGACATACGCTCATTTGTTCGATGATAAAGTTATG
AAGCAATTGAAGAGAAGAAGATACACTGGTTGGGGTAGATTGTCTAGAAAGTTGATTAATGGTATCAGAGATAAGCA
ATCTGGTAAAACAATCTTGGATTTCTTGAAGTCAGATGGTTTCGCAAACAGAAACTTCATGCAATTGATTCATGATG
ATTCATTGACTTTTAAAGAAGATATCCAAAAAGCTCAAGTTTCTGGTCAGGGTGACTCATTGCATGAACATATTGCT
AATTTGGCAGGTTCTCCAGCTATTAAGAAAGGTATCTTGCAAACAGTTAAGGTTGTTGATGAATTAGTTAAAGTTAT
GGGTAGACATAAGCCAGAAAACATCGTTATCGAAATGGCTAGAGAAAACCAAACTACACAAAAGGGTCAAAAGAATT
CAAGAGAAAGAATGAAGAGAATCGAAGAAGGTATTAAAGAATTGGGTTCTCAAATCTTGAAGGAACATCCAGTTGAA
AACACTCAATTGCAAAACGAAAAGTTGTACTTATACTACTTACAAAACGGTAGAGATATGTACGTTGATCAAGAATT
AGATATCAACAGATTGTCAGATTACGATGTTGATCATATCGTTCCACAATCATTTTTGAAGGATGATTCAATCGATA
ATAAGGTTTTGACAAGATCTGATAAGAACCGTGGTAAATCTGATAATGTTCCATCAGAAGAAGTTGTTAAGAAAATG
AAGAACTACTGGAGACAATTGTTAAATGCTAAGTTGATCACTCAAAGAAAGTTCGATAATTTGACAAAAGCTGAAAG
AGGTGGTTTGTCAGAATTAGATAAAGCAGGTTTTATTAAGAGACAATTAGTTGAAACTAGACAAATCACAAAGCATG
TTGCACAAATCTTGGATTCTAGAATGAACACTAAATATGATGAAAATGATAAATTAATTAGAGAAGTTAAAGTTATT
ACATTAAAATCTAAATTGGTTTCAGATTTTAGAAAAGATTTTCAATTCTACAAAGTTAGAGAAATTAATAACTATCA
TCATGCTCATGATGCATACTTGAATGCTGTTGTTGGTACTGCATTGATTAAGAAATACCCAAAGTTGGAATCTGAAT
TCGTTTACGGTGACTACAAGGTTTACGATGTTAGAAAGATGATCGCTAAGTCAGAACAAGAAATCGGTAAAGCTACA
GCAAAGTATTTCTTTTATTCTAACATCATGAATTTCTTTAAAACTGAAATTACATTAGCTAACGGTGAAATCAGAAA
AAGACCATTGATCGAAACTAATGGTGAAACAGGTGAAATTGTTTGGGATAAAGGTAGAGATTTCGCAACTGTTAGAA
AGGTTTTGTCAATGCCACAAGTTAACATCGTTAAGAAAACTGAAGTTCAAACAGGTGGTTTTTCTAAGGAATCAATC
TTGCCAAAGAGAAACTCTGATAAGTTGATTGCTAGAAAGAAAGATTGGGATCCAAAGAAATATGGTGGTTTTGATTC
TCCAACTGTTGCTTACTCAGTTTTAGTTGTTGCAAAGGTTGAAAAGGGTAAATCTAAGAAATTGAAATCAGTTAAAG
AATTGTTAGGTATCACAATCATGGAAAGATCTTCATTCGAAAAGAATCCAATCGATTTCTTGGAAGCAAAGGGTTAC
AAGGAAGTTAAGAAAGATTTGATTATTAAGTTGCCAAAGTACTCTTTGTTCGAATTAGAAAACGGTAGAAAAAGAAT
GTTAGCTTCAGCTGGTGAATTGCAAAAGGGTAATGAATTGGCTTTGCCATCTAAGTACGTTAATTTCTTGTATTTGG
CATCTCATTACGAAAAGTTGAAGGGTTCACCAGAAGATAATGAACAAAAACAATTGTTCGTTGAACAACATAAGCAT
TATTTGGATGAAATTATTGAACAAATTTCTGAATTTTCAAAAAGAGTTATTTTGGCTGATGCAAATTTGGATAAGGT
TTTGTCTGCTTACAATAAGCATAGAGATAAGCCAATCAGAGAACAAGCAGAAAACATCATCCATTTGTTTACTTTGA
CAAATTTGGGTGCTCCAGCTGCTTTTAAATACTTCGATACTACAATCGATAGAAAAAGATACACTTCTACAAAGGAA
GTTTTGGATGCAACATTGATCCATCAATCAATCACTGGTTTGTATGAAACAAGAATTGATTTGTCTCAATTGGGTGG
TGACTCTAGGGCAGACCCAAAGAAAAAGAGGAAAGTATAAGGCGCGCCACTTCTAAATAAGCGAATTTCTTATGATT
TATGATTTTTATTATTAAATAAGTTATAAAAAAAATAAGTGTATACAAATTTTAAAGTGACTCTTAGGTTTTAAAAC
GAAAATTCTTATTCTTGAGTAACTCTTTCCTGTAGGTCAGGTTGCTTTCTCAGGTATAGTATGAGGTCGCTCTTATT
GACCACACCTCTACCGGCAGATCCGCTAGGGATAACAGGGTAATATGCTGTTCGTGTGCGCGTCCTGGGTGACACCG
ATTATTTAAAGCTGCAGCATACGATATATATACATGTGTATATATGTATACCTATGAATGTCAGTAAGTATGTATAC
GAACAGTATGATACTGAAGATGACAAGGTAATGCATCATTCTATACGTGTCATTCTGAACGAGGCGCGCTTTCCTTT
TTTCTTTTTGCTTTTTCTTTTTTTTTCTCTTGAACTCGAGAAAAAAAATATAAAAGAGATGGAGGAACGGGAAAAAG
TTAGTTGTGGTGATAGGTGGCAAGTGGTATTCCGTAAGAACAACAAGAAAAGCATTTCATATTATGGCTGAACTGAG
CGAACAAGTGCAAAATTTAAGCATCAACGACAACAACGAGAATGGTTATGTTCCTCCTCACTTAAGAGGAAAACCAA
GAAGTGCCAGAAATAACAGTAGCAACTACAATAACAACAACGGCGGCTACAACGGTGGCCGTGGCGGTGGCAGCTTC
TTTAGCAACAACCGTCGTGGTGGTTACGGCAACGGTGGTTTCTTCGGTGGAAACAACGGTGGCAGCAGATCTAACGG
CCGTTCTGGTGGTAGATGGATCGATGGCAAACATGTCCCAGCTCCAAGAAACGAAAAGGCCGAGATCGCCATATTTG
GTGTCCCCGAGGATCCAAATTTCCAATCTTCTGGTATTAACTTCGATAACTACGATGATATTCCAGTGGACGCCTCT
GGTAAGGATGTTCCTGAACCAATCACAGAATTTACCTCACCTCCATTGGACGGATTGTTATTGGAAAACATCAAATT
GGCCCGTTTCACCAAGCCAACACCTGTGCAAAAATACTCCGTCCCTATCGTTGCCAACGGCAGAGATTTGATGGCCT
GTGCGCAGACCGGTTCTGGTAAGACTGGTGGGTTTTTATTCCCAGTGTTGTCCGAATCATTTAAGACTGGACCATCT
CCTCAACCAGAGTCTCAAGGCTCCTTTTACCAAAGAAAGGCCTACCCAACTGCTGTCATTA

(993 bp 3′ UTR to *HIS3* shown)

prSHS1::sgRNA(GFP)::SHS1(t)

prSNR52::**crRNA**::*tracrRNA*::***SUP4(t)***

(596 bp 5′ UTR to *SHS1* shown)

GGTTATCGTATTTCACTTTTTGTGGTAAACTCAGGTCAATATTATGCGACTTGAACCATTCAGTCAAAGGTTCGTTA
CCACCTTTTTCCATACGAAGAAGTTCCTCCGGTTTAAACTGATCCATAGTGATAGATCTTACAAAAGATATATGCAC
ACCAAGCCCTCTATGGATACCGGCACATTCAAGGCAAATGAAAGCTCCAAACTTAGGCGTGGCCCATTGTGGATTTG
GCGCACCACAATCCATACATTTCTTATTTGCACCAATCTTTTGCAATTGCAAAAGACGCCTGCGGGTATCTGGGTCC
ACTTTCCAATCTGACATGCTCTATAATCCGCGATAAAATTGCTCAATTGGCACCATTTAAACTCAGAATCACGTCCA
TATTTCTGCTTTCATTCTTGATATAGTTGTGCAATTTGGTTCTTGACAAAAACTGGCGTTGCTTGCGGGTAACCGCG
CGATTTTTAAAGTGCCAAACTGCGAAAAGAATATAACAAGCTTTCGAGCAAGATCAATGTACCCAGCAAGTGAAATA
ATAAAACAAGAGCCCCAAAGATCTGCTTATAATTGCTAGAAAAATATATTATTAATC**GGATCC**TCACTAA**AGGGAAC**
AAAAGCTGGAGCTTCTTTGAAAAGATAATGTATGATTATGCTTTCACTCATATTTATACAGAAACTTGATGTTTTCT
TTCGAGTATATACAAGGTGATTACATGTACGTTTGAAGTACAACTCTAGATTTTGTAGTGCCCTCTTGGGCTAGCGG
TAAAGGTGCGCATTTTTTCACACCCTACAATGTTCTGTTCAAAAGATTTTGGTCAAACGCTGTAGAAGTGAAAGTTG
GTGCGCATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGATAAATGATC**AAAGGAGAAGAACTTTTCAC**GTTTTA
GAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGGTGC**TTTT**
TTTGTTTTTTATGTCTTCGAGTCATGTAATTAGTTATGTCACGC**CTCGAG**GTTGTATCTGTACAAAATCCAAAGCTG
AGCAAATAAATAAATAAATAAATGTATAAGTTACCGAACGGGGGTATTTTTACTTTTGATCAAAAATTTATGTACCA
ACTACAAAGTTTCCTCAGCACAGCCTTCAAGAAGGGAACACACATACAAACAGTGTCAAATAATTGTAGGGATAAAT
TTAAATATGGCATAAACTAAATAAGTAGAGCATGAAAAAACTGCAAAATCCAAAAAGTAAAAACGAAGGTCAGAAAG
TAAAGCAAAAGAAAATTAATAAAGCAATACTAAATCTATCATGATTTCCCGTAACTTCCATTAAAGCTGTAACCAGA
TTTACTCCTACTGTTTGAGCCTCTAACGCCTAATGGATTTTTAGAGAAGCTCAACCTGATACCTCCTTTGTTGTTGA
GGGAAGGGCGGGGGTGAGGTAGTTGACTACCATATAATTCTGCCAATGCTCTAGTGGCAAAGCTAACATCCTC
(485 bp 3′ UTR to *SHS1* shown)

prDNL4::sgRNA(Kan)::DNL4(t)

prSNR52::**crRNA**::*tracrRNA*::***SUP4(t)***

(590 bp 5′ UTR to *DNL4* shown)
GGATGATGGGTAACATAAATAGAAGGGGTAATGGTTCGCAATCTGACACATCAGAGAGTGAGGAAAACTCAGAACAA
TCTGATTTGGAAGGCAATAATCAATGTATTGAATATGACTCTTTAGGTAATGCTATTCGTATAGATAACATGAAAAG
CAGGGAAGCGCAATCTGAGGAATCAGAAGACGAGGAAAGTGGTTCAAAAGAAAATGGAGAGCCTTTAAGTTATGACC
CCTTAGGCAATTTAATTCGATAGGTGATTAAATAGGCTGAAATCAGTGTTTAGTAACTACGTACGTTGTACATGTAA
CATTGTGATATAAATCGTAAGATTCGCCGAGTATAGATCAATAATATCGGTTTCATCACTTACGTTGTTTGTGCAGT
ACTAGAGTTAAGATCGTTTTCGATCCCTTATTTTCTTCTTTTTTCCTTTTTTTTGTTATTTTTCTCTTTTTACCTTT
TGTCACCATATTAAATCTTTAAACAAATCTAACTATGAAAAAATCCTTTAAACATATGTTAATATGTGGAAAATAAA
TACTAAAATAAAAATCTAGAACTGAAGGAAATAGTAACGGATTATTTAGGT**GGATCC**TCACTAAAGGGAACAAAAGC
TGGAGCTTCTTTGAAAAGATAATGTATGATTATGCTTTCACTCATATTTATACAGAAACTTGATGTTTTCTTTCGAG
TATATACAAGGTGATTACATGTACGTTTGAAGTACAACTCTAGATTTTGTAGTGCCCTCTTGGGCTAGCGGTAAAGG
TGCGCATTTTTTCACACCCTACAATGTTCTGTTCAAAAGATTTTGGTCAAACGCTGTAGAAGTGAAAGTTGGTGCGC
ATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGATAAATGATC**GCCATCCTATGGAACTGCCT**GTTTTAGAGCTA
GAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGGTGC**TTTTTTTGTT**
TTTTATGTCTTCGAGTCATGTAATTAGTTATGTCACGC**CTCGAG**TGGTGCGTTTTGCGGAGGCTTAATTTTTTGAAG
TTTATTTAATACTATCCTACATATGTACATTAAATACTTCCGTAACGTTTATCAATAAGAGTGGAAGATGCGCAATT
ATATTCAAAAGATTGGCCAGTCAATTAACTTAAGGAAAAAATTTACTGCGCGGCTGTGGTCCCATTTGAAACGGGGA
CGTTAGTGTTGCTTTTGTGTTTCCTTCTTCTTCTTTTTTTCTTTTCTTTAGAAGTATCTTCTTTGTCTTTCACCTCA
TCGCTGGGCGAGTTGACTTTCTTCTTCTTCTTTACGCTCAAAGGATTAGGAGCCTTTGGACCAAGCTTTCTTTTCTT
AGTGATAGATTCCTTTCCTGATCCATCACCACTTTCTTGAAGTTTTTCAATATTAGGATCATTGAGACCCTTATACA
ATTTTTGTTCTTCAGTAATCTTGCTTGCCTTGGCACTCGCCGTACTGAGAGGTTCCATAACCATAACAGACCTGGTT
AAATGGATTAATGGAACACCCGGAACCGTCCTCAGCTTCCTCCTCAAATCTATGTCCTGAGAAGCAACCACATACCT
GTGTTTATTTGCACCGC
(589 bp 3′ UTR to *DNL4* shown)

GFY-3596

prHIS3::***(u1)***::*prCDC12*::***mCherry***::***NLS***::***SHS1(t)***::*prCCW12*::***SpHIS5***::*MX(t)*::***(u1)***::*HIS3(t)*

(u1) 20 bp target and 3 bp PAM sequence

(992 bp 5′ UTR to *HIS3* shown)

GGGTCAGTTATTTCATCCAGATATAACCCGAGAGGAAACTTCTTAGCGTCTGTTTTCGTACCATAAGGCAGTTCATG
AGGTATATTTTCGTTATTGAAGCCCAGCTCGTGAATGCTTAATGCTGCTGAACTGGTGTCCATGTCGCCTAGGTACG
CAATCTCCACAGGCTGCAAAGGTTTTGTCTCAAGAGCAATGTTATTGTGCACCCCGTAATTGGTCAACAAGTTTAAT
CTGTGCTTGTCCACCAGCTCTGTCGTAACCTTCAGTTCATCGACTATCTGAAGAAATTTACTAGGAATAGTGCCATG
GTACAGCAACCGAGAATGGCAATTTCTACTCGGGTTCAGCAACGCTGCATAAACGCTGTTGGTGCCGTAGACATATT
CGAAGATAGGATTATCATTCATAAGTTTCAGAGCAATGTCCTTATTCTGGAACTTGGATTTATGGCTCTTTTGGTTT
AATTTCGCCTGATTCTTGATCTCCTTTAGCTTCTCGACGTGGGCCTTTTTCTTGCCATATGGATCCGCTGCACGGTC
CTGTTCCCTAGCATGTACGTGAGCGTATTTCCTTTTAAACCACGACGCTTTGTCTTCATTCAACGTTTCCCATTGTT
TTTTTCTACTATTGCTTTGCTGTGGGAAAAACTTATCGAAAGATGACGACTTTTTCTTAATTCTCGTTTTAAGAGCT
TGGTGAGCGCTAGGAGTCACTGCCAGGTATCGTTTGAACACGGCATTAGTCAGGGAAGTCATAACACAGTCCTTTCC
CGCAATTTTCTTTTTCTATTACTCTTGGCCTCCTCTAGTACACTCTATATTTTTTTATGCCTCGGTAATGATTTTCA
TTTTTTTTTTTCCACCTAGCGGATGACTCTTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTA
TATAAAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG**ATGACGGTG**
GACTTCGGCTACGTAGGGCGATTGGGGCAGCGCCCTGTTTTTCATTAATGTAGTCAGCAATGTCAAGATTCAACGCC
AAGTCTGGTTCAGCAAGTGACATTCTGCAAGCTCTTTGAATCTTCCTCAAAAGAGGATTGCCCAAGGCTTGAGGTTT
CCTGACGGGCAACTCAGACAAATATATGCTATGTGAGTGCGGATGGGACATGATGCAGTATCACGATTAGCAATTCA
GCTATGAGTTATGTTGCTCTTTGTTTTGTTTATGGAAATTGTCCTATGGTAAGTCTCTTTTTTTTGCAATCGTGATT
ACAGAAAAAAAACAGGGCGCTGGAAAAGTGAAGAATCCGAAATTTTTTTCGAAATCACCATTGTTTGTTTTGAGTAG
ATCAAAGTCTTGAAAGGTGCAGCAAGATATAGGATCTTGACCTGAAGAGTATTGATAACGAACTACATCACATATTG
TATCAAATA**ATG**GTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACAT
GGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCG
CCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCC
AAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCG
CGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACA
AGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCC
TCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCA
CTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCA
AGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACC
GGCGGCATGGACGAGCTGTACAAG**TCTAGGGCAGACCCAAAGAAAAAGAGGAAAGTA****TAA**GTTGTATCTGTACAAAA
TCCAAAGCTGAGCAAATAAATAAATAAATAAATGTATAAGTTACCGAACGGGGGTATTTTTACTTTTGATCAAAAAT
TTATGTACCAACTACAAAGTTTCCTCAGCACAGCCTTCAAGAAGGGAACACACATACAAACAGTGTCAAATAATTGT
AGGGATAAATTTAAATATGGCATAAACTAAATAAGTAGAGCATGAAAAAACTGCAAAATCCAAAAAGTAAAAACGAA
GGTCAGAAAGTAAAGCAAAAGAAAATTAATAAAGCAATACTAAATCTATCATGATTTCCCGTAACTTCCATTAAAGC
TGTAACCAGATTTACTCCTACTGTTTGAGCCTCTAACGCCTAATGGATTTTTAGAGAAGCTCAACCTGATACCTCCT
TTGTTGTTGAGGGAAGGGCGGGGGTGAGGTAGTTGACTACCATATAATTCTGCCAATGCTCTAGTGGCAAAGCTAAC
ATCCTCACAAAGCAAAATAAAAGAAACTTAATACGTTATGCCGTAATGAAGGGCTACCAAAAACGATAATCTCAACT
GTAAACAGGTACAATGCGGACCCTTTTGCCACAAAACATACATCATTCATTGCCGGAAAAAGAAAGAAGTGAAGACA
GCAGTGCAGCCAGCCATGTTGCGCCAATCTAATTATAGATGCTGGTGCCCTGAGGATGTATCTGGAGCCAGCCATGG
CATCATGCGCTACCGCCGGATGTAAAATCCGACACGCAAAAGAAAACCTTCGAGGTTGCGCACTTCGCCCACCCATG
AACCACACGGTTAGTCCAAAAGGGGCAGTTCAGATTCCAGATGCGGGAATTAGCTTGCTGCCACCCTCACCTCACTA
ACGCTGCGGTGTGCGGATACTTCATGCTATTTATAGACGCGCGTGTCGGAATCAGCACGCGCAAGAACCAAATGGGA
AAATCGGAATGGGTCCAGAACTGCTTTGAGTGCTGGCTATTGGCGTCTGATTTCCGTTTTGGGAATCCTTTGCCGCG
CGCCCCTCTCAAAACTCCGCACAAGTCCCAGAAAGCGGGAAAGAAATAAAACGCCACCAAAAAAAAAAAAAATAAAAG
CCAATCCTCGAAGCGTGGGTGGTAGGCCCTGGATTATCCCGTACAAGTATTTCTCAGGAGTAAAAAAACCGTTTGTT
TTGGAATTCCCCATTTCGCGGCCACCTACGCCGCTATCTTTGCAACAACTATCTGCGATAACTCAGCAAATTTTGCA
TATTCGTGTTGCAGTATTGCGATAATGGGAGTCTTACTTCCAACATAACGGCAGAAAGAAATGTGAGAAAATTTTGC
ATCCTTTGCCTCCGTTCAAGTATATAAAGTCGGCATGCTTGATAATCTTTCTTTCCATCCTACATTGTTCTAATTAT
TCTTATTCTCCTTTATTCTTTCCTAACATACCAAGAAATTAATCTTCTGTCATTCGCTTAAACACTATATCAATA**AT**
GAGGAGGGCTTTTGTAGAAAGAAATACGAACGAAACGAAAATCAGCGTTGCCATCGCTTTGGACAAAGCTCCCTTAC
CTGAAGAGTCGAATTTTATTGATGAACTTATAACTTCCAAGCATACAAACCAAAAGGGAGAACAAGTAATCCAAGTA
GACACGGGAATTGGATTCTTGGATCACATGTATCATGCACTGGCTAAACATGCAGGCTGGAGCTTACGACTTTACTC
AAGAGGTGATTTAATCATCGATGATCATCACACTGCAGAAGATACTGCTATTGCACTTGGTATTGCATTCAAGCAGG
CTATGAGTAACTTTGCCGGCGTTAAAAGATTTGGACATGCTTATTGTCCACTTGACGAAGCTCTTTCTAGAAGCGTA
GTTGACTTGTCGGGACGGCCCTATGCTGTTATCGATTTGGGATTAAAGCGTGAAAAGGTTGGGGAATTGTCCTGTGA

AATGATCCCTCACTTACTATATTCCTTTTCGGTAGCAGCTGGAATTACTTTGCATGTTACCTGCTTATATGGTAGTA
ATGACCATCATCGTGCTGAAAGCGCTTTTAAATCTCTGGCTGTTGCCATGCGCGCGGCTACTAGTCTTACTGGAAGT
TCTGAAGTCCCAAGCACGAAGGGAGTGTTGTAAAGAGTACTGACAATAAAAAGATTCTTGTTTTCAAGAACTTGTCA
TTTGTATAGTTTTTTTATATTGTAGTTGTTCTATTTTAATCAAATGTTAGCGTGATTTATATTTTTTTTCGCCTCGA
CATCATCTGCCCAGATGCGAAGTTAAGTGCGCAGAAAGTAATATCATGCGTCAATCGTATGTGAATGCTGGTCGCTA
TACTGCTGTCGATTCGATACTAACGCCGCCATCCAGTATGACGGTGGACTTCGGCTACGTAGGGCGATTTGACACCG
ATTATTTAAAGCTGCAGCATACGATATATATACATGTGTATATATGTATACCTATGAATGTCAGTAAGTATGTATAC
GAACAGTATGATACTGAAGATGACAAGGTAATGCATCATTCTATACGTGTCATTCTGAACGAGGCGCGCTTTCCTTT
TTTCTTTTTGCTTTTTCTTTTTTTTTCTCTTGAACTCGAGAAAAAAAATATAAAAGAGATGGAGGAACGGGAAAAAG
TTAGTTGTGGTGATAGGTGGCAAGTGGTATTCCGTAAGAACAACAAGAAAAGCATTTCATATTATGGCTGAACTGAG
CGAACAAGTGCAAAATTTAAGCATCAACGACAACAACGAGAATGGTTATGTTCCTCCTCACTTAAGAGGAAAACCAA
GAAGTGCCAGAAATAACAGTAGCAACTACAATAACAACAACGGCGGCTACAACGGTGGCCGTGGCGGTGGCAGCTTC
TTTAGCAACAACCGTCGTGGTGGTTACGGCAACGGTGGTTTCTTCGGTGGAAACAACGGTGGCAGCAGATCTAACGG
CCGTTCTGGTGGTAGATGGATCGATGGCAAACATGTCCCAGCTCCAAGAAACGAAAAGGCCGAGATCGCCATATTTG
GTGTCCCCGAGGATCCAAATTTCCAATCTTCTGGTATTAACTTCGATAACTACGATGATATTCCAGTGGACGCCTCT
GGTAAGGATGTTCCTGAACCAATCACAGAATTTACCTCACCTCCATTGGACGGATTGTTATTGGAAAACATCAAATT
GGCCCGTTTCACCAAGCCAACACCTGTGCAAAAATACTCCGTCCCTATCGTTGCCAACGGCAGAGATTTGATGGCCT
GTGCGCAGACCGGTTCTGGTAAGACTGGTGGGTTTTTATTCCCAGTGTTGTCCGAATCATTTAAGACTGGACCATCT
CCTCAACCAGAGTCTCAAGGCTCCTTTTACCAAAGAAAGGCCTACCCAACTGCTGTC
(993 bp 3′ UTR to *HIS3* shown)

prSHS1::***SHS1***::*GFP*::*CDC10(t)*::*prMX*::***CaURA3***::*SHS1(t)*

(596 bp 5′ UTR to *SHS1* shown), (Gly polymorphism at codon 314)
GGTTATCGTATTTCACTTTTTGTGGTAAACTCAGGTCAATATTATGCGACTTGAACCATTCAGTCAAAGGTTCGTTA
CCACCTTTTTCCATACGAAGAAGTTCCTCCGGTTTAAACTGATCCATAGTGATAGATCTTACAAAAGATATATGCAC
ACCAAGCCCTCTATGGATACCGGCACATTCAAGGCAAATGAAAGCTCCAAACTTAGGCGTGGCCCATTGTGGATTTG
GCGCACCACAATCCATACATTTCTTATTTGCACCAATCTTTTGCAATTGCAAAAGACGCCTGCGGGTATCTGGGTCC
ACTTTCCAATCTGACATGCTCTATAATCCGCGATAAAATTGCTCAATTGGCACCATTTAAACTCAGAATCACGTCCA
TATTTCTGCTTTCATTCTTGATATAGTTGTGCAATTTGGTTCTTGACAAAAACTGGCGTTGCTTGCGGGTAACCGCG
CGATTTTTAAAGTGCCAAACTGCGAAAAGAATATAACAAGCTTTCGAGCAAGATCAATGTACCCAGCAAGTGAAATA
ATAAAACAAGAGCCCCAAAGATCTGCTTATAATTGCTAGAAAAATATATTATTAATCATGAGCACTGCTTCAACACC
GCCAATTAACTTATTTCGTAGAAAGAAAGAACATAAACGTGGGATCACATACACAATGTTACTATGTGGGCCAGCAG
GTACAGGAAAGACCGCCTTTGCTAACAATCTATTGGAAACTAAGATCTTTCCGCATAAGTATCAATACGGTAAATCA
AATGCTAGTATTAGCTCTAACCCAGAAGTAAAAGTTATTGCTCCGACAAAAGTTGTTTCATTTAATTCGAAAAATGG
GATTCCATCTTATGTTTCTGAATTCGATCCAATGAGAGCCAATTTGGAACCAGGTATTACCATCACCTCCACTTCAT
TAGAACTTGGGGGCAACAAAGATCAAGGAAAGCCAGAAATGAACGAGGATGATACCGTGTTTTTCAACTTGATTATG
ACGCATGGTATAGGCGAAAACTTGGACGATTCGTTGTGTTCTGAGGAAGTTATGTCGTATTTAGAACAACAATTTGA
CATTGTTTTAGCTGAGGAAACCAGAATTAAAAGGAATCCGAGGTTTGAGGACACCAGGGTTCACGTAGCATTATATT
TTATTGAACCCACTGGACACGGTCTGAGAGAAGTCGATGTAGAGCTCATGAAAAGCATCTCCAAATACACAAATGTA
CTGCCAATAATAACAAGAGCTGACTCATTCACCAAGGAGGAGCTAACTCAATTCAGGAAAAATATTATGTTTGATGT
GGAAAGATACAACGTCCCAATTTACAAATTTGAGGTTGACCCTGAAGATGATGATTTGGAATCCATGGAAGAGAATC
AAGCCTTGGCATCCTTGCAACCATTTGCTATTATAACTTCAGATACCAGAGATAGTGAAGGTAGATACGTTAGGGAG
TATCCGTGGGGGATAATATCAATCGACGACGACAAAATTTCGGATTTGAAAGTTTTAAAAAACGTCCTGTTTGGGTTC
TCACTTACAAGAATTCAAAGACACCACGCAAAATTTGCTTTACGAGAATTACCGTTCCGAAAAACTATCGTCCGTGG
CCAACGCTGAAGAAATTGGTCCTAATTCTACAAAGAGACAGTCAAATGCTCCAAGTTTAAGCAACTTTGCCTCTTTG
ATAAGCACTGGTCAATTCAATTCTTCTCAAACTCTTGCAAACAATTTGAGAGCGGACACACCAAGAAACCAAGTAAG
TGGAAACTTTAAGGAAAACGAATACGAAGACAATGGCGAACATGATTCAGCAGAAAATGAACAGGAAATGTCTCCCG
TGAGACAGTTGGGTAGAGAAATAAAACAAGAAAATGAAAATTTGATAAGATCTATCAAAACAGAATCTTCACCAAAA
TTCTTGAACTCTCCGGACTTACCAGAGCGTACCAAGTTAAGAAATATTTCAGAAACCGTTCCATATGTCTTGAGACA
TGAAAGAATTTTAGCAAGACAACAAAAACTGGAAGAGTTAGAGGCCCAGTCAGCTAAAGAATTACAAAAAAGAATTC
AAGAATTAGAAAGAAAAGCACACGAATTGAAATTGAGGGAAAAACTAATAAATCAGAATAAACTAAACGGTTCATCA
TCTTCAATCAATTCTCTACAACAGAGCACAAGGAGCCAAATTAAAAAAAATGACACGTATACTGATTTAGCCTCTAT
TGCATCGGGTAGAGATGGTCGACGGATCCCCGGGTTAATTAACAGTAAAGGAGAAGAACTTTTCACTGGAGTTGTCC
CAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCAGTGGGGAGGGTGAAGGTGATGCAACA
TACGGAAAACTTACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTTGTCACTACTTT
GACTTATGGTGTTCAATGCTTTTCAAGATACCCAGATCATATGAAACGGCATGACTTTTTCAAGAGTGCCATGCCCG

AAGGTTATGTACAGGAAAGAACTATATTTTTCAAAGATGACGGGAACTACAAGACACGTGCTGAAGTCAAGTTTGAA
GGTGATACCCTTGTTAATAGAATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAAATT
GGAATACAACTATAACTCACACAATGTATACATCATGGCAGACAAACAAAAGAATGGAATCAAAGTTAACTTCAAAA
TTAGACACAACATTGAAGATGGAAGCGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCT
GTCCTTTTACCAGACAACCATTACCTGTCCACACAATCTGCCCTTTCGAAAGATCCCAACGAAAAGAGAGACCACAT
GGTCCTTCTTGAGTTTGTAACAGCTGCTGGGATTACACATGGCATGGATGAACTATACAAATAGATCTCATAAGAAT
GGTGGTGATTATATATCTTATGTTATTAAGAATTCTCAAATTATTCTATATGAAAACACCGTAACTTGCTTCTCTCC
TTGGTTTTACATAATGACATAATGCGATCGAAAACTAGAGGTACAGGTATTGCTGGATTGGCGAGAGTTTTTACCTT
CTTTTCTGGCGTACAGCTATCACCTTCTCGTTTGGTAAAATGAAAGAACATTTTGTTGTCTTAGCCAAATATTTAAT
CTATGAAGAAAACGGAGTTTACCGGTAATTCTAAATAAAAGTTTGGTTAGGGATTGTGCCTCATAGAGAAGCAATTG
GTACTCATCTTATTAAAGTATTACTATAAACATTAGAAAGAGTCCCTGAGCGTTGCTAATGGGAAGCTATTCGCGCT
TTTAGTAAATTATTAAAATACGCCAAAAATAAATGTAATCCGGATATACCTCTTCTTTTAACCTTCCCAGATCTGTT
TAGCTTGCCTCGTCCCCGCCGGGTCACCCGGCCAGCGACATGGAGGCCCAGAATACCCTCCTTGACAGTCTTGACGT
GCGCAGCTCAGGGGCATGATGTGACTGTCGCCCGTACATTTAGCCCATACATCCCCATGTATAATCATTTGCATCCA
TACATTTTGATGGCCGCACGGCGCGAAGCAAAAATTACGGCTCCTCGCTGCAGACCTGCGAGCAGGGAAACGCTCCC
CTCACAGACGCGTTGAATTGTCCCCACGCCGCGCCCCTGTAGAGAAATATAAAAGGTTAGGATTTGCCACTGAGGTT
CTTCTTTCATATACTTCCTTTTAAAATCTTGCTAGGATACAGTTCTCACATCACATCCGAACATAAACAACCATGAC
AGTCAACACTAAGACCTATAGTGAGAGAGCAGAAACTCATGCCTCACCAGTAGCACAACGATTATTTCGATTAATGG
AACTGAAGAAAACCAATTTATGTGCATCAATTGATGTTGATACCACTAAGGAATTCCTTGAATTAATTGATAAATTG
GGTCCTTATGTATGCTTAATCAAGACTCATATTGATATAATCAATGATTTTTCCTATGAATCCACTATTGAACCATT
ATTAGAACTTTCACGTAAACATCAATTTATGATTTTTGAAGATAGAAAATTTGCTGATATTGGTAATACCGTGAAGA
AACAATATATTGGTGGAGTTTATAAAATTAGTAGTTGGGCAGATATTACTAATGCTCATGGTGTCACTGGGAATGGA
GTAGTTGAAGGATTAAAACAGGGAGCTAAAGAAACCACCACCAACCAAGAGCCAAGAGGGTTATTGATGTTAGCTGA
ATTATCATCAGTGGGATCATTAGCATATGGAGAATATTCTCAAAAAACTGTTGAAATTGCTAAATCCGATAAGGAAT
TTGTTATTGGATTTATTGCCCAACGTGATATGGGTGGACAAGAAGAAGGATTTGATTGGCTTATTATGACACCTGGA
GTTGGATTAGATGATAAAGGTGATGGATTAGGACAACAATATAGAACTGTTGATGAAGTTGTTAGCACTGGAACTGA
TATTATCATTGTTGGTAGAGGATTGTTTGGTAAAGGAAGAGATCCAGATATTGAAGGTAAAAGGTATAGAGATGCTG
GTTGGAATGCTTATTTGAAAAAGACTGGCCAATTATAAGTTGTATCTGTACAAAATCCAAAGCTGAGCAAATAAATA
AATAAATAAATGTATAAGTTACCGAACGGGGGTATTTTTACTTTTGATCAAAAATTTATGTACCAACTACAAAGTTT
CCTCAGCACAGCCTTCAAGAAGGGAACACACATACAAACAGTGTCAAATAATTGTAGGGATAAATTTAAATATGGCA
TAAACTAAATAAGTAGAGCATGAAAAAACTGCAAAATCCAAAAAGTAAAAACGAAGGTCAGAAAGTAAAGCAAAAGA
AAATTAATAAAGCAATACTAAATCTATCATGATTTCCCGTAACTTCCATTAAAGCTGTAACCAGATTTACTCCTACT
GTTTGAGCCTCTAACGCCTAATGGATTTTTAGAGAAGCTCAACCTGATACCTCCTTTGTTGTTGAGGGAAGGGCGGG
GGTGAGGTAGTTGACTACCATATAATTCTGCCAATGCTCTAGTGGCAAAGCTAACATCCTC

(485 bp 3′ UTR to *SHS1* shown)

prDNL4::***prMX***::***Kan^{R}***::***MX(t)***::*DNL4(t)*

(590 bp 5′ UTR to *DNL4* shown)

GGATGATGGGTAACATAAATAGAAGGGGTAATGGTTCGCAATCTGACACATCAGAGAGTGAGGAAAACTCAGAACAA
TCTGATTTGGAAGGCAATAATCAATGTATTGAATATGACTCTTTAGGTAATGCTATTCGTATAGATAACATGAAAAG
CAGGGAAGCGCAATCTGAGGAATCAGAAGACGAGGAAAGTGGTTCAAAAGAAAATGGAGAGCCTTTAAGTTATGACC
CCTTAGGCAATTTAATTCGATAGGTGATTAAATAGGCTGAAATCAGTGTTTAGTAACTACGTACGTTGTACATGTAA
CATTGTGATATAAATCGTAAGATTCGCCGAGTATAGATCAATAATATCGGTTTCATCACTTACGTTGTTTGTGCAGT
ACTAGAGTTAAGATCGTTTTCGATCCCTTATTTTCTTCTTTTTTCCTTTTTTTTGTTATTTTTCTCTTTTTACCTTT
TGTCACCATATTAAATCTTTAAACAAATCTAACTATGAAAAAATCCTTTAAACATATGTTAATATGTGGAAAATAAA
TACTAAAATAAAAATCTAGAACTGAAGGAAATAGTAACGGATTATTTAGGTGTTTAGCTTGCCTCGTCCCCGCCGGG
TCACCCGGCCAGCGACATGGAGGCCCAGAATACCCTCCTTGACAGTCTTGACGTGCGCAGCTCAGGGGCATGATGTG
ACTGTCGCCCGTACATTTAGCCCATACATCCCCATGTATAATCATTTGCATCCATACATTTTGATGGCCGCACGGCG
CGAAGCAAAAATTACGGCTCCTCGCTGCAGACCTGCGAGCAGGGAAACGCTCCCCTCACAGACGCGTTGAATTGTCC
CCACGCCGCGCCCCTGTAGAGAAATATAAAAGGTTAGGATTTGCCACTGAGGTTCTTCTTTCATATACTTCCTTTTA
AAATCTTGCTAGGATACAGTTCTCACATCACATCCGAACATAAACAACCATGGGTAAGGAAAAGACTCACGTTTCGA
GGCCGCGATTAAATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGT
GCGACAATCTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAA
TGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATCC
GTACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCCGGCAAAACAGCATTCCAGGTATTAGAAGAATATCCT
GATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCC

TTTTAACAGCGATCGCGTATTTCGTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATT
TTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAGCTTTTGCCATTCTCACCGGAT
TCAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGT
TGGACGAGTCGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCAT
TACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCATTTGATGCTCGAT
GAGTTTTTCTAATCAGTACTGACAATAAAAAGATTCTTGTTTTCAAGAACTTGTCATTTGTATAGTTTTTTTATATT
GTAGTTGTTCTATTTTAATCAAATGTTAGCGTGATTTATATTTTTTTTCGCCTCGACATCATCTGCCCAGATGCGAA
GTTAAGTGCGCAGAAAGTAATATCATGCGTCAATCGTATGTGAATGCTGGTCGCTATACTGCTGTCGATTCGATACT
AACGCCGCCATCCAGTTGGTGCGTTTTGCGGAGGCTTAATTTTTTGAAGTTTATTTAATACTATCCTACATATGTAC
ATTAAATACTTCCGTAACGTTTATCAATAAGAGTGGAAGATGCGCAATTATATTCAAAAGATTGGCCAGTCAATTAA
CTTAAGGAAAAAATTTACTGCGCGGCTGTGGTCCCATTTGAAACGGGGACGTTAGTGTTGCTTTTGTGTTTCCTTCT
TCTTCTTTTTTTCTTTTCTTTAGAAGTATCTTCTTTGTCTTTCACCTCATCGCTGGGCGAGTTGACTTTCTTCTTCT
TCTTTACGCTCAAAGGATTAGGAGCCTTTGGACCAAGCTTTCTTTTCTTAGTGATAGATTCCTTTCCTGATCCATCA
CCACTTTCTTGAAGTTTTTCAATATTAGGATCATTGAGACCCTTATACAATTTTTGTTCTTCAGTAATCTTGCTTGC
CTTGGCACTCGCCGTACTGAGAGGTTCCATAACCATAACAGACCTGGTTAAATGGATTAATGGAACACCCGGAACCG
TCCTCAGCTTCCTCCTCAAATCTATGTCCTGAGAAGCAACCACATACCTGTGTTTATTTGCACCGC
(589 bp 3′ UTR to *DNL4* shown)

GFY-2383

prHIS3::***(u2)***::*prGAL1/10*::***SpCas9***::***NLS***::*ADH1(t)*::*prMX*::***Kan^{R}***::*MX(t)*::***(u2)***::*HIS3(t)*

(u2) 20 bp target and 3 bp PAM sequence

(992 bp 5′ UTR to *HIS3* shown)
GGGTCAGTTATTTCATCCAGATATAACCCGAGAGGAAACTTCTTAGCGTCTGTTTTCGTACCATAAGGCAGTTCATG
AGGTATATTTTCGTTATTGAAGCCCAGCTCGTGAATGCTTAATGCTGCTGAACTGGTGTCCATGTCGCCTAGGTACG
CAATCTCCACAGGCTGCAAAGGTTTTGTCTCAAGAGCAATGTTATTGTGCACCCCGTAATTGGTCAACAAGTTTAAT
CTGTGCTTGTCCACCAGCTCTGTCGTAACCTTCAGTTCATCGACTATCTGAAGAAATTTACTAGGAATAGTGCCATG
GTACAGCAACCGAGAATGGCAATTTCTACTCGGGTTCAGCAACGCTGCATAAACGCTGTTGGTGCCGTAGACATATT
CGAAGATAGGATTATCATTCATAAGTTTCAGAGCAATGTCCTTATTCTGGAACTTGGATTTATGGCTCTTTTGGTTT
AATTTCGCCTGATTCTTGATCTCCTTTAGCTTCTCGACGTGGGCCTTTTTCTTGCCATATGGATCCGCTGCACGGTC
CTGTTCCCTAGCATGTACGTGAGCGTATTTCCTTTTAAACCACGACGCTTTGTCTTCATTCAACGTTTCCCATTGTT
TTTTTCTACTATTGCTTTGCTGTGGGAAAAACTTATCGAAAGATGACGACTTTTTCTTAATTCTCGTTTTAAGAGCT
TGGTGAGCGCTAGGAGTCACTGCCAGGTATCGTTTGAACACGGCATTAGTCAGGGAAGTCATAACACAGTCCTTTCC
CGCAATTTTCTTTTTCTATTACTCTTGGCCTCCTCTAGTACACTCTATATTTTTTTATGCCTCGGTAATGATTTTCA
TTTTTTTTTTTCCACCTAGCGGATGACTCTTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTA
TATAAAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG**GCTGTTCGT**
GTGCGCGTCCTGGGGACAGGTTATCAGCAACAACACAGTCATATCCATTCTCAATTAGCTCTACCACAGTGTGTGAA
CCAATGTATCCAGCACCACCTGTAACCAAAACAATTTTAGAAGTACTTTCACTTTGTAACTGAGCTGTCATTTATAT
TGAATTTTCAAAAATTCTTACTTTTTTTTTTGGATGGACGCAAAGAAGTTTAATAATCATATTACATGGCATTACCAC
CATATACATATCCATATACATATCCATATCTAATCTTACTTATATGTTGTGGAAATGTAAAGAGCCCCATTATCTTA
GCCTAAAAAAACCTTCTCTTTGGAACTTTCAGTAATACGCTTAACTGCTCATTGCTATATTGAAGTACGGATTAGAA
GCCGCCGAGCGGGTGACAGCCCTCCGAAGGAAGACTCTCCTCCGTGCGTCCTCGTCTTCACCGGTCGCGTTCCTGAA
ACGCAGATGTGCCTCGCGCCGCACTGCTCCGAACAATAAAGATTCTACAATACTAGCTTTTATGGTTATGAAGAGGA
AAAATTGGCAGTAACCTGGCCCCACAAACCTTCAAATGAACGAATCAAATTAACAACCATAGGATGATAATGCGATT
AGTTTTTTAGCCTTATTTCTGGGGTAATTAATCAGCGAAGCGATGATTTTTGATCTATTAACAGATATATAAATGCA
AAAACTGCATAACCACTTTAACTAATACTTTCAACATTTTCGGTTTGTATTACTTCTTATTCAAATGTAATAAAAGT
ATCAACAAAAAATTGTTAATATACCTCTATACTTTAACGTCAAGGAGAAAAAACTATA**ATG**GATAAGAAATACTCTA
TCGGTTTGGATATTGGTACAAATTCAGTTGGTTGGGCAGTTATTACTGATGAATACAAGGTTCCATCTAAAAAGTTT
AAAGTTTTGGGTAACACTGATAGACATTCTATTAAGAAAAATTTGATTGGTGCTTTGTTATTTGATTCTGGTGAAAC
TGCTGAAGCAACAAGATTGAAAAGAACTGCAAGAAGAAGATACACAAGAAGAAAGAATAGAATCTGTTATTTGCAAG
AAATTTTCTCTAACGAAATGGCTAAGGTTGATGATTCTTTCTTTCATAGATTGGAAGAATCATTTTTAGTTGAAGAA
GATAAGAAACATGAAAGACATCCAATCTTCGGTAACATCGTTGATGAAGTTGCTTACCATGAAAAGTACCCAACAAT
CTATCATTTGAGAAAGAAATTGGTTGATTCAACTGATAAGGCAGATTTGAGATTGATATATTTGGCTTTAGCACATA
TGATCAAGTTTAGAGGTCATTTCTTGATCGAGGGTGACTTGAATCCAGATAATTCTGATGTTGATAAGTTGTTTATT

CAATTAGTTCAAACATATAATCAATTGTTTGAAGAAAATCCAATTAATGCTTCTGGTGTTGATGCTAAGGCAATCTT
GTCAGCAAGATTGTCTAAGTCAAGAAGATTGGAAAATTTGATCGCTCAATTACCAGGTGAAAAGAAAAATGGTTTGT
TCGGTAATTTGATCGCATTGTCTTTGGGTTTGACACCAAACTTCAAGTCAAACTTCGATTTGGCTGAAGATGCAAAG
TTGCAATTGTCTAAGGATACTTACGATGATGATTTGGATAATTTGTTGGCTCAAATTGGTGACCAATATGCAGATTT
GTTTTTGGCTGCTAAAAATTTGTCTGATGCTATCTTGTTGTCAGATATCTTGAGAGTTAACACTGAAATCACAAAGG
CTCCATTGTCTGCATCAATGATCAAGAGATACGATGAACATCATCAAGATTTGACTTTGTTGAAGGCATTGGTTAGA
CAACAATTACCAGAAAAGTACAAGGAAATTTTCTTTGATCAATCTAAAAATGGTTATGCTGGTTACATTGATGGTGG
TGCATCTCAAGAAGAATTCTACAAGTTTATTAAGCCAATCTTGGAAAAGATGGATGGTACAGAAGAATTGTTAGTTA
AATTGAACAGAGAAGATTTGTTAAGAAAACAAAGAACTTTCGATAACGGTTCTATCCCACATCAAATCCATTTGGGT
GAATTACATGCTATCTTGAGAAGACAAGAAGATTTCTACCCATTTTTAAAGGATAACAGAGAAAAGATTGAAAAGAT
TTTGACTTTTAGAATTCCATATTACGTTGGTCCATTAGCTCGTGGTAATTCTAGATTTGCATGGATGACTAGAAAGT
CAGAAGAAACTATCACACCATGGAATTTTGAAGAAGTTGTTGATAAAGGTGCTTCTGCACAATCTTTTATTGAAAGA
ATGACAAACTTCGATAAAAATTTGCCAAACGAAAAGGTTTTGCCAAAGCATTCATTGTTATATGAATACTTTACTGT
TTACAATGAATTGACAAAAGTTAAATATGTTACTGAGGGTATGAGAAAACCAGCATTTTTGTCTGGTGAACAAAAGA
AAGCAATCGTTGATTTGTTGTTTAAAACTAACAGAAAGGTTACAGTTAAACAATTGAAAGAAGATTACTTTAAGAAA
ATTGAATGTTTTGATTCTGTTGAAATTTCAGGTGTTGAAGATAGATTCAATGCTTCATTAGGTACTTACCATGATTT
GTTGAAGATTATTAAGGATAAAGATTTCTTGGATAATGAAGAAAATGAAGATATTTTAGAAGATATTGTTTTAACTT
TGACATTATTTGAAGATAGAGAAATGATCGAAGAAAGATTGAAGACATACGCTCATTTGTTCGATGATAAAGTTATG
AAGCAATTGAAGAGAAGAAGATACACTGGTTGGGGTAGATTGTCTAGAAAGTTGATTAATGGTATCAGAGATAAGCA
ATCTGGTAAAACAATCTTGGATTTCTTGAAGTCAGATGGTTTCGCAAACAGAAACTTCATGCAATTGATTCATGATG
ATTCATTGACTTTTAAAGAAGATATCCAAAAAGCTCAAGTTTCTGGTCAGGGTGACTCATTGCATGAACATATTGCT
AATTTGGCAGGTTCTCCAGCTATTAAGAAAGGTATCTTGCAAACAGTTAAGGTTGTTGATGAATTAGTTAAAGTTAT
GGGTAGACATAAGCCAGAAAACATCGTTATCGAAATGGCTAGAGAAAACCAAACTACACAAAAGGGTCAAAAGAATT
CAAGAGAAAGAATGAAGAGAATCGAAGAAGGTATTAAAGAATTGGGTTCTCAAATCTTGAAGGAACATCCAGTTGAA
AACACTCAATTGCAAAACGAAAAGTTGTACTTATACTACTTACAAAACGGTAGAGATATGTACGTTGATCAAGAATT
AGATATCAACAGATTGTCAGATTACGATGTTGATCATATCGTTCCACAATCATTTTTGAAGGATGATTCAATCGATA
ATAAGGTTTTGACAAGATCTGATAAGAACCGTGGTAAATCTGATAATGTTCCATCAGAAGAAGTTGTTAAGAAAATG
AAGAACTACTGGAGACAATTGTTAAATGCTAAGTTGATCACTCAAAGAAAGTTCGATAATTTGACAAAAGCTGAAAG
AGGTGGTTTGTCAGAATTAGATAAAGCAGGTTTTATTAAGAGACAATTAGTTGAAACTAGACAAATCACAAAGCATG
TTGCACAAATCTTGGATTCTAGAATGAACACTAAATATGATGAAAATGATAAATTAATTAGAGAAGTTAAAGTTATT
ACATTAAAATCTAAATTGGTTTCAGATTTTAGAAAAGATTTTCAATTCTACAAAGTTAGAGAAATTAATAACTATCA
TCATGCTCATGATGCATACTTGAATGCTGTTGTTGGTACTGCATTGATTAAGAAATACCCAAAGTTGGAATCTGAAT
TCGTTTACGGTGACTACAAGGTTTACGATGTTAGAAAGATGATCGCTAAGTCAGAACAAGAAATCGGTAAAGCTACA
GCAAAGTATTTCTTTTATTCTAACATCATGAATTTCTTTAAAACTGAAATTACATTAGCTAACGGTGAAATCAGAAA
AAGACCATTGATCGAAACTAATGGTGAAACAGGTGAAATTGTTTGGGATAAAGGTAGAGATTTCGCAACTGTTAGAA
AGGTTTTGTCAATGCCACAAGTTAACATCGTTAAGAAAACTGAAGTTCAAACAGGTGGTTTTTCTAAGGAATCAATC
TTGCCAAAGAGAAACTCTGATAAGTTGATTGCTAGAAAGAAAGATTGGGATCCAAAGAAATATGGTGGTTTTGATTC
TCCAACTGTTGCTTACTCAGTTTTAGTTGTTGCAAAGGTTGAAAAGGGTAAATCTAAGAAATTGAAATCAGTTAAAG
AATTGTTAGGTATCACAATCATGGAAAGATCTTCATTCGAAAAGAATCCAATCGATTTCTTGGAAGCAAAGGGTTAC
AAGGAAGTTAAGAAAGATTTGATTATTAAGTTGCCAAAGTACTCTTTGTTCGAATTAGAAAACGGTAGAAAAAGAAT
GTTAGCTTCAGCTGGTGAATTGCAAAAGGGTAATGAATTGGCTTTGCCATCTAAGTACGTTAATTTCTTGTATTTGG
CATCTCATTACGAAAAGTTGAAGGGTTCACCAGAAGATAATGAACAAAAACAATTGTTCGTTGAACAACATAAGCAT
TATTTGGATGAAATTATTGAACAAATTTCTGAATTTTCAAAAAGAGTTATTTTGGCTGATGCAAATTTGGATAAGGT
TTTGTCTGCTTACAATAAGCATAGAGATAAGCCAATCAGAGAACAAGCAGAAAACATCATCCATTTGTTTACTTTGA
CAAATTTGGGTGCTCCAGCTGCTTTTAAATACTTCGATACTACAATCGATAGAAAAAGATACACTTCTACAAAGGAA
GTTTTGGATGCAACATTGATCCATCAATCAATCACTGGTTTGTATGAAACAAGAATTGATTTGTCTCAATTGGGTGG
TGACTCTAGGGCAGACCCAAAGAAAAAGAGGAAAGTATAAGGCGCGCCACTTCTAAATAAGCGAATTTCTTATGATT
TATGATTTTTATTATTAAATAAGTTATAAAAAAAATAAGTGTATACAAATTTTAAAGTGACTCTTAGGTTTTAAAAC
GAAAATTCTTATTCTTGAGTAACTCTTTCCTGTAGGTCAGGTTGCTTTCTCAGGTATAGTATGAGGTCGCTCTTATT
GACCACACCTCTACCGGCAGATCCGCTAGGGATAACAGGGTAATATAGATCTGTTTAGCTTGCCTCGTCCCCGCCGG
GTCACCCGGCCAGCGACATGGAGGCCCAGAATACCCTCCTTGACAGTCTTGACGTGCGCAGCTCAGGGGCATGATGT
GACTGTCGCCCGTACATTTAGCCCATACATCCCCATGTATAATCATTTGCATCCATACATTTTGATGGCCGCACGGC
GCGAAGCAAAAATTACGGCTCCTCGCTGCAGACCTGCGAGCAGGGAAACGCTCCCCTCACAGACGCGTTGAATTGTC
CCCACGCCGCGCCCCTGTAGAGAAATATAAAAGGTTAGGATTTGCCACTGAGGTTCTTCTTTCATATACTTCCTTTT
AAAATCTTGCTAGGATACAGTTCTCACATCACATCCGAACATAAACAACCATGGGTAAGGAAAAGACTCACGTTTCG
AGGCCGCGATTAAATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGG
TGCGACAATCTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCA

ATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATC
CGTACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCCGGCAAAACAGCATTCCAGGTATTAGAAGAATATCC
TGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTC
CTTTTAACAGCGATCGCGTATTTCGTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGAT
TTTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAGCTTTTGCCATTCTCACCGGA
TTCAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATG
TTGGACGAGTCGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCA
TTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCATTTGATGCTCGA
TGAGTTTTTCTAATCAGTACTGACAATAAAAAGATTCTTGTTTTCAAGAACTTGTCATTTGTATAGTTTTTTTATAT
TGTAGTTGTTCTATTTTAATCAAATGTTAGCGTGATTTATATTTTTTTTCGCCTCGACATCATCTGCCCAGATGCGA
AGTTAAGTGCGCAGAAAGTAATATCATGCGTCAATCGTATGTGAATGCTGGTCGCTATACTGCTGTCGATTCGATAC
TAACGCCGCCATCCAGT**GCTGTTCGTGTGCGCGTCCTGGG**TGACACCGATTATTTAAAGCTGCAGCATACGATATAT
ATACATGTGTATATATGTATACCTATGAATGTCAGTAAGTATGTATACGAACAGTATGATACTGAAGATGACAAGGT
AATGCATCATTCTATACGTGTCATTCTGAACGAGGCGCGCTTTCCTTTTTTCTTTTTGCTTTTTCTTTTTTTTTCTC
TTGAACTCGAGAAAAAAAATATAAAAGAGATGGAGGAACGGGAAAAAGTTAGTTGTGGTGATAGGTGGCAAGTGGTA
TTCCGTAAGAACAACAAGAAAAGCATTTCATATTATGGCTGAACTGAGCGAACAAGTGCAAAATTTAAGCATCAACG
ACAACAACGAGAATGGTTATGTTCCTCCTCACTTAAGAGGAAAACCAAGAAGTGCCAGAAATAACAGTAGCAACTAC
AATAACAACAACGGCGGCTACAACGGTGGCCGTGGCGGTGGCAGCTTCTTTAGCAACAACCGTCGTGGTGGTTACGG
CAACGGTGGTTTCTTCGGTGGAAACAACGGTGGCAGCAGATCTAACGGCCGTTCTGGTGGTAGATGGATCGATGGCA
AACATGTCCCAGCTCCAAGAAACGAAAAGGCCGAGATCGCCATATTTGGTGTCCCCGAGGATCCAAATTTCCAATCT
TCTGGTATTAACTTCGATAACTACGATGATATTCCAGTGGACGCCTCTGGTAAGGATGTTCCTGAACCAATCACAGA
ATTTACCTCACCTCCATTGGACGGATTGTTATTGGAAAACATCAAATTGGCCCGTTTCACCAAGCCAACACCTGTGC
AAAAATACTCCGTCCCTATCGTTGCCAACGGCAGAGATTTGATGGCCTGTGCGCAGACCGGTTCTGGTAAGACTGGT
GGGTTTTTATTCCCAGTGTTGTCCGAATCATTTAAGACTGGACCATCTCCTCAACCAGAGTCTCAAGGCTCCTTTTA
CCAAAGAAAGGCCTACCCAACTGCTGTCATTA
(993 bp 3′ UTR to *HIS3* shown)

GFY-3864

*prDNL4**::DNL4(WT)::**DNL4(t)*

DNL4 Cas9 Target sites are displayed along with the immediate reading frame. Silent mutations are illustrated.

(590 bp 5′ UTR to *DNL4* shown)
GGATGATGGGTAACATAAATAGAAGGGGTAATGGTTCGCAATCTGACACATCAGAGAGTGAGGAAAACTCAGAACAA
TCTGATTTGGAAGGCAATAATCAATGTATTGAATATGACTCTTTAGGTAATGCTATTCGTATAGATAACATGAAAAG
CAGGGAAGCGCAATCTGAGGAATCAGAAGACGAGGAAAGTGGTTCAAAAGAAAATGGAGAGCCTTTAAGTTATGACC
CCTTAGGCAATTTAATTCGATAGGTGATTAAATAGGCTGAAATCAGTGTTTAGTAACTACGTACGTTGTACATGTAA
CATTGTGATATAAATCGTAAGATTCGCCGAGTATAGATCAATAATATCGGTTTCATCACTTACGTTGTTTGTGCAGT
ACTAGAGTTAAGATCGTTTTCGATCCCTTATTTTCTTCTTTTTTCCTTTTTTTTTGTTATTTTTCTCTTTTTACCTTT
TGTCACCATATTAAATCTTTAAACAAATCTAACTATGAAAAAATCCTTTAAACATATGTTAATATGTGGAAAATAAA
TACTAAAATAAAAATCTAGAACTGAAGGAAATAGTAACGGATTATTTAGGT**ATG**ATATCAGCACTAGATTCTATACC
CGAGCCCCAAAACTTTGCGCCTAGTCCAGATTTCAAATGGCTTTGTGAAGAGCTATTTGTGAAGATACATGAAGTTC
AAATTAATGGAACGGCCGGCACTGGCAAATCAAGGTCTTTCAAGTACTATGAAATAATATCGAATTTCGTCGAAATG
TGGAGAAAAACCGTGGGAAATAATATATATCCTGCACTGGTTCTTGCTCTTCCCTACCGCGATAGACGAATCTATAA
TATTAAGGATTATGTATTAATAAGAACTATATGCTCTTACTTGAAGTTGCCAAAAAATTCTGCAACAGAGCAGCGGT
TAAAAGATTGGAAACAGCGTGTCGGTAAAGGTGGGAATCTTTCTTCTCTTCTTGTGGAAGAAATTGCTAAAAGAAGG
GCTGAACCTAGCTCAAAAGCGATTACAATTGATAACGTCAATCACTATCTGGATAGTTTGAGTGGAGACAGGTTCGC
TTCCGGACGAGGATTTAAGAGTCTTGTCAAGTCCAAACCTTTCCTGCACTGTGTGGAGAATATGAGTTTCGTCGAAT
TAAAATACTTCTTTGATATCGTGCTTAAAAATAGAGTAATAGGAGGTCAAGAGCACAAATTGCTAAACTGCTGGCAT
CCTGATGCTCAGGATTATCTTAGCGTGATATCTGATTTAAAGGTGGTAACTTCAAAACTTTATGATCCAAAAGTTCG
TCTAAAGGATGATGATTTGAGTATAAAAGTTGGCTTTGCATTCGCCCCCCAATTAGCCAAAAAAGTGAATCTTTCTT
ATGAGAAAATATGCCGTACACTACATGATGATTTTTTGGTAGAAGAAAAAATGGATGGAGAACGAATTCAAGTTCAT
TATATGAATTATGGTGAATCCATAAAATTTTTTAGTAGACGGGGCATCGACTATACCTATTTGTACGGAGCGAGCTT
ATCATCAGGAACTATATCTCAACATTTGAGGTTTACAGATAGTGTTAAAGAATGTGTTTTAGATGGAGAAATGGTGA

CGTTTGATGCAAAAAGACGGGTGATTCTTCCATTCGGTCTTGTTAAAGGAAGTGCAAAGGAAGCGCTATCTTTTAAT
AGTATAAATAATGTTGACTTTCACCCCTTATATATGGTGTTTGATCTGTTATACCTGAATGGGACTTCGTTGACACC
ATTACCCCTTCATCAAAGGAAGCAATATCTGAACAGCATTTTAAGTCCCTTGAAAAATATTGTAGAAATAGTACGAT
CTTCTAGATGTTATGGTGTGGAGTCAATCAAAAAGTCTTTAGAAGTTGCAATCTCACTGGGTTCAGAAGGAGTTGTT
TTGAAATATTATAATTCAAGTTATAATGTCGCCAGTCGAAACAACAACTGGATCAAGGTAAAACCTGAATATTTGGA
GGAATTTGGAGAGAATTTAGACTTAATAGTAATAGGCAGAGATTCTGGGAAAAAAGATTCTTTTATGCTAGGGTTAC
TTGTGCTAGATGAAGAAGAGTATAAAAAGCACCAAGGAGACTCCTCTGAAATTGTAGACCACTCAAGCCAAGAAAAA
CACATACAAAATTCAAGAAGAAGGGTGAAAAAAATACTTTCATTCTGTTCTATCGCAAACGGTATATCTCAAGAAGA
ATTCAAAGAAATCGACCGCAAAACGAGAGGACATTGGAAAAGAACCTCCGAAGTTGCTCCCCCTGCTTCAATTTTAG
AATTTGGCTCAAAAATACCTGCCGAATGGATTGACCCCAGTGAATCAATTGTTCTAGAAATAAAATCACGGTCTTTG
GATAACACAGAAACGAATATGCAGAAGTACGCTACCAATTGTACTTTGTACGGTGGCTATTGTAAAAGAATACGGTA
CGATAAAGAATGGACAGATTGTTACACACTTAACGACTTATACGAAAGTAGGACGGTTAAATCTAACCCCAGCTATC
AAGCGGAAAGGTCACAGCTTGGATTGATACGGAAAAAGAGAAAGAGAGTACTTATTTCAGACAGCTTTCACCAAAAC
AGGAAACAACTGCCAATTTCAAACATCTTTGCCGGATTACTTTTTTATGTTCTCTCTGACTATGTCACTGAAGATAC
TGGAATACGGATTACACGGGCAGAACTTGAAAAAACTATTGTGGAACATGGTGGTAAACTGATATATAATGTAATTT
TAAAACGTCATTCAATTGGGGACGTTCGGTTAATCAGCTGTAAAACTACCACGGAATGCAAGGCTTTAATAGATCGA
GGATATGATATATTGCACCCAAATTGGGTACTCGATTGTATAGCATATAAGAGGCTCATCCTGATCGAGCCCAATTA
TTGCTTTAACGTCTCTCAAAAAATGAGAGCCGTCGCTGAAAAAAGGGTAGATTGTTTGGGTGATAGTTTTGAAAATG
ACATTTCGGAAACCAAACTGTCATCATTGTATAAATCACAACTAAGTCTACCACCGATGGGGGAACTCGAGATAGAT
TCTGAGGTTCGGCGGTTTCCATTATTTTTATTCTCCAACAGGATTGCATACGTACCACGTCGCAAAATTAGCACAGA
AGATGACATTATAGAAATGAAAATTAAGTTGTTTGGTGGAAAAATAACAGATCAACAGTCACTTTGTAACTTAATAA
TTATACCATATACTGATCCTATTTTGAGGAAAGACTGCATGAATGAGGTACACGAAAAAATAAAAGAACAAATAAAG
GCTTCTGATACTATACCGAAAATAGCCAGGGTCGTTGCCCCTGAATGGGTGGATCATTCTATTAATGAAAACTGTCA
AGTGCCTGAGGAGGATTTCCCCGTAGTCAACTACTGATGGTGCGTTTTGCGGAGGCTTAATTTTTTGAAGTTTATTT
AATACTATCCTACATATGTACATTAAATACTTCCGTAACGTTTATCAATAAGAGTGGAAGATGCGCAATTATATTCA
AAAGATTGGCCAGTCAATTAACTTAAGGAAAAAATTTACTGCGCGGCTGTGGTCCCATTTGAAACGGGGACGTTAGT
GTTGCTTTTGTGTTTCCTTCTTCTTCTTTTTTTCTTTTCTTTAGAAGTATCTTCTTTGTCTTTCACCTCATCGCTGG
GCGAGTTGACTTTCTTCTTCTTCTTTACGCTCAAAGGATTAGGAGCCTTTGGACCAAGCTTTCTTTTCTTAGTGATA
GATTCCTTTCCTGATCCATCACCACTTTCTTGAAGTTTTTCAATATTAGGATCATTGAGACCCTTATACAATTTTTG
TTCTTCAGTAATCTTGCTTGCCTTGGCACTCGCCGTACTGAGAGGTTCCATAACCATAACAGACCTGGTTAAATGGA
TTAATGGAACACCCGGAACCGTCCTCAGCTTCCTCCTCAAATCTATGTCCTGAGAAGCAACCACATACCTGTGTTTA
TTTGCACCGC

(589 bp 3′ UTR to *DNL4* shown)

GFY-3875

prDNL4::***DNL4(WT)***::***CDC11(t)***::***sgRNA(Kan)***::*DNL4(t)*

prSNR52::***crRNA***::*tracrRNA*::***SUP4(t)***

(590 bp 5′ UTR to *DNL4* shown)

GGATGATGGGTAACATAAATAGAAGGGGTAATGGTTCGCAATCTGACACATCAGAGAGTGAGGAAAACTCAGAACAA
TCTGATTTGGAAGGCAATAATCAATGTATTGAATATGACTCTTTAGGTAATGCTATTCGTATAGATAACATGAAAAG
CAGGGAAGCGCAATCTGAGGAATCAGAAGACGAGGAAAGTGGTTCAAAAGAAAATGGAGAGCCTTTAAGTTATGACC
CCTTAGGCAATTTAATTCGATAGGTGATTAAATAGGCTGAAATCAGTGTTTAGTAACTACGTACGTTGTACATGTAA
CATTGTGATATAAATCGTAAGATTCGCCGAGTATAGATCAATAATATCGGTTTCATCACTTACGTTGTTTGTGCAGT
ACTAGAGTTAAGATCGTTTTCGATCCCTTATTTTCTTCTTTTTTCCTTTTTTTTTGTTATTTTTCTCTTTTTACCTTT
TGTCACCATATTAAATCTTTAAACAAATCTAACTATGAAAAAATCCTTTAAACATATGTTAATATGTGGAAAATAAA
TACTAAAATAAAAATCTAGAACTGAAGGAAATAGTAACGGATTATTTAGGT**ATG**ATATCAGCACTAGATTCTATACC
CGAGCCCCAAAACTTTGCGCCTAGTCCAGATTTCAAATGGCTTTGTGAAGAGCTATTTGTGAAGATACATGAAGTTC
AAATTAATGGAACGGCCGGCACTGGCAAATCAAGGTCTTTCAAGTACTATGAAATAATATCGAATTTCGTCGAAATG
TGGAGAAAAACCGTGGGAAATAATATATATCCTGCACTGGTTCTTGCTCTTCCCTACCGCGATAGACGAATCTATAA
TATTAAGGATTATGTATTAATAAGAACTATATGCTCTTACTTGAAGTTGCCAAAAAATTCTGCAACAGAGCAGCGGT
TAAAAGATTGGAAACAGCGTGTCGGTAAAGGTGGGAATCTTTCTTCTCTTCTTGTGGAAGAAATTGCTAAAAGAAGG
GCTGAACCTAGCTCAAAAGCGATTACAATTGATAACGTCAATCACTATCTGGATAGTTTGAGTGGAGACAGGTTCGC
TTCCGGACGAGGATTTAAGAGTCTTGTCAAGTCCAAACCTTTCCTGCACTGTGTGGAGAATATGAGTTTCGTCGAAT

TAAAATACTTCTTTGATATCGTGCTTAAAAATAGAGTAATAGGAGGTCAAGAGCACAAATTGCTAAACTGCTGGCAT
CCTGATGCTCAGGATTATCTTAGCGTGATATCTGATTTAAAGGTGGTAACTTCAAAACTTTATGATCCAAAAGTTCG
TCTAAAGGATGATGATTTGAGTATAAAAGTTGGCTTTGCATTCGCCCCCCAATTAGCCAAAAAAGTGAATCTTTCTT
ATGAGAAAATATGCCGTACACTACATGATGATTTTTTGGTAGAAGAAAAAATGGATGGAGAACGAATTCAAGTTCAT
TATATGAATTATGGTGAATCCATAAAATTTTTTAGTAGACGGGGCATCGACTATACCTATTTGTACGGAGCGAGCTT
ATCATCAGGAACTATATCTCAACATTTGAGGTTTACAGATAGTGTTAAAGAATGTGTTTTAGATGGAGAAATGGTGA
CGTTTGATGCAAAAAGACGGGTGATTCTTCCATTCGGTCTTGTTAAAGGAAGTGCAAAGGAAGCGCTATCTTTTAAT
AGTATAAATAATGTTGACTTTCACCCCTTATATATGGTGTTTGATCTGTTATACCTGAATGGGACTTCGTTGACACC
ATTACCCCTTCATCAAAGGAAGCAATATCTGAACAGCATTTTAAGTCCCTTGAAAAATATTGTAGAAATAGTACGAT
CTTCTAGATGTTATGGTGTGGAGTCAATCAAAAAGTCTTTAGAAGTTGCAATCTCACTGGGTTCAGAAGGAGTTGTT
TTGAAATATTATAATTCAAGTTATAATGTCGCCAGTCGAAACAACAACTGGATCAAGGTAAAACCTGAATATTTGGA
GGAATTTGGAGAGAATTTAGACTTAATAGTAATAGGCAGAGATTCTGGGAAAAAAGATTCTTTTATGCTAGGGTTAC
TTGTGCTAGATGAAGAAGAGTATAAAAAGCACCAAGGAGACTCCTCTGAAATTGTAGACCACTCAAGCCAAGAAAAA
CACATACAAAATTCAAGAAGAAGGGTGAAAAAAATACTTTCATTCTGTTCTATCGCAAACGGTATATCTCAAGAAGA
ATTCAAAGAAATCGACCGCAAAACGAGAGGACATTGGAAAAGAACCTCCGAAGTTGCTCCCCCTGCTTCAATTTTAG
AATTTGGCTCAAAAATACCTGCCGAATGGATTGACCCCAGTGAATCAATTGTTCTAGAAATAAAATCACGGTCTTTG
GATAACACAGAAACGAATATGCAGAAGTACGCTACCAATTGTACTTTGTACGGTGGCTATTGTAAAAGAATACGGTA
CGATAAAGAATGGACAGATTGTTACACACTTAACGACTTATACGAAAGTAGGACGGTTAAATCTAACCCCAGCTATC
AAGCGGAAAGGTCACAGCTTGGATTGATACGGAAAAAGAGAAAGAGAGTACTTATTTCAGACAGCTTTCACCAAAAC
AGGAAACAACTGCCAATTTCAAACATCTTTGCCGGATTACTTTTTTATGTTCTCTCTGACTATGTCACTGAAGATAC
TGGAATACGGATTACACGGGCAGAACTTGAAAAAACTATTGTGGAACATGGTGGTAAACTGATATATAATGTAATTT
TAAAACGTCATTCAATTGGGGACGTTCGGTTAATCAGCTGTAAAACTACCACGGAATGCAAGGCTTTAATAGATCGA
GGATATGATATATTGCACCCAAATTGGGTACTCGATTGTATAGCATATAAGAGGCTCATCCTGATCGAGCCCAATTA
TTGCTTTAACGTCTCTCAAAAAATGAGAGCCGTCGCTGAAAAAAGGGTAGATTGTTTGGGTGATAGTTTTGAAAATG
ACATTTCGGAAACCAAACTGTCATCATTGTATAAATCACAACTAAGTCTACCACCGATGGGGGAACTCGAGATAGAT
TCTGAGGTTCGGCGGTTTCCATTATTTTTATTCTCCAACAGGATTGCATACGTACCACGTCGCAAAATTAGCACAGA
AGATGACATTATAGAAATGAAAATTAAGTTGTTTGGTGGAAAAATAACAGATCAACAGTCACTTTGTAACTTAATAA
TTATACCATATACTGATCCTATTTTGAGGAAAGACTGCATGAATGAGGTACACGAAAAAATAAAAGAACAAATAAAG
GCTTCTGATACTATACCGAAAATAGCCAGGGTCGTTGCCCCTGAATGGGTGGATCATTCTATTAATGAAAACTGTCA
AGTGCCTGAGGAGGATTTCCCCGTAGTCAACTACTGAGTCCGCTTTTGGCTTCCTCACTTATTTCTTCTTTCTCTAT
ATATATAAAGAGTGAGTGTTGTATATAAGTAAATACATCTGGTATATTATTTTTTTTTTTCTTCATTCTTAAAAAG
TATTAATATCGATCAGCAAAAAAAAATTAACAAAAAAGTTCCTTATTATATCTGCGTAGAAGTACTTATTTCTGCTC
CACCTTTGGAGTATTTTTCCAAAATTGTGATGCCAAATGAGTAATGAAATAGAATTTCTTGTGTGGATCGTCATTAT
CGATTTGGTTTGGACAGTTAAAGGAGAATTTTGGAAAGACCAGAACATCACTTAGCGGATCCTCACTAAAGGGAACA
AAAGCTGGAGCTTCTTTGAAAAGATAATGTATGATTATGCTTTCACTCATATTTATACAGAAACTTGATGTTTTCTT
TCGAGTATATACAAGGTGATTACATGTACGTTTGAAGTACAACTCTAGATTTTGTAGTGCCCTCTTGGGCTAGCGGT
AAAGGTGCGCATTTTTTCACACCCTACAATGTTCTGTTCAAAAGATTTTGGTCAAACGCTGTAGAAGTGAAAGTTGG
TGCGCATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGATAAATGATCGCCATCCTATGGAACTGCCTGTTTTAG
AGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGGTGCTTTTT
TTGTTTTTTTATGTCTTCGAGTCATGTAATTAGTTATGTCACGCCTCGAGTGGTGCGTTTTGCGGAGGCTTAATTTTT
TGAAGTTTATTTAATACTATCCTACATATGTACATTAAATACTTCCGTAACGTTTATCAATAAGAGTGGAAGATGCG
CAATTATATTCAAAAGATTGGCCAGTCAATTAACTTAAGGAAAAAATTTACTGCGCGGCTGTGGTCCCATTTGAAAC
GGGGACGTTAGTGTTGCTTTTGTGTTTCCTTCTTCTTCTTTTTTTCTTTTCTTTAGAAGTATCTTCTTTGTCTTTCA
CCTCATCGCTGGGCGAGTTGACTTTCTTCTTCTTCTTTACGCTCAAAGGATTAGGAGCCTTTGGACCAAGCTTTCTT
TTCTTAGTGATAGATTCCTTTCCTGATCCATCACCACTTTCTTGAAGTTTTTCAATATTAGGATCATTGAGACCCTT
ATACAATTTTTGTTCTTCAGTAATCTTGCTTGCCTTGGCACTCGCCGTACTGAGAGGTTCCATAACCATAACAGACC
TGGTTAAATGGATTAATGGAACACCCGGAACCGTCCTCAGCTTCCTCCTCAAATCTATGTCCTGAGAAGCAACCACA
TACCTGTGTTTATTTGCACCGC
(589 bp 3′ UTR to *DNL4* shown)

GFY-3856

The *HIS3* locus is identical to GFY-2383.

*prDNL4::prMX::**SpHIS5**::MX(t)::DNL4(t)*
[Additional sequence from original yeast genome deletion strain library]

(590 bp 5′ UTR to *DNL4* shown)

GGATGATGGGTAACATAAATAGAAGGGGTAATGGTTCGCAATCTGACACATCAGAGAGTGAGGAAAACTCAGAACAA
TCTGATTTGGAAGGCAATAATCAATGTATTGAATATGACTCTTTAGGTAATGCTATTCGTATAGATAACATGAAAAG
CAGGGAAGCGCAATCTGAGGAATCAGAAGACGAGGAAAGTGGTTCAAAAGAAAATGGAGAGCCTTTAAGTTATGACC
CCTTAGGCAATTTAATTCGATAGGTGATTAAATAGGCTGAAATCAGTGTTTAGTAACTACGTACGTTGTACATGTAA
CATTGTGATATAAATCGTAAGATTCGCCGAGTATAGATCAATAATATCGGTTTCATCACTTACGTTGTTTGTGCAGT
ACTAGAGTTAAGATCGTTTTCGATCCCTTATTTTCTTCTTTTTTCCTTTTTTTTGTTATTTTTCTCTTTTTACCTTT
TGTCACCATATTAAATCTTTAAACAAATCTAACTATGAAAAAATCCTTTAAACATATGTTAATATGTGGAAAATAAA
TACTAAAATAAAAATCTAGAACTGAAGGAAATAGTAACGGATTATTTAGGTATGGATGTCCACGAGGTCTCTCGCAC
GATTACAACATCTGACGTACGCTGCAGGTCGACGGATCCCCGGGTTAATTAAGGCGCGCCAGATCTGTTTAGCTTGC
CTCGTCCCCGCCGGGTCACCCGGCCAGCGACATGGAGGCCCAGAATACCCTCCTTGACAGTCTTGACGTGCGCAGCT
CAGGGGCATGATGTGACTGTCGCCCGTACATTTAGCCCATACATCCCCATGTATAATCATTTGCATCCATACATTTT
GATGGCCGCACGGCGCGAAGCAAAAATTACGGCTCCTCGCTGCAGACCTGCGAGCAGGGAAACGCTCCCCTCACAGA
CGCGTTGAATTGTCCCCACGCCGCGCCCCTGTAGAGAAATATAAAAGGTTAGGATTTGCCACTGAGGTTCTTCTTTC
ATATACTTCCTTTTAAAATCTTGCTAGGATACAGTTCTCACATCACATCCGAACATAAACAACC**ATG**GGTAGGAGGG
CTTTTGTAGAAAGAAATACGAACGAAACGAAAATCAGCGTTGCCATCGCTTTGGACAAAGCTCCCTTACCTGAAGAG
TCGAATTTTATTGATGAACTTATAACTTCCAAGCATACAAACCAAAAGGGAGAACAAGTAATCCAAGTAGACACGGG
AATTGGATTCTTGGATCACATGTATCATGCACTGGCTAAACATGCAGGCTGGAGCTTACGACTTTACTCAAGAGGTG
ATTTAATCATCGATGATCATCACACTGCAGAAGATACTGCTATTGCACTTGGTATTGCATTCAAGCAGGCTATGAGT
AACTTTGCCGGCGTTAAAAGATTTGGACATGCTTATTGTCCACTTGACGAAGCTCTTTCTAGAAGCGTAGTTGACTT
GTCGGGACGGCCCTATGCTGTTATCGATTTGGGATTAAAGCGTGAAAAGGTTGGGGAATTGTCCTGTGAAATGATCC
CTCACTTACTATATTCCTTTTCGGTAGCAGCTGGAATTACTTTGCATGTTACCTGCTTATATGGTAGTAATGACCAT
CATCGTGCTGAAAGCGCTTTTAAATCTCTGGCTGTTGCCATGCGCGCGGCTACTAGTCTTACTGGAAGTTCTGAAGT
CCCAAGCACGAAGGGAGTGTTGTAAAGAGTACTGACAATAAAAAGATTCTTGTTTTCAAGAACTTGTCATTTGTATA
GTTTTTTTATATTGTAGTTGTTCTATTTTAATCAAATGTTAGCGTGATTTATATTTTTTTTCGCCTCGACATCATCT
GCCCAGATGCGAAGTTAAGTGCGCAGAAAGTAATATCATGCGTCAATCGTATGTGAATGCTGGTCGCTATACTGCTG
TCGATTCGATACTAACGCCGCCATCCAGTGTCGAAAACAAGCTCGAATTCATCGATTGTCTGATTGTGCGTGACCCC
TACGAGACCGACACCGTGATGGTGCGTTTTGCGGAGGCTTAATTTTTTGAAGTTTATTTAATACTATCCTACATATG
TACATTAAATACTTCCGTAACGTTTATCAATAAGAGTGGAAGATGCGCAATTATATTCAAAAGATTGGCCAGTCAAT
TAACTTAAGGAAAAAATTTACTGCGCGGCTGTGGTCCCATTTGAAACGGGGACGTTAGTGTTGCTTTTGTGTTTCCT
TCTTCTTCTTTTTTTCTTTTCTTTAGAAGTATCTTCTTTGTCTTTCACCTCATCGCTGGGCGAGTTGACTTTCTTCT
TCTTCTTTACGCTCAAAGGATTAGGAGCCTTTGGACCAAGCTTTCTTTTCTTAGTGATAGATTCCTTTCCTGATCCA
TCACCACTTTCTTGAAGTTTTTCAATATTAGGATCATTGAGACCCTTATACAATTTTTGTTCTTCAGTAATCTTGCT
TGCCTTGGCACTCGCCGTACTGAGAGGTTCCATAACCATAACAGACCTGGTTAAATGGATTAATGGAACACCCGGAA
CCGTCCTCAGCTTCCTCCTCAAATCTATGTCCTGAGAAGCAACCACATACCTGTGTTTATTTGCACCGC

(589 bp 3′ UTR to *DNL4* shown)

pGF-V809

pRS425 + sgRNA(u2)

prSNR52::***crRNA***::*tracrRNA*::***SUP4(t)***

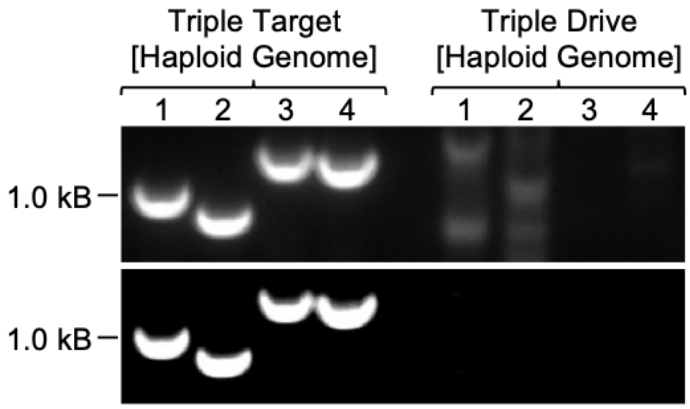
GGATCCtcactaaagggaacaaaagctggagcttctttgaaaagataatgtatgattatgctttcactcatatttat
acagaaacttgatgttttctttcgagtatatacaaggtgattacatgtacgtttgaagtacaactctagattttgta
gtgccctcttgggctagcggtaaaggtgcgcatttttcacaccctacaatgttctgttcaaaagattttggtcaaa
cgctgtagaagtgaaagttggtgcgcatgtttcggcgttcgaaacttctccgcagtgaaagataaatgatc**GCTGTT**
CGTGTGCGCGTCCTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGC
ACCGAGTCGGTGGTGC**TTTTTTTGTTTTTTATGTCT**tcgagtcatgtaattagttatgtcacgc**CTCGAG**

pGF-V1220

pRS425 + sgRNA(u1)

prSNR52::***crRNA***::*tracrRNA*::***SUP4(t)***

GGATCCtcactaaagggaacaaaagctggagcttctttgaaaagataatgtatgattatgctttcactcatatttat
acagaaacttgatgttttctttcgagtatatacaaggtgattacatgtacgtttgaagtacaactctagattttgta
gtgccctcttgggctagcggtaaaggtgcgcatttttcacaccctacaatgttctgttcaaaagattttggtcaaa
cgctgtagaagtgaaagttggtgcgcatgtttcggcgttcgaaacttctccgcagtgaaagataaatgatc**CGGTGG**
ACTTCGGCTACGTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGC
ACCGAGTCGGTGGTGC**TTTTTTTGTTTTTTATGTCT**tcgagtcatgtaattagttatgtcacgc**CTCGA**



PCR		Primers	Expected Size
1	Target 1	F4/R4	935 bp
2	Target 2	F5/R5	720 bp
3	Target 3	F9/R6	1424 bp
4	Target 3	F3/R6	1348 bp

Figure B. S3. Analysis of haploid genomes of the triple drive and triple target yeast strains by diagnostic PCRs.

Purified chromosomal preparations of GFY-3596 (targets) and GFY-3675 (drives) haploid yeast strains were analyzed by PCR at each locus (Targets 1-3). Agarose (1%) gels were imaged (*left*) with equal loads of PCR samples. Two images (*top*, unedited) of the same DNA gel are presented. Oligonucleotides used can be found in Supplemental Table S5 and the primers correspond to the gene drive system presented in Fig. 2A. The expected fragments sizes are illustrated (*right*). For the chosen PCRs at each target locus, identical reactions using the haploid triple drive genome as a template does not result in amplification of the expected fragment. All PCR reactions were performed using identical conditions.

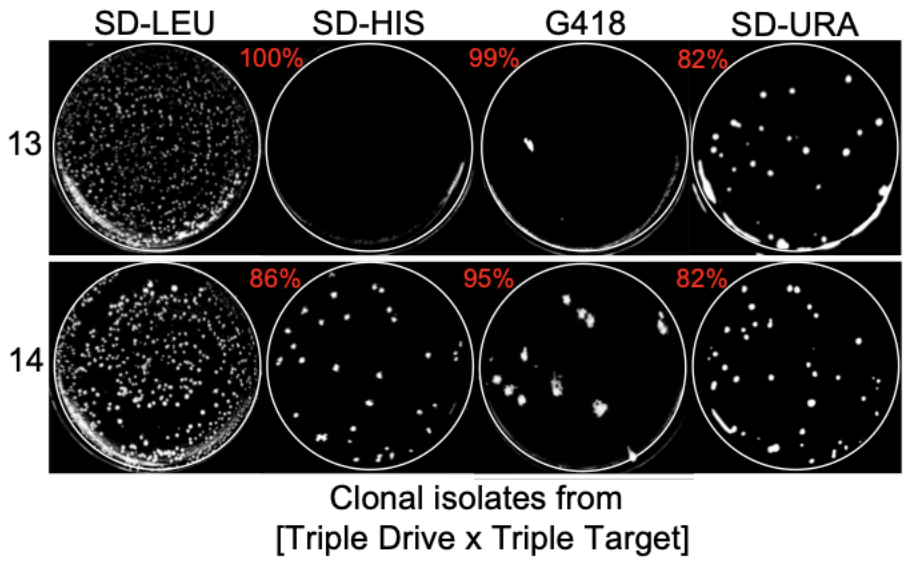


Figure B. S4. Re-activation of gene drive system in clonal isolates displaying incomplete initial drive activity.

Two isolates (13,14) from the [triple drive x triple target] cross displayed imperfect drive activity by both growth and diagnostic PCR (Fig. 2F). Both clonal isolates were confirmed as diploids as previously described[1]. Isolate-13 had maintained the sgRNA(u1) plasmid (marked with *LEU2*), and had activated Drive 1 at the *HIS3* locus (sensitivity on SD-HIS). Isolate-14 had lost the sgRNA(u1) plasmid when re-tested on SD-LEU medium; the plasmid was transformed back into this strain for a second round of activation. As in Fig. 2B, Cas9 expression was induced by culturing in galactose for 5 hr followed by plating onto SD-LEU medium for 2 days. Next, yeast were transferred by replica plating to SD-HIS, G418, and SD-URA medium and incubated for an additional 24 hr prior to imaging. The percentage of yeast colonies sensitive to each growth condition is displayed (red text).

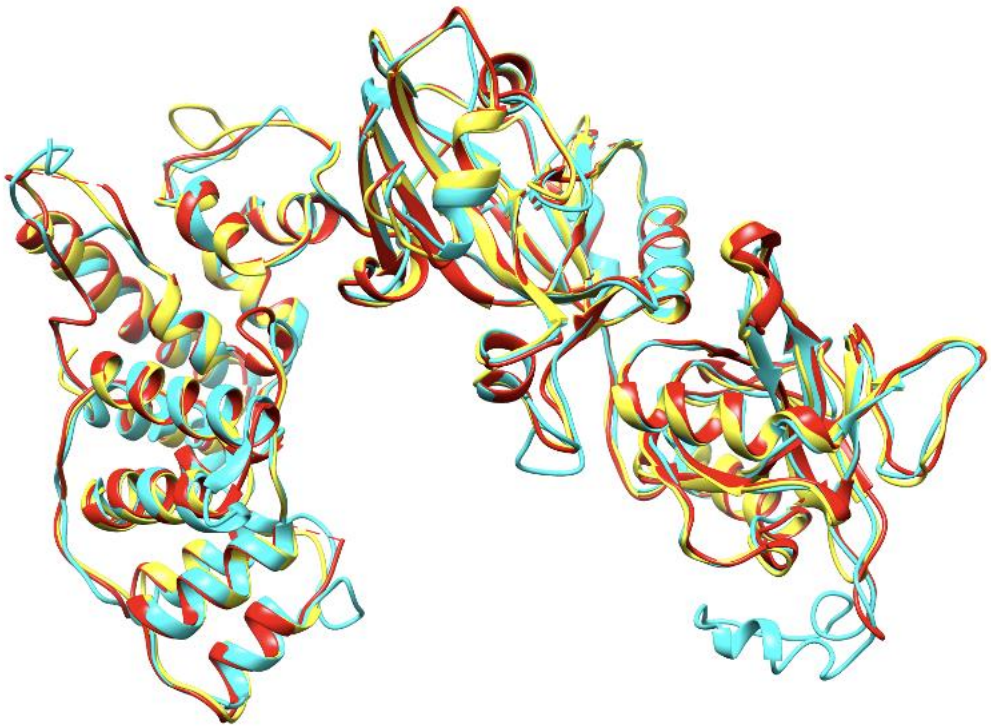
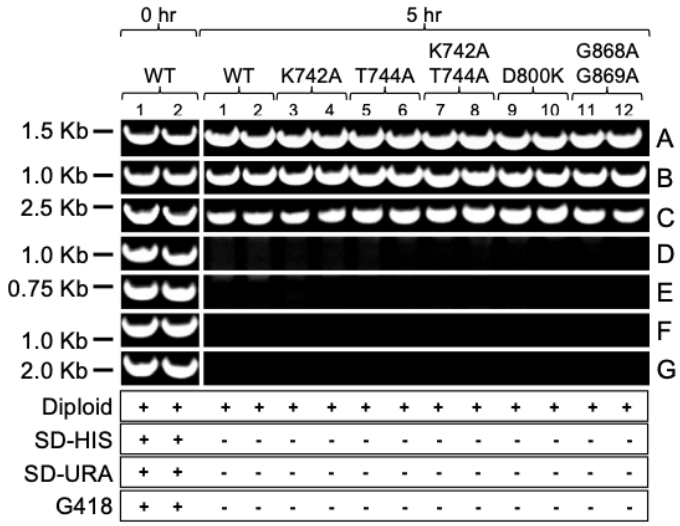


Figure B.S5. Conservation of the N-terminal domain of DNA Ligase IV.

The primary sequences of the *S. cerevisiae* Dnl4 (1-648, teal), and *A. gambiae* Ligase IV (1-591, yellow) were modeled against the crystal structure of the N-terminus of human Lig4 (1-605, red) (PDB:3W1B) using I-TASSER[2] and aligned using MatchMaker in Chimera[3]. Each I-TASSER model was individually aligned against the human Lig4 N-terminal structure.



PCR		Primers	Expected Size
A	Drive 1	F1/R1	1409 bp
B	Drive 2	F2/R2	1147 bp
C	Drive 3	F8/R3	2344 bp
D	Target 1	F4/R4	935 bp
E	Target 2	F5/R5	720 bp
F	Target 3	F9/R6	1424 bp
G	Target 3	F9/R8	2094 bp

Figure B.S6. Further analysis of clonal isolates from the triple gene drive harboring *dnl4* replacement alleles.

Clonal isolates (2 per genotype) were obtained from the gene drive analysis from Fig. 4F. All yeast were re-tested on each media type for growth and ploidy status was confirmed as diploid (*below*). The WT control (GFY-3875) was mated to the triple target strain (GFY-3596), diploids were selected (three consecutive rounds), and chromosomal DNA was isolated from clonal samples (no galactose or raffinose/sucrose treatment). For all other drives (1-12), a 5 hr galactose induction was used (Fig. 4F). Diagnostic PCRs were performed on isolated genomic DNA for each locus for samples before (0 hr) and after (5 hr) drive activation. For the *DNL4* locus, one PCR (C) was used to examine the *dnl4::CDC11(t)::sgRNA(Kan)* construct whereas two PCRs (F and G) were used to assay for the presence or absence of the *dnl4*Δ*::Kan*^{R} target cassette. The oligonucleotides used can be found in Supplemental Table S5. These data illustrate that the three target loci have been lost following gene drive activation. Images of DNA bands were cropped from larger gels or from independent gels (separated by white lines). The unedited gel images can be found in Supplemental Fig. S9.

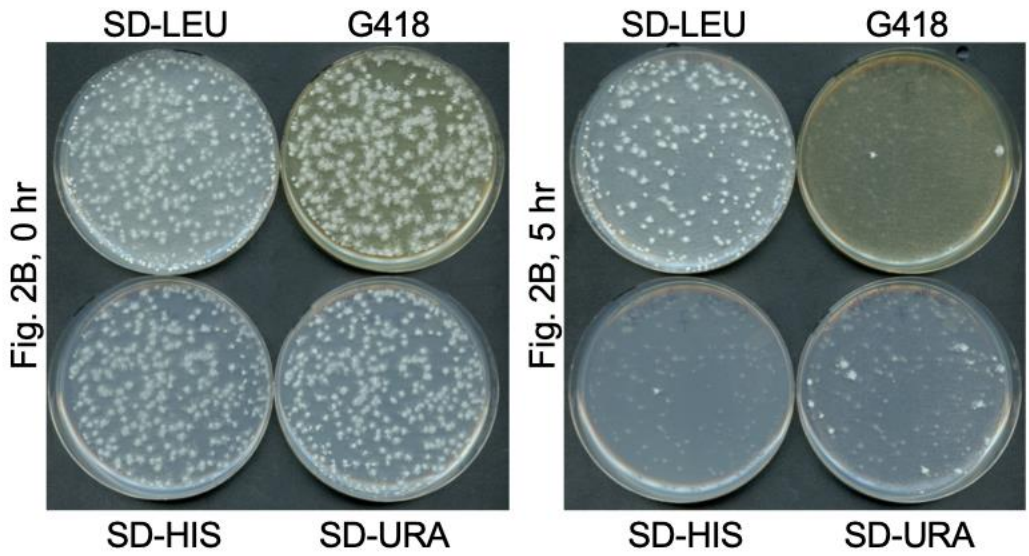


Figure B.S7. Original images of yeast agar plates used within this study.

The two sets of plates used in Fig. 2B are included, unedited. Plates were scanned and not processed further.

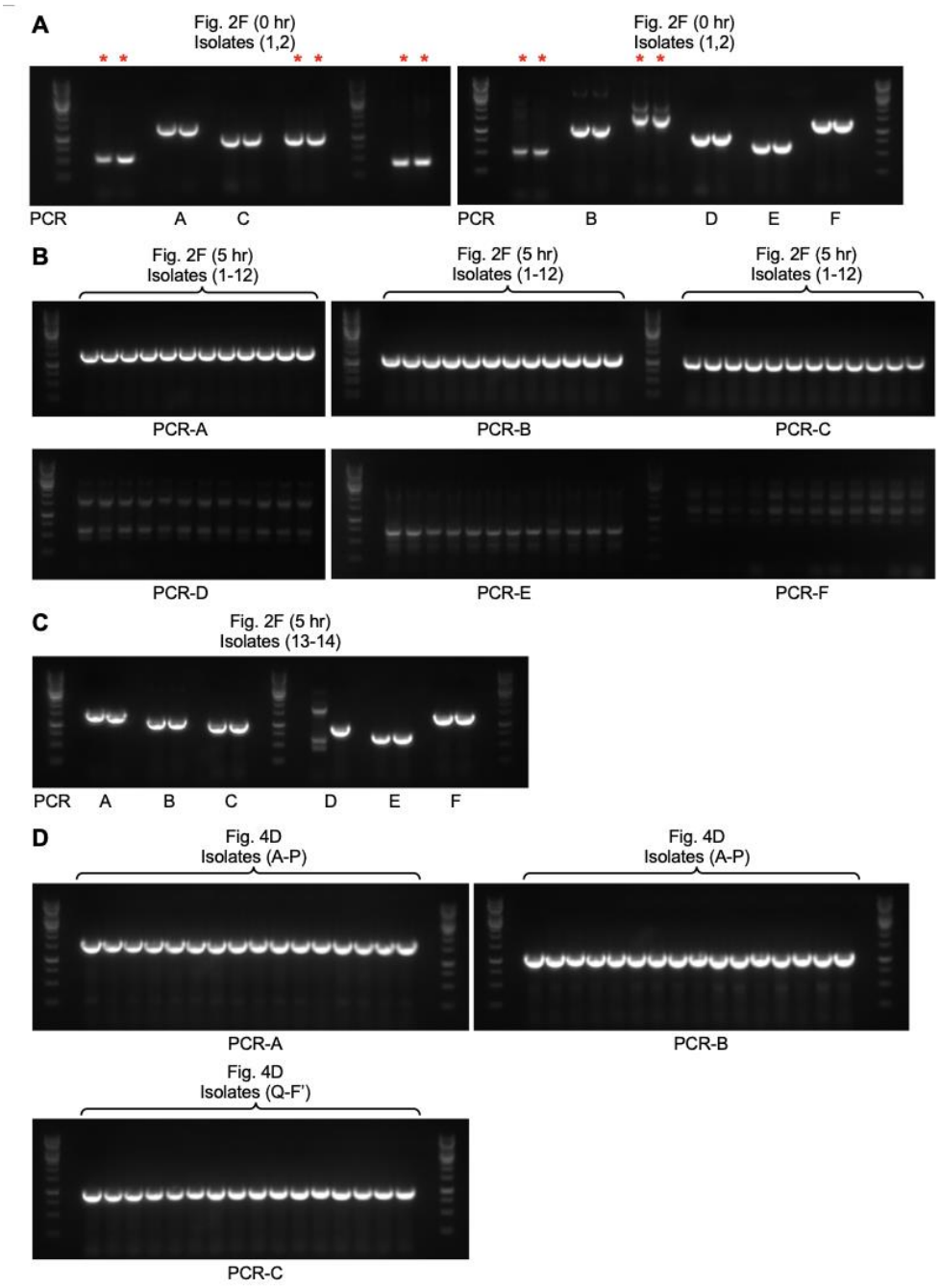


Figure B.S8. Original images of DNA agarose gels used for PCRs from Fig. 2.2F and Fig. 2.4D are included with molecular markers.

(A) PCRs from the 0 hr time point (2 isolates) are included on unedited gels. The red asterisk designates PCR lanes not included within the final figures. (B) PCRs from the 5 hr time point (12 isolates) are included on unedited gels. (C) A gel containing the PCRs for isolates 13 and 14 from Fig. 2.2F. (D) PCRs from Fig. 2.4D are included on three unedited DNA gels. Isolates (A-P) are included on two gels whereas isolates (Q-F') are included on a third gel. All gel images shown were collected from the Invitrogen E-Gel™ Imager (ThermoFisher Scientific), cropped for clarity, but were not processed by any method.

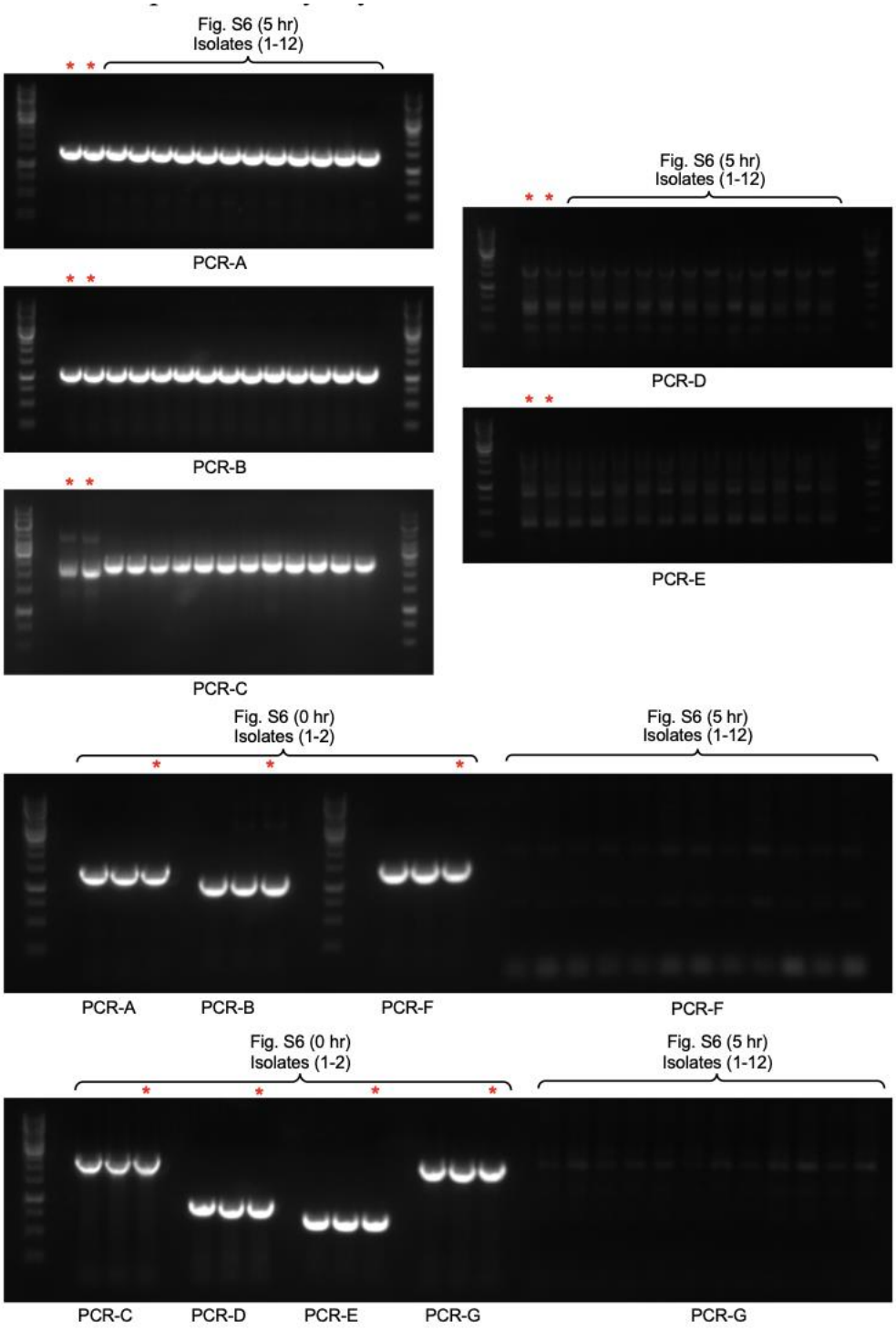


Figure B.S9. Original DNA gels from Supplemental Fig. B.S6.

Molecular markers are included. PCRs from both 0 hr and 5 hr time points are included. Red asterisk, PCR lanes not included in the final figure. All gel images shown were collected as in Fig. B.S8.

Table B.S1. Yeast strains used in this study.

Strain	Genotype	Reference
BY4741	*MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*	4
BY4742	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*	4
GFY-3675[a]	BY4741; *his3*Δ*::prHIS3:(u2)::prGAL1/10::SpCas9::NLS::ADH1(t):(u2)::HIS3(t)* *shs1*Δ*::prSHS1::sgRNA(GFP)::SHS1(t)* *dnl4*Δ*::prDNL4::sgRNA(Kan)::DNL4(t)*	This study
GFY-3206[b]	BY4742; *his3*Δ*::prHIS3::(u1)::prCDC12::mCherry::NLS::SHS1(t)::prCCW12::SpHIS5::MX(t)::(u1)::HIS3(t)*	1
GFY-3593[c]	BY4742; *shs1*Δ*::prSHS1::SHS1::GFP::CDC10(t)::prMX::CaURA3::SHS1(t)*	This study
GFY-3264b[d]	BY4742; *dnl4*Δ*::prDNL4::prMX::Kan^R::MX(t)::DNL4(t)*	4
GFY-3578	BY4742; *his3*Δ*::prHIS3::(u1)::prCDC12::mCherry::NLS::SHS1(t)::prCCW12::SpHIS5::MX(t)::(u1)::HIS3(t)* *shs1*Δ*::prSHS1::SHS1::GFP::CDC10(t)::prMX::CaURA3::SHS1(t)*	This study
GFY-3594	BY4742; *his3*Δ*::prHIS3::(u1)::prCDC12::mCherry::NLS::SHS1(t)::prCCW12::SpHIS5::MX(t)::(u1)::HIS3(t)* *dnl4*Δ*::prDNL4::prMX::Kan^R::MX(t)::DNL4(t)*	This study

GFY-3623	BY4742; *shs1Δ::prSHS1::SHS1::GFP::CDC10(t)::prMX::CaURA3::SHS1(t) dnl4Δ::prDNL4::prMX::Kan^R::MX(t)::DNL4(t)*	This study
GFY-3596[e]	BY4742; *his3Δ::prHIS3::(u1)::prCDC12::mCherry::NLS::SHS1(t)::prCCW12::SpHIS5::MX(t)::(u1)::HIS3(t) shs1Δ::prSHS1::SHS1::GFP::CDC10(t)::prMX::CaURA3::SHS1(t) dnl4Δ::prDNL4::prMX::Kan^R::MX(t)::DNL4(t)*	This study
GFY-3611	BY4741; *his3Δ::prHIS3:(u2)::prGAL1/10::SpCas9::NLS::ADH1(t):(u2)::HIS3(t) shs1Δ::prSHS1::sgRNA(GFP)::SHS1(t) DNL4*	This study
GFY-2383	BY4741; *his3Δ::prHIS3:(u2)::prGAL1/10::SpCas9::NLS::ADH1(t):prMX::Kan^R::MX(t)::(u2)::HIS3(t)*	1
GFY-3850[f]	BY4741; *his3Δ::prHIS3:(u2)::prGAL1/10::SpCas9::NLS::ADH1(t):prMX::Kan^R::MX(t)::(u2)::HIS3(t) dnl4Δ::prDNL4::dnl4(K742A)::DNL4(t)*	This study
GFY-3851	BY4741; *his3Δ::prHIS3:(u2)::prGAL1/10::SpCas9::NLS::ADH1(t):prMX::Kan^R::MX(t)::(u2)::HIS3(t) dnl4Δ::prDNL4::dnl4(T744A)::DNL4(t)*	This study
GFY-3852	BY4741; *his3Δ::prHIS3:(u2)::prGAL1/10::SpCas9::NLS::ADH1(t):prMX::Kan^R::MX(t)::(u2)::HIS3(t) dnl4Δ::prDNL4::dnl4(K742A T744A)::DNL4(t)*	This study
GFY-3853	BY4741; *his3Δ::prHIS3:(u2)::prGAL1/10::SpCas9::NLS::ADH1(t):prMX::Kan^R::MX(t)::(u2)::HIS3(t) dnl4Δ::prDNL4::dnl4(L750::STOP)::DNL4(t)*	This study
GFY-3854	BY4741; *his3Δ::prHIS3:(u2)::prGAL1/10::SpCas9::NLS::ADH1(t):prMX::Kan^R::MX(t)::(u2)::HIS3(t) dnl4Δ::prDNL4::dnl4(D800K)::DNL4(t)*	This study

GFY-3855	BY4741; *his3Δ::prHIS3:(u2)::prGAL1/10::SpCas9::NLS::ADH1(t):prMX::Kan^R::MX(t)::(u2)::HIS3(t) dnl4Δ::prDNL4::dnl4(G868A G869A)::DNL4(t)*	This study
GFY-3864	BY4741; *his3Δ::prHIS3:(u2)::prGAL1/10::SpCas9::NLS::ADH1(t):prMX::Kan^R::MX(t)::(u2)::HIS3(t) dnl4Δ::prDNL4::DNL4(WT)::DNL4(t)*	This study
GFY-3856[g]	BY4741; *his3Δ::prHIS3:(u2)::prGAL1/10::SpCas9::NLS::ADH1(t):prMX::Kan^R::MX(t)::(u2)::HIS3(t) dnl4Δ::prDNL4::prMX::SpHIS5::MX(t)::DNL4(t)*	This study
GFY-3865[h]	BY4741; *his3Δ::prHIS3:(u2)::prGAL1/10::SpCas9::NLS::ADH1(t):(u2)::HIS3(t) shs1Δ::prSHS1::sgRNA(GFP)::SHS1(t) dnl4Δ::prDNL4::dnl4(K742A)::CDC11(t)::sgRNA(Kan)::DNL4(t)*	This study
GFY-3866	BY4741; *his3Δ::prHIS3:(u2)::prGAL1/10::SpCas9::NLS::ADH1(t):(u2)::HIS3(t) shs1Δ::prSHS1::sgRNA(GFP)::SHS1(t) dnl4Δ::prDNL4::dnl4(T744A)::CDC11(t)::sgRNA(Kan)::DNL4(t)*	This study
GFY-3867	BY4741; *his3Δ::prHIS3:(u2)::prGAL1/10::SpCas9::NLS::ADH1(t):(u2)::HIS3(t) shs1Δ::prSHS1::sgRNA(GFP)::SHS1(t) dnl4Δ::prDNL4::dnl4(K742A T744A)::CDC11(t)::sgRNA(Kan)::DNL4(t)*	This study
GFY-3871	BY4741; *his3Δ::prHIS3:(u2)::prGAL1/10::SpCas9::NLS::ADH1(t):(u2)::HIS3(t) shs1Δ::prSHS1::sgRNA(GFP)::SHS1(t) dnl4Δ::prDNL4::dnl4(D800K)::CDC11(t)::sgRNA(Kan)::DNL4(t)*	This study

GFY-3872	BY4741; *his3*Δ*::prHIS3:(u2)::prGAL1/10::SpCas9::NLS::ADH1(t):(u2)::HIS3(t)* *shs1*Δ*::prSHS1::sgRNA(GFP)::SHS1(t)* *dnl4*Δ*::prDNL4::dnl4(G868A G869A)::CDC11(t)::sgRNA(Kan)::DNL4(t)*	This study
GFY-3875	BY4741; *his3*Δ*::prHIS3:(u2)::prGAL1/10::SpCas9::NLS::ADH1(t):(u2)::HIS3(t)* *shs1*Δ*::prSHS1::sgRNA(GFP)::SHS1(t)* *dnl4*Δ*::prDNL4::DNL4(WT)::CDC11(t)::sgRNA(Kan)::DNL4(t)*	This study

[a]Strain GFY-3675 was derived from GFY-2383. First, *S. pyogenes* Cas9 expression was activated as previously described[1] followed by transformation of the sgRNA(Kan) plasmid (pGF-V1642) and an amplified PCR fragment: *ADH1(t)::(u2)::HIS3(t)* which included 168 bp of the *ADH1* terminator, the 23 bp unique (u2) sequence (5' GCTGTTCGTGTGCGCGTCCT**GGG**_3') where the PAM is in bold[5], and 263 bp of the *HIS3* terminator. Yeast were plated on SD-LEU medium. DSB formation followed by genomic repair via HR removed the Kan^R cassette entirely. Editing was followed by strain propagation on rich media (dextrose) to inhibit Cas9 expression and allow for loss of the high-copy sgRNA-containing plasmid. Second, *SHS1* was deleted using an amplified knock-out cassette (from pGF-V170) and 500 bp of flanking UTR. Next, Cas9-based editing was performed using the sgRNA(Kan) plasmid and repair DNA: *prSHS1::sgRNA(GFP)::SHS1(t)*. Third, *DNL4* was deleted using the knock-out cassette from strain GFY-3264b. Cas9 editing allowed for integration of *prDNL4::sgRNA(Kan)::DNL4(t)* (amplified from pGF-IVL1498) at the native locus. All strains and intermediates were confirmed by diagnostic PCR and DNA sequencing of all manipulated loci. The final strain was propagated for multiple weeks ensuring complete loss of both residual Cas9 and high copy plasmid used for construction.

[b]The unique (u1) sequence (5' ATGACGGTGGACTTCGGCTACGTA**GGG**CGATT 3') includes a PAM (bold) and the 20 bp target site for *Sp*Cas9 (underlined)[5]. The *HIS5* gene (functional equivalent to *S. cerevisiae HIS3*) is from *Schizosaccharomyces pombe.*

[c]Strain GFY-3593 was constructed by direct transformation and integration of a PCR fragment: *prSHS1::SHS1::GFP::CDC10(t)::prMX::CaURA3::SHS1(t)*. The *Candida albicans URA3* gene (from plasmid JT-2868) does not include the standard *MX(t)* sequence and, instead, uses the native *SHS1* 3' UTR.

[d]The haploid (MATα) *dnl4*Δ*::*Kan^R strain was confirmed using multiple diagnostic PCR. The isogenic isolate was from the yeast haploid genome deletion collection (UC Berkeley).

[e]Strain GFY-3596 was derived from GFY-3206 and required multiple rounds of HR-based integration at the *SHS1* and *DNL4* loci.

[f]Strains GFY-3850 to 3855, and GFY-3864 were constructed by first generating a plasmid construct containing the last 1559 bp of the *DNL4* coding sequence and 589 bp of the *DNL4* 3' UTR sequence. Consecutive rounds of a modified PCR mutagenesis protocol[6] introduced 6 silent polymorphisms into the *DNL4* gene; these occur within codons 699, 700, 701, 936, 937 and 938. These were designed within two separate native Cas9 target sites (labeled as "A" and "B") within *DNL4*: (5' GACTATGTCACTGAAGATAC**TGG** 3') and (5' **CCT**GAGGAGGATTTCCCCGTAGT 3') where the PAM (bold) and target sequences (underlined) are marked. Additional mutation(s) were then added to *DNL4*. Second, strain GFY-2383 was induced for Cas9 expression and transformed with two separate high-copy plasmids (pYY-DNL4(A) and pYY-DNL4(B), marked with *LEU2/URA3*, respectively) to target *DNL4* at the dual targets (A and B) as well as repair DNA (594 bp of upstream homology of the cleavage site (A) and 589 bp of downstream homology within the terminator) PCR amplified from pYY-IVL6 to pYY-IVL10, pYY-IVL12, and pYY-IVL14 to yield the seven yeast strains. Viable yeast were selected on SD-LEU-URA medium and confirmed for subsequent loss of both sgRNA-containing plasmids by growth on rich medium.

[g]Strain GFY-3856 was generated by first switching the selectable marker in GFY-3264b from Kan^R to *SpHIS5*. Second, the entire *dnl4*Δ*::SpHIS5* locus was PCR amplified and transformed into GFY-2383.

[h]Strains GFY-3865 to GFY-3867, GFY-3871, GFY-3872, and GFY-3875 were generated using a similar methodology to GFY-3850 with several modifications. First, Cas9-based editing was performed in GFY-3611 as the parent strain. Second, the integrating construct included the *DNL4* coding sequence followed by 327 bp of the *CDC11* terminator, the 455 bp sgRNA(Kan) expression construct, and 589 bp of the *DNL4* 3' UTR. Third, donor PCRs were amplified from pYY-IVL1 to pYY-IVL3, pYY-IVL5, pYY-IVL11, and pYY-IVL13 to generate appropriate strains.

Table B.S2. Plasmids used in this study.

Plasmid	Description	Reference
pRS425	2μ, *LEU2*	7
pRS426	2μ, *URA3*	7
pGF-V1220[a]	pRS425; *prSNR52::Sp-sgRNA(u1)::SUP4(t)*	1
pGF-V809[b]	pRS425; *prSNR52::Sp-sgRNA(u2)::SUP4(t)*	1
pYY-DNL4(A)[c]	pRS425; *prSNR52::Sp-sgRNA(u2)::SUP4(t)*	This study
pYY-DNL4(B)[d]	pRS426; *prSNR52::Sp-sgRNA(u2)::SUP4(t)*	This study

[a]The sgRNA(u1) sequence is 5' CGGTGGACTTCGGCTACGTA 3'[5]. All sgRNA constructs include 269 bp of the *SNR52* promoter, the 79 bp tracrRNA, and the 20 bp *SUP4* terminator, as modeled from[8].
[b]The sgRNA(u1) sequence is 5' GCTGTTCGTGTGCGCGTCCT 3'.
[c]The sgRNA(DNL4-A) sequence is 5' GACTATGTCACGGAGGACAC 3'.
[d]The sgRNA(DNL4-B) sequence is 5' ACTACGGGGAAGTCTTCTTC 3'. The target is present on the non-coding strand at the 3' end of the *DNL4* gene.

Table B.S3. Species used for alignments and phylogeny of DNA Ligase IV.

Species Name (Strain) [Protein Name]	**Amino Acids**	**Reference ID**
Saccharomyces cerevisiae (S288C) [Dnl4]	944	NCBI Ref Seq: NP_014647.1
Candida glabrata [DNA ligase 4]	946	GenBank: KTB12236.1
Candida albicans (P75016) [DNA ligase 4]	928	GenBank: KHC72222.1
Schizosaccharomyces pombe [Lig4]	913	NCBI Ref Seq: NP_587888.2
Neurospora crassa (OR74A) [Mus53]	1050	NCBI Ref Seq: XP_962868.2
Yarrowia lipolytica [hypothetical protein YALI1_D27044g]	956	GenBank: AOW04412.1
Papilio machaon [DNA ligase 4, predicted]	880	NCBI Ref Seq: XP_014356497.1
Drosophila melanogaster [Ligase4]	918	NCBI Ref Seq: NP_572907.1
Anopheles gambiae (str. PEST) [AGAP000623-PA]	914	NCBI Ref Seq: XP_310455.4
Aedes aegypti [AAEL017365-PA, partial]	925	GenBank: EJY57630.1
Zootermopsis nevadensis [DNA ligase 4]	899	NCBI Ref Seq: XP_021923135.1
Danio rerio [DNA ligase 4]	909	NCBI Ref Seq: NP_001096593.1
Xenopus laevis [DNA ligase IV]	911	GenBank: AAL56554.1
Gallus gallus [DNA ligase IV]	912	GenBank: BAB68506.1
Mus musculus [DNA ligase 4]	911	NCBI Ref Seq: NP_795927.2
Rattus norvegicus [DNA ligase 4]	911	NCBI Ref Seq: NP_001099565.1

Homo sapiens [DNA ligase 4 isoform 1]	911	NCBI Ref Seq: NP_002303.2
Bos taurus [DNA ligase 4]	911	NCBI Ref Seq: NP_001178055.1

Table B.S4. Accuracy metrics for I-TASSER modeling of DNA Ligase IV structure.

I-TASSER Template:

Human Lig4 NT (1-605) (PDB:3W1B):

I-TASSER Models	**C-score**[a]	**TM-score**[b]	**RMSD**[c]
Yeast Dnl4 NT (1-648)	0.51	0.78 ± 0.10	8.8 ± 4.0Å
Mosquito Lig4 NT (1-591)	1.92	0.99 ± 0.04	3.7 ± 2.5Å

I-TASSER Template:

Yeast Dnl4 CT (683-939) (PDB:1Z56):

I-TASSER Models	**C-score**	**TM-score**	**RMSD**
Human Lig4 CT (656-911)	0.27	0.75 ± 0.10	5.3 ± 3.4Å
Mosquito Lig4 CT (645-914)	0.10	0.73 ± 0.11	5.8 ± 3.6Å

[a]The C-score is a metric for the confidence of the quality of the predicted I-TASSER model. This score usually ranges from -5 to 2 (higher value, higher confidence)[2].
[b]The TM-score is a metric for structural similarity and scores > 0.5 indicate high confidence [9].
[c]Root-mean-square deviation of atomic positions.

Table B.S5. Oligonucleotides used for diagnostic PCRs.

Oligonucleotide Name	**DNA Sequence (5' to 3')**
F1: *prHIS3* +196 F	GGCCTCCTCTAGTACACTCTATATTTTTTTATGC
R1: Int *S.p.Cas9* +373 R	CATCAACGATGTTACCGAAGATTGGATGTC
F2: *prSHS1* +497 F	GTTCCTCCGGTTTAAACTGATCCATAGTGATAG
R2: *SHS1(t)* -192 R	GCCATATTTAAATTTATCCCTACAATTATTTGACACTGTTTG
F3: *prDNL4* +590 F	GGATGATGGGTAACATAAATAGAAGGGGTAATG
R3: sgRNA clone out R1	CTCGAGGCGTGACATAACTAATTACATGACTC
F4: Int *SpHIS5* F1	GGGAGAACAAGTAATCCAAGTAGACACGGG
R4: *HIS3(term)* -151 R	CGCCTCGTTCAGAATGACACGTATAGAATG
F5: Int GFP F new	GGTGTTCAATGCTTTTCAAGATACCCAGATC
R5: *CDC10(term)* -201 R	CAAACGAGAAGGTGATAGCTGTACGCCAG
R6: Internal Kan R	GAACACTGCCAGCGCATCAACAATATTTTC
F6: *prHIS3* +500 F	TTCTCGACGTGGGCCTTTTTCTTGCCATAT
F7: Internal Kan F	CGGTTGCATTCGATTCCTGTTTGTAATTGTCC
R7: *HIS3(term)* -497 R	GTAACCACCACGACGGTTGTTGCTAAAGAA
F8: Int *DNL4* +1308 F	GTGTGGAGTCAATCAAAAAGTCTTTAGAAGTTGCA
F9: *prDNL4* +666 F	GCTGAAGAGGAATGGTTGCAAAAATTAGAAAAGG
R8: MX clone out R2	ACTGGATGGCGGCGTTAGTATCGAATCGA

References

1 Roggenkamp, E. *et al.* Tuning CRISPR-Cas9 Gene Drives in Saccharomyces cerevisiae. *G3 (Bethesda, Md.)* **8**, 999-1018, doi:10.1534/g3.117.300557 (2018).

2 Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols* **5**, 725-738, doi:10.1038/nprot.2010.5 (2010).

3 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).

4 Brachmann, C. B. *et al.* Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast (Chichester, England)* **14**, 115-132, doi:10.1002/(sici)1097-0061(19980130)14:2<115::aid-yea204>3.0.co;2-2 (1998).

5 Finnigan, G. C. & Thorner, J. mCAL: a new approach for versatile multiplex action of Cas9 using one sgRNA and loci flanked by a programmed target sequence. *G3 (Bethesda, Md.)* **6**, 2147-2156, doi:10.1534/g3.116.029801 (2016).

6 Zheng, L., Baumann, U. & Reymond, J. L. An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic acids research* **32**, e115, doi:10.1093/nar/gnh110 (2004).

7 Christianson, T. W., Sikorski, R. S., Dante, M., Shero, J. H. & Hieter, P. Multifunctional yeast high-copy-number shuttle vectors. *Gene* **110**, 119-122 (1992).

8 DiCarlo, J. E. *et al.* Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic acids research* **41**, 4336-4343, doi:10.1093/nar/gkt135 (2013).

9 Zhang, Y. & Skolnick, J. Scoring function for automated assessment of protein structure template quality. *Proteins* **57**, 702-710, doi:10.1002/prot.20264 (2004).

Appendix C - Supplementary Information for Analysis of CRISPR gene drive design in budding yeast

Yao Yan[1] and Gregory C. Finnigan[1*]

[1]Department of Biochemistry and Molecular Biophysics, Kansas State University, 141 Chalmers Hall, Manhattan, KS 66506 USA

*Correspondence to: Gregory C. Finnigan, Phone: (785) 532-6939; FAX; (785) 532-7278;

E-mail: gfinnigan@ksu.edu

Table C.S1. Yeast strains used in this study.

Strain	Genotype	Reference
BY4741	*MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 LYS2*	1
BY4742	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 MET15*	1
GFY-150[a]	BY4742; *cdc11*Δ*::prMX::Kan^R::MX(t)* + pJT1520 (*prCDC11::CDC11*)	This study
GFY-153[a]	BY4741; *cdc11*Δ*::prMX::Kan^R::MX(t)* + pJT1520 (*prCDC11::CDC11*)	This study
GFY-2383	BY4741; *his3*Δ*::prHIS3:(u2)::prGAL1/10::SpCas9::NLS::ADH1(t):prMX::Kan^R:: MX(t)::(u2)::HIS3(t)*	2
GFY-3733[b]	BY4742; *his3*Δ*::prHIS3::(u1')::prCDC12::mCherry::SHS1(t)::prCCW12:: SpHIS5::MX(t)::(u1')::HIS3(t)*	This study
GFY-4325, GFY-4326[c]	BY4741; *his3*Δ*::prHIS3:(u2')::prGAL1/10::SpCas9::NLS::ADH1(t)::sgRNA(u1) : :prMX::CaURA3::MX(t)::(u2')::HIS3(t)*	This study

[a]Strains GFY-150 and GFY-153 were generated from a previous study[3]. GFY-153 was the parental strain to generate GFY-163. Briefly, the *CDC11* locus was replaced by the Kan^R deletion cassette in BY4741 or BY4742 WT yeast harboring a *URA3*-based covering vector (pJT1520[4]) that expresses a copy of *CDC11* under its native promoter. These strains were used in mating tests (Fig. C.S3).

[b]Strain GFY-3733 is similar to GFY-3207. Artificial CRISPR sites[5] (u1') are positioned flanking the entire cassette with the sequence 5'-T**TTTC**CGGTGGACTTCGGCTACGTA**GGG**AGT-3'. The bold and underlined sequences include the PAM sites for Cas12a/Cpf1 (TTTV at the 5' end) and *S. pyogenes* Cas9 (NGG at the 3' end) with a common target site. Strain GFY-3733 was generated from integration of the cassette from the pGF-IVL1511 vector at the *HIS3* locus in BY4742 yeast. The *HIS5* gene is from fission yeast *S. pombe* and is the functional equivalent of *S. cerevisiae HIS3*. The 5' T for the upstream (u1') site has been artificially added. For the downstream (u1') site, the T already existed within the MX(t) sequence. For the 3' T within the upstream (u1') site, this base was artificially added. For the downstream (u1') site, this T already existed as part of the *HIS3*(t) sequence.

[c]Strains GFY-4325 and GFY-4326 were two separate isolates created in an identical manner. First, BY4741 yeast were transformed with six overlapping PCR fragments that were assembled *in vivo* through selection on rich medium containing G418 (the initial integration contained the prMX-Kan^{R}-MX(t) drug resistance cassette). Second, CRISPR-based editing was performed by activating Cas9 expression (galactose metabolism) and co-transformation of the pGF-V1642 plasmid expressing the sgRNA(Kan^{R}) cassette with a PCR fragment, prMX-*CaURA3*-MX(t) (amplified from plasmid JT2869), to serve as donor DNA. The *URA3* gene is from *C. albicans*. Colonies were selected that grew on SD-LEU. Clonal isolates were tested for loss of G418 resistance and also survival on SD-URA plates. Third, yeast were maintained on non-selection media (multiple plates) to allow for loss of the high-copy sgRNA(Kan^{R}) plasmid (and sensitivity on SD-LEU plates). The final strains were confirmed by diagnostic PCR and Sanger DNA sequencing. This strain includes a 431 bp expression cassette for the *S. pyogenes* sgRNA to target (u1') immediately following the *ADH1*(t) sequence. Additionally, a modified (u2') site was included (similar to the (u2) sites present within strain GFY-2383) with the artificial sequence 5'-T**TTTC**GCTGTTCGTGTGCGCGTCCT**GGG**AGT-3'. Dual sequences for *F. novicida* Cas12a/Cpf1 and *S. pyogenes* Cas9 (PAMs) are in bold and underlined text, respectively.

Table C.S2. Plasmids used in this study.

Plasmid	Description	Reference
pRS425	2μ, *LEU2*	6
pRS426	2μ, *URA3*	6
pRS313	CEN, *HIS3*	7
pRS315	CEN, *LEU2*	7
pGF-V1220[a]	pRS425; *prSNR52::Sp-sgRNA(u1)::SUP4(t)*	2
pGF-V809[b]	pRS425; *prSNR52::Sp-sgRNA(u2)::SUP4(t)*	2
pGF-V1642[c]	pRS425; *prSNR52::Sp-sgRNA(Kan^R)::SUP4(t)*	8
pGF-V2152[d]	pRS425; *prSNR52::Sp-sgRNA(SpHIS5)::SUP4(t)*	This study
pGF-V2153	pRS426; *prSNR52::Sp-sgRNA(SpHIS5)::SUP4(t)*	This study
pGF-V2158[e]	pRS425; *prSNR52::Sp-sgRNA(mCherry)::SUP4(t)*	This study
pGF-V2159	pRS426; *prSNR52::Sp-sgRNA(mCherry)::SUP4(t)*	This study

[a]The (u1/u1') target sequence is 5'-CGGTGGACTTCGGCTACGTA-3'. The guide RNA constructs were modeled from a previous study[9] and contain 269 bp of the *SNR52* promoter, a 79 bp tracrRNA, and the *SUP4* terminator (20 bp) along with variable flanking sequence (approximately 25-35 bp). The sgRNA(u1) is able to target both (u1) and (u1') sequences.

[b]The sgRNA(u2) target sequence is 5'-GCTGTTCGTGTGCGCGTCCT-3'. The sgRNA(u2) is able to target both (u2) and (u2') sequences.

[c]The sgRNA(Kan^R) target sequence is 5'-GCCATCCTATGGAACTGCCT-3'. An alternative name to this vector is "pGF-425+1275."

[d]The sgRNA(*SpHIS5*) target sequence is 5'-ACAAGTAATCCAAGTAGACA-3'. The entire cassette was synthesized (GenScript) as a custom gene into a pUC57 (Kan^R) vector and sub-cloned to pRS425/pRS426 using unique flanking restriction sites.

[e]The sgRNA(mCherry) target sequence is 5'-CAAGGAGTTCATGCGCTTCA-3'. The expression cassette was generated similar to pGF-V1642 using *in vivo* plasmid assembly[10] and unique overlapping oligonucleotides to generate the unique 20 bp guide sequence.[8] Second, the entire cassette was PCR amplified (from pGF-IVL1277) and cloned into the pCR™Blunt II-TOPO® vector (Life Technologies). Third, the cassette was sub-cloned to pRS425/pRS426 using unique restriction sites.

Table S3. Oligonucleotides used in this study.

Oligonucleotide Name	**DNA Sequence (5' to 3')**
F1: Int pr*GAL1/10* +192 F	GGGGTAATTAATCAGCGAAGCGATGATTTTTG
R1: Int *S.p.Cas9* +373 R	CATCAACGATGTTACCGAAGATTGGATGTC
F2: Int *S.p.Cas9* +3653 F	CGGTAGAAAAAGAATGTTAGCTTCAGCTGG
R2: Int Kan R	GAACACTGCCAGCGCATCAACAATATTTTC
F3: pr*HIS3* +196 F	GGCCTCCTCTAGTACACTCTATATTTTTTTATGC
R3: *SHS1*(t) -192 R	GCCATATTTAAATTTATCCCTACAATTATTTGACACTGTTTG
F4: Int *SpHIS5* F1	GGGAGAACAAGTAATCCAAGTAGACACGGG
R4: *HIS3*(t) -151 R	CGCCTCGTTCAGAATGACACGTATAGAATG
F5: pr*LYS2* +258 F	CAATAGTTTTGCCAGCGGAATTCCACTTGC
R5: *LYS2* Int +629 R	GTTATGCAATTGGATGGATCGCTTAGCGC
F6: *LYS2* Int +3796 F	GACTACTTGTATACTTTACACGATTATGGTTACGATG
R6: *LYS2*(t) -367 R	TTCATTTTAGACCCATGGTGGAACCCTAGTG
F7: pr*LYS2* +636 F	GGTAAGTATGCTCATCAATCGTTCGGACTC
R7: *LYS2*(t) -755 R	CGGGCTAAGTATCGATTTGTCTCAACCTGC
F8: *MET15* Int +1253 F	TCTGGTGTTACCAAGGACTTAATTCGTGTCTC
R8: *MET15*(t) -469 R	CGATGGAATTCCAACAGCTTTACTAATCTTTACTTG
F9: pr*MET15* +1066 F	CCACAAAGCTACGAAAATTTGAAGAAAGGTTCC
R9: *MET15*(t) -897 R	CATCTTATAGGACATATTAAACTATGACGACATTGTTGC
F10: Int pr*CCW12* F	CGTACAAGTATTTCTCAGGAGTAAAAAAACCGTTTG
R10: Int *ADH1*(t) R new	CCTGACCTACAGGAAAGAGTTACTCAAGAATAAG

F11: pr*CDC12* +377 F	TGACATTCTGCAAGCTCTTTGAATCTTCCTCAAAA
R11: Ca*URA3* clone out R	TTATAATTGGCCAGTCTTTTTCAAATAAGCATTCCAAC
F12: pr*CDC12* +276 F	GATGGGACATGATGCAGTATCACGATTAGCAA

Figure C.S1. Yeast strains used in this study.

***LYS2* (from BY4741 background)**
prLYS2::***LYS2(WT)***::*LYS2(t)*

TACGGTGGTACCTTTTTGAACTTCGTCTCTTATGCTGCAGGAGCCGTTAGGTTAGCCGCCTTGTCTGGTAATCCAGT
CATTTGGGTTGCAACACATGACTCTATCGGGCTTGGTGAGGATGGTCCAACGCACCAACCTATTGAAACTCTGGCTC
ACTTGAGGGCTATTCCAAACATGCATGTATGGAGACCTGCTGATGGTAACGAAACTTCTGCTGCGTATTATTCTGCT
ATCAAATCTGGTCGAACACCATCTGTTGTGGCTTTATCACGACAGAATCTTCCTCAATTGGAGCATTCCTCTTTTGA
AAAAGCCTTGAAGGGTGGCTATGTGATCCATGACGTGGAGAATCCTGATATTATCCTGGTGTCAACAGGATCAGAAG
TCTCCATTTCTATAGATGCAGCCAAAAAATTGTACGATACTAAAAAAATCAAAGCAAGAGTTGTTTCCCTGCCAGAC
TTTTATACTTTTGACAGGCAAAGTGAAGAATACAGATTCTCTGTTCTACCAGACGGTGTTCCGATCATGTCCTTTGA
AGTATTGGCTACTTCAAGCTGGGGTAAGTATGCTCATCAATCGTTCGGACTCGACGAATTTGGTCGTTCAGGCAAGG
GGCCTGAAATTTACAAATTGTTCGATTTCACAGCGGACGGTGTTGCGTCAAGGGCTGAAAAGACAATCAATTACTAC
AAAGGAAAGCAGTTGCTTTCTCCTATGGGAAGAGCTTTCTAAGTCTGAAGAAGTAAACAGTTCTTTGCTATTTCACA
CTTCCTGGTTGATGGTCACTTGCTGCCTGAAATATATATATATGTATGACATATGTACTTGTTTTCTTTTTTGTGCC
TTTGTTACGTCTATATTCATTGAAACTGATTATTCGATTTTCTTCTTGCTGACCGCTTCTAGAGGCATCGCACAGTT
TTAGCGAGGAAAACTCTTCAATAGTTTTGCCAGCGGAATTCCACTTGCAATTACATAAAAAATTCCGGCGGTTTTTC
GCGTGTGACTCAATGTCGAAATACCTGCCTAATGAACATGAACATCGCCCAAATGTATTTGAAGACCCGCTGGGAGA
AGTTCAAGATATATAAGTAACAAGCAGCCAATAGTATAAAAAAAAAATCTGAGTTTATTACCTTTCCTGGAATTTCAG
TGAAAAACTGCTAATTATAGAGAGATATCACAGAGTTACTCACTA**ATG**ACTAACGAAAAGGTCTGGATAGAGAAGTT
GGATAATCCAACTCTTTCAGTGTTACCACATGACTTTTTACGCCCACAACAAGAACCTTATACGAAACAAGCTACAT
ATTCGTTACAGCTACCTCAGCTCGATGTGCCTCATGATAGTTTTTCTAACAAATACGCTGTCGCTTTGAGTGTATGG
GCTGCATTGATATATAGAGTAACCGGTGACGATGATATTGTTCTTTATATTGCGAATAACAAAATCTTAAGATTCAA
TATTCAACCAACGTGGTCATTTAATGAGCTGTATTCTACAATTAACAATGAGTTGAACAAGCTCAATTCTATTGAGG
CCAATTTTTCCTTTGACGAGCTAGCTGAAAAAATTCAAAGTTGCCAAGATCTGGAAAGGACCCCTCAGTTGTTCCGT
TTGGCCTTTTTGGAAAACCAAGATTTCAAATTAGACGAGTTCAAGCATCATTTAGTGGACTTTGCTTTGAATTTGGA
TACCAGTAATAATGCGCATGTTTTGAACTTAATTTATAACAGCTTACTGTATTCGAATGAAAGAGTAACCATTGTTG
CGGACCAATTTACTCAATATTTGACTGCTGCGCTAAGCGATCCATCCAATTGCATAACTAAAATCTCTCTGATCACC
GCATCATCCAAGGATAGTTTACCTGATCCAACTAAGAACTTGGGCTGGTGCGATTTCGTGGGGTGTATTCACGACAT
TTTCCAGGACAATGCTGAAGCCTTCCCAGAGAGAACCTGTGTTGTGGAGACTCCAACACTAAATTCCGACAAGTCCC
GTTCTTTCACTTATCGCGACATCAACCGCACTTCTAACATAGTTGCCCATTATTTGATTAAAACAGGTATCAAAAGA
GGTGATGTAGTGATGATCTATTCTTCTAGGGGTGTGGATTTGATGGTATGTGTGATGGGTGTCTTGAAAGCCGGCGC
AACCTTTTCAGTTATCGACCCTGCATATCCCCCAGCCAGACAAACCATTTACTTAGGTGTTGCTAAACCACGTGGGT
TGATTGTTATTAGAGCTGCTGGACAATTGGATCAACTAGTAGAAGATTACATCAATGATGAATTGGAGATTGTTTCA
AGAATCAATTCCATCGCTATTCAAGAAAATGGTACCATTGAAGGTGGCAAATTGGACAATGGCGAGGATGTTTTGGC
TCCATATGATCACTACAAAGACACCAGAACAGGTGTTGTAGTTGGACCAGATTCCAACCCAACCCTATCTTTCACAT
CTGGTTCCGAAGGTATTCCTAAGGGTGTTCTTGGTAGACATTTTTCCTTGGCTTATTATTTCAATTGGATGTCCAAA
AGGTTCAACTTAACAGAAAATGATAAATTCACAATGCTGAGCGGTATTGCACATGATCCAATTCAAAGAGATATGTT
TACACCATTATTTTTAGGTGCCCAATTGTATGTCCCTACTCAAGATGATATTGGTACACCGGGCCGTTTAGCGGAAT
GGATGAGTAAGTATGGTTGCACAGTTACCCATTTAACACCTGCCATGGGTCAATTACTTACTGCCCAAGCTACTACA
CCATTCCCTAAGTTACATCATGCGTTCTTTGTGGGTGACATTTTAACAAAACGTGATTGTCTGAGGTTACAAACCTT
GGCAGAAAATTGCCGTATTGTTAATATGTACGGTACCACTGAAACACAGCGTGCAGTTTCTTATTTCGAAGTTAAAT
CAAAAAATGACGATCCAAACTTTTTGAAAAAATTGAAAGATGTCATGCCTGCTGGTAAAGGTATGTTGAACGTTCAG
CTACTAGTTGTTAACAGGAACGATCGTACTCAAATATGTGGTATTGGCGAAATAGGTGAGATTTATGTTCGTGCAGG
TGGTTTGGCCGAAGGTTATAGAGGATTACCAGAATTGAATAAAGAAAAATTTGTGAACAACTGGTTTGTTGAAAAAG
ATCACTGGAATTATTTGGATAAGGATAATGGTGAACCTTGGAGACAATTCTGGTTAGGTCCAAGAGATAGATTGTAC
AGAACGGGTGATTTAGGTCGTTATCTACCAAACGGTGACTGTGAATGTTGCGGTAGGGCTGATGATCAAGTTAAAAT
TCGTGGGTTCAGAATCGAATTAGGAGAAATAGATACGCACATTTCCCAACATCCATTGGTAAGAGAAAACATTACTT
TAGTTCGCAAAAATGCCGACAATGAGCCAACATTGATCACATTTATGGTCCCAAGATTTGACAAGCCAGATGACTTG
TCTAAGTTCCAAAGTGATGTTCCAAAGGAGGTTGAAACTGACCCTATAGTTAAGGGCTTAATCGGTTACCATCTTTT
ATCCAAGGACATCAGGACTTTCTTAAAGAAAAGATTGGCTAGCTATGCTATGCCTTCCTTGATTGTGGTTATGGATA
AACTACCATTGAATCCAAATGGTAAAGTTGATAAGCCTAAACTTCAATTCCCAACTCCCAAGCAATTAAATTTGGTA
GCTGAAAATACAGTTTCTGAAACTGACGACTCTCAGTTTACCAATGTTGAGCGCGAGGTTAGAGACTTATGGTTAAG
TATATTACCTACCAAGCCAGCATCTGTATCACCAGATGATTCGTTTTTCGATTTAGGTGGTCATTCTATCTTGGCTA
CCAAAATGATTTTTACCTTAAAGAAAAAGCTGCAAGTTGATTTACCATTGGGCACAATTTTCAAGTATCCAACGATA
AAGGCCTTTGCCGCGGAAATTGACAGAATTAAATCATCGGGTGGATCATCTCAAGGTGAGGTCGTCGAAAATGTCAC

TGCAAATTATGCGGAAGACGCCAAGAAATTGGTTGAGACGCTACCAAGTTCGTACCCCTCTCGAGAATATTTTGTTG
AACCTAATAGTGCCGAAGGAAAAACAACAATTAATGTGTTTGTTACCGGTGTCACAGGATTTCTGGGCTCCTACATC
CTTGCAGATTTGTTAGGACGTTCTCCAAAGAACTACAGTTTCAAAGTGTTTGCCCACGTCAGGGCCAAGGATGAAGA
AGCTGCATTTGCAAGATTACAAAAGGCAGGTATCACCTATGGTACTTGGAACGAAAAATTTGCCTCAAATATTAAAG
TTGTATTAGGCGATTTATCTAAAAGCCAATTTGGTCTTTCAGATGAGAAGTGGATGGATTTGGCAAACACAGTTGAT
ATAATTATCCATAATGGTGCGTTAGTTCACTGGGTTTATCCATATGCCAAATTGAGGGATCCAAATGTTATTTCAAC
TATCAATGTTATGAGCTTAGCCGCCGTCGGCAAGCCAAAGTTCTTTGACTTTGTTTCCTCCACTTCTACTCTTGACA
CTGAATACTACTTTAATTTGTCAGATAAACTTGTTAGCGAAGGGAAGCCAGGCATTTTAGAATCAGACGATTTAATG
AACTCTGCAAGCGGGCTCACTGGTGGATATGGTCAGTCCAAATGGGCTGCTGAGTACATCATTAGACGTGCAGGTGA
AAGGGGCCTACGTGGGTGTATTGTCAGACCAGGTTACGTAACAGGTGCCTCTGCCAATGGTTCTTCAAACACAGATG
ATTTCTTATTGAGATTTTTGAAAGGTTCAGTCCAATTAGGTAAGATTCCAGATATCGAAAATTCCGTGAATATGGTT
CCAGTAGATCATGTTGCTCGTGTTGTTGTTGCTACGTCTTTGAATCCTCCCAAAGAAAATGAATTGGCCGTTGCTCA
AGTAACGGGTCACCCAAGAATATTATTCAAAGACTACTTGTATACTTTACACGATTATGGTTACGATGTCGAAATCG
AAAGCTATTCTAAATGGAAGAAATCATTGGAGGCGTCTGTTATTGACAGGAATGAAGAAAATGCGTTGTATCCTTTG
CTACACATGGTCTTAGACAACTTACCTGAAAGTACCAAAGCTCCGGAACTAGACGATAGGAACGCCGTGGCATCTTT
AAAGAAAGACACCGCATGGACAGGTGTTGATTGGTCTAATGGAATAGGTGTTACTCCAGAAGAGGTTGGTATATATA
TTGCATTTTTAAACAAGGTTGGATTTTTACCTCCACCAACTCATAATGACAAACTTCCACTGCCAAGTATAGAACTA
ACTCAAGCGCAAATAAGTCTAGTTGCTTCAGGTGCTGGTGCTCGTGGAAGCTCCGCAGCAGCTTAAGGTTGAGCATT
ACGTATGATATGTCCATGTACAATAATTAAATATGAATTAGGAGAAAGACTTAGCTTCTTTTCGGGTGATGTCACTT
AAAAACTCCGAGAATAATATATAATAAGAGAATAAAATATTAGTTATTGAATAAGAACTGTAAATCAGCTGGCGTTA
GTCTGCTAATGGCAGCTTCATCTTGGTTTATTGTAGCATGAATCATATTTGCCTTTTTTTCCTGTAATTCAATGATT
CTTGCTTCTATACTATCCTCAATGCAAAACCTTGTGATCTTCACAGGTCGATACTGACCAATTCTATGAACTCTATC
ACCACTTTGCCATTCAACACTAGGGTTCCACCATGGGTCTAAAATGAATACTTGCGAAGCTTCACAAAGATTCAAAG
CAACACCGCCCGCCTTTAAACTGACCAAGAAAACCTCGCATTGAATGTTGTTCATGAAATACTTGATGGTTTCATCT
CTTTGCGTCGGTGACATACTACCCTGAAGCTTCACTGTTTGAAATCCAGCTCTTTTCAATCTCCACTCTACCAGATC
CAGCATACTGGTAAACTGGGAAAACACAATGGATTTAATCGTTCTCTTGTTGCTTCTCAGTTTGTATAGTTCTTCCA
CAAGTGCTTCGATTTTCGTTGATGATTGCCACTTGCCACTCATGTTTAGACGGCTAACAATACTTTGCTTTTTGAAG
GAATCAAGGTCCACTTCCAAAGCAGGTTGAGACAAATCGATACTTAGCCCGATATGACAAACAGGACAAGTAAGTTT
ATTGTTGTTTTCCATGAAAGATTCCACATATTCTTTGATGCATAAACGACAGAACTTATGGTGACATTTAGATTCAA
TGGGCTCCTCAGCTTCATCGTTACATAATTGGCAGATCACGACGCCGATATCATCGCCAGGAAAATTGTTTAATCTT
TTCAAAACTAAATCAGGATGATCTGCCAGTTGCCTCATTCTTGTGATTAGGGTGAAAATGTTTGCATAATTGTTTAG
AACAACACCCTCCTCAACAAAGGAATTATACTTCCTTTTAGAATCTGTGTATAAACTTCTGTAAAGATCTTTTTCCT
CTTCATTGAAGAAGTCTCTCCTCACGGTAACAATTCTGGGCGGTAGACCCAAGTCATCCGCTCTTTCCACTTTAGTT
CTTCGCAGCATGATGTTTTTCAATAATGTCTGAA

***lys2*Δ0 (from BY4742 background)**
*prLYS2::***Added Sequence***::LYS2(t)*

TACGGTGGTACCTTTTTGAACTTCGTCTCTTATGCTGCAGGAGCCGTTAGGTTAGCCGCCTTGTCTGGTAATCCAGT
CATTTGGGTTGCAACACATGACTCTATCGGGCTTGGTGAGGATGGTCCAACGCACCAACCTATTGAAACTCTGGCTC
ACTTGAGGGCTATTCCAAACATGCATGTATGGAGACCTGCTGATGGTAACGAAACTTCTGCTGCGTATTATTCTGCT
ATCAAATCTGGTCGAACACCATCTGTTGTGGCTTTATCACGACAGAATCTTCCTCAATTGGAGCATTCCTCTTTTGA
AAAAGCCTTGAAGGGTGGCTATGTGATCCATGACGTGGAGAATCCTGATATTATCCTGGTGTCAACAGGATCAGAAG
TCTCCATTTCTATAGATGCAGCCAAAAAATTGTACGATACTAAAAAAATCAAAGCAAGAGTTGTTTCCCTGCCAGAC
TTTTATACTTTTGACAGGCAAAGTGAAGAATACAGATTCTCTGTTCTACCAGACGGTGTTCCGATCATGTCCTTTGA
AGTATTGGCTACTTCAAGCTGGGGTAAGTATGCTCATCAATCGTTCGGACTCGACGAATTTGGTCGTTCAGGCAAGG
GGCCTGAAATTTACAAATTGTTCGATTTCACAGCGGACGGTGTTGCGTCAAGGGCTGAAAAGACAATCAATTACTAC
AAAGGAAAGCAGTTGCTTTCTCCTATGGGAAGAGCTTTCTAAGTCTGAAGAAGTAAACAGTTCTTTGCTATTTCACA
CTTCCTGGTTGATGGTCACTTGCTGCCTGAAATATATATATATGTATGACATATGTACTTGTTTTCTTTTTTGTGCC
TTTGTTACGTCTATATTCATTGAAACTGATTATTCGATTTCTTCTTGCTGACC**TCTTCTGGATCC**TCCATGTACAAT
AATTAAATATGAATTAGGAGAAAGACTTAGCTTCTTTTCGGGTGATGTCACTTAAAAACTCCGAGAATAATATATAA
TAAGAGAATAAAATATTAGTTATTGAATAAGAACTGTAAATCAGCTGGCGTTAGTCTGCTAATGGCAGCTTCATCTT
GGTTTATTGTAGCATGAATCATATTTGCCTTTTTTTCCTGTAATTCAATGATTCTTGCTTCTATACTATCCTCAATG
CAAAACCTTGTGATCTTCACAGGTCGATACTGACCAATTCTATGAACTCTATCACCACTTTGCCATTCAACACTAGG
GTTCCACCATGGGTCTAAAATGAATACTTGCGAAGCTTCACAAAGATTCAAAGCAACACCGCCCGCCTTTAAACTGA
CCAAGAAAACCTCGCATTGAATGTTGTTCATGAAATACTTGATGGTTTCATCTCTTTGCGTCGGTGACATACTACCC
TGAAGCTTCACTGTTTGAAATCCAGCTCTTTTCAATCTCCACTCTACCAGATCCAGCATACTGGTAAACTGGGAAAA

CACAATGGATTTAATCGTTCTCTTGTTGCTTCTCAGTTTGTATAGTTCTTCCACAAGTGCTTCGATTTTCGTTGATG
ATTGCCACTTGCCACTCATGTTTAGACGGCTAACAATACTTTGCTTTTTGAAGGAATCAAGGTCCACTTCCAAAGCA
GGTTGAGACAAATCGATACTTAGCCCGATATGACAAACAGGACAAGTAAGTTTATTGTTGTTTTCCATGAAAGATTC
CACATATTCTTTGATGCATAAACGACAGAACTTATGGTGACATTTAGATTCAATGGGCTCCTCAGCTTCATCGTTAC
ATAATTGGCAGATCACGACGCCGATATCATCGCCAGGAAAATTGTTTAATCTTTTCAAAACTAAATCAGGATGATCT
GCCAGTTGCCTCATTCTTGTGATTAGGGTGAAAATGTTTGCATAATTGTTTAGAACAACACCCTCCTCAACAAAGGA
ATTATACTTCCTTTTAGAATCTGTGTATAAACTTCTGTAAAGATCTTTTTCCTCTTCATTGAAGAAGTCTCTCCTCA
CGGTAACAATTCTGGGCGGTAGACCCAAGTCATCCGCTCTTTCCACTTTAGTTCTTCGCAGCATGATGTTTTTCAAT
AATGTCTGAA

Promoter region: 299 bp of immediate 5’ UTR deleted and T at +313 upstream of start codon was deleted. “TCTTCTGGATCC” sequence added in place of *LYS2* gene. Terminator region: 23 bps of immediate 3’ UTR was deleted.

***MET15* (from BY4742 background)**
prMET15::***MET15(WT)***::*MET15(t)*

CCTTTTTACCTCATTGCACTAATAAAAAAATTCTACAGAATCTCCGAAAAAGAAAATCCAGCTTACTCTTTTTGTTT
TCTTCTTCACACGTGAGCTTTTCCGCCGGCATACGTTCCGTTCCGTGTCGTCTTGCATAAAATTTCCGAATCACATG
TTCGTAAAACAACCGGAAGTGCCCCGAATATAAAGTCAATTCTCACCGCTGTTGTAACTGGAGCTTTAAGGTGTTAT
CTAAGGAAGGATAAAAGAACTTAAACAACAACAAACAAAAAATTAAGTTACAATGCCAGCATTATTAAAAAGATTAT
TGTTTCAAGTGGGTCCTCATCCAAACGAAAGAACATTCACCTTATCCTCTGTTTCAACTGATGGGCACTATATTTCC
TTGAGACCATTTGTTAAGCCAAGCGGTGATGAGTTATCTTTTCCTTTCGAATGGGCCTTTGCCGGTACAAACGAAAC
AGTTAAAGCTAATGATCAAGGAAACGGTGTCGTTACTCAAGATTTCAATTTCTGGTTGGATACAAATGTGTACTTGA
ACGTTCCAAACACCCATCGTGGCGAAGTGAACACCACTTGGAAAAATTGGGATTCTGGTTGTGTCGAGGAAACAGGC
GCTGTTTACCCATTCGGTGCCGACAAAGAAAGCGTCTCTTTCAGAGAATTGTGGCAACCAGTTGACCCATCAAGAGA
AGATCTAGTCATCGTCTCACCAAACAATGAGAAGTTCTCGTCGAATGCTAGGTCAATTGTCCTCAAAGTTACTGACG
AAGCTTATGATGGTTTGGTTATTGTTATTGGTAGATGGATTCAAGGGTTTTTGTCCCAAAAGAATAATAACACTATT
GAAGGCTTGAACTTCATCAGATTACTTGAAAAAGATTCAGGTAAATCTGAGTTCTTATTAAGCTACGGTAAGGAAGT
AAACAAAATTCCACAAAGCTACGAAAATTTGAAGAAAGGTTCCACTGTAACCAGCAATGGGTTGAACTGGGAAGTTA
TTGAATATCACGCTTAATAAAGGAGAATAAATCGTTTTCTACTTTCTTCTGCTGCTATAATAAGCACCTATGGGATC
TATATAGTATTTTTATAACGATAGACTTTATAAAAGAAAATACCTAAGTGAAAATTTGGTGAATTTTGAGATAATTG
TTGGGATTCCATTTTTAATAAGGCAATAATATTAGGTATGTAGAATATACTAGAAGTTCTCCTCGAGGATTTAGGAA
TCCATAAAAGGGAATCTGCAATTCTACACAATTCTATAAATATTATTATCATCGTTTTATATGTTAATATTCATTGA
TCCTATTACATTATCAATCCTTGCGTTTCAGCTTCCACTAATTTAGATGACTATTTCTCATCATTTGCGTCATCTTC
TAACACCGTATATGATAATATACTAGTAACGTAAATACTAGTTAGTAGATGATAGTTGATTTTTATTCCAACACTAA
GAAATAATTTCGCCATTTCTTGAATGTATTTAAAGATATTTAATGCTATAATAGACATTTAAATCCAATTCTTCCAA
CATACAATGGGAGTTTGGCCGAGTGGTTTAAGGCGTCAGATTTAGGTGGATTTAACCTCTAAAATCTCTGATATCTT
CGGATGCAAGGGTTCGAATCCCTTAGCTCTCATTATTTTTTGCTTTTTCTCTTGAGGTCACATGATCGCAAAATGGC
AAATGGCACGTGAAGCTGTCGATATTGGGGAACTGTGGTGGTTGGCAAATGACTAATTAAGTTAGTCAAGGCGCCAT
CCTCATGAAAACTGTGTAACATAATAACCGAAGTGTCGAAAAGGTGGCACCTTGTCCAATTGAACACGCTCGATGAA
AAAAATAAGATATATATAAGGTTAAGTAAAGCGTCTGTTAGAAAGGAAGTTTTTCCTTTTTCTTGCTCTCTTGTCTT
TTCATCTACTATTTCCTTCGTGTAATACAGGGTCGTCAGATACATAGATACAATTCTATTACCCCCATCCATACAAT
GCCATCTCATTTCGATACTGTTCAACTACACGCCGGCCAAGAGAACCCTGGTGACAATGCTCACAGATCCAGAGCTG
TACCAATTTACGCCACCACTTCTTATGTTTTCGAAAACTCTAAGCATGGTTCGCAATTGTTTGGTCTAGAAGTTCCA
GGTTACGTCTATTCCCGTTTCCAAAACCCAACCAGTAATGTTTTGGAAGAAAGAATTGCTGCTTTAGAAGGTGGTGC
TGCTGCTTTGGCTGTTTCCTCCGGTCAAGCCGCTCAAACCCTTGCCATCCAAGGTTTGGCACACACTGGTGACAACA
TCGTTTCCACTTCTTACTTATACGGTGGTACTTATAACCAGTTCAAAATCTCGTTCAAAAGATTTGGTATCGAGGCT
AGATTTGTTGAAGGTGACAATCCAGAAGAATTCGAAAAGGTCTTTGATGAAAGAACCAAGGCTGTTTATTTGGAAAC
CATTGGTAATCCAAAGTACAATGTTCCGGATTTTGAAAAAATTGTTGCAATTGCTCACAAACACGGTATTCCAGTTG
TCGTTGACAACACATTTGGTGCCGGTGGTTACTTCTGTCAGCCAATTAAATACGGTGCTGATATTGTAACACATTCT
GCTACCAAATGGATTGGTGGTCATGGTACTACTATCGGTGGTATTATTGTTGACTCTGGTAAGTTCCCATGGAAGGA
CTACCCAGAAAAGTTCCCTCAATTCTCTCAACCTGCCGAAGGATATCACGGTACTATCTACAATGAAGCCTACGGTA

ACTTGGCATACATCGTTCATGTTAGAACTGAACTATTAAGAGATTTGGGTCCATTGATGAACCCATTTGCCTCTTTC
TTGCTACTACAAGGTGTTGAAACATTATCTTTGAGAGCTGAAAGACACGGTGAAAATGCATTGAAGTTAGCCAAATG
GTTAGAACAATCCCCATACGTATCTTGGGTTTCATACCCTGGTTTAGCATCTCATTCTCATCATGAAAATGCTAAGA
AGTATCTATCTAACGGTTTCGGTGGTGTCTTATCTTTCGGTGTAAAAGACTTACCAAATGCCGACAAGGAAACTGAC
CCATTCAAACTTTCTGGTGCTCAAGTTGTTGACAATTTAAAGCTTGCCTCTAACTTGGCCAATGTTGGTGATGCCAA
GACCTTAGTCATTGCTCCATACTTCACTACCCACAAACAATTAAATGACAAAGAAAAGTTGGCATCTGGTGTTACCA
AGGACTTAATTCGTGTCTCTGTTGGTATCGAATTTATTGATGACATTATTGCAGACTTCCAGCAATCTTTTGAAACT
GTTTTCGCTGGCCAAAAACCATGAGTGTGCGTAATGAGTTGTAAAATTATGTATAAACCTACTTTCTCTCACAAGTA
CTATACTTTTATAAAACGAACTTTATTGAAATGAATATCCTTTTTTTCCCTTGTTACATGTCGTGACTCGTACTTTG
AACCTAAATTGTTCTAACATCAAAGAACAGTGTTAATTCGCAGTCGAGAAGAAAAATATGGTGAACAAGACTCATCT
ACTTCATGAGACTACTTTACGCCTCCTATAAAGCTGTCACACTGGATAAATTTATTGTAGGACCAAGTTACAAAAGA
GGATGATGGAGGTTTCTTTACAATAAAGAAGCACATGTGTGTTAACGTTTTTAGTATTTGCTTGTTATGTAAATCAG
GAAAACTTCGCGGGATTTGGTTGGATGCTACTTTCCATACAATAAATATTATAGATCTAAAAAGCCAAATTACAAGT
AAAGATTAGTAAAGCTGTTGGAATTCCATCGTTGATAAAAATGTTAGTTATTAAATATAAAAGTCAGAATAGGTGAA
CTTGGATTTAATTGTTGGCATTTCGTTGCTGCTAGAGGCCATAATATTAGATAGCCAGGACATACTAGTTCTCCTCG
TGGTATAGGAATCCATAAAATGGAATTGGTGATTCTATGTGATATATTCACATTCTTACTACATTATCAATCCTTGC
ACTTCAGCTTCCTCTAACCTCGATGACATCTTCTCATAACTTATGTCATCATCTAACGCCGTCTATTATAATATATT
GATAGTATAAGTATTAGTTGATAGACAATAGTGGATTTTTATTCCAACAGTGTCTTTGTTCGTCTCAGATATAGTCG
GATTGCCCTTTTAAGCAATCAATAGTGTTTTATTTGCAACAATGTCGTCATAGTTTAATATGTCCTATAAGATGTTA
ACTTGCTCAACATTCAACAAAGTTTGGTCTCTTGGCCCAGTTGGTTAAGGCACCGTGCTAATAACGCGGGGATCAGC
GGTTCGATCCCGCTAGAGACCATTTATTTTTAAATTCGACCGTCTGACAGACGGTGCACTTATTACTTAATTTTTTT
TTTTTTTTAATCCTCGTTAAACTGAGTACATAATCAGCAACAATATAATATAAAAGTTTTGCTTTTTTATGTACGTA
CAATAGGGGTACTTAAATAAATAAATAAACAATAAAACAATCAATAAATAAATAATATACGATCATAAATTACGAGA
TATATTTCATAGTCTTGACTTCTCCAGCCGAACGTCATTAAATTTAAGGCAGAAAAGATAAACAAGGGTACGGTGTA
CAGAAGACAAAATAAATAATAACCTTTAAAAGATTATAACAATGTTTTCATTATTTCTTCTCATCGGCCTTAATTTT
ATTTAAGGCAGAACCATATTTGAACCATTCAATTTGCTCATCGTTGAAAGTATGGGTCAACACAGCATCCCATGGCT
TACCATTCTTTGGATGAACTCTCATTGTTACAGGCTTACCTGGAGCCAATTCAGCTAGACCCAGAATATCGATTCTG
TCATCAGGGTTGATCTTGTCATAGTCAGCTGGGTTCTTGAAGTTCAATGGCAATAGACCTTGTTTTTTCAAGTTAGT
TTCATGGATACGAGCGAAAGACTTTGTGATGATAGCGAAACCGCCCAAGAATCTTGGTTCCAAAGCAGCGTGTTCAC
GAGAGGAACCTTCACCAAAGTTTTCATCACCAATAACAACCCACTTGATACCTTGGTCTCTGTAATCTCTAGCAGTG
TCTGGAACACCTTTGTATTCACCAGTATATACATTTTTAACACAGTTAGCCTTCTTGTTTTCAGCATTAATAGCACC
AATCATATAGTTATTAGAAATGTTTTCTAAATGACCTCTGTATTTCAACCATGGACCAGCCATAGAAATATGATCAG
TAGTTGTCTTACCGACGGCCTTAATCAAGATTGGCATGTCTTTAGCATCCTTACCATCCCAAGGTTTGAATGGTTTC
AACAGTTGTAGACGGTCTGAAG

***met15*Δ0 (from BY4741 background)**
prMET15::***Added Sequence***::*MET15(t)*

CCACAAAGCTACGAAAATTTGAAGAAAGGTTCCACTGTAACCAGCAATGGGTTGAACTGGGAAGTTATTGAATATCA
CGCTTAATAAAGGAGAATAAATCGTTTTCTACTTTCTTCTGCTGCTATAATAAGCACCTATGGGATCTATATAGTAT
TTTTATAACGATAGACTTTATAAAAGAAAATACCTAAGTGAAAATTTGGTGAATTTTGAGATAATTGTTGGGATTCC
ATTTTTAATAAGGCAATAATATTAGGTATGTAGAATATACTAGAAGTTCTCCTCGAGGATTTAGGAATCCATAAAAG
GGAATCTGCAATTCTACACAATTCTATAAATATTATTATCATCGTTTTATATGTTAATATTCATTGATCCTATTACA
TTATCAATCCTTGCGTTTCAGCTTCCACTAATTTAGATGACTATTTCTCATCATTTGCGTCATCTTCTAACACCGTA
TATGATAATATACTAGTAACGTAAATACTAGTTAGTAGATGATAGTTGATTTTTATTCCAACACTAAGAAATAATTT
CGCCATTTCTTGAATGTATTTAAAGATATTTAATGCTATAATAGACATTTAAATCCAATTCTTCCAACATACAATGG
GAGTTTGGCCGAGTGGTTTAAGGCGTCAGATTTAGGTGGATTTAACCTCTAAAATCTCTGATATCTTCGGATGCAAG
GGTTCGAATCCCTTAGCTCTCATTATTTTTTGCTTTTTCTCTTGAGGTCACATGATCGCAAAATGGCAAATGGCACG
TGAAGCTGTCGATATTGGGGAACTGTGGTGGTTGGCA**GGATC**CTCAGATATAGTCGGATTGCCCTTTTAAGCAATCA
ATAGTGTTTTATTTGCAACAATGTCGTCATAGTTTAATATGTCCTATAAGATGTTAACTTGCTCAACATTCAACAAA
GTTTGGTCTCTTGGCCCAGTTGGTTAAGGCACCGTGCTAATAACGCGGGGATCAGCGGTTCGATCCCGCTAGAGACC
ATTTATTTTTAAATTCGACCGTCTGACAGACGGTGCACTTATTACTTAATTTTTTTTTTTTTTAATCCTCGTTAAA
CTGAGTACATAATCAGCAACAATATAATATAAAAGTTTTGCTTTTTTATGTACGTACAATAGGGGTACTTAAATAAA
TAAATAAACAATAAAACAATCAATAAATAAATAATATACGATCATAAATTACGAGATATATTTCATAGTCTTGACTT
CTCCAGCCGAACGTCATTAAATTTAAGGCAGAAAAGATAAACAAGGGTACGGTGTACAGAAGACAAAATAAATAATA
ACCTTTAAAAGATTATAACAATGTTTTCATTATTTCTTCTCATCGGCCTTAATTTTATTTAAGGCAGAACCATATTT

GAACCATTCAATTTGCTCATCGTTGAAAGTATGGGTCAACACAGCATCCCATGGCTTACCATTCTTTGGATGAACTC
TCATTGTTACAGGCTTACCTGGAGCCAATTCAGCTAGACCCAGAATATCGATTCTGTCATCAGGGTTGATCTTGTCA
TAGTCAGCTGGGTTCTTGAAGTTCAATGGCAATAGACCTTGTTTTTTCAAGTTAGTTTCATGGATACGAGCGAAAGA
CTTTGTGATGATAGCGAAACCGCCCAAGAATCTTGGTTCCAAAGCAGCGTGTTCACGAGAGGAACCTTCACCAAAGT
TTTCATCACCAATAACAACCCACTTGATACCTTGGTCTCTGTAATCTCTAGCAGTGTCTGGAACACCTTTGTATTCA
CCAGTATATACATTTTTAACACAGTTAGCCTTCTTGTTTTCAGCATTAATAGCACCAATCATATAGTTATTAGAAAT
GTTTTCTAAATGACCTCTGTATTTCAACCATGGACCAGCCATAGAAATATGATCAGTAGTTGTCTTACCGACGGCCT
TAATCAAGATTGGCATGTCTTTAGCATCCTTACCATCCCAAGGTTTGAATGGTTTCAACAGTTGTAGACGGTCTGAA
G

Promoter region: 259 bp of immediate 5’ UTR deleted. “**GGATC”** sequence added in place of

MET15. Terminator region: 809 bp of immediate 3’ UTR was deleted.

First Generation Yeast CRISPR Gene Drive

prHIS3-[u2]-prGAL1/10-SpCas9-NLS-ADH1(t)-prMX-Kan^{R}-MX(t)-[u2]-HIS3(t)
(Yeast Strain Name, GFY-2383)

GGGTCAGTTATTTCATCCAGATATAACCCGAGAGGAAACTTCTTAGCGTCTGTTTTCGTACCATAAGGCAGTTCATG
AGGTATATTTTCGTTATTGAAGCCCAGCTCGTGAATGCTTAATGCTGCTGAACTGGTGTCCATGTCGCCTAGGTACG
CAATCTCCACAGGCTGCAAAGGTTTTGTCTCAAGAGCAATGTTATTGTGCACCCCGTAATTGGTCAACAAGTTTAAT
CTGTGCTTGTCCACCAGCTCTGTCGTAACCTTCAGTTCATCGACTATCTGAAGAAATTTACTAGGAATAGTGCCATG
GTACAGCAACCGAGAATGGCAATTTCTACTCGGGTTCAGCAACGCTGCATAAACGCTGTTGGTGCCGTAGACATATT
CGAAGATAGGATTATCATTCATAAGTTTCAGAGCAATGTCCTTATTCTGGAACTTGGATTTATGGCTCTTTTGGTTT
AATTTCGCCTGATTCTTGATCTCCTTTAGCTTCTCGACGTGGGCCTTTTTCTTGCCATATGGATCCGCTGCACGGTC
CTGTTCCCTAGCATGTACGTGAGCGTATTTCCTTTTAAACCACGACGCTTTGTCTTCATTCAACGTTTCCCATTGTT
TTTTTCTACTATTGCTTTGCTGTGGGAAAAACTTATCGAAAGATGACGACTTTTTCTTAATTCTCGTTTTAAGAGCT
TGGTGAGCGCTAGGAGTCACTGCCAGGTATCGTTTGAACACGGCATTAGTCAGGGAAGTCATAACACAGTCCTTTCC
CGCAATTTTCTTTTTCTATTACTCTTGGCCTCCTCTAGTACACTCTATATTTTTTTATGCCTCGGTAATGATTTTCA
TTTTTTTTTTTCCACCTAGCGGATGACTCTTTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTA
TATAAAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG**GCTGTTCGT**
GTGCGCGTCCTGGGGACAGGTTATCAGCAACAACACAGTCATATCCATTCTCAATTAGCTCTACCACAGTGTGTGAA
CCAATGTATCCAGCACCACCTGTAACCAAAACAATTTTAGAAGTACTTTCACTTTGTAACTGAGCTGTCATTTATAT
TGAATTTTCAAAAATTCTTACTTTTTTTTTTGGATGGACGCAAAGAAGTTTAATAATCATATTACATGGCATTACCAC
CATATACATATCCATATACATATCCATATCTAATCTTACTTATATGTTGTGGAAATGTAAAGAGCCCCATTATCTTA
GCCTAAAAAAACCTTCTCTTTGGAACTTTCAGTAATACGCTTAACTGCTCATTGCTATATTGAAGTACGGATTAGAA
GCCGCCGAGCGGGTGACAGCCCTCCGAAGGAAGACTCTCCTCCGTGCGTCCTCGTCTTCACCGGTCGCGTTCCTGAA
ACGCAGATGTGCCTCGCGCCGCACTGCTCCGAACAATAAAGATTCTACAATACTAGCTTTTATGGTTATGAAGAGGA
AAAATTGGCAGTAACCTGGCCCCACAAACCTTCAAATGAACGAATCAAATTAACAACCATAGGATGATAATGCGATT
AGTTTTTTAGCCTTATTTCTGGGGTAATTAATCAGCGAAGCGATGATTTTTGATCTATTAACAGATATATAAATGCA
AAAACTGCATAACCACTTTAACTAATACTTTCAACATTTTCGGTTTGTATTACTTCTTATTCAAATGTAATAAAAGT
ATCAACAAAAAATTGTTAATATACCTCTATACTTTAACGTCAAGGAGAAAAAACTATA**ATG**GATAAGAAATACTCTA
TCGGTTTGGATATTGGTACAAATTCAGTTGGTTGGGCAGTTATTACTGATGAATACAAGGTTCCATCTAAAAAGTTT
AAAGTTTTGGGTAACACTGATAGACATTCTATTAAGAAAAATTTGATTGGTGCTTTGTTATTTGATTCTGGTGAAAC
TGCTGAAGCAACAAGATTGAAAAGAACTGCAAGAAGAAGATACACAAGAAGAAAGAATAGAATCTGTTATTTGCAAG
AAATTTTCTCTAACGAAATGGCTAAGGTTGATGATTCTTTCTTTCATAGATTGGAAGAATCATTTTTAGTTGAAGAA
GATAAGAAACATGAAAGACATCCAATCTTCGGTAACATCGTTGATGAAGTTGCTTACCATGAAAAGTACCCAACAAT
CTATCATTTGAGAAAGAAATTGGTTGATTCAACTGATAAGGCAGATTTGAGATTGATATATTTGGCTTTAGCACATA
TGATCAAGTTTAGAGGTCATTTCTTGATCGAGGGTGACTTGAATCCAGATAATTCTGATGTTGATAAGTTGTTTATT
CAATTAGTTCAAACATATAATCAATTGTTTGAAGAAAATCCAATTAATGCTTCTGGTGTTGATGCTAAGGCAATCTT
GTCAGCAAGATTGTCTAAGTCAAGAAGATTGGAAAATTTGATCGCTCAATTACCAGGTGAAAAGAAAAATGGTTTGT
TCGGTAATTTGATCGCATTGTCTTTGGGTTTGACACCAAACTTCAAGTCAAACTTCGATTTGGCTGAAGATGCAAAG
TTGCAATTGTCTAAGGATACTTACGATGATGATTTGGATAATTTGTTGGCTCAAATTGGTGACCAATATGCAGATTT
GTTTTTGGCTGCTAAAAATTTGTCTGATGCTATCTTGTTGTCAGATATCTTGAGAGTTAACACTGAAATCACAAAGG

CTCCATTGTCTGCATCAATGATCAAGAGATACGATGAACATCATCAAGATTTGACTTTGTTGAAGGCATTGGTTAGA
CAACAATTACCAGAAAAGTACAAGGAAATTTTCTTTGATCAATCTAAAAATGGTTATGCTGGTTACATTGATGGTGG
TGCATCTCAAGAAGAATTCTACAAGTTTATTAAGCCAATCTTGGAAAAGATGGATGGTACAGAAGAATTGTTAGTTA
AATTGAACAGAGAAGATTTGTTAAGAAAACAAAGAACTTTCGATAACGGTTCTATCCCACATCAAATCCATTTGGGT
GAATTACATGCTATCTTGAGAAGACAAGAAGATTTCTACCCATTTTTAAAGGATAACAGAGAAAAGATTGAAAAGAT
TTTGACTTTTAGAATTCCATATTACGTTGGTCCATTAGCTCGTGGTAATTCTAGATTTGCATGGATGACTAGAAAGT
CAGAAGAAACTATCACACCATGGAATTTTGAAGAAGTTGTTGATAAAGGTGCTTCTGCACAATCTTTTATTGAAAGA
ATGACAAACTTCGATAAAAATTTGCCAAACGAAAAGGTTTTGCCAAAGCATTCATTGTTATATGAATACTTTACTGT
TTACAATGAATTGACAAAAGTTAAATATGTTACTGAGGGTATGAGAAAACCAGCATTTTTGTCTGGTGAACAAAAGA
AAGCAATCGTTGATTTGTTGTTTAAAACTAACAGAAAGGTTACAGTTAAACAATTGAAAGAAGATTACTTTAAGAAA
ATTGAATGTTTTGATTCTGTTGAAATTTCAGGTGTTGAAGATAGATTCAATGCTTCATTAGGTACTTACCATGATTT
GTTGAAGATTATTAAGGATAAAGATTTCTTGGATAATGAAGAAAATGAAGATATTTTAGAAGATATTGTTTTAACTT
TGACATTATTTGAAGATAGAGAAATGATCGAAGAAAGATTGAAGACATACGCTCATTTGTTCGATGATAAAGTTATG
AAGCAATTGAAGAGAAGAAGATACACTGGTTGGGGTAGATTGTCTAGAAAGTTGATTAATGGTATCAGAGATAAGCA
ATCTGGTAAAACAATCTTGGATTTCTTGAAGTCAGATGGTTTCGCAAACAGAAACTTCATGCAATTGATTCATGATG
ATTCATTGACTTTTAAAGAAGATATCCAAAAAGCTCAAGTTTCTGGTCAGGGTGACTCATTGCATGAACATATTGCT
AATTTGGCAGGTTCTCCAGCTATTAAGAAAGGTATCTTGCAAACAGTTAAGGTTGTTGATGAATTAGTTAAAGTTAT
GGGTAGACATAAGCCAGAAAACATCGTTATCGAAATGGCTAGAGAAAACCAAACTACACAAAAGGGTCAAAAGAATT
CAAGAGAAAGAATGAAGAGAATCGAAGAAGGTATTAAAGAATTGGGTTCTCAAATCTTGAAGGAACATCCAGTTGAA
AACACTCAATTGCAAAACGAAAAGTTGTACTTATACTACTTACAAAACGGTAGAGATATGTACGTTGATCAAGAATT
AGATATCAACAGATTGTCAGATTACGATGTTGATCATATCGTTCCACAATCATTTTTGAAGGATGATTCAATCGATA
ATAAGGTTTTGACAAGATCTGATAAGAACCGTGGTAAATCTGATAATGTTCCATCAGAAGAAGTTGTTAAGAAAATG
AAGAACTACTGGAGACAATTGTTAAATGCTAAGTTGATCACTCAAAGAAAGTTCGATAATTTGACAAAAGCTGAAAG
AGGTGGTTTGTCAGAATTAGATAAAGCAGGTTTTATTAAGAGACAATTAGTTGAAACTAGACAAATCACAAAGCATG
TTGCACAAATCTTGGATTCTAGAATGAACACTAAATATGATGAAAATGATAAATTAATTAGAGAAGTTAAAGTTATT
ACATTAAAATCTAAATTGGTTTCAGATTTTAGAAAAGATTTTCAATTCTACAAAGTTAGAGAAATTAATAACTATCA
TCATGCTCATGATGCATACTTGAATGCTGTTGTTGGTACTGCATTGATTAAGAAATACCCAAAGTTGGAATCTGAAT
TCGTTTACGGTGACTACAAGGTTTACGATGTTAGAAAGATGATCGCTAAGTCAGAACAAGAAATCGGTAAAGCTACA
GCAAAGTATTTCTTTTATTCTAACATCATGAATTTCTTTAAAACTGAAATTACATTAGCTAACGGTGAAATCAGAAA
AAGACCATTGATCGAAACTAATGGTGAAACAGGTGAAATTGTTTGGGATAAAGGTAGAGATTTCGCAACTGTTAGAA
AGGTTTTGTCAATGCCACAAGTTAACATCGTTAAGAAAACTGAAGTTCAAACAGGTGGTTTTTCTAAGGAATCAATC
TTGCCAAAGAGAAACTCTGATAAGTTGATTGCTAGAAAGAAAGATTGGGATCCAAAGAAATATGGTGGTTTTGATTC
TCCAACTGTTGCTTACTCAGTTTTAGTTGTTGCAAAGGTTGAAAAGGGTAAATCTAAGAAATTGAAATCAGTTAAAG
AATTGTTAGGTATCACAATCATGGAAAGATCTTCATTCGAAAAGAATCCAATCGATTTCTTGGAAGCAAAGGGTTAC
AAGGAAGTTAAGAAAGATTTGATTATTAAGTTGCCAAAGTACTCTTTGTTCGAATTAGAAAACGGTAGAAAAAGAAT
GTTAGCTTCAGCTGGTGAATTGCAAAAGGGTAATGAATTGGCTTTGCCATCTAAGTACGTTAATTTCTTGTATTTGG
CATCTCATTACGAAAAGTTGAAGGGTTCACCAGAAGATAATGAACAAAAACAATTGTTCGTTGAACAACATAAGCAT
TATTTGGATGAAATTATTGAACAAATTTCTGAATTTTCAAAAAGAGTTATTTTGGCTGATGCAAATTTGGATAAGGT
TTTGTCTGCTTACAATAAGCATAGAGATAAGCCAATCAGAGAACAAGCAGAAAACATCATCCATTTGTTTACTTTGA
CAAATTTGGGTGCTCCAGCTGCTTTTAAATACTTCGATACTACAATCGATAGAAAAAGATACACTTCTACAAAGGAA
GTTTTGGATGCAACATTGATCCATCAATCAATCACTGGTTTGTATGAAACAAGAATTGATTTGTCTCAATTGGGTGG
TGACTCTAGGGCAGACCCAAAGAAAAAGAGGAAAGTATAAGGCGCGCCATTCTAAATAAGCGAATTTCTTATGATTT
ATGATTTTTATTATTAAATAAGTTATAAAAAAAATAAGTGTATACAAATTTTAAAGTGACTCTTAGGTTTTAAAACG
AAAATTCTTATTCTTGAGTAACTCTTTCCTGTAGGTCAGGTTGCTTTCTCAGGTATAGTATGAGGTCGCTCTTATTG
ACCACACCTCTACCGGCAGATCCGCTAGGGATAACAGGGTAATATAGATCTGTTTAGCTTGCCTCGTCCCCGCCGGG
TCACCCGGCCAGCGACATGGAGGCCCAGAATACCCTCCTTGACAGTCTTGACGTGCGCAGCTCAGGGGCATGATGTG
ACTGTCGCCCGTACATTTAGCCCATACATCCCCATGTATAATCATTTGCATCCATACATTTTGATGGCCGCACGGCG
CGAAGCAAAAATTACGGCTCCTCGCTGCAGACCTGCGAGCAGGGAAACGCTCCCCTCACAGACGCGTTGAATTGTCC
CCACGCCGCGCCCCTGTAGAGAAATATAAAAGGTTAGGATTTGCCACTGAGGTTCTTCTTTCATATACTTCCTTTTA
AAATCTTGCTAGGATACAGTTCTCACATCACATCCGAACATAAACAACCATGGGTAAGGAAAAGACTCACGTTTCGA
GGCCGCGATTAAATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGT
GCGACAATCTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAA
TGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATCC
GTACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCCGGCAAAACAGCATTCCAGGTATTAGAAGAATATCCT
GATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCC
TTTTAACAGCGATCGCGTATTTCGTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATT
TTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAGCTTTTGCCATTCTCACCGGAT

TCAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGT
TGGACGAGTCGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCAT
TACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCATTTGATGCTCGAT
GAGTTTTTCTAATCAGTACTGACAATAAAAAGATTCTTGTTTTCAAGAACTTGTCATTTGTATAGTTTTTTTATATT
GTAGTTGTTCTATTTTAATCAAATGTTAGCGTGATTTATATTTTTTTTCGCCTCGACATCATCTGCCCAGATGCGAA
GTTAAGTGCGCAGAAAGTAATATCATGCGTCAATCGTATGTGAATGCTGGTCGCTATACTGCTGTCGATTCGATACT
AACGCCGCCATCCAGT**GCTGTTCGTGTGCGCGTCCTGGG**TGACACCGATTATTTAAAGCTGCAGCATACGATATATA
TACATGTGTATATATGTATACCTATGAATGTCAGTAAGTATGTATACGAACAGTATGATACTGAAGATGACAAGGTA
ATGCATCATTCTATACGTGTCATTCTGAACGAGGCGCGCTTTCCTTTTTTCTTTTTGCTTTTTCTTTTTTTTTCTCT
TGAACTCGAGAAAAAAAATATAAAAGAGATGGAGGAACGGGAAAAAGTTAGTTGTGGTGATAGGTGGCAAGTGGTAT
TCCGTAAGAACAACAAGAAAAGCATTTCATATTATGGCTGAACTGAGCGAACAAGTGCAAAATTTAAGCATCAACGA
CAACAACGAGAATGGTTATGTTCCTCCTCACTTAAGAGGAAAACCAAGAAGTGCCAGAAATAACAGTAGCAACTACA
ATAACAACAACGGCGGCTACAACGGTGGCCGTGGCGGTGGCAGCTTCTTTAGCAACAACCGTCGTGGTGGTTACGGC
AACGGTGGTTTCTTCGGTGGAAACAACGGTGGCAGCAGATCTAACGGCCGTTCTGGTGGTAGATGGATCGATGGCAA
ACATGTCCCAGCTCCAAGAAACGAAAAGGCCGAGATCGCCATATTTGGTGTCCCCGAGGATCCAAATTTCCAATCTT
CTGGTATTAACTTCGATAACTACGATGATATTCCAGTGGACGCCTCTGGTAAGGATGTTCCTGAACCAATCACAGAA
TTTACCTCACCTCCATTGGACGGATTGTTATTGGAAAACATCAAATTGGCCCGTTTCACCAAGCCAACACCTGTGCA
AAAATACTCCGTCCCTATCGTTGCCAACGGCAGAGATTTGATGGCCTGTGCGCAGACCGGTTCTGGTAAGACTGGTG
GGTTTTTATTCCCAGTGTTGTCCGAATCATTTAAGACTGGACCATCTCCTCAACCAGAGTCTCAAGGCTCCTTTTAC
CAAAGAAAGGCCTACCCAACTGCTGTCATTA

Second Generation Yeast CRISPR Gene Drive

prHIS3-[u2']-prGAL1/10-SpCas9-NLS-ADH1(t)-sgRNA(u1)-prMX-CaURA3-MX(t)-[u2']-HIS3(t)
(Yeast Strain Names, GFY-4325/GFY-4326)

GGGTCAGTTATTTCATCCAGATATAACCCGAGAGGAAACTTCTTAGCGTCTGTTTTCGTACCATAAGGCAGTTCATG
AGGTATATTTTCGTTATTGAAGCCCAGCTCGTGAATGCTTAATGCTGCTGAACTGGTGTCCATGTCGCCTAGGTACG
CAATCTCCACAGGCTGCAAAGGTTTTGTCTCAAGAGCAATGTTATTGTGCACCCCGTAATTGGTCAACAAGTTTAAT
CTGTGCTTGTCCACCAGCTCTGTCGTAACCTTCAGTTCATCGACTATCTGAAGAAATTTACTAGGAATAGTGCCATG
GTACAGCAACCGAGAATGGCAATTTCTACTCGGGTTCAGCAACGCTGCATAAACGCTGTTGGTGCCGTAGACATATT
CGAAGATAGGATTATCATTCATAAGTTTCAGAGCAATGTCCTTATTCTGGAACTTGGATTTATGGCTCTTTTGGTTT
AATTTCGCCTGATTCTTGATCTCCTTTAGCTTCTCGACGTGGGCCTTTTTCTTGCCATATGGATCCGCTGCACGGTC
CTGTTCCCTAGCATGTACGTGAGCGTATTTCCTTTTAAACCACGACGCTTTGTCTTCATTCAACGTTTCCCATTGTT
TTTTTCTACTATTGCTTTGCTGTGGGAAAAACTTATCGAAAGATGACGACTTTTTCTTAATTCTCGTTTTAAGAGCT
TGGTGAGCGCTAGGAGTCACTGCCAGGTATCGTTTGAACACGGCATTAGTCAGGGAAGTCATAACACAGTCCTTTCC
CGCAATTTTCTTTTTCTATTACTCTTGGCCTCCTCTAGTACACTCTATATTTTTTTATGCCTCGGTAATGATTTTCA
TTTTTTTTTTTCCACCTAGCGGATGACTCTTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTA
TATAAAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG**TTTTCGCTG**
TTCGTGTGCGCGTCCTGGGAGTGACAGGTTATCAGCAACAACACAGTCATATCCATTCTCAATTAGCTCTACCACAG
TGTGTGAACCAATGTATCCAGCACCACCTGTAACCAAAACAATTTTAGAAGTACTTTCACTTTGTAACTGAGCTGTC
ATTTATATTGAATTTTCAAAAATTCTTACTTTTTTTTTGGATGGACGCAAAGAAGTTTAATAATCATATTACATGGC
ATTACCACCATATACATATCCATATACATATCCATATCTAATCTTACTTATATGTTGTGGAAATGTAAAGAGCCCCA
TTATCTTAGCCTAAAAAAACCTTCTCTTTGGAACTTTCAGTAATACGCTTAACTGCTCATTGCTATATTGAAGTACG
GATTAGAAGCCGCCGAGCGGGTGACAGCCCTCCGAAGGAAGACTCTCCTCCGTGCGTCCTCGTCTTCACCGGTCGCG
TTCCTGAAACGCAGATGTGCCTCGCGCCGCACTGCTCCGAACAATAAAGATTCTACAATACTAGCTTTTATGGTTAT
GAAGAGGAAAAATTGGCAGTAACCTGGCCCCACAAACCTTCAAATGAACGAATCAAATTAACAACCATAGGATGATA
ATGCGATTAGTTTTTTAGCCTTATTTCTGGGGTAATTAATCAGCGAAGCGATGATTTTTGATCTATTAACAGATATA
TAAATGCAAAAACTGCATAACCACTTTAACTAATACTTTCAACATTTTCGGTTTGTATTACTTCTTATTCAAATGTA
ATAAAAGTATCAACAAAAAATTGTTAATATACCTCTATACTTTAACGTCAAGGAGAAAAAACTATA**ATG**GATAAGAA
ATACTCTATCGGTTTGGATATTGGTACAAATTCAGTTGGTTGGGCAGTTATTACTGATGAATACAAGGTTCCATCTA
AAAAGTTTAAAGTTTTGGGTAACACTGATAGACATTCTATTAAGAAAAATTTGATTGGTGCTTTGTTATTTGATTCT
GGTGAAACTGCTGAAGCAACAAGATTGAAAAGAACTGCAAGAAGAAGATACACAAGAAGAAAGAATAGAATCTGTTA
TTTGCAAGAAATTTTCTCTAACGAAATGGCTAAGGTTGATGATTCTTTCTTTCATAGATTGGAAGAATCATTTTTAG
TTGAAGAAGATAAGAAACATGAAAGACATCCAATCTTCGGTAACATCGTTGATGAAGTTGCTTACCATGAAAAGTAC

CCAACAATCTATCATTTGAGAAAGAAATTGGTTGATTCAACTGATAAGGCAGATTTGAGATTGATATATTTGGCTTT
AGCACATATGATCAAGTTTAGAGGTCATTTCTTGATCGAGGGTGACTTGAATCCAGATAATTCTGATGTTGATAAGT
TGTTTATTCAATTAGTTCAAACATATAATCAATTGTTTGAAGAAAATCCAATTAATGCTTCTGGTGTTGATGCTAAG
GCAATCTTGTCAGCAAGATTGTCTAAGTCAAGAAGATTGGAAAATTTGATCGCTCAATTACCAGGTGAAAAGAAAAA
TGGTTTGTTCGGTAATTTGATCGCATTGTCTTTGGGTTTGACACCAAACTTCAAGTCAAACTTCGATTTGGCTGAAG
ATGCAAAGTTGCAATTGTCTAAGGATACTTACGATGATGATTTGGATAATTTGTTGGCTCAAATTGGTGACCAATAT
GCAGATTTGTTTTTGGCTGCTAAAAATTTGTCTGATGCTATCTTGTTGTCAGATATCTTGAGAGTTAACACTGAAAT
CACAAAGGCTCCATTGTCTGCATCAATGATCAAGAGATACGATGAACATCATCAAGATTTGACTTTGTTGAAGGCAT
TGGTTAGACAACAATTACCAGAAAAGTACAAGGAAATTTTCTTTGATCAATCTAAAAATGGTTATGCTGGTTACATT
GATGGTGGTGCATCTCAAGAAGAATTCTACAAGTTTATTAAGCCAATCTTGGAAAAGATGGATGGTACAGAAGAATT
GTTAGTTAAATTGAACAGAGAAGATTTGTTAAGAAAACAAAGAACTTTCGATAACGGTTCTATCCCACATCAAATCC
ATTTGGGTGAATTACATGCTATCTTGAGAAGACAAGAAGATTTCTACCCATTTTTAAAGGATAACAGAGAAAAGATT
GAAAAGATTTTGACTTTTAGAATTCCATATTACGTTGGTCCATTAGCTCGTGGTAATTCTAGATTTGCATGGATGAC
TAGAAAGTCAGAAGAAACTATCACACCATGGAATTTTGAAGAAGTTGTTGATAAAGGTGCTTCTGCACAATCTTTTA
TTGAAAGAATGACAAACTTCGATAAAAATTTGCCAAACGAAAAGGTTTTGCCAAAGCATTCATTGTTATATGAATAC
TTTACTGTTTACAATGAATTGACAAAAGTTAAATATGTTACTGAGGGTATGAGAAAACCAGCATTTTTGTCTGGTGA
ACAAAAGAAAGCAATCGTTGATTTGTTGTTTAAAACTAACAGAAAGGTTACAGTTAAACAATTGAAAGAAGATTACT
TTAAGAAAATTGAATGTTTTGATTCTGTTGAAATTTCAGGTGTTGAAGATAGATTCAATGCTTCATTAGGTACTTAC
CATGATTTGTTGAAGATTATTAAGGATAAAGATTTCTTGGATAATGAAGAAAATGAAGATATTTTAGAAGATATTGT
TTTAACTTTGACATTATTTGAAGATAGAGAAATGATCGAAGAAAGATTGAAGACATACGCTCATTTGTTCGATGATA
AAGTTATGAAGCAATTGAAGAGAAGAAGATACACTGGTTGGGGTAGATTGTCTAGAAAGTTGATTAATGGTATCAGA
GATAAGCAATCTGGTAAAACAATCTTGGATTTCTTGAAGTCAGATGGTTTCGCAAACAGAAACTTCATGCAATTGAT
TCATGATGATTCATTGACTTTTAAAGAAGATATCCAAAAAGCTCAAGTTTCTGGTCAGGGTGACTCATTGCATGAAC
ATATTGCTAATTTGGCAGGTTCTCCAGCTATTAAGAAAGGTATCTTGCAAACAGTTAAGGTTGTTGATGAATTAGTT
AAAGTTATGGGTAGACATAAGCCAGAAAACATCGTTATCGAAATGGCTAGAGAAAACCAAACTACACAAAAGGGTCA
AAAGAATTCAAGAGAAAGAATGAAGAGAATCGAAGAAGGTATTAAAGAATTGGGTTCTCAAATCTTGAAGGAACATC
CAGTTGAAAACACTCAATTGCAAAACGAAAAGTTGTACTTATACTACTTACAAAACGGTAGAGATATGTACGTTGAT
CAAGAATTAGATATCAACAGATTGTCAGATTACGATGTTGATCATATCGTTCCACAATCATTTTTGAAGGATGATTC
AATCGATAATAAGGTTTTGACAAGATCTGATAAGAACCGTGGTAAATCTGATAATGTTCCATCAGAAGAAGTTGTTA
AGAAAATGAAGAACTACTGGAGACAATTGTTAAATGCTAAGTTGATCACTCAAAGAAAGTTCGATAATTTGACAAAA
GCTGAAAGAGGTGGTTTGTCAGAATTAGATAAAGCAGGTTTTATTAAGAGACAATTAGTTGAAACTAGACAAATCAC
AAAGCATGTTGCACAAATCTTGGATTCTAGAATGAACACTAAATATGATGAAAATGATAAATTAATTAGAGAAGTTA
AAGTTATTACATTAAAATCTAAATTGGTTTCAGATTTTAGAAAAGATTTTCAATTCTACAAAGTTAGAGAAATTAAT
AACTATCATCATGCTCATGATGCATACTTGAATGCTGTTGTTGGTACTGCATTGATTAAGAAATACCCAAAGTTGGA
ATCTGAATTCGTTTACGGTGACTACAAGGTTTACGATGTTAGAAAGATGATCGCTAAGTCAGAACAAGAAATCGGTA
AAGCTACAGCAAAGTATTTCTTTTATTCTAACATCATGAATTTCTTTAAAACTGAAATTACATTAGCTAACGGTGAA
ATCAGAAAAAGACCATTGATCGAAACTAATGGTGAAACAGGTGAAATTGTTTGGGATAAAGGTAGAGATTTCGCAAC
TGTTAGAAAGGTTTTGTCAATGCCACAAGTTAACATCGTTAAGAAAACTGAAGTTCAAACAGGTGGTTTTTCTAAGG
AATCAATCTTGCCAAAGAGAAACTCTGATAAGTTGATTGCTAGAAAGAAAGATTGGGATCCAAAGAAATATGGTGGT
TTTGATTCTCCAACTGTTGCTTACTCAGTTTTAGTTGTTGCAAAGGTTGAAAAGGGTAAATCTAAGAAATTGAAATC
AGTTAAAGAATTGTTAGGTATCACAATCATGGAAAGATCTTCATTCGAAAAGAATCCAATCGATTTCTTGGAAGCAA
AGGGTTACAAGGAAGTTAAGAAAGATTTGATTATTAAGTTGCCAAAGTACTCTTTGTTCGAATTAGAAAACGGTAGA
AAAAGAATGTTAGCTTCAGCTGGTGAATTGCAAAAGGGTAATGAATTGGCTTTGCCATCTAAGTACGTTAATTTCTT
GTATTTGGCATCTCATTACGAAAAGTTGAAGGGTTCACCAGAAGATAATGAACAAAAACAATTGTTCGTTGAACAAC
ATAAGCATTATTTGGATGAAATTATTGAACAAATTTCTGAATTTTCAAAAAGAGTTATTTTGGCTGATGCAAATTTG
GATAAGGTTTTGTCTGCTTACAATAAGCATAGAGATAAGCCAATCAGAGAACAAGCAGAAAACATCATCCATTTGTT
TACTTTGACAAATTTGGGTGCTCCAGCTGCTTTTAAATACTTCGATACTACAATCGATAGAAAAAGATACACTTCTA
CAAAGGAAGTTTTGGATGCAACATTGATCCATCAATCAATCACTGGTTTGTATGAAACAAGAATTGATTTGTCTCAA
TTGGGTGGTGACTCTAGGGCAGACCCAAAGAAAAAGAGGAAAGTATAAGGCGCGCCATTCTAAATAAGCGAATTTCT
TATGATTTATGATTTTTATTATTAAATAAGTTATAAAAAAAATAAGTGTATACAAATTTTAAAGTGACTCTTAGGTT
TTAAAACGAAAATTCTTATTCTTGAGTAACTCTTTCCTGTAGGTCAGGTTGCTTTCTCAGGTATAGTATGAGGTCGC
TCTTATTGACCACACCTCTACCGGCAGATCCGCTAGGGATAACAGGGTAATATAGGGAACAAAAGCTGGAGCTTCTT
TGAAAAGATAATGTATGATTATGCTTTCACTCATATTTATACAGAAACTTGATGTTTTCTTTCGAGTATATACAAGG
TGATTACATGTACGTTTGAAGTACAACTCTAGATTTTGTAGTGCCCTCTTGGGCTAGCGGTAAAGGTGCGCATTTTT
TCACACCCTACAATGTTCTGTTCAAAAGATTTTGGTCAAACGCTGTAGAAGTGAAAGTTGGTGCGCATGTTTCGGCG
TTCGAAACTTCTCCGCAGTGAAAGATAAATGATCCGGTGGACTTCGGCTACGTAGTTTTAGAGCTAGAAATAGCAAG
TTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGGTGCTTTTTTTGTTTTTTATGTCTT

CGAGTCATGTAATTAGTTATGTGTTTAGCTTGCCTCGTCCCCGCCGGGTCACCCGGCCAGCGACATGGAGGCCCAGA
ATACCCTCCTTGACAGTCTTGACGTGCGCAGCTCAGGGGCATGATGTGACTGTCGCCCGTACATTTAGCCCATACAT
CCCCATGTATAATCATTTGCATCCATACATTTTGATGGCCGCACGGCGCGAAGCAAAAATTACGGCTCCTCGCTGCA
GACCTGCGAGCAGGGAAACGCTCCCCTCACAGACGCGTTGAATTGTCCCCACGCCGCGCCCCTGTAGAGAAATATAA
AAGGTTAGGATTTGCCACTGAGGTTCTTCTTTCATATACTTCCTTTTAAAATCTTGCTAGGATACAGTTCTCACATC
ACATCCGAACATAAACAACC**ATG**ACAGTCAACACTAAGACCTATAGTGAGAGAGCAGAAACTCATGCCTCACCAGTA
GCACAACGATTATTTCGATTAATGGAACTGAAGAAAACCAATTTATGTGCATCAATTGATGTTGATACCACTAAGGA
ATTCCTTGAATTAATTGATAAATTGGGTCCTTATGTATGCTTAATCAAGACTCATATTGATATAATCAATGATTTTT
CCTATGAATCCACTATTGAACCATTATTAGAACTTTCACGTAAACATCAATTTATGATTTTTGAAGATAGAAAATTT
GCTGATATTGGTAATACCGTGAAGAAACAATATATTGGTGGAGTTTATAAAATTAGTAGTTGGGCAGATATTACTAA
TGCTCATGGTGTCACTGGGAATGGAGTAGTTGAAGGATTAAAACAGGGAGCTAAAGAAACCACCACCAACCAAGAGC
CAAGAGGGTTATTGATGTTAGCTGAATTATCATCAGTGGGATCATTAGCATATGGAGAATATTCTCAAAAAACTGTT
GAAATTGCTAAATCCGATAAGGAATTTGTTATTGGATTTATTGCCCAACGTGATATGGGTGGACAAGAAGAAGGATT
TGATTGGCTTATTATGACACCTGGAGTTGGATTAGATGATAAAGGTGATGGATTAGGACAACAATATAGAACTGTTG
ATGAAGTTGTTAGCACTGGAACTGATATTATCATTGTTGGTAGAGGATTGTTTGGTAAAGGAAGAGATCCAGATATT
GAAGGTAAAAGGTATAGAGATGCTGGTTGGAATGCTTATTTGAAAAAGACTGGCCAATTATAA**TCAGTACTGACAAT**
AAAAAGATTCTTGTTTTCAAGAACTTGTCATTTGTATAGTTTTTTTATATTGTAGTTGTTCTATTTTAATCAAATGT
TAGCGTGATTTATATTTTTTTTCGCCTCGACATCATCTGCCCAGATGCGAAGTTAAGTGCGCAGAAAGTAATATCAT
GCGTCAATCGTATGTGAATGCTGGTCGCTATACTGCTGTCGATTCGATACTAACGCCGCCATCCAGT**TTTC**GCTGTT
CGTGTGCGCGTCCTGGGAGTGACACCGATTATTTAAAGCTGCAGCATACGATATATATACATGTGTATATATGTATA
CCTATGAATGTCAGTAAGTATGTATACGAACAGTATGATACTGAAGATGACAAGGTAATGCATCATTCTATACGTGT
CATTCTGAACGAGGCGCGCTTTCCTTTTTTCTTTTTGCTTTTTCTTTTTTTTTCTCTTGAACTCGAGAAAAAAAATA
TAAAAGAGATGGAGGAACGGGAAAAAGTTAGTTGTGGTGATAGGTGGCAAGTGGTATTCCGTAAGAACAACAAGAAA
AGCATTTCATATTATGGCTGAACTGAGCGAACAAGTGCAAAATTTAAGCATCAACGACAACAACGAGAATGGTTATG
TTCCTCCTCACTTAAGAGGAAAACCAAGAAGTGCCAGAAATAACAGTAGCAACTACAATAACAACAACGGCGGCTAC
AACGGTGGCCGTGGCGGTGGCAGCTTCTTTAGCAACAACCGTCGTGGTGGTTACGGCAACGGTGGTTTCTTCGGTGG
AAACAACGGTGGCAGCAGATCTAACGGCCGTTCTGGTGGTAGATGGATCGATGGCAAACATGTCCCAGCTCCAAGAA
ACGAAAAGGCCGAGATCGCCATATTTGGTGTCCCCGAGGATCCAAATTTCCAATCTTCTGGTATTAACTTCGATAAC
TACGATGATATTCCAGTGGACGCCTCTGGTAAGGATGTTCCTGAACCAATCACAGAATTTACCTCACCTCCATTGGA
CGGATTGTTATTGGAAAACATCAAATTGGCCCGTTTCACCAAGCCAACACCTGTGCAAAAATACTCCGTCCCTATCG
TTGCCAACGGCAGAGATTTGATGGCCTGTGCGCAGACCGGTTCTGGTAAGACTGGTGGGTTTTTATTCCCAGTGTTG
TCCGAATCATTTAAGACTGGACCATCTCCTCAACCAGAGTCTCAAGGCTCCTTTTACCAAAGAAAGGCCTACCCAAC
TGCTGTCATTA

Second Generation Target Strain

prHIS3-[u1’]-prCDC12-mCherry-SHS1(t)-prCCW12-SpHIS5-MX(t)-[u1’]-HIS3(t)
(Yeast Strain Name, GFY-3733)

GGGTCAGTTATTTCATCCAGATATAACCCGAGAGGAAACTTCTTAGCGTCTGTTTTCGTACCATAAGGCAGTTCATG
AGGTATATTTTCGTTATTGAAGCCCAGCTCGTGAATGCTTAATGCTGCTGAACTGGTGTCCATGTCGCCTAGGTACG
CAATCTCCACAGGCTGCAAAGGTTTTGTCTCAAGAGCAATGTTATTGTGCACCCCGTAATTGGTCAACAAGTTTAAT
CTGTGCTTGTCCACCAGCTCTGTCGTAACCTTCAGTTCATCGACTATCTGAAGAAATTTACTAGGAATAGTGCCATG
GTACAGCAACCGAGAATGGCAATTTCTACTCGGGTTCAGCAACGCTGCATAAACGCTGTTGGTGCCGTAGACATATT
CGAAGATAGGATTATCATTCATAAGTTTCAGAGCAATGTCCTTATTCTGGAACTTGGATTTATGGCTCTTTTGGTTT
AATTTCGCCTGATTCTTGATCTCCTTTAGCTTCTCGACGTGGGCCTTTTTCTTGCCATATGGATCCGCTGCACGGTC
CTGTTCCCTAGCATGTACGTGAGCGTATTTCCTTTTAAACCACGACGCTTTGTCTTCATTCAACGTTTCCCATTGTT
TTTTTCTACTATTGCTTTGCTGTGGGAAAAACTTATCGAAAGATGACGACTTTTTCTTAATTCTCGTTTTAAGAGCT
TGGTGAGCGCTAGGAGTCACTGCCAGGTATCGTTTGAACACGGCATTAGTCAGGGAAGTCATAACACAGTCCTTTCC
CGCAATTTTCTTTTTCTATTACTCTTGGCCTCCTCTAGTACACTCTATATTTTTTTATGCCTCGGTAATGATTTTCA
TTTTTTTTTTTCCACCTAGCGGATGACTCTTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATACATTA
TATAAAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAGT**TTTC**CGGT
GGACTTCGGCTACGTAGGGAGTGGGGCAGCGCCCTGTTTTTCATTAATGTAGTCAGCAATGTCAAGATTCAACGCCA
AGTCTGGTTCAGCAAGTGACATTCTGCAAGCTCTTTGAATCTTCCTCAAAAGAGGATTGCCCAAGGCTTGAGGTTTC
CTGACGGGCAACTCAGACAAATATATGCTATGTGAGTGCGGATGGGACATGATGCAGTATCACGATTAGCAATTCAG
CTATGAGTTATGTTGCTCTTTGTTTTGTTTATGGAAATTGTCCTATGGTAAGTCTCTTTTTTTTGCAATCGTGATTA

CAGAAAAAAAACAGGGCGCTGGAAAAGTGAAGAATCCGAAATTTTTTTCGAAATCACCATTGTTTGTTTTGAGTAGA
TCAAAGTCTTGAAAGGTGCAGCAAGATATAGGATCTTGACCTGAAGAGTATTGATAACGAACTACATCACATATTGT
ATCAAATA**ATG**GTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATG
GAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGC
CAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCA
AGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGC
GTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAA
GGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCT
CCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCAC
TACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAA
GTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCG
GCGGCATGGACGAGCTGTACAAG**TAA**GTTGTATCTGTACAAAATCCAAAGCTGAGCAAATAAATAAATAAATAAATG
TATAAGTTACCGAACGGGGGTATTTTTACTTTTGATCAAAAATTTATGTACCAACTACAAAGTTTCCTCAGCACAGC
CTTCAAGAAGGGAACACACATACAAACAGTGTCAAATAATTGTAGGGATAAATTTAAATATGGCATAAACTAAATAA
GTAGAGCATGAAAAAACTGCAAAATCCAAAAAGTAAAAACGAAGGTCAGAAAGTAAAGCAAAAGAAAATTAATAAAG
CAATACTAAATCTATCATGATTTCCCGTAACTTCCATTAAAGCTGTAACCAGATTTACTCCTACTGTTTGAGCCTCT
AACGCCTAATGGATTTTTAGAGAAGCTCAACCTGATACCTCCTTTGTTGTTGAGGGAAGGGCGGGGGTGAGGTAGTT
GACTACCATATAATTCTGCCAATGCTCTAGTGGCAAAGCTAACATCCTCACAAAGCAAAATAAAAGAAACTTAATAC
GTTATGCCGTAATGAAGGGCTACCAAAAACGATAATCTCAACTGTAAACAGGTACAATGCGGACCCTTTTGCCACAA
AACATACATCATTCATTGCCGGAAAAAGAAAGAAGTGAAGACAGCAGTGCAGCCAGCCATGTTGCGCCAATCTAATT
ATAGATGCTGGTGCCCTGAGGATGTATCTGGAGCCAGCCATGGCATCATGCGCTACCGCCGGATGTAAAATCCGACA
CGCAAAAGAAAACCTTCGAGGTTGCGCACTTCGCCCACCCATGAACCACACGGTTAGTCCAAAAGGGGCAGTTCAGA
TTCCAGATGCGGGAATTAGCTTGCTGCCACCCTCACCTCACTAACGCTGCGGTGTGCGGATACTTCATGCTATTTAT
AGACGCGCGTGTCGGAATCAGCACGCGCAAGAACCAAATGGGAAAATCGGAATGGGTCCAGAACTGCTTTGAGTGCT
GGCTATTGGCGTCTGATTTCCGTTTTGGGAATCCTTTGCCGCGCGCCCCTCTCAAAACTCCGCACAAGTCCCAGAAA
GCGGGAAAGAAATAAAACGCCACCAAAAAAAAAAAAAATAAAAGCCAATCCTCGAAGCGTGGGTGGTAGGCCCTGGAT
TATCCCGTACAAGTATTTCTCAGGAGTAAAAAAACCGTTTGTTTTGGAATTCCCCATTTCGCGGCCACCTACGCCGC
TATCTTTGCAACAACTATCTGCGATAACTCAGCAAATTTTGCATATTCGTGTTGCAGTATTGCGATAATGGGAGTCT
TACTTCCAACATAACGGCAGAAAGAAATGTGAGAAAATTTTGCATCCTTTGCCTCCGTTCAAGTATATAAAGTCGGC
ATGCTTGATAATCTTTCTTTCCATCCTACATTGTTCTAATTATTCTTATTCTCCTTTATTCTTTCCTAACATACCAA
GAAATTAATCTTCTGTCATTCGCTTAAACACTATATCAATA**ATG**AGGAGGGCTTTTGTAGAAAGAAATACGAACGAA
ACGAAAATCAGCGTTGCCATCGCTTTGGACAAAGCTCCCTTACCTGAAGAGTCGAATTTTATTGATGAACTTATAAC
TTCCAAGCATACAAACCAAAAGGGAGAACAAGTAATCCAAGTAGACACGGGAATTGGATTCTTGGATCACATGTATC
ATGCACTGGCTAAACATGCAGGCTGGAGCTTACGACTTTACTCAAGAGGTGATTTAATCATCGATGATCATCACACT
GCAGAAGATACTGCTATTGCACTTGGTATTGCATTCAAGCAGGCTATGAGTAACTTTGCCGGCGTTAAAAGATTTGG
ACATGCTTATTGTCCACTTGACGAAGCTCTTTCTAGAAGCGTAGTTGACTTGTCGGGACGGCCCTATGCTGTTATCG
ATTTGGGATTAAAGCGTGAAAAGGTTGGGGAATTGTCCTGTGAAATGATCCCTCACTTACTATATTCCTTTTCGGTA
GCAGCTGGAATTACTTTGCATGTTACCTGCTTATATGGTAGTAATGACCATCATCGTGCTGAAAGCGCTTTTAAATC
TCTGGCTGTTGCCATGCGCGCGGCTACTAGTCTTACTGGAAGTTCTGAAGTCCCAAGCACGAAGGGAGTGTTG**TAA**A
GAGTACTGACAATAAAAAGATTCTTGTTTTCAAGAACTTGTCATTTGTATAGTTTTTTTATATTGTAGTTGTTCTAT
TTTAATCAAATGTTAGCGTGATTTATATTTTTTTTCGCCTCGACATCATCTGCCCAGATGCGAAGTTAAGTGCGCAG
AAAGTAATATCATGCGTCAATCGTATGTGAATGCTGGTCGCTATACTGCTGTCGATTCGATACTAACGCCGCCATCC
AGT**TTTC**CGGTGGACTTCGGCTACGTA**GGGAG**TGACACCGATTATTTAAAGCTGCAGCATACGATATATATACATGT
GTATATATGTATACCTATGAATGTCAGTAAGTATGTATACGAACAGTATGATACTGAAGATGACAAGGTAATGCATC
ATTCTATACGTGTCATTCTGAACGAGGCGCGCTTTCCTTTTTTCTTTTTGCTTTTTCTTTTTTTTTCTCTTGAACTC
GAGAAAAAAAATATAAAAGAGATGGAGGAACGGGAAAAAGTTAGTTGTGGTGATAGGTGGCAAGTGGTATTCCGTAA
GAACAACAAGAAAAGCATTTCATATTATGGCTGAACTGAGCGAACAAGTGCAAAATTTAAGCATCAACGACAACAAC
GAGAATGGTTATGTTCCTCCTCACTTAAGAGGAAAACCAAGAAGTGCCAGAAATAACAGTAGCAACTACAATAACAA
CAACGGCGGCTACAACGGTGGCCGTGGCGGTGGCAGCTTCTTTAGCAACAACCGTCGTGGTGGTTACGGCAACGGTG
GTTTCTTCGGTGGAAACAACGGTGGCAGCAGATCTAACGGCCGTTCTGGTGGTAGATGGATCGATGGCAAACATGTC
CCAGCTCCAAGAAACGAAAAGGCCGAGATCGCCATATTTGGTGTCCCCGAGGATCCAAATTTCCAATCTTCTGGTAT
TAACTTCGATAACTACGATGATATTCCAGTGGACGCCTCTGGTAAGGATGTTCCTGAACCAATCACAGAATTTACCT
CACCTCCATTGGACGGATTGTTATTGGAAAACATCAAATTGGCCCGTTTCACCAAGCCAACACCTGTGCAAAAATAC
TCCGTCCCTATCGTTGCCAACGGCAGAGATTTGATGGCCTGTGCGCAGACCGGTTCTGGTAAGACTGGTGGGTTTTT
ATTCCCAGTGTTGTCCGAATCATTTAAGACTGGACCATCTCCTCAACCAGAGTCTCAAGGCTCCTTTTACCAAAGAA
AGGCCTACCCAACTGCTGTC

sgRNAs Expression Cassettes

prSNR52::***crRNA***::*tracrRNA*::***SUP4(t)***

pRS425 + sgRNA(u1) (Vector name, pGF-V1220)

GGATCCTCACTAAAGGGAACAAAAGCTGGAGCTTCTTTGAAAAGATAATGTATGATTATGCTTTCACTCATATTTAT
ACAGAAACTTGATGTTTTCTTTCGAGTATATACAAGGTGATTACATGTACGTTTGAAGTACAACTCTAGATTTTGTA
GTGCCCTCTTGGGCTAGCGGTAAAGGTGCGCATTTTTTCACACCCTACAATGTTCTGTTCAAAAGATTTTGGTCAAA
CGCTGTAGAAGTGAAAGTTGGTGCGCATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGATAAATGATCCGGTGG
ACTTCGGCTACGTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGC
ACCGAGTCGGTGGTGCTTTTTTTGTTTTTTATGTCTTCGAGTCATGTAATTAGTTATGTCACGCCTCGAG

pRS425 + sgRNA(u2) (Vector name, pGF-V809)

GGATCCTCACTAAAGGGAACAAAAGCTGGAGCTTCTTTGAAAAGATAATGTATGATTATGCTTTCACTCATATTTAT
ACAGAAACTTGATGTTTTCTTTCGAGTATATACAAGGTGATTACATGTACGTTTGAAGTACAACTCTAGATTTTGTA
GTGCCCTCTTGGGCTAGCGGTAAAGGTGCGCATTTTTTCACACCCTACAATGTTCTGTTCAAAAGATTTTGGTCAAA
CGCTGTAGAAGTGAAAGTTGGTGCGCATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGATAAATGATCGCTGTT
CGTGTGCGCGTCCTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGC
ACCGAGTCGGTGGTGCTTTTTTTGTTTTTTATGTCTTCGAGTCATGTAATTAGTTATGTCACGCCTCGAG

pRS425 + sgRNA(*SpHIS5*) (Vector name, pGF-V2152)
pRS426 + sgRNA(*SpHIS5*) (Vector name, pGF-V2153)

GCGGCCGCTCACTAAAGGGAACAAAAGCTGGAGCTTCTTTGAAAAGATAATGTATGATTATGCTTTCACTCATATTT
ATACAGAAACTTGATGTTTTCTTTCGAGTATATACAAGGTGATTACATGTACGTTTGAAGTACAACTCTAGATTTTG
TAGTGCCCTCTTGGGCTAGCGGTAAAGGTGCGCATTTTTTCACACCCTACAATGTTCTGTTCAAAAGATTTTGGTCA
AACGCTGTAGAAGTGAAAGTTGGTGCGCATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGATAAATGATCACAA
GTAATCCAAGTAGACAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTG
GCACCGAGTCGGTGGTGCTTTTTTTGTTTTTTATGTCTTCGAGTCATGTAATTAGTTATGTAACACTAGT

pRS425 + sgRNA(mCherry) (Vector name, pGF-V2158)
pRS426 + sgRNA(mCherry) (Vector name, pGF-V2159)

GGATCCTCACTAAAGGGAACAAAAGCTGGAGCTTCTTTGAAAAGATAATGTATGATTATGCTTTCACTCATATTTAT
ACAGAAACTTGATGTTTTCTTTCGAGTATATACAAGGTGATTACATGTACGTTTGAAGTACAACTCTAGATTTTGTA
GTGCCCTCTTGGGCTAGCGGTAAAGGTGCGCATTTTTTCACACCCTACAATGTTCTGTTCAAAAGATTTTGGTCAAA
CGCTGTAGAAGTGAAAGTTGGTGCGCATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGATAAATGATCCAAGGA
GTTCATGCGCTTCAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGC
ACCGAGTCGGTGGTGCTTTTTTTGTTTTTTATGTCTTCGAGTCATGTAATTAGTTATGTCACGCCTCGAG

pRS425 + sgRNA(Kan^{R}) (Vector name, pGF-V1642)

GGATCCTCACTAAAGGGAACAAAAGCTGGAGCTTCTTTGAAAAGATAATGTATGATTATGCTTTCACTCATATTTAT
ACAGAAACTTGATGTTTTCTTTCGAGTATATACAAGGTGATTACATGTACGTTTGAAGTACAACTCTAGATTTTGTA
GTGCCCTCTTGGGCTAGCGGTAAAGGTGCGCATTTTTTCACACCCTACAATGTTCTGTTCAAAAGATTTTGGTCAAA
CGCTGTAGAAGTGAAAGTTGGTGCGCATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGATAAATGATCGCCATC
CTATGGAACTGCCTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGC
ACCGAGTCGGTGGTGCTTTTTTTGTTTTTTATGTCTTCGAGTCATGTAATTAGTTATGTCACGCCTCGAG

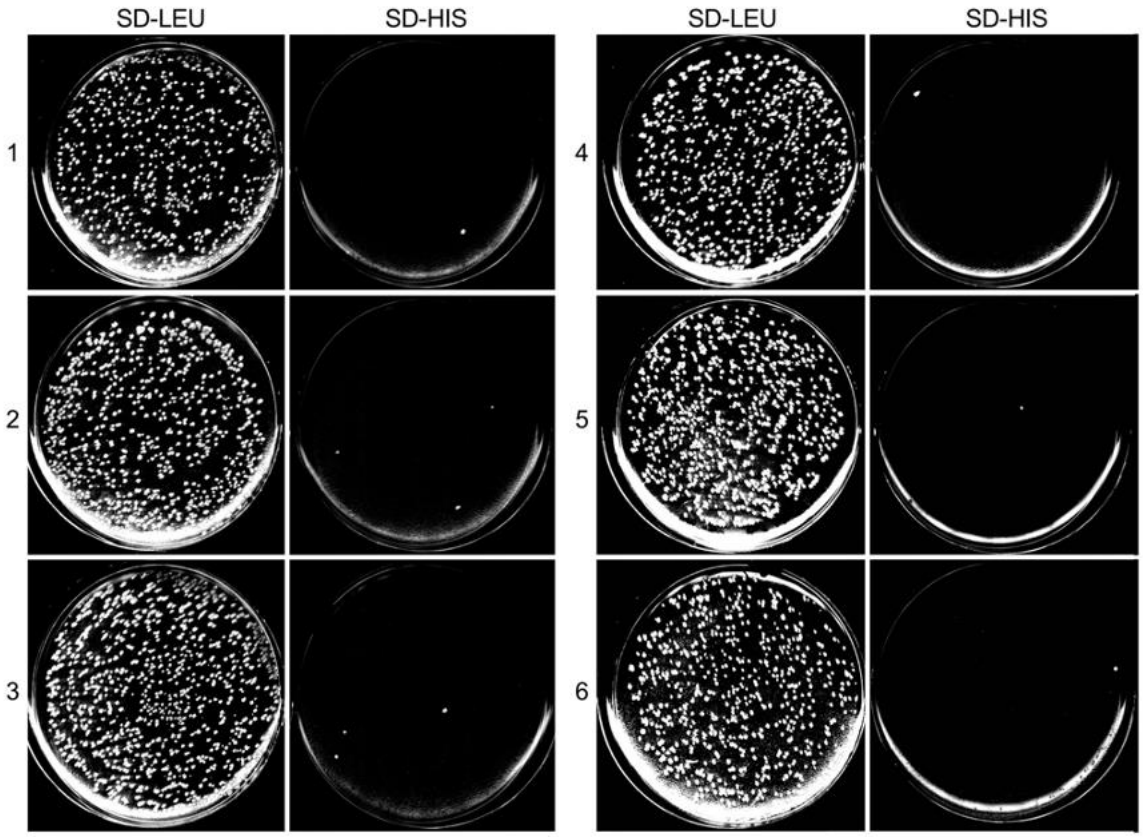


Figure C.S2. Gene drive growth assays for first-generation (GD1) system in budding yeast.

From Fig. 3.1C, additional examples of diploid gene drive strains tested on SD-LEU (control plate to select for sgRNA(u1)-containing plasmid) and SD-HIS (experimental condition to assay for the presence of the *S. pombe HIS5* gene within the target allele). GD1 diploid strains were generated through independent mating of the parental haploid drive (GFY-2383) and target (GFY-3733) yeast and subsequent diploid selection steps. All examples shown (as well as multiple repetitions) display a near 100% drive activity with few surviving colonies on the SD-HIS condition. Images of individual plates were edited for contrast (each plate separately). Horizontal and vertical white lines denote separate plates.

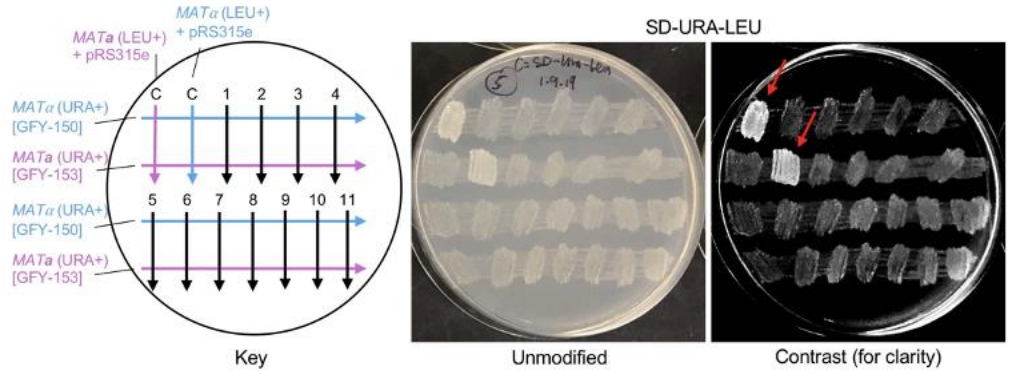


Figure C.S3. Determination of ploidy status through a mating and diploid selection assay.

Following gene drive activation, strains were selected as clonal isolates on permissive media (for example, SD-LEU) and tested against known haploid control strains (including GFY-150/GFY-153 or BY4742/BY4741). Haploids were mated on YPD plates for 24 h and then transferred by sterile velvets to selection plates for 24 h prior to imaging. Selection plates (such as SD-URA-LEU) provided a growth challenge where only diploid strains harboring both selection markers (one from each haploid strain) would survive. The unmodified and edited (for contrast) plates shown are identical. For this example, control strains (GFY-150 and GFY-153) harbored a *URA3*-based *CEN* plasmid. The haploid positive controls (labeled "C") included BY4741 and BY4742 containing an empty *CEN*-based pRS315 plasmid (marked with *LEU2*). The clonal samples (labeled 1-11) selected after activation of the gene drive assay (Fig. 3.1) all contained a *LEU2*-based plasmid (harboring the sgRNA) but no *URA3*-based marker. For control combinations of a *MAT**a*** mated to a *MATα* strain, diploids were able to survive on media testing for the presence of both *LEU2*+ and *URA3*+ plasmids (red arrows). The GD1 diploid yeast, unable to mate to either of the *URA3*-containing strains, were unable to grow on this selection plate. This ploidy growth test was also repeated in an independent assay using the pRS313-containing WT strains (*HIS3*) to mate against the clonal diploids from the gene drive assay (that had lost *SpHIS5*) with a similar procedure (selection on SD-HIS-LEU plates).

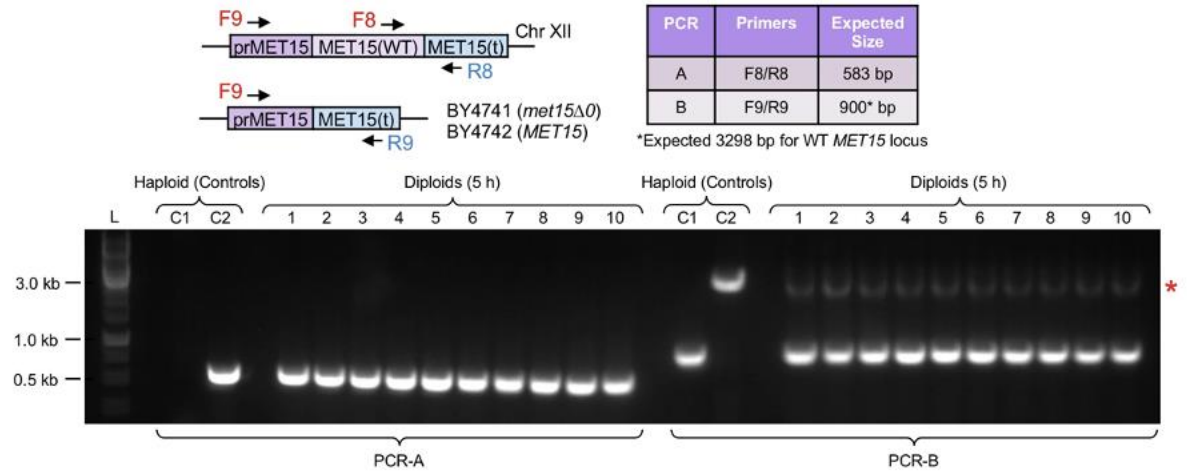


Figure C.S4. PCR amplification of the *MET15* locus to assay ploidy status.

As an independent test (compared to *LYS2*) of whether yeast strains following gene drive activation were haploid or diploid, we examined the *MET15* (also termed *MET17* or *MET25*) locus on chromosome XII. BY4741 haploid yeast are *met15Δ0* whereas BY4742 haploid yeast are *MET15*. We included the two haploid parental strains harboring the drive (GFY-2383) and target constructs (GFY-3733) at the *HIS3* locus (labeled "C1" and "C2," respectively). We chose ten isolates (1-10) from the gene drive assay (Fig. 3.1) that were tested as diploids using the mating test (described in Fig. C.S3). One set of oligonucleotides (PCR A) tested for the presence of the *MET15* coding sequence. The second set (PCR B) amplified the entire *MET15* locus from within the promoter and terminator regions. For the first PCR (*left*), the BY4742 (target) haploid and all 10 isolates displayed a fragment at the expected size of 583 bp. For the second PCR, two fragment sizes were expected: (i) *met15Δ0* would yield a product size of 900 bp whereas the *MET15* locus would yield a band of 3,298 bp. These amplified fragments were seen for haploid controls. However, for the GD1 isolates, two bands were observed at both sizes (red asterisk marks the larger band). Note, PCR reactions were optimized for generation of the 900 bp fragment. These data support that these gene drive strains included both the *met15Δ0* and *MET15* alleles and were diploid.

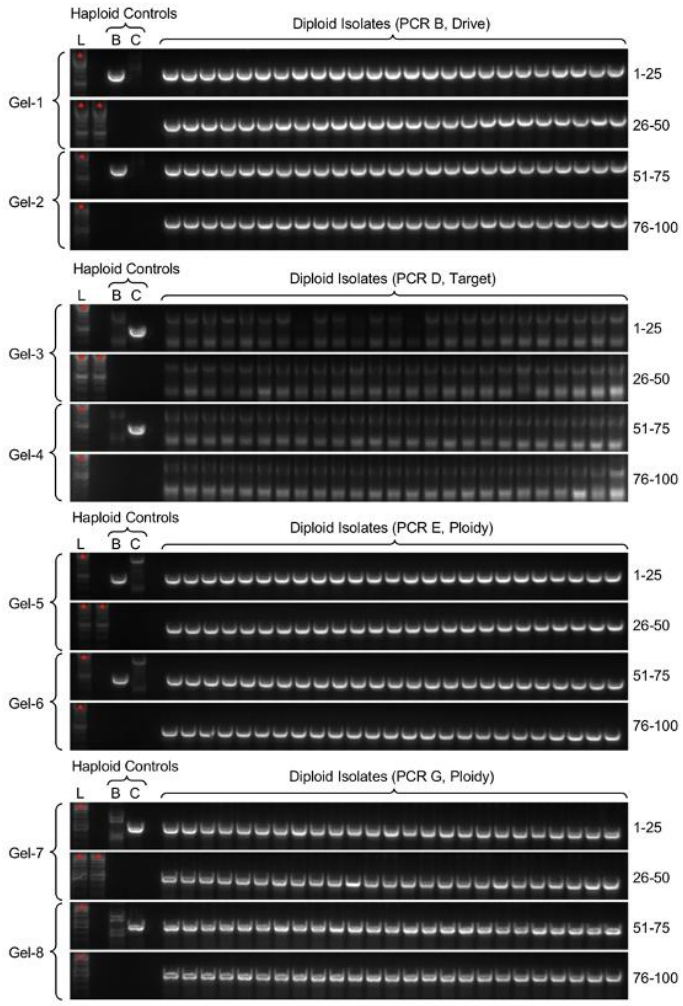
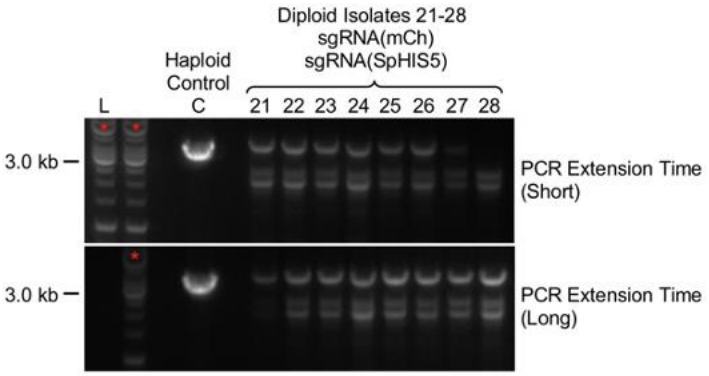


Figure C.S5. PCR amplification of the *HIS3* and *LYS2* loci for 100 separate gene drive diploid isolates.

Samples 1-100 correspond to the same clones from Fig. 3.2C (isolated after 5 h activation of the drive from SD-LEU plates). Note, these 100 samples also include the 20 isolates displayed in Fig. 3.1D (20 + 80 new isolates); the PCRs have been repeated with the entire set of 100 chromosomal DNA preparations using similar conditions. The drive allele was confirmed (PCR B) using oligonucleotides F2/R2 (PCRs include identical labels from Fig. 3.1D). The target allele was amplified (PCR D) using primers F4/R4. The *LYS2* locus was amplified (PCRs E,G) using primers F5/R5 and F7/R7, respectively. Red asterisks, positions of included DNA ladders ("L"). Two haploid controls (labeled "B" and "C") are the original drive haploid strain (GFY-2383) and target haploid strain (GFY-3733), as in Fig. 3.1D. White lines indicate separate DNA gels. Note, the first and second gels were run together on a single agarose gel (two separate rows); the third and fourth gels were run in a similar fashion. One set of haploid controls was run per complete gel (for example, the first and third gels). No additional image processing or editing (aside from cropping for clarity) has been done. The isolate number is included on the far right of each image. These data demonstrate these 100 isolates were diploids, contained the drive allele, and had lost the target allele.



PCR	Primers	Expected Size
Target	F12/R4	3531* bp

Figure C.S6. PCR amplification of the target allele for gene drive isolates harboring two guide RNAs to target mCherry and *S. pombe HIS5*.

NHEJ-based repair did not occur for any examined isolates (21-28) from Fig. 3.3C. PCRs were performed on GD1 diploid isolates following drive activation; the haploid target strain (GFY-3733) served as a control ("C"). The expected product size for the full-length target allele was 3,531 bp. Assuming dual cleavage at both the mCherry and *SpHIS5* sites and exacting repair via NHEJ, the expected PCR fragment size would be 1,240 bp. For all tested isolates (21-28), neither band was observed, despite two independent PCR reactions using extension times optimized for either the shorter fragment (*top*) or longer fragment (*bottom*). Red asterisks, DNA ladders ("L"). Gel images have not been processed aside from cropping for clarity.

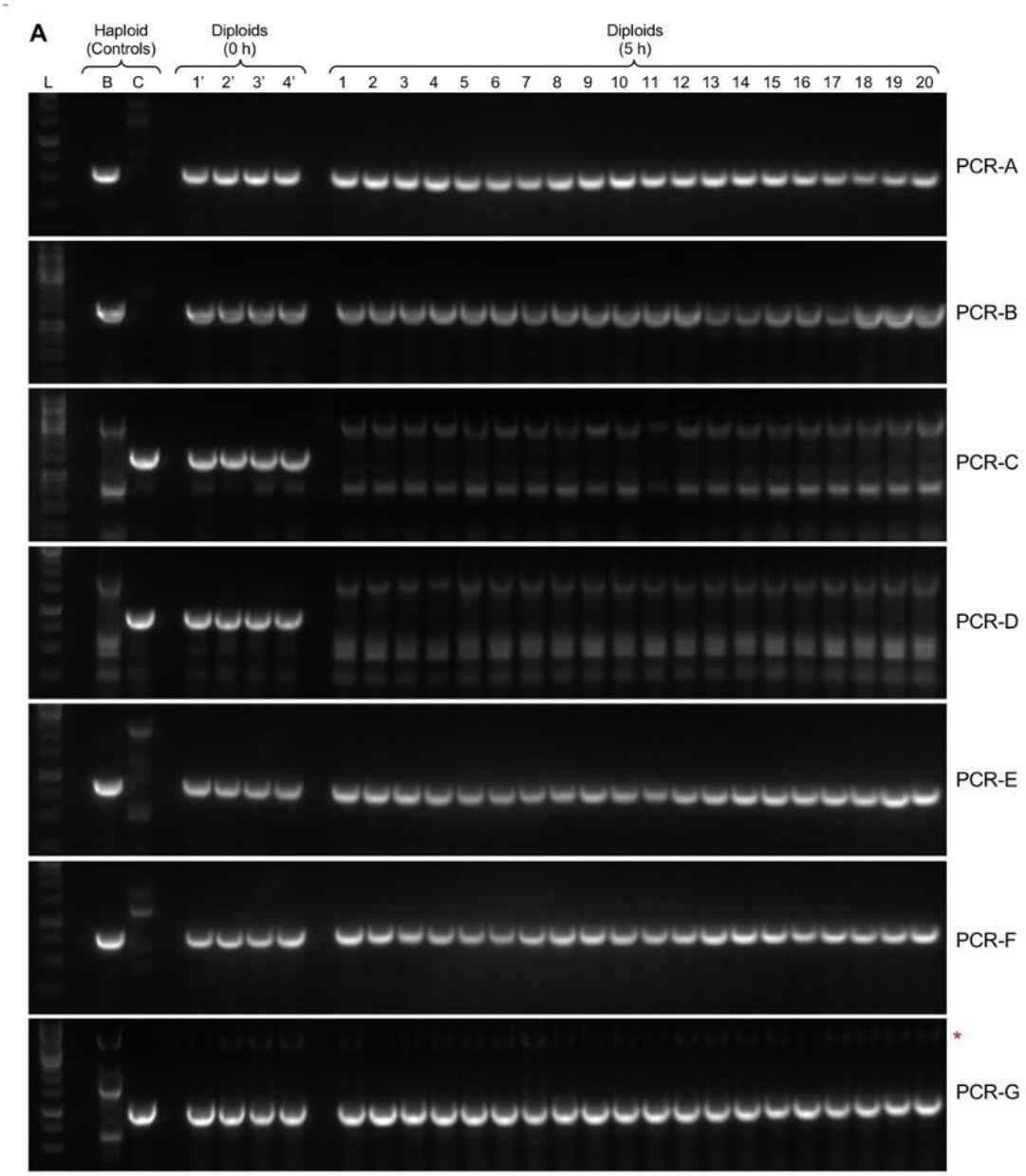
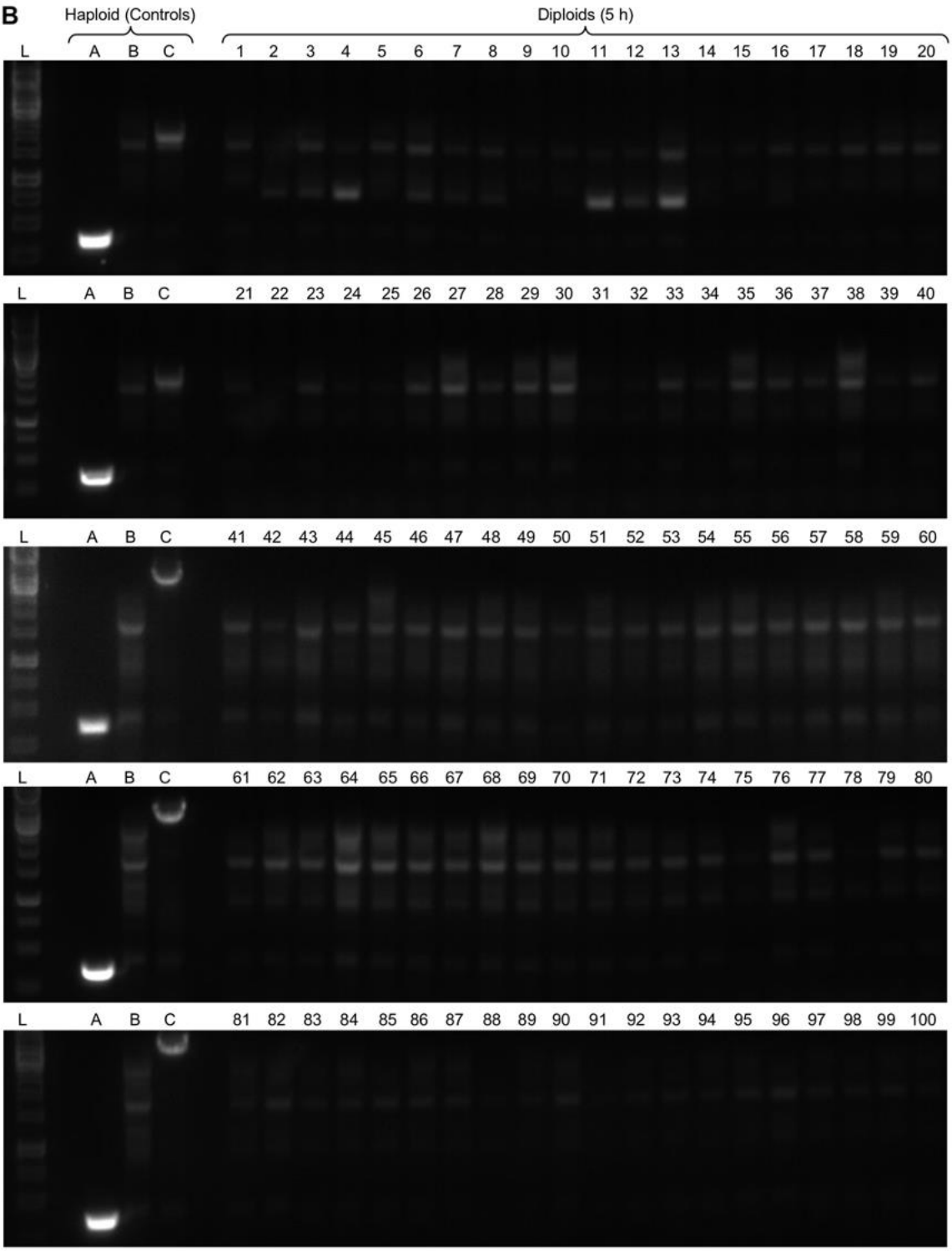
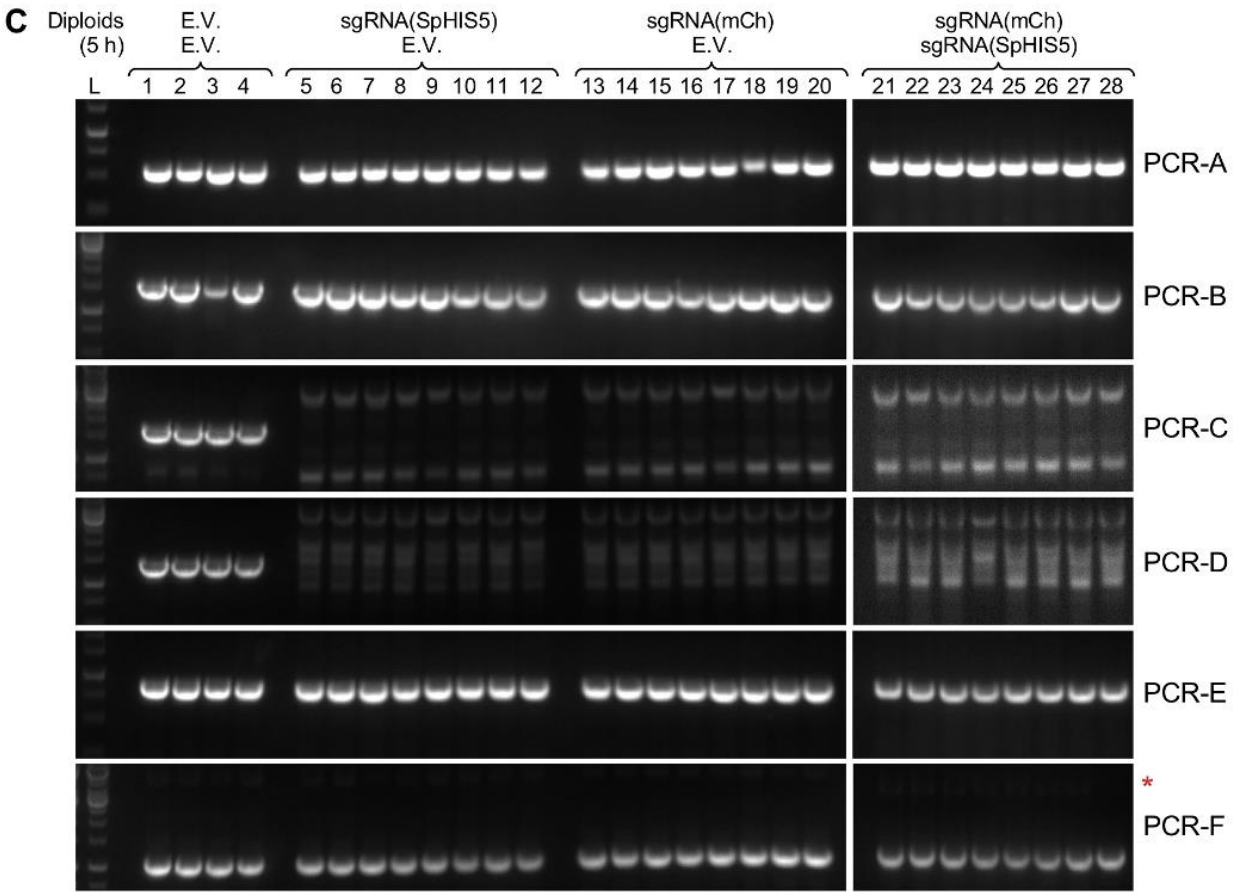
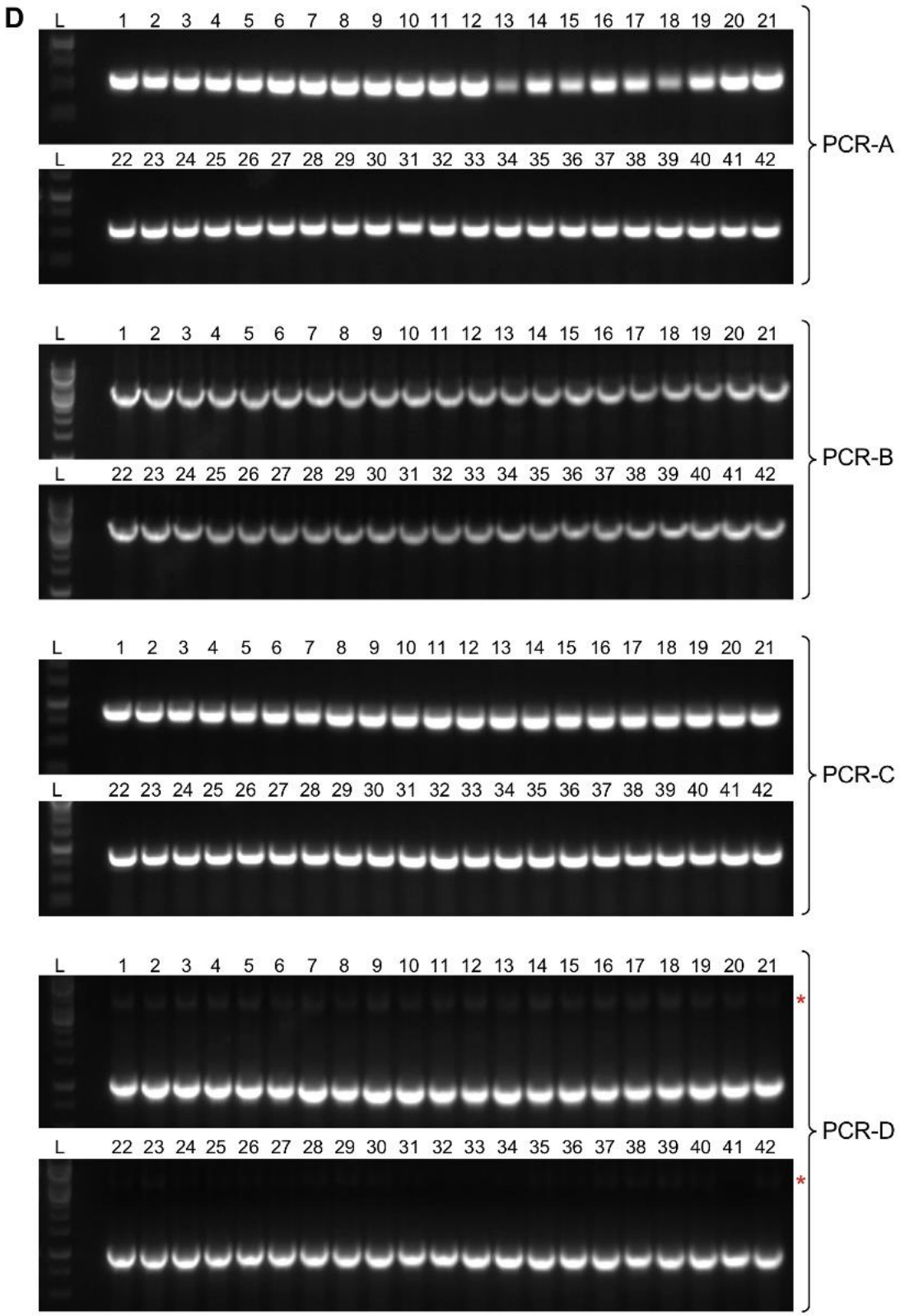


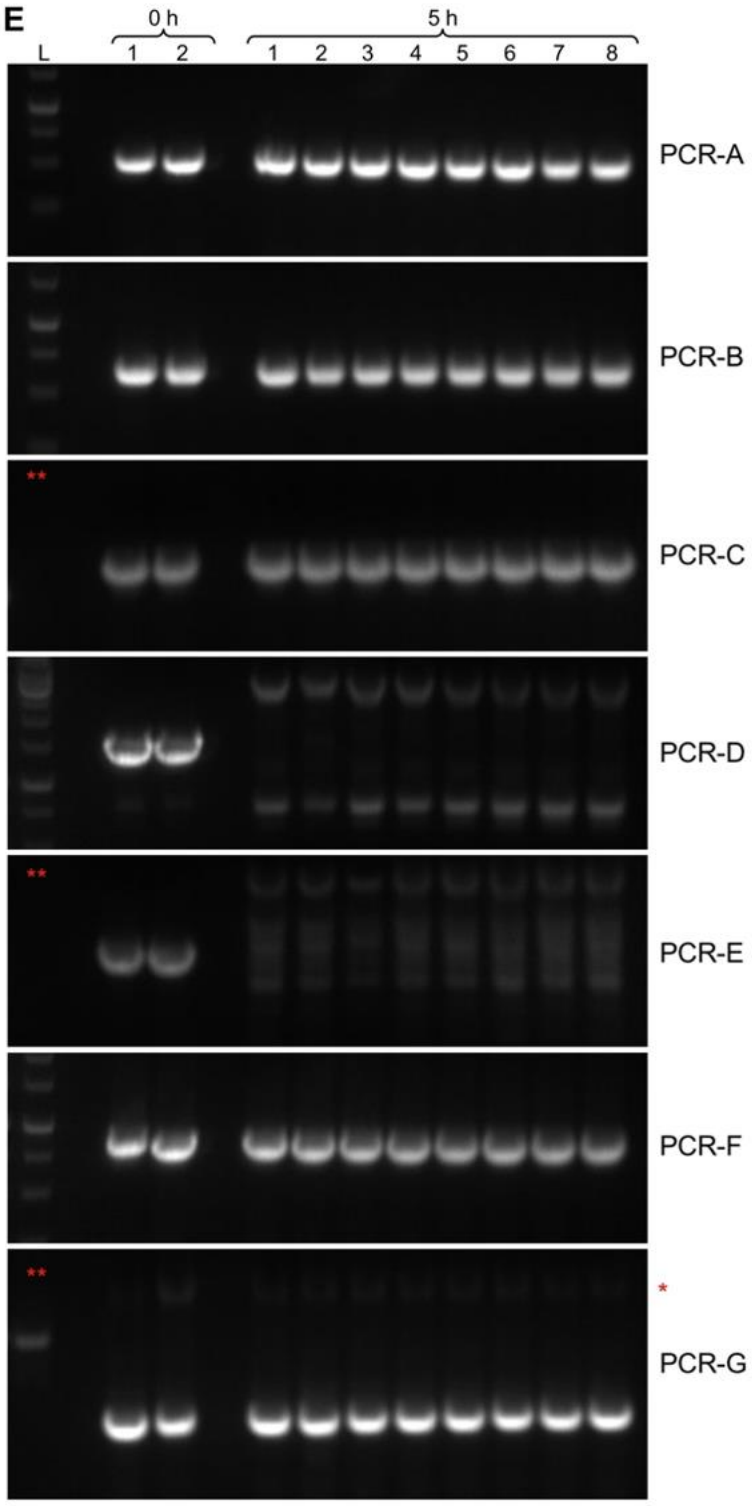
Figure C.S7. Original DNA agarose gel images used to visualize PCRs from Figs. 3.1-3.5.

All images shown were obtained using an Invitrogen E-Gel™ Imager (ThermoFisher Scientific) and were cropped for positioning and clarity, but were not altered by other methods. DNA molecular ladders (labeled "L"). Red asterisk, faint PCR band (5,573 bp) corresponding to the WT *LYS2* locus. Horizontal and vertical white lines denote separate DNA gels. (A) Gels from Fig. 3.1. (B) Gels from Fig. 3.2. In some gels, a band for Control-C can be seen at approximately 3,930 bp. (C) Gels from Fig. 3.3. (D) Gels from Fig. 3.4. (E) Gels from Fig. 3.5. Double red asterisk, no ladder was included within nearby lanes as this image was cropped from a larger gel. Figures C.S7, B-E are found below.


B
Haploid (Controls)
Diploids (5 h)
L A B C 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
L A B C 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
L A B C 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
L A B C 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80
L A B C 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100


C
Diploids (5 h)
E.V. E.V.
sgRNA(SpHIS5) E.V.
sgRNA(mCh) E.V.
sgRNA(mCh) sgRNA(SpHIS5)
L 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28
PCR-A
PCR-B
PCR-C
PCR-D
PCR-E
*
PCR-F


D
L 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
L 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42
PCR-A
L 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
L 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42
PCR-B
L 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
L 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42
PCR-C
L 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
L 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42
PCR-D
*
*


E
0 h
5 h
L
1 2
1 2 3 4 5 6 7 8
PCR-A
PCR-B
**
PCR-C
PCR-D
**
PCR-E
PCR-F
**
*
PCR-G

References

1 Brachmann, C. B. *et al.* Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast (Chichester, England)* **14**, 115-132, doi:10.1002/(sici)1097-0061(19980130)14:2<115::aid-yea204>3.0.co;2-2 (1998).

2 Roggenkamp, E. *et al.* Tuning CRISPR-Cas9 Gene Drives in *Saccharomyces cerevisiae*. *G3 (Bethesda, Md.)* **8**, 999-1018, doi:10.1534/g3.117.300557 (2018).

3 Finnigan, G. C., Takagi, J., Cho, C. & Thorner, J. Comprehensive Genetic Analysis of Paralogous Terminal Septin Subunits Shs1 and Cdc11 in *Saccharomyces cerevisiae*. *Genetics* **200**, 821-841, doi:10.1534/genetics.115.176495 (2015).

4 Versele, M. *et al.* Protein-protein interactions governing septin heteropentamer assembly and septin filament organization in *Saccharomyces cerevisiae*. *Molecular biology of the cell* **15**, 4568-4583, doi:10.1091/mbc.E04-04-0330 (2004).

5 Finnigan, G. C. & Thorner, J. mCAL: a new approach for versatile multiplex action of Cas9 using one sgRNA and loci flanked by a programmed target sequence. *G3 (Bethesda, Md.)* **6**, 2147-2156, doi:10.1534/g3.116.029801 (2016).

6 Christianson, T. W., Sikorski, R. S., Dante, M., Shero, J. H. & Hieter, P. Multifunctional yeast high-copy-number shuttle vectors. *Gene* **110**, 119-122 (1992).

7 Sikorski, R. S. & Hieter, P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* **122**, 19-27 (1989).

8 Roggenkamp, E. *et al.* CRISPR-UnLOCK: multipurpose Cas9-based strategies for Conversion of yeast libraries and strains. *Frontiers in microbiology* **8**, 1773, doi:10.3389/fmicb.2017.01773 (2017).

9 DiCarlo, J. E. *et al.* Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic acids research* **41**, 4336-4343, doi:10.1093/nar/gkt135 (2013).

10 Finnigan, G. C. & Thorner, J. Complex *in vivo* ligation using homologous recombination and high-efficiency plasmid rescue from *Saccharomyces cerevisiae*. *Bio-protocol* **5**, e1521. http://www.bio-protocol.org/e1521 (2015).

Appendix D - Analysis of a Cas12a-based gene drive system in budding yeast

This chapter has been published in *Access Microbiology*.

Isabel Lewis[1,2], Yao Yan[1], and Gregory C. Finnigan[1†]

[1]Department of Biochemistry and Molecular Biophysics,

Kansas State University, Manhattan, KS 66506 USA

Running Title: Cas12a gene drive in yeast

Keywords: CRISPR, Cas12a, gene drive, yeast

[2]Current Address: School of Medicine, University of Texas Medical Branch, Galveston, TX, 77555 USA

[†]Correspondence to:

Gregory C. Finnigan

Dept. of Biochemistry & Molecular Biophysics

Kansas State University

141 Chalmers Hall, 1711 Claflin Rd.

Manhattan, KS 66506 USA

Phone: (785) 532-6939; FAX; (785) 532-7278

E-mail: gfinnigan@ksu.edu

Abbreviations

CRISPR, clustered regularly interspaced short palindromic repeat; DSB, double strand break; GD, gene drive; HDR, homology-directed repair; NHEJ, non-homologous end joining; PAM, protospacer adjacent motif; gRNA, guide RNA; pr, promoter; t, terminator; NLS, nuclear localization signal; SD, synthetic drop-out with dextrose; (u1′)/(u2′), unique artificial CRISPR target sites programmed within the yeast genome; *Sp*Cas9, *Streptococcus pyogenes* Cas9; *Fn*Cas12a, *Francisella novicida* Cas12a (formally known as Cpf1); *SpHIS5*, *Schizosaccharomyces pombe HIS5* (functional equivalent of budding yeast *HIS3*); *CaURA3, Candida albicans URA3*; dCas9, an enzymatically dead variant of Cas9 containing D10A and H840A mutations.

Abstract

The discovery and adaptation of CRISPR/Cas systems within molecular biology has provided advances across biological research, agriculture, and human health. Genomic manipulation through use of a CRISPR nuclease and programmed guide RNAs has become a common and widely accessible practice. The identification and introduction of new engineered variants and orthologs of Cas9 as well as alternative CRISPR systems such as the type V group have provided additional molecular options for editing. These include distinct PAM requirements, staggered DNA double strand break formation, and the ability to multiplex guide RNAs from a single expression construct. Use of CRISPR/Cas has allowed for the construction and testing of a powerful genetic architecture known as a gene drive within eukaryotic model systems. Our previous work developed a drive within budding yeast using *S. pyogenes* Cas9. Here, we installed the type V *F. novicida* Cas12a (Cpf1) nuclease gene and its corresponding guide RNA to power a highly efficient artificial gene drive in diploid yeast. We examined the consequence of altering guide length or introduction of individual mutational substitutions to the crRNA sequence. Cas12a-dependent gene drive function required a guide RNA of at least 18 bp and could not tolerate most changes within the 5' end of the crRNA.

Introduction

The discovery of a diverse set of clustered regularly interspaced short palindromic repeat (CRISPR) systems in microbes has led to the development of many molecular tools for genomic editing in basic and medical research. For the most well studied and utilized system, the type II *S. pyogenes* Cas9 (*Sp*Cas9) has provided major advances across many types of organisms [1]. Use of this system for genomic manipulation requires: (i) expression of the Cas9 nuclease, (ii) a single guide RNA (gRNA) fragment, and (iii) an intended DNA target that also contains a 3' protospacer adjacent motif (PAM) with the sequence 5'-NGG-3' (where N represents any nucleotide) [2,3]. The power of this system lies within the programmability of the gRNA sequence—inclusion of nearly any 20 base pair (bp) combination can be used with the universal structural RNA component (tracrRNA) and co-expressed nuclease. Together, the Cas9/gRNA complex will (i) search and bind to the corresponding DNA sequence that is complementary to the intended gRNA sequence included within the crRNA, (ii) recognize the 3' PAM sequence within the DNA target, and (iii) create a double strand break (DSB) at a precise position three bp upstream of the 5' end of the PAM [4]. Eukaryotic DNA repair systems including non-homologous end joining (NHEJ) and homology directed repair (HDR) allow for subsequent editing. In the absence of any added donor DNA source, end joining causes the severed ends to be repaired, typically with an insertion or deletion occurring at the site of cleavage. On the other hand, HDR systems will allow for the inclusion of exogenous DNA (either a single base, or entire genes or constructs) provided there is sufficient flanking homology surrounding the DSB [4]. This simple, yet highly efficient methodology now allows for routine genomic alteration and editing within plants, animals, fungi, and microbes.

Since the adoption of *Sp*Cas9, there has been interest in the discovery and characterization of other orthologous CRISPR systems. Many groups have pursued (i) Cas9 nucleases from different bacterial species, (ii) engineered variants developed in the lab, and (iii) independent classes of CRISPR systems [5-10]. Briefly, Cas9 orthologs seem to require a unique PAM sequence and guide RNA sequence and structure [8,11]. Modifications to existing Cas nucleases have focused on both reduction of overall protein size as well as altering PAM specificity and/or changes in potential off-target effects [6,12]. Interestingly, the characterization of non-Cas9 nucleases has provided a new suite of molecular tools and opportunities for editing. The type V Cas12a nuclease (formally known as Cpf1) includes several major differences compared to Cas9. First, the PAM requirement is on the 5' end of the DNA target for Cas12a and is T-rich (5'-TTN-3') [9]. Second, the guide RNA specific for Cas12a does not include a 3' tracrRNA structure. Rather, crRNA is flanked by short direct repeats and the nuclease is able to self-process its own guide RNA such that multiple independent (mature) guides will result from a single RNA fragment [9,13]. Third, Cas12a creates a staggered DNA break as opposed to a blunt ended cleavage event for Cas9 [9] and this may aid in shifting a preference for DNA repair by HDR over NHEJ [14]. These and other differences provide the opportunity for potential advantages over Cas9-based editing including (i) the ability to multiplex from one gRNA expression construct, (ii) editing within AT-rich genomic regions, and (iii) combinatorial applications using both type II and type V nucleases and their corresponding guide RNAs.

A unique application of CRISPR/Cas systems is within a "gene drive" (GD). This biotechnology could one day be used to aid in controlling biological populations of pests, parasites, and invasive species that could have profound effects on human health, agriculture, and at-risk ecosystems [15-17]. The basic architecture of a CRISPR-based gene drive includes installation of the

gene cassettes for the nuclease and the guide RNA at a particular genomic locus. What makes this arrangement unique is that the programmed DNA target of Cas9/gRNA will be the WT allele on the opposing homologous chromosome within a heterozygous diploid genome. In this way, a DSB will be introduced opposite of the drive locus; activation of HDR systems will use the homologous, intact chromosome to copy the entirety of the gene drive in place of the severed WT allele, thus converting the GD/WT heterozygous cell into a GD/GD homozygous cell [18-20]. This super Mendelian mechanism could theoretically allow for propagation of a drive through a global population with high speed and in a small number of generations. The intended purpose of this strategy would be to either (i) control the population through bias of sex determination causing populations to crash and be reduced or eliminated and/or (ii) to install genetic cargo proximal to the drive itself that could modify the organism (for instance, providing resistance to pathogens or altering other phenotypes). To date, these systems have been developed and studied in fungi, metazoans, bacteria, and viruses under laboratory conditions [18,19,21-23]. Our previous work has developed a safe and "artificial" drive system in budding yeast for study of various aspects of CRISPR/Cas editing such as nucleocytoplasmic trafficking, guide RNA specificity, and anti-CRISPR inhibition [24-26].

In this study, we have modified our gene drive design in *S. cerevisiae* to include the *F. novicida* Cas12a (*Fn*Cas12a) nuclease gene, a corresponding guide RNA, and modified target allele system using non-native DNA sequences. We demonstrate activation of Cas12a/gRNA allows for highly efficient and successful gene drive propagation within diploid yeast cells using a triple selection methodology. Additionally, we tested alterations to both the guide RNA length and mutational substitutions within the programmed crRNA to determine the effect on overall gene drive success. Our findings highlight that *Fn*Cas12a can edit at high levels with a variety of guide

lengths, displays sensitivity to single changes within the 5' end of the guide RNA sequence, and appears to have both positional and sequence dependent effects with regards to single nucleotide changes in the guide RNA.

Materials and Methods

Yeast strains and plasmids

Saccharomyces cerevisiae strains and plasmids used in this study can be found in Supplementary Tables S1 and S2. Haploid yeast were constructed by a combination of *in vivo* plasmid ligation [27] and subsequent chromosomal integration(s) using amplified linear PCR fragments and CRISPR/Cas12a-based editing. The design of the *Fn*Cas12a-based drive and target alleles were modeled after past iterations of *S. pyogenes* Cas9-based drives in yeast [25,28] including a custom synthesized *Fn*Cas12a gene with a yeast codon bias (Genscript). Several alterations were also included compared to past drive systems. For instance, altered (u1) and (u2) sites [29], termed (u1') and (u2') were used for inclusion of the 5' Cas12a PAM site, and a terminator sequence from the *CDC11* locus downstream of the *Schizosaccharomyces pombe HIS5* (*SpHIS5*) selection marker within the target alleles was also included (Supplemental Table S1). For inclusion of the *Candida albicans URA3* (*CaURA3*) marker within the gene drive allele strains, an identical parental strain was first constructed containing the Kan^R cassette at the 3' end of the *HIS3* locus. A *Fn*Cas12a guide RNA construct (pGF-V2149) was generated that targeted a sequence within the Kan^R gene coding strand (Supplemental Fig. S1). Briefly, yeast were co-transformed with (i) the gRNA(Kan^R) high-copy plasmid and (ii) an amplified MX-based *CaURA3* fragment followed by cell recovery within YP+GAL liquid (to activate expression of Cas12a) at 30°C and final plating onto SD-LEU medium. Clonal isolates were assessed for (i) loss of G418 resistance (loss of the Kan^R allele), (ii)

growth on SD-URA (inclusion of *CaURA3* allele), and (iii) sensitivity to SD-LEU (loss of guide RNA plasmid) to generate the final yeast strains. PCRs of isolated chromosomal preparations (Supplemental Fig. S2) and DNA sequencing confirmed all modified loci in haploids (Supplemental Fig. S1). Plasmids for this study included modifications to a parental DNA construct expressing the "WT" (u1') guide RNA sequence for *Fn*Cas12a (pGF-V1895). Site-directed mutagenesis [30] or custom gene synthesis produced the collection of expression constructs for the Cas12a guide RNAs of varying lengths and mismatches (Supplemental Table S2, Supplemental Fig. S1). Briefly, unique restriction sites (*SpeI*/*NotI*) were included within gRNA constructs and obtained by custom synthesis on a pUC57-based vector. The gRNA cassette was then subcloned to pRS425 (digestion of both constructs using *SpeI*/*NotI*, separation by gel electrophoresis, gel extraction and purification, *in vitro* DNA ligation, transformation into *E. coli*, plasmid extraction, and DNA sequencing for confirmation). For DNA amplification via PCR, KOD Hot Start DNA Polymerase (EMB Millipore) was used with synthetic oligonucleotides (IDT). Isolation of plasmid DNA from *E. coli* was performed using a commercially available GeneJET plasmid DNA miniprep kit (Thermo Scientific); isolation of plasmid DNA from budding yeast utilized a modified protocol [27] and the same DNA extraction materials.

Culture conditions

Yeast were grown on rich medium (YPD consisting of 2% peptone, 1% yeast extract, and 2% dextrose) or synthetic minimal medium (S+DEX consisting of a yeast nitrogen base, ammonium sulfate, and amino acids) at 30°C. Drop-out mixtures removed one or more amino acids (e.g., SD-URA-LEU). Raffinose and galactose were used at 2%, while sucrose was used at

0.2% (all sugars were filtered, not autoclaved). G418 was included in YPD plates at a concentration of 240 μg/mL.

Gene Drive Assays and Biosafety

Haploid strains harboring the Cas12a drive were first transformed with the guide RNA plasmid of interest (or empty vector control) and propagated on SD-LEU-URA plates; haploids harboring the target allele were maintained on SD-HIS medium. For gene drive activation within diploids, haploid strains were first mated on YPD medium at 30°C for 24 hr. Next, diploids were selected by replica-plating to SD-LEU-URA-HIS plates and incubating at 30°C for 48 hr. Two additional diploid selection steps were performed on the same media type and incubated at 30°C for 24 hr. Overnight cultures of diploids were grown in S+RAFF/SUC-LEU-URA-HIS liquid medium at 30°C with constant shaking followed by back-dilution into YP+GAL liquid for between 0-7.5 hr. Cell were harvested by centrifugation, diluted into sterile water, spread onto SD-LEU-URA plates at an approximate density of 100-500 cells per plate, and incubated at 30°C for 48 hr. Finally, colonies were transferred by replica-plating (sterile velvet transfer) to a new SD-LEU-URA plate and a SD-HIS plate and incubated at 30°C for 24 hr before imaging. Experiments were performed in at least triplicate. Quantification of surviving yeast colonies was done in a single-blind fashion as previously described [25,28]. Depending on the total number of colonies per agar plate, a random sector was chosen from the SD-LEU-URA plate. The gene drive efficiency was calculated as the percentage of colonies able to grow on SD-LEU-URA that were also sensitive to SD-HIS medium following gene drive activation. Error was presented as the standard deviation across separate experimental trials (also see Supplemental Table S4).

Our previous work has described a number of safety mechanisms built into our CRISPR/Cas9-based gene drives in budding yeast systems [24,25]. The same principles have been adopted in this study with use of Cas12a in place of Cas9. Briefly, these include the following strategies. First, target DNA sequences and the corresponding guide RNAs are programmed to cleave non-native sequences (such as the (u1') motif) to the *S. cerevisiae* genome to both reduce the potential for off-target effects and potential action of the drive within native populations. Second, the expression cassette for the guide RNA is present on a high-copy plasmid; this has been shown to be spontaneously lost in the absence of selection. Third, expression of the Cas12a nuclease is under control of the inducible *GAL1/10* promoter and is repressed during growth on dextrose-containing medium. Fourth, the drive locus is flanked by (u2') sequences that can be used to self-cleave (and remove) the entire drive cassette; this can also be accomplished by targeting other non-yeast DNA within the locus (such as Kan^R) along with donor DNA containing adequate *HIS3* flanking sequence and HDR-based repair. Fifth, there now exist a suite of anti-CRISPR proteins that evolved to directly counter and inhibit action of CRISPR/Cas nucleases such as Cas9 [24,31,32] and Cas12a [33,34]. We have confirmed that several of the AcrVA proteins, when expressed in budding yeast, can inhibit *Fn*Cas12a-based gene drives (our unpublished data). Sixth, all strains, DNA, and materials used with gene drive yeast were autoclaved and sterilized following experimentation.

Results

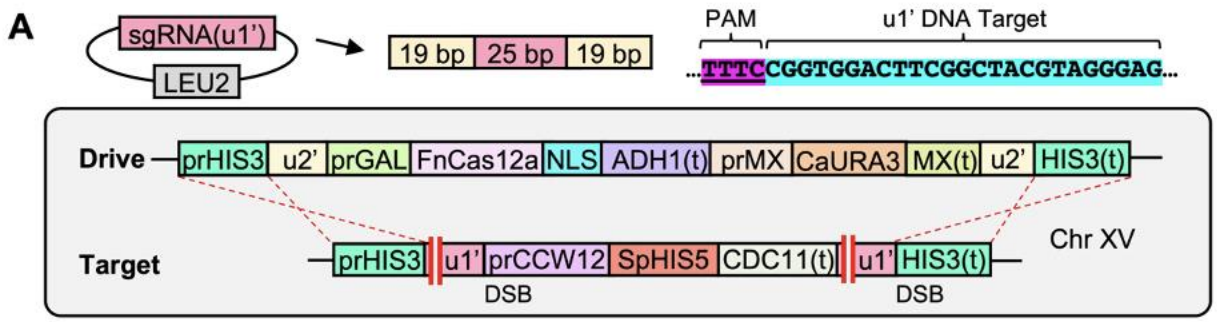
Design of CRISPR gene drive system in *S. cerevisiae* using *F. novicida* Cas12a

Our previous work in yeast has employed use of the type II *S. pyogenes* Cas9 to power gene drive systems within diploid cells [24-26,28]. Given the discovery and widespread availability of

other CRISPR nucleases and engineered variants, we predicted that use of alternative systems may also allow for successful gene drive (GD) activity *in vivo*. Therefore, we altered our previous GD constructs to include (i) the type V *Francisella novicida* Cas12a (formally Cpf1) nuclease, (ii) a corresponding guide RNA (gRNA), and (iii) appropriate target DNA sequences (Fig. D.1). The *Fn*Cas12a gene (translated to 1,300 residues) was codon optimized for expression in yeast including a C-terminal nuclear localization signal (NLS). The guide RNA for Cas12a included several important differences compared to the guide RNA for Cas9. First, there is no extended tracrRNA sequence at the 3' end of the crRNA sequence; rather, a short repeat is found at the 5' end of the crRNA. Previous work with Cas12a including studies in yeast [35] determined that inclusion of a flanking direct repeat of 19 bp placed on either side of the crRNA is sufficient for successful editing *in vivo*. Moreover, Cas12a is involved in processing of pre-crRNA fragments resulting in a 5' direct repeat upstream of the crRNA for subsequent editing of target DNA. Finally, rather than requiring a PAM sequence of 5'-NGG-3' at the 3' end of the DNA target (such as for *Sp*Cas9), the *Fn*Cas12a nuclease required a PAM sequence 5' of the target DNA with the sequence 5'-TTTV-3' (V is any nucleotide other than thymine) [35].

To accommodate these differences, we modified the existing architecture of our previous gene drive system in yeast to include the sequence 5'-TTTC-3' directly upstream of the unique (u2) or (u1) sites flanking the nuclease expression cassette or target allele, respectively, to create (u2') and (u1') (Fig. D.1; Supplemental Fig. S1). We have previously demonstrated the utility and importance of targeting artificial DNA sequences when studying gene drives *in vivo* [25]. Briefly, use of a programmed non-native sequence [29] allows for (i) multiplexing with one guide RNA, (ii) complete excision of target genes (including selection markers), (iii) potential reduction of off-target effects, (iv) rapid self-excision of the nuclease gene if needed as a safeguard, and (v)

prevention of inappropriate activation of the gene drive in case of theoretical release since the target sequence(s) do not exist in native *S. cerevisiae* genomes. We chose to synthesize a guide RNA expression cassette (RNA Pol III) developed by previous groups [36] that included 19 bp direct repeats flanking a crRNA of 25 bp. We also modified the target allele(s) at the *HIS3* locus to include a much shorter gene cassette with a unique terminator sequence (compared to previous iterations). Finally, our gene drive design standardized all selection markers within the drive (*URA3*), guide RNA (*LEU2*), and target (*HIS3*) to be compatible auxotrophic markers. Prior to expression of Cas12a using galactose medium, our system maintained active repression of the nuclease, constitutive expression of the guide RNA, and maintenance of both the drive and target alleles at the *HIS3* locus through the diploid selection process (Fig. D.1B).



B

Strain Ploidy	CRISPR Component(s)	Selection Marker(s)	gRNA	FnCas12a Expression
Haploid	Gene Drive + gRNA	*CaURA3 (URA3) LEU2*	ON	Dextrose (OFF)
Haploid	Target	*SpHIS5 (HIS3)*	ON	Dextrose (OFF)
Diploid	Gene Drive + gRNA + Target	*(URA3) LEU2 (HIS3)*	ON	Dextrose (OFF) Raff/Suc (OFF) Galactose (ON)

Figure D.1. Design of a CRISPR/Cas12a-based gene drive system in *S. cerevisiae.*

(A) Artificial (u1') and (u2') sites were used as flanking DNA within the Cas12a drive system at the *HIS3* locus (Chromosome XV). The selection marker for the gene drive was *Candida albicans URA3*; for the target allele, *Schizosaccharomyces pombe HIS5* (functional equivalent of yeast *HIS3*) was used. An expression cassette for the *Fn*Cas12a guide RNA included 19 bp repeats flanking a crRNA of 25 bp. This high-copy plasmid included a *LEU2* selection marker. The

*Fn*Cas12a gene (containing a C-terminal NLS) was under control of the inducible *GAL1/10* promoter whereas the target allele was under control of the constitutive *CCW12* promoter. (B) Table illustrating the generation and propagation of parental haploid (drive or target) strains versus final diploid (drive + target) yeast. Expression of Cas12a (unlike the guide RNA) was restricted to the diploid strain (galactose). Also, selection was constantly maintained for both the guide RNA plasmid (*LEU2*) and the drive allele (*URA3*) until activation of the nuclease in rich medium with galactose.

An efficient and successful Cas12a-based gene drive system in yeast

Our gene drive system included placement of the nuclease (expression inhibited) and the guide RNA-containing plasmid within one haploid yeast strain and the intended target DNA at the same locus within a second haploid yeast strain of the opposite mating type. Activation of the system involved mating the strains together and selecting for only diploid cells (also harboring the guide RNA plasmid). Finally, cultures of diploids were incubated in galactose for varying amounts of time to activate expression of Cas12a (along with constitutive expression of the guide RNA). Cells were then plated onto SD-URA-LEU medium and incubated, allowing for growth of individual visible colonies. Finally, yeast were transferred onto both a SD-URA-LEU and a separate SD-HIS plate. Diploid cells that (i) expressed Cas12a, (ii) targeted the (u1') site(s) flanking the target allele, and (iii) copied the drive locus in place of the target locus would be sensitive to the SD-HIS condition. This experiment was performed for drive-containing haploids with an empty plasmid or the guide RNA to the target sites (Fig. D.2A, Supplemental Fig. S3). Yeast with the empty plasmid displayed robust growth on SD-HIS medium following galactose culturing. However, inclusion of the guide RNA plasmid caused a distinct and similar pattern of growth as seen with Cas9-containing yeast gene drives. Increasing the time cultured in galactose resulted in a marked decrease in surviving colonies on SD-HIS (Fig. D.2A). The experiment was repeated in triplicate with two variations of the target allele (Supplemental Fig. S1); we observed that five hours in galactose resulted in nearly 100% of the colonies sensitive to the SD-HIS

condition, suggesting a loss of the *SpHIS5* target allele and replacement by the drive allele through HDR.

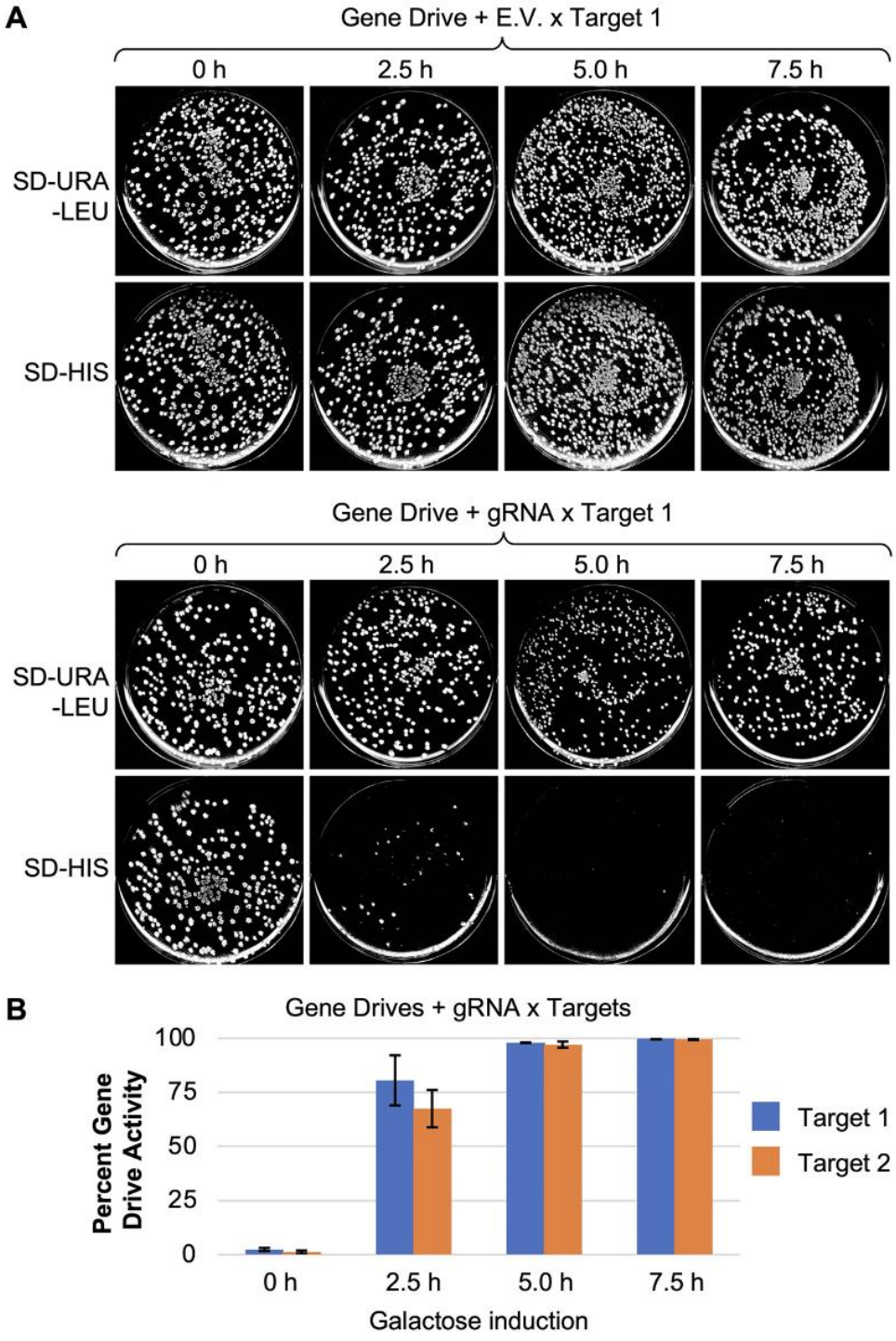
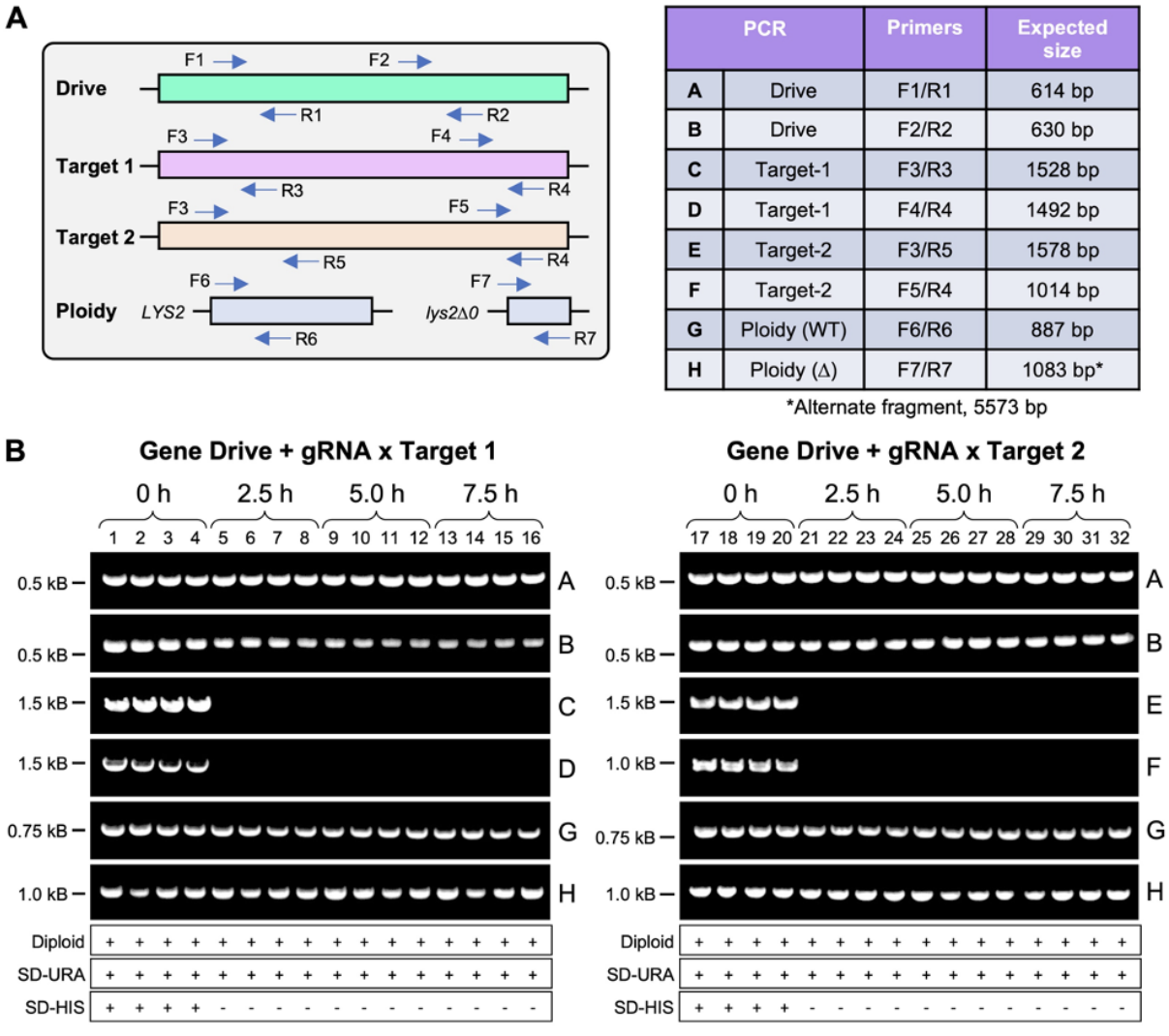


Figure D.2. An efficient Cas12a-based gene drive *in vivo*.

(A) The haploid strain GFY-4625 (gene drive) was transformed with a plasmid containing the guide RNA plasmid (pGF-V1895) to target the (u1') DNA sites or an empty vector (pRS425). The gene drive assay included (i) mating the drive strains with haploids of GFY-4424 (target), (ii) selecting diploids, (iii) culturing diploids in S+RAFF/SUC-LEU-URA-HIS liquid overnight, (iv) diluting cells in YP+GAL for between 0-7.5 hr, and (v) plating yeast onto SD-LEU-URA plates. Finally, viable colonies were transferred to SD-LEU-URA and SD-HIS plates by replica-plating and imaged after 24 hr of growth (also see Supplemental Fig. S3 for unmodified images). (B) Gene drive activation was performed for three independent drive strains (GFY-4625, GFY-4626, and GFY-4627) and two separate haploids containing targets (Target 1, GFY-4424; Target 2, GFY-4425) for a similar time course as performed in (A). Surviving colonies were quantified to illustrate the percent gene drive activity (the fraction of colonies viable on SD-LEU-URA but sensitive to the SD-HIS condition). The average of three separate drive strains is shown (n>100 colonies per sample). Error, SD.

To verify that the gene drive plating assay was accurately illustrating Cas12a-based editing of the target allele followed by replacement of the target locus using the drive allele, we utilized a collection of diagnostic PCRs to assay both the *HIS3* locus (drive and/or target) as well as the *LYS2* locus. The parental haploid strains used in this study (BY4741, *MAT***a** and BY4742**,** *MAT*α) differ at both the *MET15* and *LYS2* loci (Supplemental Table S1). PCR amplification of the *LYS2* locus served as a convenient method to determine if isolated clones from the gene drive assay (presumably diploids from the selection protocols) contained both a WT and a deleted copy of *LYS2* (haploids would only harbor one or the other allele, not both). For both the *HIS3* and *LYS2* loci, two independent PCRs were performed using unique oligonucleotide pairs (Fig. D.3A, Supplemental Table S3). Following gene drive activation, clonal isolates were randomly selected from the SD-URA-LEU plate for all conditions (Fig. D.2B) and retested for growth on either SD-URA (drive selection) or SD-HIS (target selection) (Fig. D.3B). We performed PCRs on isolated and purified chromosomal DNA of four random clones from each time point for experiments using the Target 1 strain (Fig. D.3B, *left*) or Target 2 strain (Fig. D.3B, *right*). Prior to induction in galactose, DNA from isolates allowed for the amplification of all six PCR products, suggesting the presence of both the drive and target alleles as well as both versions of the *LYS2* locus (Fig.

D.3B). Following activation of Cas12a, all isolates tested were unable to successfully amplify fragments of the expected size at the target *HIS3* locus and this correlated with sensitivity to the SD-HIS condition (Fig. D.3B, Supplemental Fig. S4). Moreover, all tested yeast appeared to be diploids and allowed for amplification of both alleles of *LYS2*. Together, these data suggest that Cas12a and its guide RNA allow for a highly efficient gene drive system in budding yeast diploid cells.



	PCR	Primers	Expected size
A	Drive	F1/R1	614 bp
B	Drive	F2/R2	630 bp
C	Target-1	F3/R3	1528 bp
D	Target-1	F4/R4	1492 bp
E	Target-2	F3/R5	1578 bp
F	Target-2	F5/R4	1014 bp
G	Ploidy (WT)	F6/R6	887 bp
H	Ploidy (Δ)	F7/R7	1083 bp*

*Alternate fragment, 5573 bp

Figure D.3. Characterization of gene drive success within diploids using diagnostic PCR.

Following gene drive activation of diploids (Fig. D.2), four random clonal isolates were obtained from each time point for the gene drive strain crossed to either haploid target 1 or target 2 for

further analysis. (A) *Left,* Chromosomal DNA was isolated and purified for each diploid strain and tested by PCR across both the *HIS3* and *LYS2* loci. Oligonucleotides used in this study can be found in Supplemental Table S3. *Right,* The expected product size(s) for each diagnostic PCR. (B) Strains were also retested on SD-URA and SD-HIS plates for survival. Diploid status was determined through amplification of *LYS2* (PCRs G, H). PCRs were separated by DNA gel electrophoresis and visualized. Distinct DNA gels are separated by white lines; the nearest molecular marker is illustrated on the left. PCRs C and D were only performed for target 1 (*left*) whereas PCRs E and F were only performed for target 2 (*right*). Unmodified images can be found in Supplemental Fig. S4.

Robustness of Cas12a-based gene drives to alterations in guide RNA sequence

Numerous studies have demonstrated that editing efficiency with CRISPR/Cas is dependent on guide RNA sequence, length, mismatches, structure, and even chemical modifications [9,37-39]. We chose to examine how alteration of the crRNA sequence length used to target the flanking (u1') sites within our system would affect overall gene drive efficiency. We tested guide RNAs of 16-27 bp (extending or reducing the sequence at the 3' end) and found nearly 100% drive activity for guide RNAs of lengths 18-27 (Fig. D.4, *top*). There was a low level of activity for a guide RNA length of 17 bp, whereas no activity was seen for a guide length of 16 bp. This appears similar to other findings [9] where a range of guide lengths could provide efficient editing. Of note, all guide RNA expression cassettes maintained identical flanking 19 bp direct repeats (Supplemental Fig. S1).

Next, we examined the consequence of a single base mismatch within the crRNA sequence of a guide RNA of 25 bp in total length. For completeness, we mutated each base within the guide RNA expression cassette to all possible nucleotides for positions 1-17 (counting from the 5' end of the crRNA). For positions 18-25, only a single substitution was tested. These experiments were done in triplicate, using three independent drive strains and performed at the same time as the guide length experiments. Including a negative control strain harboring an empty vector, these 72 separate gene drive haploids were mated to the target strain, selected for diploids, activated by

culturing in galactose, plated, and quantified for activity (Fig. D.4, *bottom*, Supplemental Table S4). We observed several trends that defined regions within the crRNA sequence. First, substitutions within positions 1-8 resulted in a near total loss of activity apart from the C1G and A7G mutants (high activity) and all C8 substitutions (low activity). Second, there appeared to be little effect from any substitution within positions 9-11. Third, for positions 12-17, an interesting pattern seemed to repeat—two substitutions provided relatively high editing whereas one substitution caused a marked decrease or low editing. It may include sequence specific changes as both G12C and G13C caused reduced editing as well as both C14A and C17A; it remains unclear if the same trend exists for uracil or adenine mutations such as U15A and A16C, both of which had the strongest reduction on overall editing and drive activity. Finally, no change was seen for mutations within positions 18-25. These data illustrate that (i) critical nucleotides within the crRNA include the first eight at the 5' end and this appears independent of the identity of the substitution, (ii) a central region of the crRNA (defined by positions 9-11) may not be critical for editing within the context of a 25 bp guide, (iii) positions 12-17 may represent a region where Cas12a/gRNA function is dependent on the nature of the substitution, and (iv) the 3' region consisting of positions 18-25 have little individual contribution to overall editing success and this is supported by our observations of guide RNAs of varying lengths.

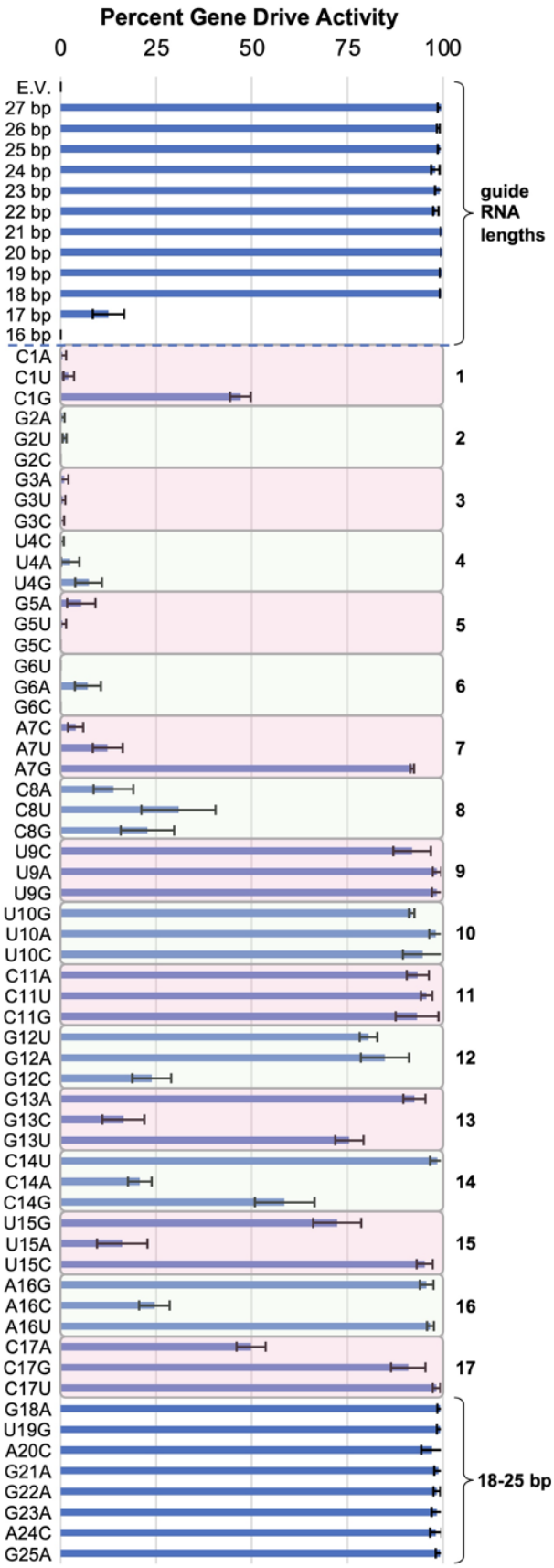

Percent Gene Drive Activity
0
25
50
75
100
E.V.
27 bp
26 bp
25 bp
24 bp
23 bp
22 bp
21 bp
20 bp
19 bp
18 bp
17 bp
16 bp
guide RNA lengths
C1A
C1U
C1G
1
G2A
G2U
G2C
2
G3A
G3U
G3C
3
U4C
U4A
U4G
4
G5A
G5U
G5C
5
G6U
G6A
G6C
6
A7C
A7U
A7G
7
C8A
C8U
C8G
8
U9C
U9A
U9G
9
U10G
U10A
U10C
10
C11A
C11U
C11G
11
G12U
G12A
G12C
12
G13A
G13C
G13U
13
C14U
C14A
C14G
14
U15G
U15A
U15C
15
A16G
A16C
A16U
16
C17A
C17G
C17U
17
G18A
U19G
A20C
G21A
G22A
G23A
A24C
G25A
18-25 bp

Figure D.4. Tolerance of the Cas12a gene drive system to alterations in guide RNA length and RNA/DNA mismatches.

Three gene drive strains (GFY-4625, GFY-4626, GFY-4627) were transformed with an identical sets of guide RNA or control plasmids (72 total; see Supplemental Table S2) and crossed with haploid GFY-4424 to generate diploids. Following Cas12a activation, strains were quantified for the percent gene drive activity as in Fig. D.2B (also see Supplemental Table S4). The average of the three strains is presented; error is standard deviation. *Top*, the guide RNA plasmid was altered to present a final crRNA size of 16-27 bp (flanked by the 19 bp repeat). Note, the guide RNA length of 27 bp only contains a full RNA/DNA match (27/27) to the downstream (u1') target site as the upstream site contains a 3' mismatch (26/27). *Bottom*, a guide RNA of 25 bp targeting the (u1') sequence was used. Nucleotide position is numbered 1-25 beginning at the 5' end of the crRNA sequence. For the 17 nucleotides at the 5' end, each base was mutated to one of the three possible alternatives (RNA change is displayed). For positions 18-25, only a single bp change was tested. For most conditions, >100 colonies were quantified.

Discussion

Alternative CRISPR nucleases to power gene drives

Current gene drive biotechnology utilizes a nuclease (most commonly, *S. pyogenes* Cas9) to create a double strand break at a particular location within the genome. Subsequent DNA repair via HDR allows the non-targeted homologous chromosome to serve as the donor DNA and copies the entirety of the drive locus itself (and any associated cargo DNA), thus converting a heterozygous condition into a homozygous condition for the allele(s) of interest. Since the primary purpose of the (active) nuclease, when paired with appropriate guide RNA(s), is to introduce a DSB, it was expected that a drive system could include the type V *Fn*Cas12a nuclease and allow for proper drive conversion *in vivo*. Use of alternative nucleases and engineered variants of CRISPR nucleases (including Cas12a from other species) will allow for the development of more complex drive systems in the future. For instance, due to the difference in PAM specificity between Cas9 (5'-NGG-3') and Cas12a (5'-TTN-3'), the choice of which nuclease to include within a gene drive may depend on the GC-content of either the locus of interest, or the prevalence across a genome. Moreover, we envision that complex drives could co-exist within the same genome—our

previous work has demonstrated that a triple-locus drive is possible with three separate guide RNAs [40]. However, in that arrangement, a single copy of *Sp*Cas9 was expressed and used to independently target each locus with a unique guide RNA [39]. A modified system might include two or more nucleases, each with a unique guide RNA construct, to selectively target one or more loci within a genome. This would provide additional control over selectivity of drive function and modulation or inhibition (if needed) using, for instance, competing "anti-drives" [41] or introduction or induction of anti-CRISPRs [24]. The ability to selectively activate (or repress) multiple independently functioning nucleases might also be applied to use of enzymatically "dead" variants (such as dCas9) [42] in order to modify transcription of distal gene targets unrelated to the primary function of DSB introduction to propagate the drive itself. Finally, use of two or more nucleases might provide additional redundancy to ensure DSB formation at a particular locus with independent targets within a single gene of interest. Along these lines, recent work has successfully engineered a "fusion guide RNA" construct that can be used by both Cas9 and Cas12a [43].

Guide RNA mismatch specificity

Our analysis of substitutions to the guide RNA for *Fn*Cas12a illustrate the importance of nucleotides at the 5' end; other studies of Cas12a have also suggested the first few bases of the "seed" region (PAM proximal bases required for full editing) are critical for function [9,13,44-47]. While numerous studies have investigated the effects of mutational substitutions (or additions or deletions) within the crRNA guide sequence for a variety of CRISPR nucleases, it is important to highlight that our assay is distinct from previous efforts for a number of reasons: (i) there are dual identical (u1') sites at the target locus and (ii) the measurement of "gene drive activity" corresponds to loss of the selection marker and is not a direct measure of DNA cleavage. Our

previous work has demonstrated that in a yeast drive system, NHEJ may occur at an extremely low level (if at all) [28]. Therefore, we attribute the "success" of the drive (via loss of the target locus) to the combination of Cas12a/gRNA binding, recruitment to the DNA target(s), DSB formation, and finally, repair via HDR within a diploid genome. Therefore, our findings for guide RNA requirements may differ from either *in vitro* or *in vivo* studies (typically done through plasmid DNA linearization or NHEJ-based repair and detection, respectively). For technical reasons, we chose to introduce the mutational substitution within the guide RNA expression cassette and not within the yeast genome. Our approach created a gRNA(mutated)/DNA(WT) pairing. Future work will be required to determine if similar results would be obtained through the same substitutions within the genome to create a gRNA(WT)/DNA(mutated) combination.

We cannot rule out the possibility that guide RNA sequence alterations reduced RNA stability, mature crRNA processing, binding to Cas12a, and/or RNA/DNA hybridization at the target sequence. Previous work has suggested that sequences within the programmable crRNA may disrupt secondary structure formation of the direct repeat and subsequent processing of the guide RNA or binding to Cas12a [48]. Moreover, a similar observation was reported for Cas9 where interactions between bases within the crRNA could impede nuclease function [49].

Our findings from this study suggest that mismatches within the *Fn*Cas12a guide RNA to the target genome may be both positional and sequence dependent. Our experiments did not investigate whether additional alterations to the guide RNA (consecutive mutations, deletions, additions, etc.) would exacerbate the effects of a single substitution, as has been performed by other groups [44,46,50]; commonly, the presence of multiple mutations greatly disrupts editing. Along these lines, our analysis was limited to only the programmed artificial [u1'] target sites; it remains to be seen whether similar results will be found at different genomic loci and/or using distinct

DNA sequences as targets. Extensive studies on *S. pyogenes* Cas9 guide RNA substitutions demonstrated variability across independent loci [37]; however, it may be difficult to separate the effects of genomic position versus sequence without use of an identical target sequence that either naturally exists across loci or has been engineered to include identical targets (as has been done in our yeast system). While our study has provided a preliminary investigation of guide RNA requirements for *Fn*Cas12a, future work will be required to validate whether positions within the guide, specific nucleotide substitutions, and/or potential RNA secondary structure(s) can be predicted to reduce or eliminate editing given the chosen sequence and genomic context.

Gene drives and evolved resistance

It has been well documented that there are multiple sources that can directly inhibit successful action of a gene drive [51-55]. Briefly, this corresponds to alteration of the target DNA sequence through various mechanisms: (i) natural variation within a population that includes mismatches to the guide RNA, (ii) *de novo* mutation at the target DNA site, and/or (iii) destruction or elimination of the target site through NHEJ-based repair following successful DSB formation. Therefore, different strategies are being employed to potentially reduce and/or eliminate drive resistance. These include use of multiplexed guide RNAs targeting separate DNA positions, choice of the gene target, choice of conserved DNA sites, and novel drive architectures [56-58]. Understanding how DNA polymorphisms might affect CRISPR/Cas-based DSB formation at intended target site(s) may provide additional resources to combat potential drive resistance. This could include installation of a small gRNA library harboring potential substitutions for a high priority DNA target to buffer against possible mutation or unknown genetic variability at the site. A more complete understanding of Cas/gRNA recognition, sequence preferences, and mismatch

tolerance would assist in the design and study of safe, controllable, and effective gene drive systems.

Acknowledgements

We would like to thank members of the Finnigan laboratory for useful comments, suggestions, and laboratory assistance. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institute of Health or the United States Department of Agriculture. I.C.L and G.C.F. built all constructs, performed all experiments, and analyzed data. G.C.F. generated figures and wrote the manuscript. Y.Y. assisted in data analysis, figure preparation, and writing the manuscript.

Funding Information

This research was supported by an Institutional Development Award (IDeA) P20 GM103418 to Kansas State Univ. (G.C.F. and I.C.L.) from the National Institute of General Medical Sciences (GM) of the National Institute of Health (NIH), the Johnson Cancer Research Center and the College of Arts and Sciences at Kansas State University (I.C.L.), and by the USDA National Institute of Food and Agriculture, Hatch Project 1013520 (G.C.F).

Conflict of Interest

G.C.F. has applied for a patent on April 20, 2017, by the Univ. of California, Berkeley, “Methods and Compositions for Genomic Editing” International Application No. PCT/US2017/028676 and published as No. WO 2017/189336 A1 on November 2, 2017, for the

artificial CRISPR target DNA sites (u1')/(u2') and corresponding guide RNAs. I.C.L. and Y.Y declare no financial conflicts of interest. All authors declare no non-financial conflicts of interest.

Ethical Statement

This work did not involve any human or animal subjects of any kind.

Data Availability Statement

Yeast strains and/or DNA plasmids used in this study will be made available for research or educational purposes upon reasonable request.

References

1. Asmamaw M, Zawdie B. Mechanism and applications of CRISPR/Cas-9-mediated genome editing. *Biologics* 2021; 15:353-361 DOI: 10.2147/BTT.S326422
2. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 2012; 337:816-821 DOI: 10.1126/science.1225829
3. Jinek M, East A, Cheng A, Lin S, Ma E et al. RNA-programmed genome editing in human cells. *Elife* 2013; 2:e00471 DOI: 10.7554/eLife.00471
4. Sternberg SH, Doudna JA. Expanding the biologist's toolkit with CRISPR-Cas9. *Mol Cell* 2015; 58:568-574 DOI: 10.1016/j.molcel.2015.02.032
5. Hou Z, Zhang Y, Propson NE, Howden SE, Chu, LF et al. Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. *Proc Natl Acad Sci* USA 2013; 110(39):15644-15649 DOI: 10.1073/pnas.1313587110

6. Kleinstiver BP, Prew MS, Tsai SQ, Topkar VV, Nguyen NT et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* 2015; 523(7561): 481-485 DOI: 10.1038/nature14592

7. Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS et al. In vivo genome editing using Staphylococcus aureus Cas9. *Nature* 2015; 520(7546): 186-191 DOI: 10.1038/nature14299

8. Shmakov S, Abudayyeh OO, Makarova KS, Wolf, YI, Gootenberg JS et al. Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. *Mol Cell* 2015; 60(3): 385-397 DOI: 10.1016/j.molcel.2015.10.008

9. Zetsche B, Gootenberg JS, Abudayyeh, Slaymaker IM, Makarova KS et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 2015; 163(3): 759-771 DOI: 10.1016/j.cell.2015.09.038

10. Abudayyeh OO, Gootenberg JS, Essletzbichler P, Han S, Joung J et al. RNA targeting with CRISPR-Cas13. *Nature* 2017; 550(7675): 280-284 DOI: 10.1038/nature24049

11. Koonin EV, Makarova KS, Zhang F. Diversity, classification and evolution of CRISPR-Cas systems. *Curr Opin Microbiol* 2017; 37: 67-78. DOI: 10.1016/j.mib.2017.05.008

12. Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX et al. Rationally engineered Cas9 nucleases with improved specificity. *Science* 2016; 351(6268): 84-88 DOI: 10.1126/science.aad5227

13. Fonfara I, Richter H, Bratovič M, Le Rhun A, Charpentier E. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* 2016; 532(7600): 517-521 DOI: 10.1038/nature17945

14. Paul B, Montoya G. CRISPR-Cas12a: Functional overview and applications. *Biomed J* 2020; 43(1): 8-17 DOI: 10.1016/j.bj.2019.10.005

15. Godfray HCJ, North A, Burt A. How driving endonuclease genes can be used to combat pests and disease vectors. *BMC Biol* 2017; 15(1): 81 DOI: 10.1186/s12915-017-0420-4

16. Hammond AM, Galizi R. Gene drives to fight malaria: current state and future directions. *Pathog Glob Health* 2017; 111(8): 412-423 DOI: 10.1080/20477724.2018.1438880

17. Bier E. Gene drives gaining speed. *Nat Rev Genet* 2021; 1-18 DOI: 10.1038/s41576-021-00386-0

18. DiCarlo JE, Chavez A, Dietz SL, Esvelt KM, Church GM. Safeguarding CRISPR-Cas9 gene drives in yeast. *Nat Biotechnol* 2015; 33:1250-1255 DOI: 10.1038/nbt.3412

19. Gantz VM, Jasinskiene N, Tatarenkova O, Fazekas A, Macias VM et al. Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi. *Proc Natl Acad Sci* USA 2015; 112(49): E6736-E6743 DOI: 10.1073/pnas.1521077112

20. Hammond A, Galizi R, Kyrou K, Simoni A, Siniscalchi C et al. A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae. *Nat Biotechnol* 2016; 34(1): 78-83 DOI: 10.1038/nbt.3439

21. Grunwald HA, Gantz VM, Poplawski G, Xu XRS, Bier E et al. Super-Mendelian inheritance mediated by CRISPR-Cas9 in the female mouse germline. *Nature* 2019; 566(7742): 105-109 DOI: 10.1038/s41586-019-0875-2

22. Valderrama JA, Kulkarni SS, Nizet V, Bier E. A bacterial gene-drive system efficiently edits and inactivates a high copy number antibiotic resistance locus. *Nat Commun* 2019; 10(1): 5726 DOI: 10.1038/s41467-019-13649-6

23. Walter M, Verdin E. Viral gene drive in herpesviruses. *Nat Commun* 2020; 11(1): 4884 DOI: 10.1038/s41467-020-18678-0

24. Basgall EM, Goetting SC, Goeckel ME, Giersch RM, Roggenkamp E et al. Gene drive inhibition by the anti-CRISPR proteins AcrIIA2 and AcrIIA4 in Saccharomyces cerevisiae. *Microbiology* 2018; 164:464-474 DOI: 10.1099/mic.0.000635

25. Roggenkamp E, Giersch RM, Schrock MN, Turnquist E, Halloran M et al. Tuning CRISPR-Cas9 gene drives in Saccharomyces cerevisiae. *G3* 2018; 8:999–1018 DOI: 10.1534/g3.117.300557

26. Goeckel ME, Basgall EM, Lewis IC, Goetting SC, Yan Y et al. Modulating CRISPR gene drive activity through nucleocytoplasmic localization of Cas9 in S. cerevisiae . *Fungal Biol Biotechnol* 2019; 6:2 DOI: 10.1186/s40694-019-0065-x

27. Finnigan GC, Thorner J. Complex in vivo ligation using homologous recombination and high-efficiency plasmid rescue from Saccharomyces cerevisiae . *Bio Protoc* 2015; 5:e1521 DOI: 10.21769/bioprotoc.1521

28. Yan Y, Finnigan GC. Analysis of CRISPR gene drive design in budding yeast. *Access Microbiol* 2019; 1(9): e000059 DOI: 10.1099/acmi.0.000059

29. Finnigan GC, Thorner J. mCAL: a new approach for versatile multiplex action of Cas9 using one sgRNA and loci flanked by a programmed target sequence. *G3* 2016; 6:2147-2156 DOI: 10.1534/g3.116.029801

30. Zheng L, Baumann U, Reymond JL. An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res* 2004; 32(14): e115 DOI: 10.1093/nar/gnh110

31. Pawluk A, Amrani N, Zhang Y, Garcia B, Hidalgo-Reyes Y et al. Naturally occurring off-switches for CRISPR-Cas9. *Cell* 2016; 167(7): 1829-1838.e9 DOI: 10.1016/j.cell.2016.11.017

32. Rauch BJ, Silvis MR, Hultquist JF, Waters CS, McGregor MJ et al. Inhibition of CRISPR-Cas9 with bacteriophage proteins. *Cell* 2017; 168(1-2): 150-158.e10 DOI: 10.1016/j.cell.2016.12.009

33. Marino ND, Zhang JY, Borges AL, Sousa AA, Leon LM et al. Discovery of widespread type I and type V CRISPR-Cas inhibitors. *Science* 2018; 362(6411): 240-242 DOI: 10.1126/science.aau5174

34. Watters KE, Fellmann C, Bai HB, Ren SM, Doudna JA. Systematic discovery of natural CRISPR-Cas12a inhibitors. *Science* 2018; 362(6411): 236-239 DOI: 10.1126/science.aau5138

35. Swiat MA, Dashko S, den Ridder M, Wijsman M, van der Oost J et al. FnCpf1: a novel and efficient genome editing tool for Saccharomyces cerevisiae. *Nucleic Acids Res* 2017; 45(21): 12585-12598 DOI: 10.1093/nar/gkx1007

36. DiCarlo JE, Norville JE, Mali P, Rios X, Aach J et al. Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic Acids Res* 2013; 41:4336-4343 DOI: 10.1093/nar/gkt135

37. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* 2013; 31(9): 827-832 DOI: 10.1038/nbt.2647

38. Le K, Mackley VA, Rao A, Chong AT, Dewitt MA et al. Synthetically modified guide RNA and donor DNA are a versatile platform for CRISPR-Cas9 engineering. *Elife* 2017; 6 DOI: 10.7554/eLife.25312

39. Zheng T, Hou Y, Zhang P, Zhang Z, Xu Y et al. Profiling single-guide RNA specificity reveals a mismatch sensitive core sequence. *Sci Rep* 2017; 7: 40638 DOI: 10.1038/srep40638

40. Yan Y, Finnigan GC. Development of a multi-locus CRISPR gene drive system in budding yeast. *Sci Rep* 2018; 8:17277 DOI: 10.1038/s41598-018-34909-3

41. Vella MR, Gunning CE, Lloyd AL, Gould F. Evaluating strategies for reversing CRISPR-Cas9 gene drives. *Sci Rep* 2017; 7(1): 11038 DOI: 10.1038/s41598-017-10633-2

42. Gilbert LA, Larson MH, Morsut L, Brar GA, Torres SE et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 2013; 154(2): 442-451 DOI: 10.1016/j.cell.2013.06.044

43. Kweon J, Jang AH, Kim DE, Yang JW, Yoon M et al. Fusion guide RNAs for orthogonal gene manipulation with Cas9 and Cpf1. *Nat Commun* 2017 ; 8(1): 1723 DOI: 10.1038/s41467-017-01650-w

44. Kleinstiver BP, Tsai SQ, Prew MS, Nguyen NT, Welch MM et al. Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nat Biotechnol* 2016; 34(8): 869-874 DOI: 10.1038/nbt.3620

45. Swarts DC, van der Oost J, Jinek M. Structural basis for guide RNA processing and seed-dependent DNA targeting by CRISPR-Cas12a. *Mol Cell* 2017; 66(2): 221-233.e4 DOI: 10.1016/j.molcel.2017.03.016

46. Murugan K, Seetharam AS, Severin AJ, Sashital DG. CRISPR-Cas12a has widespread off-target and dsDNA-nicking effects. *J Biol Chem* 2020; 295(17): 5538-5553 DOI: 10.1074/jbc.RA120.012933

47. Wang Y, Wang M, Zheng T, Hou Y, Zhang P et al. Specificity profiling of CRISPR system reveals greatly enhanced off-target gene editing. *Sci Rep* 2020; 10(1): 2269 DOI: 10.1038/s41598-020-58627-x

48. Creutzburg SCA, Wu WY, Mohanraju P, Swartjes T, Alkan F, et al. Good guide, bad guide: spacer sequence-dependent cleavage efficiency of Cas12a. *Nucleic Acids Res* 2020; 48(6): 3228-3243 DOI: 10.1093/nar/gkz1240

49. Thyme SB, Akhmetova L, Montague TG, Valen E, Schier AF. Internal guide RNA interactions interfere with Cas9-mediated cleavage. *Nat Commun* 2016; 7: 11750 DOI: 10.1038/ncomms11750

50. Kim D, Kim J, Hur JK, Wook Been K, Yoon SH et al. Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat Biotechnol* 2016; 34(8): 863-868 DOI: 10.1038/nbt.3609

51. Champer J, Reeves R, Yeon Oh S, Liu C, Liu J et al. Novel CRISPR/Cas9 gene drive constructs reveal insights into mechanisms of resistance allele formation and drive efficiency in genetically diverse populations. *PLoS Genet* 2017; 13(7): e1006796 DOI: 10.1371/journal.pgen.1006796

52. Hammond AM, Kyrou K, Bruttini M, North A, Galizi R et al. The creation and selection of mutations resistant to a gene drive over multiple generations in the malaria mosquito. *PLoS Genet* 2017; 13(10): e1007039 DOI: 10.1371/journal.pgen.1007039

53. Noble C, Olejarz J, Esvelt KM, Church GM, Nowak MA. Evolutionary dynamics of CRISPR gene drives. *Sci Adv* 2017; 3(4): e1601964 DOI: 10.1126/sciadv.1601964

54. Unckless RL, Clark AG, Messer PW. Evolution of resistance against CRISPR/Cas9 gene drive. *Genetics* 2017; 205(2): 827-841 DOI: 10.1534/genetics.116.197285

55. KaramiNejadRanjbar M, Eckermann KN, Ahmed HMM, Sánchez C HM, Dippel S et al. Consequences of resistance evolution in a Cas9-based sex conversion-suppression gene drive for insect pest management. *Proc Natl Acad Sci USA* 2018; 115(24): 6189-6194 DOI: 10.1073/pnas.1713825115

56. Champer J, Liu J, Yeon Oh S, Reeves R, Luthra A et al. Reducing resistance allele formation in CRISPR gene drive. *Proc Natl Acad Sci USA* 2018; 115(21): 5522-5527 DOI: 10.1073/pnas.1720354115

57. Kyrou K, Hammond AM, Galizi R, Kranjc N, Burt A et al. A CRISPR-Cas9 gene drive targeting doublesex causes complete population suppression in caged Anopheles gambiae mosquitoes. *Nat Biotechnol* 2018. 36(11): 1062-1066 DOI: 10.1038/nbt.4245

58. Champer J, Lee E, Yang E, Liu C, Clark AG et al. A toxin-antidote CRISPR gene drive system for regional population modification. *Nat Commun* 2020; 11(1): 1082 DOI: 10.1038/s41467-020-14960-3

Appendix E - Modulating CRISPR gene drive activity through nucleocytoplasmic localization of Cas9 in *S. cerevisiae*

This chapter has been published in *Fungal Biology and Biotechnology*.

Megan E. Goeckel[1†], Erianna M. Basgall[1†], Isabel C. Lewis[1†], Samantha C. Goetting[1†], Yao Yan[1], Megan Halloran[1], and Gregory C. Finnigan[1*]

[1]Department of Biochemistry and Molecular Biophysics, 141 Chalmers Hall, Kansas State University, Manhattan, KS 66506 USA

[†]Authors contributed equally

*Correspondence to: Gregory C. Finnigan, Phone: (785) 532-6939; FAX; (785) 532-7278; E-mail: gfinnigan@ksu.edu

Running title: Control of Cas9 nuclear transport

Keywords: CRISPR, Cas9, gene drive, biotechnology, nucleocytoplasmic trafficking, yeast

Abbreviations

CRISPR, clustered regularly interspaced short palindromic repeats; GD, gene drive; DSB, double stranded break; HR, homologous recombination; HDR, homology directed repair; NHEJ, non-homologous end joining; sgRNA, single guide RNA fragment; NLS, nuclear localization sequence; NES, nuclear export signal; GPCR, G-protein coupled receptor; DIC, differential interference contrast.

Abstract

The bacterial CRISPR/Cas genome editing system has provided a major breakthrough in molecular biology. One use of this technology is within a nuclease-based gene drive. This type of system can install a genetic element within a population at unnatural rates. Combatting of vector-borne diseases carried by metazoans could benefit from a delivery system that bypasses traditional Mendelian laws of segregation. Recently, laboratory studies in fungi, insects, and even mice, have demonstrated successful propagation of CRISPR gene drives and the potential utility of this type of mechanism. However, current gene drives still face challenges including evolved resistance, containment, and the consequences of application in wild populations. Additional research into molecular mechanisms that would allow for control, titration, and inhibition of drive systems is needed. In this study, we use an artificial gene drive system in budding yeast to explore mechanisms to modulate nuclease activity of Cas9 through its nucleocytoplasmic localization. We examine non-native nuclear localization sequences (both NLS and NES) on Cas9 fusion proteins *in vivo* through fluorescence microscopy and genomic editing. Our results demonstrate that mutational substitutions to nuclear signals and combinatorial fusions can both modulate the level of gene drive activity within a population of cells. These findings have implications for control of traditional nuclease-dependent editing and use of gene drive systems within other organisms. For instance, initiation of an nuclear export mechanism to Cas9 could serve as a molecular safeguard within an active gene drive to reduce or eliminate editing.

Appendix F - Characterization of Bud3 domains sufficient for bud neck targeting in *S. cerevisiaeC*

This chapter has been accepted to publish in *Access Microbiology*.

Madison N. Schrock[1,2]*, Yao Yan[1]*, Megan E. Goeckel[1,3], Erianna M. Basgall[1,4], Isabel C. Lewis[1,5], Katherine G. Leonard[1,6], Megan Halloran[1,7], and Gregory C. Finnigan[1†]

[1]Department of Biochemistry and Molecular Biophysics,

Kansas State University, 141 Chalmers Hall, Manhattan, KS 66506 USA

[2]Current Address: School of Biological Sciences, University of Utah, Salt Lake City, UT, 84112 USA

[3]Current Address: Department of Cell Biology and Physiology, Washington University in St. Louis, School of Medicine, St. Louis, MO, 63110 USA

[4]Current Address: Department of Neurobiology, School of Medicine, University of Utah, Salt Lake City, UT, 84112 USA

[5]Current Address: School of Medicine, University of Texas Medical Branch, Galveston, TX, 77555 USA

[6]Current Address: Memorial Sloan Kettering Cancer Center, New York, NY, 10065 USA

[7]Current Address: Department of Psychology, University of Kentucky, Lexington, KY, 40506 USA

*These authors contributed equally to this work.

Keywords: Bud3, septins, bud neck, fluorescence microscopy

†Correspondence to:

Gregory C. Finnigan

Dept. of Biochemistry & Molecular Biophysics

Kansas State University

141 Chalmers Hall, 1711 Claflin Rd.

Manhattan, KS 66506 USA

Phone: (785) 532-6939; FAX; (785) 532-7278;

E-mail: gfinnigan@ksu.edu

Abbreviations

AH, amphipathic helix; B4ID: Bud4-interacting domain; DH/RhoGEF: Dbl homology domain/Rho guanine-nucleotide-exchange factor; LactC2, C2 domain of bovine lactadherin protein that binds phosphatidylserine on the plasma membrane; GFP, green fluorescent protein; mCherry, monomeric red fluorescent protein derivative; PCR, polymerase chain reaction; PM, plasma membrane, SD, synthetic drop-out medium with dextrose; YPD, yeast extract peptone dextrose medium; DIC, differential interference contrast; fungal species [*Cg, Candida glabrata*; *Ka, Kazachstania Africana; Ke, Kazachstania exigua; Lf, Lachancea fermentati; Nd, Naumovozyma dairenensis; Sc, Saccharomyces cerevisiae; Td, Torulaspora delbrueckii; Vp, Vanderwaltozyma polyspora; Zm, Zygotorulaspora mrakii; Sp Schizosaccharomyces pombe*].

Abstract

The cytoskeleton serves a diverse set of functions in both multi- and unicellular organisms including movement, transport, morphology, cell division, and cell signaling. The septin family of cytoskeletal proteins are found within all fungi and metazoans and can generate three-dimensional scaffolds *in vivo* that promote membrane curvature, serve as physical barriers, and coordinate cell cycle checkpoints. In budding yeast, the septins organize into polymerized filaments that decorate the division site between mother and daughter cells during mitosis; assembly of this structure at the "bud neck" is critical for completion of cytokinesis and execution of numerous other cellular events. One such pathway includes bud site selection and the recruitment of proteins such as Bud4 and Bud3 that are responsible for promoting an axial budding pattern in haploid yeast. While Bud4 appears to be recruited to the septins independent of the presence of Bud3, it is likely that Bud3 can localize to the bud neck using both Bud4-dependent and Bud4-independent mechanisms. Furthermore, it remains unclear the precise domain(s) within Bud3 that are both necessary and sufficient for optimal association at the septin structure. In this study, we examined the localization of GFP-Bud3 constructs in otherwise WT haploid yeast cells expressing Cdc10-mCherry using fluorescence microscopy; we tested a collection of N- and C-terminal truncations and fusions of separate Bud3 protein elements to identify the smallest domain(s) responsible for bud neck localization. We found that the coordinate action of the central amphipathic helix (residues 847-865) and a partially conserved C-terminal motif (residues 1172-1273) was sufficient to promote bud neck recruitment in the presence of endogenous Bud3. This domain is considerably smaller than the previously characterized C-terminal portion required to physically interact with Bud4 (1221-1636) and utilizes a similar mechanism of pairing membrane association with a separate

localization domain similar to other non-septin proteins targeted to the division site during cell division.

Introduction

The cytoskeleton within fungi and metazoans includes a component termed the septins that have a diverse set of roles at the cellular level[1,2]. These proteins polymerize into longer filaments and superstructures that (i) influence membrane curvature, (ii) serve as a physical barrier between membrane-bound compartments, and (iii) function as a three-dimensional scaffold on which scores of other non-septin proteins can bind and carry out information exchange[3-5]. In the well-studied budding yeast species *Saccharomyces cerevisiae*, there are seven total septin genes—five expressed during mitosis (*CDC10, CDC3, CDC12, CDC11*, and *SHS1*) and two exclusively expressed during sporulation (*SPR28* and *SPR3*)[6,7]. The essential mitotic septin subunits (with a "CDC" designation) assemble into a core octameric structure consisting of a two-fold axis of symmetry with the arrangement Cdc11-Cdc12-Cdc3-Cdc10-Cdc10-Cdc3-Cdc12-Cdc11[8]. The non-essential Shs1 subunit can assemble into separate octamers and replaces the terminal Cdc11 protein to generate Shs1-Cdc12-Cdc3-Cdc10-Cdc10-Cdc3-Cdc12-Shs1 assemblies[9]. Septin octamers then bind laterally with other protomers and polymerize end-on-end to create long filaments and this assembly is critical for function; filaments can be arranged into numerous geometric shapes and structures including arcs, gauzes, spirals, rings, and hour-glass shapes[10-12].

During the mitotic cell cycle, septins localize at the site of the emerging bud and decorate the "bud neck" between mother and daughter cells; during cytokinesis, the septin collar splits into a double ring structure prior to the complete separation of the two cells[4]. Throughout this process, numerous non-septin proteins are recruited to the bud neck and septin superstructure. These include proteins from cell signaling pathways such as Hsl1, a kinase involved in the G2/M morphogenesis cell cycle checkpoint[13], Bni5, an adaptor protein linking the septins with the actomyosin contractile ring[14], components of the mitotic exit network such as Cdc5[15], and a host

of other protein modifying enzymes[3]. A diverse set of post-translational modifications of septin subunits or their associated binding partners is critical for regulation of septin assembly, physical geometry, and information exchange among these cellular pathways[16].

One pathway that utilizes the septin structure at the bud neck is bud site selection; in haploid cells, an axial budding pattern occurs—choice of the next division site proximal to the previous site[17]. Previous studies have demonstrated that proteins Axl1, Axl2, Bud3, and Bud4 are responsible for determination of the bud site through activation of Cdc42[18-22]. Genetic and biochemical data have demonstrated that Bud3 and Bud4 are recruited to and interact with the septin cytoskeleton[23-26]. Bud4 appears to have multiple domains sufficient for recruitment of the septin collar that may also act independent of Bud3[24,25]. The C-terminal regions of Bud3 and Bud4 appear to be responsible for their interaction[24,27]. The complete mechanisms regulating Bud3 localization to the septin collar appear to be complex, and both Bud4-dependent and Bud4-independent targeting have been observed[27].

The precise domain(s) within Bud3 responsible for interaction with septin subunits and/or Bud4 (or other factors at the bud neck) remain unclear. A previous study suggested that three separate fragments (1-946, 674-1220, and 1221-1636) of Bud3 might independently provide targeting information to the bud neck[23]. However, when Bud3 fragments were expressed *in vivo*, they appeared to only weakly associate at the bud neck[23]. While it is possible that optimal Bud3 recruitment to the septin collar might require both Bud4-dependent *and* Bud4-independent modes, we suspect that, like other large multi-domain proteins at the bud neck, Bud3 may require the contribution of a membrane-association domain. For example, our previous work demonstrated that Hsl1 requires both a central septin binding domain as well as a C-terminal KA1 membrane-binding domain to assist in bud neck targeting[13]. Within Bud3, a central amphipathic helix (AH)

is defined by residues 850-858 that is sufficient for membrane association[23]. Finally, limited work has been done on the C-terminal (1221-1636) domain of Bud3—yeast two-hybrid experiments indirectly suggested a smaller region (1221-1466) may be sufficient for Bud4 interaction *in vivo*[25].

Therefore, in order to examine the molecular determinant(s) within the Bud3 protein responsible for bud neck targeting, we tested whether a collection of Bud3 fragments and fusion proteins were necessary and/or sufficient for its localization pattern. This included expression of N-terminally GFP-tagged Bud3 within yeast expressing Cdc10-mCherry and endogenous levels of both Bud3 and Bud4. A collection of N-terminal and C-terminal truncations were tested for co-localization with the septins using fluorescence microscopy and revealed that the N-terminus (residues 1-846) was dispensable for bud neck targeting. Additional constructs and fusions *in trans* demonstrated that coordinate action of the central AH domain (847-865) paired with a C-terminal fragment (1172-1273) was sufficient to promote localization to the septin collar. Finally, the contribution of the AH domain likely involved its ability to bind membranes, rather than septin association, as replacement with the C2 domain from bovine lactadherin was able to replace this region. Together, these results highlight a conserved motif found within the C-terminal Bud4-association domain and suggest this smaller region of Bud3 (residues 1172-1273) is sufficient for targeting in the presence of both WT Bud3 and Bud4.

Materials and Methods

Yeast strains and plasmids

The *Saccharomyces cerevisiae* strain and DNA plasmids used in this study are described in Supplemental Table E.S1. Creation of vectors expressing Bud3 fusions followed a general strategy utilizing *in vivo* plasmid assembly, lithium acetate transformation, and selection[28]. Briefly,

a parental vector was first constructed containing the promoter of *CDC11*, GFP coding sequence, a *SpeI* restriction site, the *ADH1* terminator, and the hygromycin resistance cassette (Hyg^R). Yeast were co-transformed with (i) the aforementioned plasmid linearized with *SpeI* enzyme and (ii) PCR amplified fragment(s) of the *BUD3* gene containing flanking homology to both GFP (upstream) and the *ADH1* terminator (downstream) using a high-fidelity polymerase. Confirmation of proper assembly included diagnostic PCRs on clonal isolates and Sanger DNA sequencing (Genscript). For several vectors (Supplemental Table G.S1), a similar strategy was used with a parental plasmid containing only the *CDC11* promoter; in this case, all additional components were introduced on separate PCR fragments such as GFP and the 3' drug resistance cassette.

Culture conditions

Two types of nutritional medium were used to grow yeast: rich YPD included 2% peptone, 1% yeast extract, and 2% dextrose whereas minimal drop-out synthetic based mixtures included a yeast nitrogen base, ammonium sulfate, and amino acids. Solutions and sugars were filter sterilized prior to use. Yeast cultures were grown with constant circular shaking in a temperature-controlled unit at 30°C; agar plates were incubated at 30°C.

Fluorescence microscopy

Yeast cultures were grown overnight in synthetic liquid medium (SD-LEU) overnight at 30°C, back-diluted into YPD, grown for an additional 4 h, harvested by centrifugation, washed with sterile water, and prepared onto glass microscope slides with a coverslip. Live cells were imaged within 1 h of preparation. A Leica DMI6000 fluorescence microscope (Leica

Microsystems, Buffalo Grove, IL) with a 100x objective lens was used. Fluorescence filters included both GFP and mCherry (Semrock, GFP-4050B-LDKM-ZERO and mCherry-C LDMK-ZERO). The Leica Microsystems Application Suite AF software was used to capture images. All images within a single experimental trial were captured with identical exposure times. For clarity of presentation (and ease of visualizing cell morphology), the differential interference contrast (DIC) image contrast was adjusted prior to being merged with the mCherry channel. The fluorescence images were not altered except for cropping the image to illustrate a representative sampling of cells. Our analysis focused on the localization and pattern of GFP fluorescence within dividing cells and co-localization to mcherry (septin) signal. It did not address potential changes in cell morphology and/or budding pattern of dividing cells. Finally, images may contain some of the boundary outside of the GFP/mCherry filter set; however, these areas typically did not contain any cells within the image and were on the periphery of the chosen image. All experiments were performed in at least triplicate.

Results

Bud3 truncations reveal domains required for efficient bud neck localization

We began our investigation of cellular Bud3 localization and recruitment to the division site in yeast by testing expression of the full-length protein (Fig. F.1A) *in vivo*. We expressed GFP-Bud3(1-1636) under control of the *CDC11* promoter (plasmid) in yeast containing the septin subunit Cdc10 tagged with mCherry (integrated) (Fig. F.1B). In cells undergoing mitosis (displaying a prominent septin structure at the bud neck), GFP-Bud3(1-1636) co-localized with mCherry signal at the division site along with diffuse cytosolic signal and some weak GFP signal at the cell periphery that appeared as a weak continuous signal or small puncta (Fig. F.1B). The

detection of WT Bud3 protein at the plasma membrane is likely due to the experimental setup; our system included expression of GFP-Bud3 on a plasmid with a native copy of *BUD3* within the genome of haploid yeast. Our reasoning for analyzing GFP-Bud3 (or various mutants or fusions) localization within a *BUD3*-containing cell was (i) to not require the Bud3 variant of interest to complement the function(s) of WT Bud3, (ii) standardize the genetic background in which all experiments were being performed in terms of growth rate and cell morphology, (iii) to require Bud3 variants to "compete" with the natural quantity of full-length untagged Bud3 also present in cells, and (iv) allow for the possibility of homotypic interaction between tagged and untagged WT Bud3 protein. Using these parameters, our goal was to perform a systematic analysis of the protein structure of Bud3 to identify domain(s) that were either necessary and/or sufficient for maintaining recruitment to the septin collar in dividing cells.

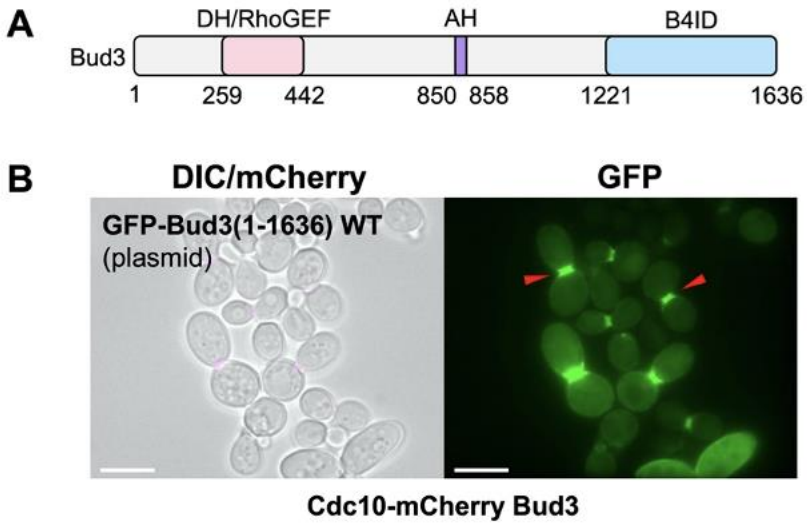


Figure F.1. Bud3 localizes at the bud neck during cell division with the septin structure.

(A) primary structure of Bud3 protein; previously characterized features include the Dbl homology domain/Rho guanine-nucleotide-exchange factor (DH/RhoGEF), amphipathic helix (AH), Bud4-interacting domain (B4ID) (modeled after[27]). (B) Yeast strain GFY-42 expressing Cdc10-mCherry from the endogenous *CDC10* locus was transformed with a plasmid expressing GFP-tagged full-length Bud3 (pGF-IVL-1631). Of note, haploid yeast also contain a WT copy of untagged *BUD3* within the genome. Cells were incubated overnight in SD-LEU at 30°C, back-diluted into YPD, grown for an additional 4 h at 30°C, and visualized by fluorescence microscopy. The DIC and

mCherry channels were merged; for clarity, the contrast of the DIC image was adjusted. Representative cells are presented. Scale bar, 7.5 μm. Red triangles denote the position of GFP signal at the bud neck in cells also displaying mCherry signal.

We began our analysis by testing GFP-Bud3 variants with increasing truncations at the C-terminus (Fig. F.2) positioned within various previously defined structures across the full-length protein (Fig. F.1A). Expression of GFP-Bud3(1-1273) still displayed GFP signal at the bud neck, albeit a weaker signal when compared to the WT control (Fig. F.2). However, the localization pattern changed dramatically for GFP-Bud3(1-1191) which showed GFP signal within the cytosol and cell periphery and almost none concentrated at the bud neck; it was unclear if the faint fluorescence at or near the bud neck was due to recruitment to the septin structure or association with the membrane (Fig. F.2). This same pattern was also observed for GFP-Bud3(1-1064), GFP-Bud3(1-964), and GFP-Bud3(1-865) (Fig. F.2). Expression of GFP-Bud3(1-747) or GFP-Bud3(1-442) only resulted in cytosolic GFP signal; previous work [23] has demonstrated that a central AH domain (850-858) promotes membrane association for Bud3 and neither of these fragments contained this region (Fig. F.2). Finally, a number of constructs displayed a cellular pattern that is likely to be the yeast nucleus undergoing division. Since this was not observed for WT Bud3 in our experiments, it is likely due to presentation of one or more nuclear localization signals (NLS) not readily accessible in the full-length protein but became exposed within truncated variants. These data suggest that the bud neck localization of GFP-Bud3 depends on a domain defined by residues 1192-1273.

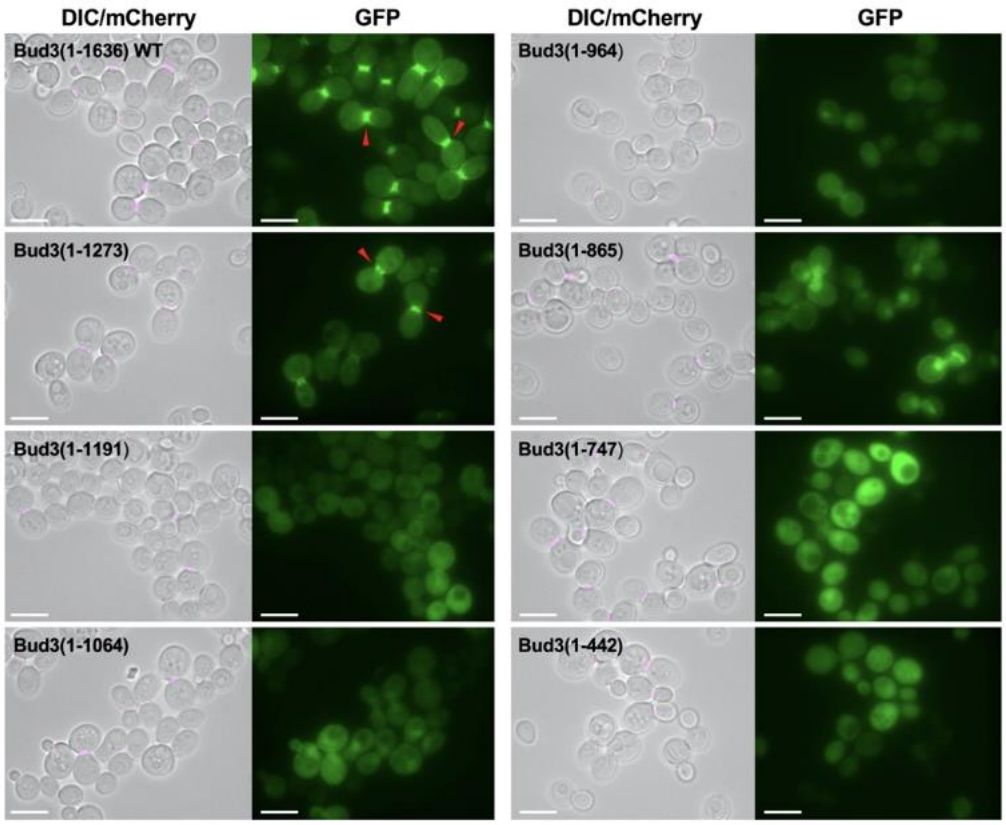


Figure F.2. The subcellular localization of GFP-Bud3 fusions truncated at the C-terminus.

Yeast containing an integrated Cdc10-mCherry construct (GFY-42) were transformed with plasmids (pGF-IVL-1631, pGF-IVL-11-124, pGF-IVL-11-121, and pGF-IVL-11-137 to pGF-IVL-11-141). Cells were visualized as in Fig. F.1. The DIC and mCherry channels were merged for clarity. Scale bar, 7.5 μm. Red triangles denote the position of GFP signal at the bud neck of dividing yeast cells.

Next, we performed a complementary analysis by generating truncations at the N-terminus of Bud3 while still including GFP at the N-terminus of the fusion protein (Fig. F.3). Removal of the first 846 amino acids within the GFP-Bud3(847-1636) construct had little overall effect on the localization pattern which phenocopied the WT control (Fig. F.3). However, extending the truncation beyond the central AH domain within GFP-Bud3(859-1636) caused a marked reduction in GFP at the bud neck (Fig. F.3). Cells displayed a low but reproducible amount of fluorescence at the division site and loss of any signal at the cell periphery; this same pattern was also seen for

the GFP-Bud3(1224-1636) construct (Fig. F.3). Finally, expression of either GFP-Bud3(1326-1636) or GFP-Bud3(1421-1636) resulted in no GFP signal at the bud neck and only robust cytosolic localization (Fig. F.3). Together, these experiments suggest a number of findings. First, the AH domain (residues 847-858, as defined by our construct) is essential for promoting robust bud neck localization of the C-terminal portion (residues 859-1636) of Bud3. Second, a small domain appears necessary for bud neck recruitment (albeit, very weak) within the region of 1224-1325 although this may also include surrounding residues. Third, the extreme C-terminal domain (residues 1326-1636) is not sufficient to promote bud neck localization when expressed alone. Comparison of the data from the C-terminal (Fig. F.2) and N-terminal (Fig. F.3) truncations reveals that a likely candidate for a bud neck localization domain within Bud3 is centered around residues 1192-1325.

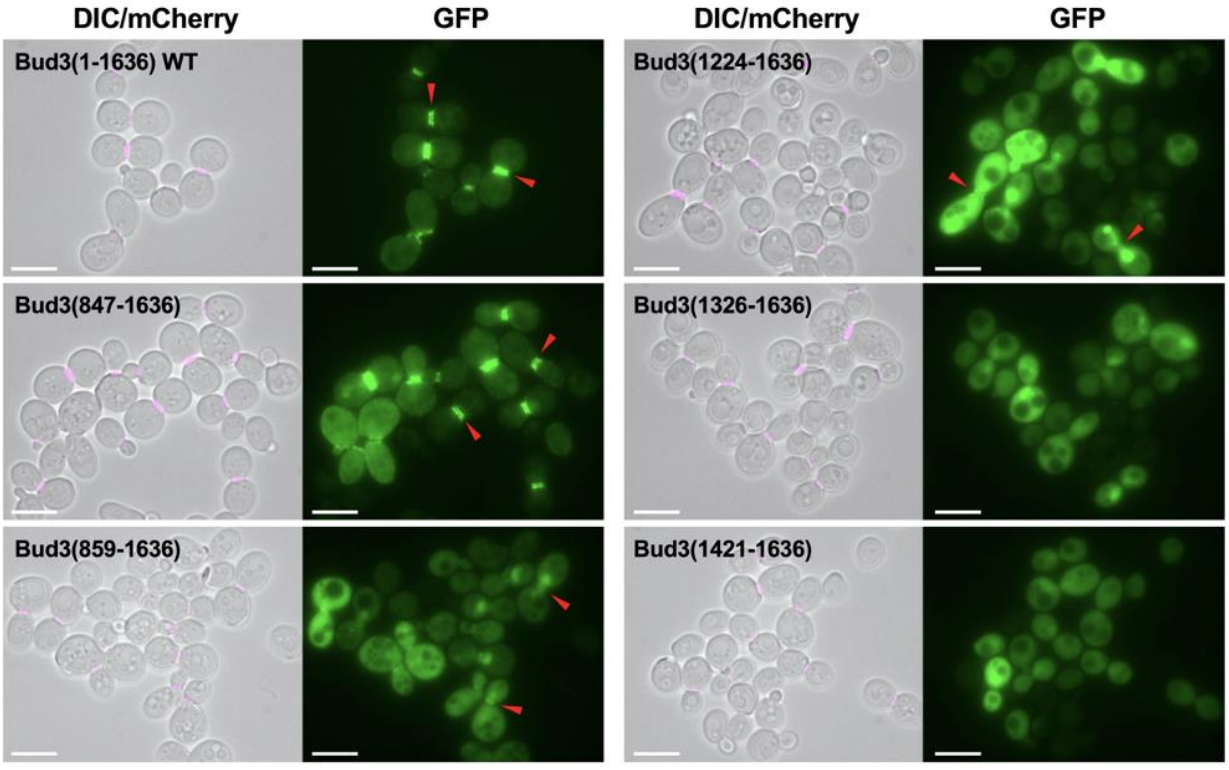


Figure F.3. The subcellular localization of GFP-Bud3 fusions truncated at the N-terminus.

Yeast containing an integrated Cdc10-mCherry construct (GFY-42) were transformed with plasmids (pGF-IVL-1631, pGF-IVL-11-6, pGF-IVL-11-12, pGF-IVL-11-15, pGF-IVL-11-18, and pGF-IVL-11-21) and were visualized as in Fig. F.1. The DIC and mCherry channels were merged; scale bar, 7.5 μm. Red triangles denote the position of GFP signal at the bud neck of dividing yeast cells.

Bud3 fusion proteins reveal a conserved motif sufficient for bud neck localization

Since removal of the first 846 amino acids had little effect on the localization pattern of GFP-Bud3 *in vivo*, we generated a set of C-terminal truncations while maintaining 847 as the first amino acid of Bud3 to ensure inclusion of the AH domain (Fig. F.4). We observed that deleting progressively larger fragments from the C-terminus did not cause any change in localization phenotype for GFP-Bud3(847-1536), GFP-Bud3(847-1436), GFP-Bud3(847-1375), GFP-Bud3(847-1325), and GFP-Bud3(847-1273) (Fig. F.4). These constructs all displayed similar GFP patterns at the bud neck. However, expression of GFP-Bud3(847-1220) or smaller fragments such as GFP-Bud3(847-1111) or GFP-Bud3(847-1064) displayed a total loss of bud neck localization (Fig. F.4). These data suggest that critical localization information residues within (or near) residues 1221-1273. Moreover, we conclude that residues 1274-1325 do not provide a strong contribution to this cellular pattern, at least in the context of this truncated Bud3 construct.

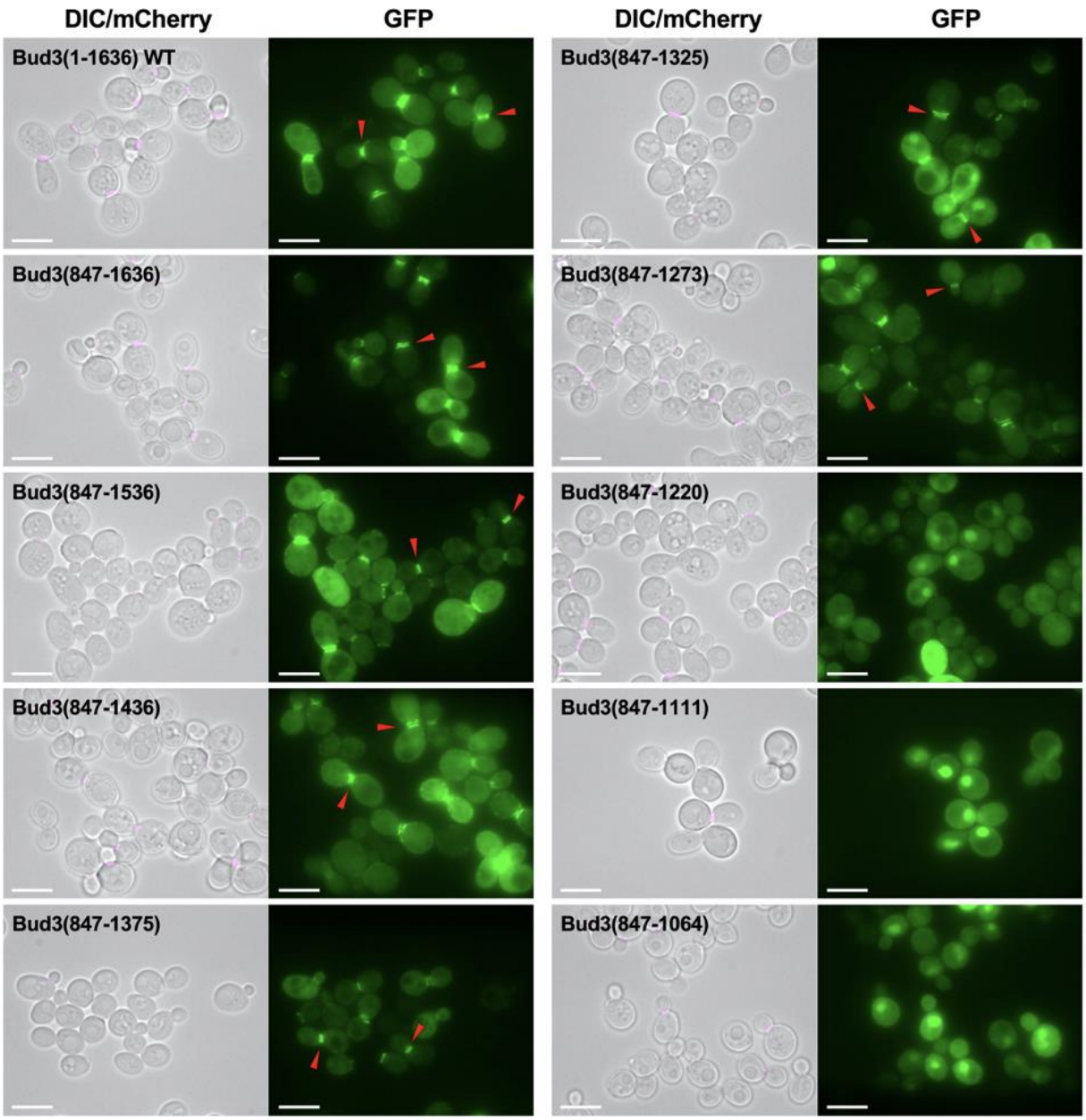


Figure F.4. A minimal continuous Bud3 fragment sufficient for bud neck localization *in vivo*.

Yeast (GFY-42) were transformed with plasmids (pGF-IVL-1631, pGF-IVL-11-1 through pGF-IVL-11-6, and pGF-IVL-11-31 through pGF-IVL-11-33) and visualized as in Fig. F.1. The DIC and mCherry channels were merged. Scale bar, 7.5 μm. Red triangles denote the position of GFP signal at the bud neck of dividing yeast cells.

Thus far, our analysis narrowed the focus on Bud3 residues between 1192 and 1273 as necessary in the context of specific truncations from either terminus. Therefore, we built a series

of fusion proteins that included the AH domain (847-865) paired with smaller fragments *in trans* that all ended at position 1273 (Fig. F.5). Our primary interest was to test whether residues immediately proximal to position 1192 might also contribute to a localization signal. Expression of GFP-Bud3(847-865; 1112-1273), GFP-Bud3(847-865; 1132-1273), GFP-Bud3(847-865; 1152-1273), and GFP-Bud3(847-865; 1172-1273) all displayed a similar pattern of GFP signal at the bud neck in dividing cells (Fig. F.5). However, there was a noticeable shift in the overall localization for GFP-Bud3(847-865; 1192-1273) with more robust fluorescence present at the cell periphery (Fig. F.5). Finally, shortening the construct within GFP-Bud3(847-865; 1224-1273) resulted in a loss of bud neck localization (Fig. F.5).

To investigate whether there was an explanation for the difference in phenotype between Bud3 fusions that included fragments beginning at 1172, 1192, or 1224, we performed a search for other Bud3 orthologs in related fungal species (Fig. F.5B, Supplemental Fig. G.S2). An alignment of proteins similar to Bud3 revealed several unique features when examining the *S. cerevisae* region from 1142-1286 (Fig. F.5B). First, the presence of a conserved motif appears within yeast Bud3 at residues 1207-1260 (Fig. F.5B). This region includes thirteen amino acids that are identical across 9/9 species and twenty additional residues that are conserved in at least 5/9 fungi (Fig. F.5B). This is in stark contrast to the surrounding regions (1142-1206 or 1261-1286) where there is almost no conservation of amino acids (Fig. F.5B). Given that the GFP-Bud3(847-865; 1192-1273) displayed bud neck localization whereas the GFP-Bud3(847-865; 1224-1273) construct did not, it may be that (i) the signal requires a minimal number of essential residues within the motif and/or (ii) there needs to be adequate physical distance in the form of a linker sequence extending between the AH domain contacting the membrane and bud neck association. Our data illustrate two separate reductions in bud neck localization as the Bud3

domain was shortened from position 1112 to 1224 (Fig. F.5). The GFP-Bud3(847-865; 1172-1273) construct displayed less GFP signal at the cell periphery compared to GFP-Bud3(847-865; 1192-1273); we suspect that this may be due to decreased binding and/or suboptimal positioning of the C-terminal fragment to the bud neck (Fig. F.5). Second, we observed total loss of recruitment to the division site once the fragment was shortened to residue 1224 and this may be due to proximity to the evolutionarily conserved motif present between residues 1207-1260 (Fig. F.5B).

To examine whether the conserved Bud3 motif was sufficient to promote bud neck localization in an independent context, we generated additional fusions using a separate protein domain that associates with membranes (Fig. F.6). We replaced the AH (847-865) domain with the 158-residue C2 domain from bovine lactadherin (LactC2) which binds to phosphatidylserine on the PM [29,30]. Compared to GFP-Bud3(847-865; 1065-1273), the GFP-Bud3(LactC2; 1065-1273) fusion displayed robust localization to the bud neck as well as an increased level at the plasma membrane (Fig. F.6). This effect was also evident when examining GFP-Bud3(847-865; 1192-1273) against GFP-Bud3(LactC2; 1192-1273) (Fig. F.6). In both cases, these additional constructs illustrated that (i) membrane association, rather than the identity of the AH domain residues, was critical to promote bud neck localization, and (ii) a non-native protein domain separating GFP from the central domain of Bud3 (1065-1273 or 1192-1273) still allowed for robust fluorescence signal at the division site suggesting that membrane association and the bud neck signal work in concert but are separable.

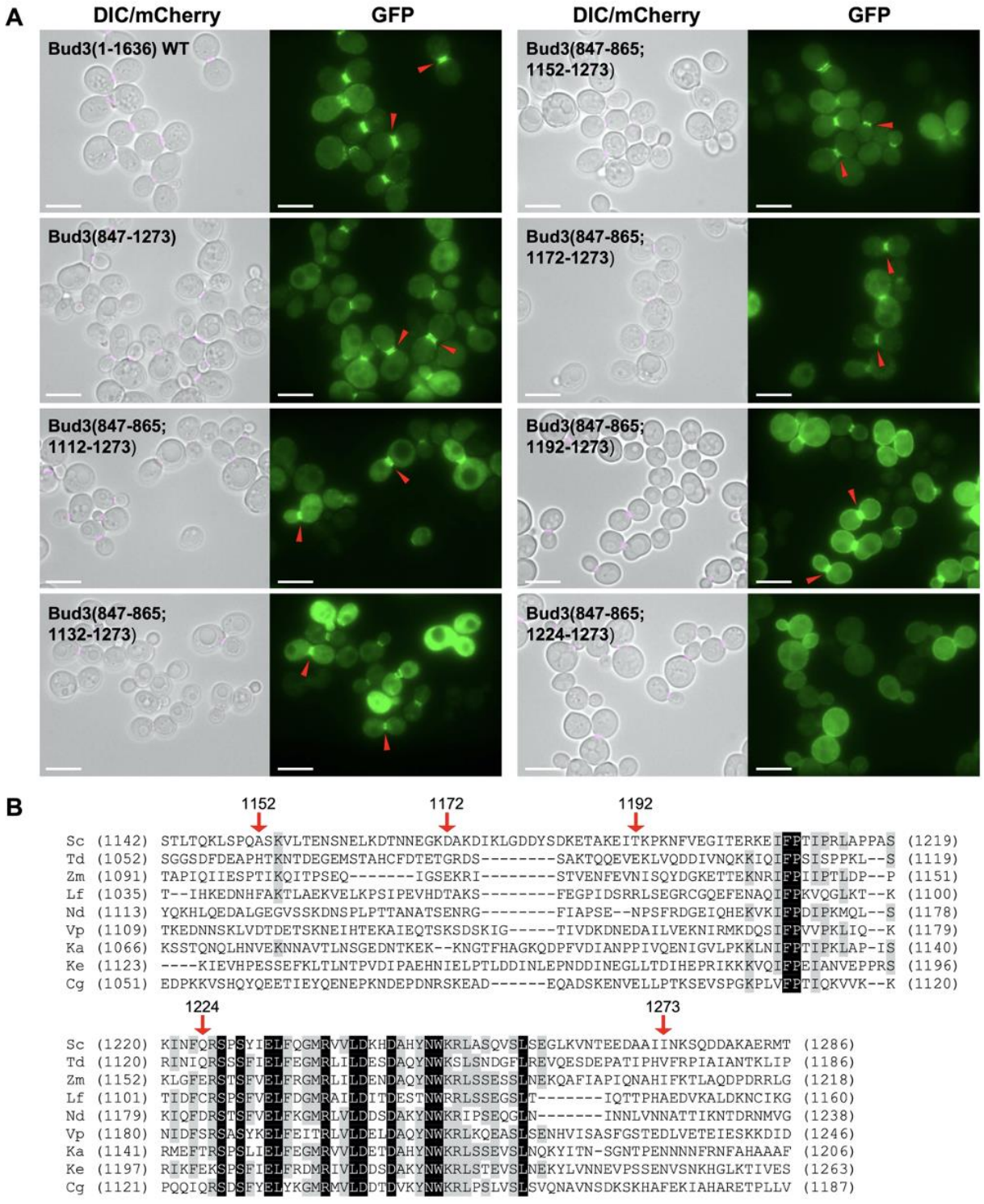


Figure F.5. Direct fusions between the central amphipathic helix and C-terminal fragments of Bud3 reveals a minimal construct for bud neck localization.

(A) Yeast (GFY-42) were transformed with plasmids (pGF-IVL-1631, pGF-IVL-11-31, pGF-IVL-11-41, pGF-IVL-11-69, and pGF-IVL-11-76 through pGF-IVL-11-79) and visualized as in Fig.

D.1. The DIC and mCherry channels were merged. Scale bar, 7.5 μm. Red triangles denote the position of GFP signal at the bud neck of dividing yeast cells. (B) Protein sequence alignment of putative Bud3 orthologs from related fungal species. Full-length Bud3(1-1636) from *S. cerevisiae* was used as a query sequence to identify similar proteins within the fungal kingdom using BLAST (NCBI). Full-length proteins from eight species (Supplemental Fig. G.S2) were aligned against *S. cerevisiae* Bud3 using CLUSTAL-W [31]. The alignment for residues 1142-1286 (numbering from budding yeast Bud3) is displayed. Residues identical across all nine species are colored white against a black background; residues identical across at least five of the nine species are colored black against a grey background. Red arrows illustrate specific residue positions in *S. cerevisiae* Bud3.

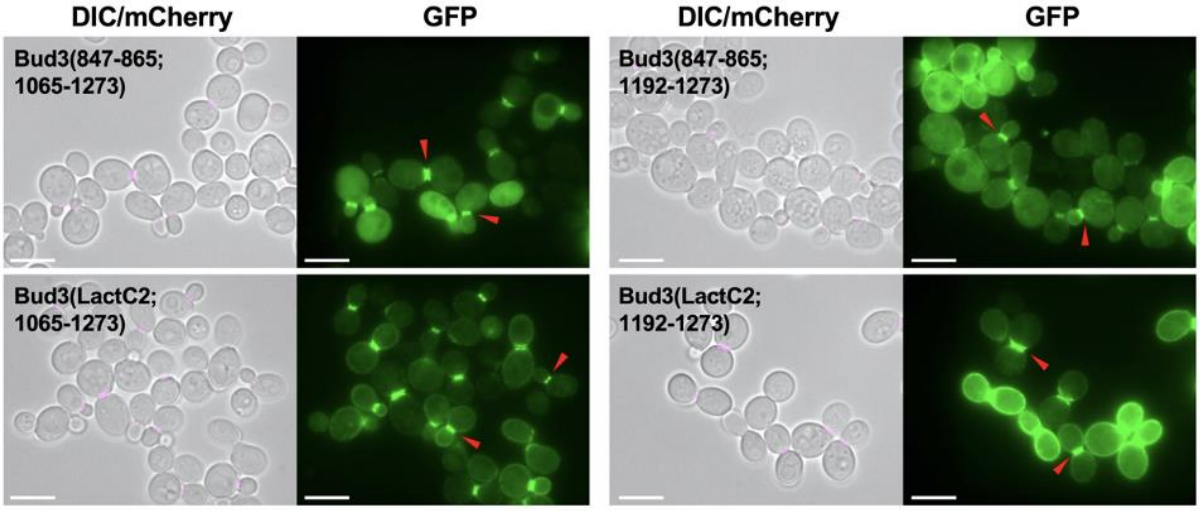


Figure F.6. Replacement of the Bud3 amphipathic helix with LactC2 domain.

Yeast (GFY-42) containing plasmids (pGF-IVL-11-68, pGF-IVL-11-79, pGF-IVL-11-107, and pGF-IVL-11-110) were cultured and visualized as in Fig. F.1. The DIC and mCherry channels were merged. Scale bar, 7.5 μm. Red triangles denote the position of GFP signal at the bud neck of dividing yeast cells.

Discussion

Using a series of truncations and fusion proteins, our study has defined two small domains that coordinate *in trans* to promote optimal localization of the Bud3 protein to the bud neck. This includes the central AH domain (847-865) and a fragment at the C-terminus (1172-1273) with a small, conserved region at residues 1207-1260. A previous study reported that three separate regions spanning the full-length Bud3 (1-946, 674-1220, and 1221-1636) could each promote bud

neck localization [23]. A number of critical differences exist with our experimental approach including the choice of the *CDC11* promoter driving expression of GFP-Bud3 as opposed to *MET25* as well as the use of a *BUD3*-expressing WT background rather than a *bud3*Δ strain. It remains possible that the combination of overexpression of Bud3 fragments (such as 1-946 or 674-1220) could localize to the division site in the absence of competing WT Bud3. If any such targeting signal(s) exists within the N-terminal region of Bud3, it is extremely weak compared to the C-terminal information we have identified in this study; along these lines, the previous report observed that localization of Bud3(1-946) was markedly lower than that of WT protein [23]. Furthermore, numerous studies have suggested both Bud4-dependent and Bud4-independent mechanisms for bud neck recruitment of Bud3 [23,27]; therefore, it is highly likely that numerous targeting signals exist within the protein.

Interestingly, previous work on Bud3 has provided limited information on the nature of the bud neck targeting domain within the C-terminal region defined by residues 1221-1636, an arbitrarily designated fragment from a previous study [23]. Yeast two-hybrid experiments illustrated that Bud3(948-1466) and Bud3(1221-1614) were both able to interact with the TD2 domain of Bud4 (residues 1067-1447) [25]. The conclusion, given the overlap between both Bud3 fragments, was that the region of 1221-1466 was sufficient to interact with the C-terminus of Bud4, although this exact fragment was not experimentally tested [25]. Furthermore, the ability of Bud3(1221-1636) to localize to the bud neck was compromised in *bud4*Δ and *bud3*Δ *bud4*Δ strains, supporting a model where this region of Bud3 requires Bud4 for recruitment [25]. Other groups have also demonstrated that loss of the C-terminal region of Bud3 (1221-1636) causes a (partial) loss of recruitment to the division site similar to a loss of Bud4 [27]. Our findings suggest that a much smaller domain within this region (1221-1636) may be sufficient for bud neck recruitment *in vivo*

even in the presence of endogenous Bud3. We propose that the domain may coincidently begin near residue 1221 yet only extend to residue 1273 (smaller than the 1466 proposed from yeast two-hybrid data). From our Bud3 fusion constructs, we observed improved bud neck localization when the minimal fragment was extended from 1192 to 1172. Given that none of these residues were conserved even between closely related species, it seems plausible that this represents physical separation from the fused AH domain to potentially allow proper association at the bud neck (and possibly physical interaction with the C-terminus of Bud4). Future biochemical interaction studies (with the TD2 domain of Bud4 or septins) and/or a mutational analysis of this smaller fragment and the highly conserved residues found near position 1221 will provide additional information about this targeting region within Bud3.

Our study also highlighted that like other large multi-domain proteins localized to the septin collar and bud neck, a membrane-association domain acts to optimize and promote efficient localization *in vivo*. Our previous work demonstrated that for Hsl1, a septin-interacting kinase responsible for the G2/M checkpoint, utilizes a central septin-binding domain with a C-terminal membrane binding KA1 domain [13]. Here, we illustrated bud neck recruitment of the C-terminal domain of Bud3 when it was combined with the central AH domain either *in cis* or *in trans*; additionally, use of the non-native membrane-associating LactC2 domain also had the same effect.

Similar to previous findings that deletion of the Bud3(1221-1636) Bud4-interacting domain caused a reduction (but not total loss) of bud neck targeting [27], we also found that a full-length Bud3 construct lacking only the central 1192-1273 domain was still able to localize to the division site, albeit with much weaker signal (our unpublished data). Additional experimentation will be required to identify the potential Bud4-independent mechanisms allowing for bud neck recruitment using variants of Bud3 lacking 1192-1273 (or 1221-1636) and/or in *bud4*Δ cells. It

would be of interest to elucidate whether region(s) of Bud3 can associate directly with the septin superstructure or one or more of the many other proteins targeted to the bud neck during cell division [3].

Funding Information

This project was supported by an Institutional Development Award (IDeA) from the National Institute of General Medical Science of the National Institute of Health under grant number P20 GM103418 (G.C.F.). This work was also supported by the USDA National Institute of Food and Agriculture, Hatch Project 1013520 (G.C.F.). The Kansas State Univ. College of Art & Sciences and Johnson Cancer Research Center also supported this work.

Acknowledgements

We would like to thank previous members of the Finnigan laboratory for useful comments, suggestions, and laboratory assistance. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institute of Health or the United States Department of Agriculture. M.N.S, M.H., and G.C.F. designed and built the DNA constructs used in this study. M.N.S., Y.Y., M.E.G., E.M.B., I.C.L., K.G.L., and G.C.F. performed the experiments. M.N.S., Y.Y. and G.C.F. analyzed the data. Y.Y. and G.C.F. generated the table and figures and wrote the manuscript.

Conflict of Interest

All authors declare no financial or non-financial conflicts of interest.

Ethical Statement

This work did not involve any human or animal subjects.

Reagent Availability Statement

Yeast and/or plasmids will be made available for educational or research purposes upon reasonable request.

References

1 Nishihama, R., Onishi, M. & Pringle, J. R. New insights into the phylogenetic distribution and evolutionary origins of the septins. *Biological chemistry* **392**, 681-687, doi:10.1515/bc.2011.086 (2011).

2 Auxier, B., Dee, J., Berbee, M. L. & Momany, M. Diversity of opisthokont septin proteins reveals structural constraints and conserved motifs. *BMC evolutionary biology* **19**, 4, doi:10.1186/s12862-018-1297-8 (2019).

3 Perez, A. M., Finnigan, G. C., Roelants, F. M. & Thorner, J. Septin-Associated Protein Kinases in the Yeast Saccharomyces cerevisiae. *Frontiers in cell and developmental biology* **4**, 119, doi:10.3389/fcell.2016.00119 (2016).

4 Spiliotis, E. T. & McMurray, M. A. Masters of asymmetry - lessons and perspectives from 50 years of septins. *Molecular biology of the cell* **31**, 2289-2297, doi:10.1091/mbc.E19-11-0648 (2020).

5 Woods, B. L. & Gladfelter, A. S. The state of the septin cytoskeleton from assembly to function. *Current opinion in cell biology* **68**, 105-112, doi:10.1016/j.ceb.2020.10.007 (2021).

6 Cid, V. J. *et al.* Cell integrity and morphogenesis in a budding yeast septin mutant. *Microbiology (Reading, England)* **144 (Pt 12)**, 3463-3474, doi:10.1099/00221287-144-12-3463 (1998).

7 Neiman, A. M. Sporulation in the budding yeast Saccharomyces cerevisiae. *Genetics* **189**, 737-765, doi:10.1534/genetics.111.127126 (2011).

8 Bertin, A. *et al.* Saccharomyces cerevisiae septins: supramolecular organization of heterooligomers and the mechanism of filament assembly. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 8274-8279, doi:10.1073/pnas.0803330105 (2008).

9 Garcia, G., 3rd *et al.* Subunit-dependent modulation of septin assembly: budding yeast septin Shs1 promotes ring and gauze formation. *The Journal of cell biology* **195**, 993-1004, doi:10.1083/jcb.201107123 (2011).

10 McMurray, M. A. *et al.* Septin filament formation is essential in budding yeast. *Developmental cell* **20**, 540-549, doi:10.1016/j.devcel.2011.02.004 (2011).

11 Bertin, A. *et al.* Three-dimensional ultrastructure of the septin filament network in Saccharomyces cerevisiae. *Molecular biology of the cell* **23**, 423-432, doi:10.1091/mbc.E11-10-0850 (2012).

12 Ong, K., Wloka, C., Okada, S., Svitkina, T. & Bi, E. Architecture and dynamic remodelling of the septin cytoskeleton during the cell cycle. *Nature communications* **5**, 5698, doi:10.1038/ncomms6698 (2014).

13 Finnigan, G. C. *et al.* Coordinate action of distinct sequence elements localizes checkpoint kinase Hsl1 to the septin collar at the bud neck in Saccharomyces cerevisiae. *Molecular biology of the cell* **27**, 2213-2233, doi:10.1091/mbc.E16-03-0177 (2016).

14 Finnigan, G. C., Booth, E. A., Duvalyan, A., Liao, E. N. & Thorner, J. The Carboxy-Terminal Tails of Septins Cdc11 and Shs1 Recruit Myosin-II Binding Factor Bni5 to the Bud Neck in Saccharomyces cerevisiae. *Genetics* **200**, 843-862, doi:10.1534/genetics.115.176503 (2015).

15 Sakchaisri, K. *et al.* Coupling morphogenesis to mitotic entry. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 4124-4129, doi:10.1073/pnas.0400641101 (2004).

16 Hernández-Rodríguez, Y. & Momany, M. Posttranslational modifications and assembly of septin heteropolymers and higher-order structures. *Current opinion in microbiology* **15**, 660-668, doi:10.1016/j.mib.2012.09.007 (2012).

17 Chiou, J. G., Balasubramanian, M. K. & Lew, D. J. Cell Polarity in Yeast. *Annual review of cell and developmental biology* **33**, 77-101, doi:10.1146/annurev-cellbio-100616-060856 (2017).

18 Fujita, A. *et al.* A yeast gene necessary for bud-site selection encodes a protein similar to insulin-degrading enzymes. *Nature* **372**, 567-570, doi:10.1038/372567a0 (1994).

19 Chant, J., Mischke, M., Mitchell, E., Herskowitz, I. & Pringle, J. R. Role of Bud3p in producing the axial budding pattern of yeast. *The Journal of cell biology* **129**, 767-778, doi:10.1083/jcb.129.3.767 (1995).

20 Roemer, T., Madden, K., Chang, J. & Snyder, M. Selection of axial growth sites in yeast requires Axl2p, a novel plasma membrane glycoprotein. *Genes & development* **10**, 777-793, doi:10.1101/gad.10.7.777 (1996).

21 Sanders, S. L. & Herskowitz, I. The BUD4 protein of yeast, required for axial budding, is localized to the mother/BUD neck in a cell cycle-dependent manner. *The Journal of cell biology* **134**, 413-427, doi:10.1083/jcb.134.2.413 (1996).

22 Kang, P. J., Angerman, E., Jung, C. H. & Park, H. O. Bud4 mediates the cell-type-specific assembly of the axial landmark in budding yeast. *Journal of cell science* **125**, 3840-3849, doi:10.1242/jcs.103697 (2012).

23 Guo, J., Gong, T. & Gao, X. D. Identification of an amphipathic helix important for the formation of ectopic septin spirals and axial budding in yeast axial landmark protein Bud3p. *PloS one* **6**, e16744, doi:10.1371/journal.pone.0016744 (2011).

24 Kang, P. J., Hood-DeGrenier, J. K. & Park, H. O. Coupling of septins to the axial landmark by Bud4 in budding yeast. *Journal of cell science* **126**, 1218-1226, doi:10.1242/jcs.118521 (2013).

25 Wu, H., Guo, J., Zhou, Y. T. & Gao, X. D. The anillin-related region of Bud4 is the major functional determinant for Bud4's function in septin organization during bud growth and axial bud site selection in budding yeast. *Eukaryotic cell* **14**, 241-251, doi:10.1128/ec.00268-14 (2015).

26 Finnigan, G. C., Duvalyan, A., Liao, E. N., Sargsyan, A. & Thorner, J. Detection of protein-protein interactions at the septin collar in Saccharomyces cerevisiae using a tripartite split-GFP system. *Molecular biology of the cell* **27**, 2708-2725, doi:10.1091/mbc.E16-05-0337 (2016).

27 Chen, X., Wang, K., Svitkina, T. & Bi, E. Critical Roles of a RhoGEF-Anillin Module in Septin Architectural Remodeling during Cytokinesis. *Current biology : CB* **30**, 1477-1490.e1473, doi:10.1016/j.cub.2020.02.023 (2020).

28 Finnigan, G. C. & Thorner, J. Complex in vivo Ligation Using Homologous Recombination and High-efficiency Plasmid Rescue from Saccharomyces cerevisiae. *Bio-protocol* **5**, doi:10.21769/bioprotoc.1521 (2015).

29 Andersen, M. H., Graversen, H., Fedosov, S. N., Petersen, T. E. & Rasmussen, J. T. Functional analyses of two cellular binding domains of bovine lactadherin. *Biochemistry* **39**, 6200-6206, doi:10.1021/bi992221r (2000).

30 Shao, C., Novakovic, V. A., Head, J. F., Seaton, B. A. & Gilbert, G. E. Crystal structure of lactadherin C2 domain at 1.7A resolution with mutational and computational analyses of its

membrane-binding motif. *The Journal of biological chemistry* **283**, 7230-7241, doi:10.1074/jbc.M705195200 (2008).

31 Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research* **22**, 4673-4680, doi:10.1093/nar/22.22.4673 (1994).

Appendix G - Supplementary Tables and Figures for Characterization of Bud3 domains sufficient for bud neck targeting in S. cerevisiae

Madison N. Schrock[1,2]*, Yao Yan[1]*, Megan E. Goeckel[1,3], Erianna M. Basgall[1,4], Isabel C. Lewis[1,5], Katherine G. Leonard[1,6], Megan Halloran[1,7], and Gregory C. Finnigan[1†]

[1]Department of Biochemistry and Molecular Biophysics,

Kansas State University, 141 Chalmers Hall, Manhattan, KS 66506 USA

[2]Current Address: School of Biological Sciences, University of Utah, Salt Lake City, UT, 84112 USA

[3]Current Address: Department of Cell Biology and Physiology, Washington University in St. Louis, School of Medicine, St. Louis, MO, 63110 USA

[4]Current Address: Department of Neurobiology, School of Medicine, University of Utah, Salt Lake City, UT, 84112 USA

[5]Current Address: School of Medicine, University of Texas Medical Branch, Galveston, TX, 77555 USA

[6]Current Address: Memorial Sloan Kettering Cancer Center, New York, NY, 10065 USA

[7]Current Address: Department of Psychology, University of Kentucky, Lexington, KY, 40506 USA

*These authors contributed equally to this work.

Keywords: Bud3, septins, bud neck, fluorescence microscopy

†Correspondence to:

Gregory C. Finnigan

Dept. of Biochemistry & Molecular Biophysics

Kansas State University

141 Chalmers Hall, 1711 Claflin Rd.

Manhattan, KS 66506 USA

Phone: (785) 532-6939; FAX; (785) 532-7278;

E-mail: gfinnigan@ksu.edu

Supplemental Table G.S1. Yeast and DNA plasmids used in this study.

Strain	**Genotype**	**Reference**
GFY-42[a,b]	*MAT***a** *cdc10Δ::S.c.CDC10::mCherry::ADH1(t)::S.p.HIS5*	[1]
Plasmid	**Description**	**Reference**
pRS315	*CEN, LEU2*	[2]
pGF-IVL-1631[c,d,e]	pRS315; *prCDC11::GFP::Bud3(1-1636)::ADH1(t)::Hyg^R*	This study
pGF-IVL-11-1	pRS315; *prCDC11::GFP::Bud3(847-1064)::ADH1(t)::Hyg^R*	This study
pGF-IVL-11-2	pRS315; *prCDC11::GFP::Bud3(847-1111)::ADH1(t)::Hyg^R*	This study
pGF-IVL-11-3	pRS315; *prCDC11::GFP::Bud3(847-1220)::ADH1(t)::Hyg^R*	This study
pGF-IVL-11-4	pRS315; *prCDC11::GFP::Bud3(847-1436)::ADH1(t)::Hyg^R*	This study
pGF-IVL-11-5	pRS315; *prCDC11::GFP::Bud3(847-1536)::ADH1(t)::Hyg^R*	This study
pGF-IVL-11-6	pRS315; *prCDC11::GFP::Bud3(847-1636)::ADH1(t)::Hyg^R*	This study
pGF-IVL-11-12	pRS315; *prCDC11::GFP::Bud3(859-1636)::ADH1(t)::Hyg^R*	This study
pGF-IVL-11-15	pRS315; *prCDC11::GFP::Bud3(1224-1636)::ADH1(t)::Hyg^R*	This study
pGF-IVL-11-18	pRS315; *prCDC11::GFP::Bud3(1326-1636)::ADH1(t)::Hyg^R*	This study
pGF-IVL-11-21	pRS315; *prCDC11::GFP::Bud3(1421-1636)::ADH1(t)::Hyg^R*	This study
pGF-IVL-11-31	pRS315; *prCDC11::GFP::Bud3(847-1273)::ADH1(t)::Hyg^R*	This study
pGF-IVL-11-32	pRS315; *prCDC11::GFP::Bud3(847-1325)::ADH1(t)::Hyg^R*	This study
pGF-IVL-11-33	pRS315; *prCDC11::GFP::Bud3(847-1375)::ADH1(t)::Hyg^R*	This study
pGF-IVL-11-42	pRS315; *prCDC11::GFP::Bud3(847-865; 1224-1273)::ADH1(t)::Hyg^R*	This study

pGF-IVL-11-68	pRS315; *prCDC11::GFP::Bud3(847-865; 1065-1273)::ADH1(t)::Hyg^R*	This study
pGF-IVL-11-69	pRS315; *prCDC11::GFP::Bud3(847-865; 1112-1273)::ADH1(t)::Hyg^R*	This study
pGF-IVL-11-76	pRS315; *prCDC11::GFP::Bud3(847-865; 1132-1273)::ADH1(t)::Hyg^R*	This study
pGF-IVL-11-77	pRS315; *prCDC11::GFP::Bud3(847-865; 1152-1273)::ADH1(t)::Hyg^R*	This study
pGF-IVL-11-78	pRS315; *prCDC11::GFP::Bud3(847-865; 1172-1273)::ADH1(t)::Hyg^R*	This study
pGF-IVL-11-79	pRS315; *prCDC11::GFP::Bud3(847-865; 1192-1273)::ADH1(t)::Hyg^R*	This study
pGF-IVL-11-107[f]	pRS315; *prCDC11::GFP::LactC2(1-158)::Bud3(1065-1273)::ADH1(t)::Hyg^R*	This study
pGF-IVL-11-110	pRS315; *prCDC11::GFP::LactC2(1-158)::Bud3(1192-1273)::ADH1(t)::Hyg^R*	This study
pGF-IVL-11-121[g]	pRS315; *prCDC11::GFP::Bud3(1-1064)::ADH1(t)::Kan^R*	This study
pGF-IVL-11-124	pRS315; *prCDC11::GFP::Bud3(1-1191)::ADH1(t)::Kan^R*	This study
pGF-IVL-11-137	pRS315; *prCDC11::GFP::Bud3(1-442)::ADH1(t)::Hyg^R*	This study
pGF-IVL-11-138	pRS315; *prCDC11::GFP::Bud3(1-747)::ADH1(t)::Hyg^R*	This study
pGF-IVL-11-139	pRS315; *prCDC11::GFP::Bud3(1-865)::ADH1(t)::Hyg^R*	This study
pGF-IVL-11-140	pRS315; *prCDC11::GFP::Bud3(1-964)::ADH1(t)::Hyg^R*	This study
pGF-IVL-11-141	pRS315; *prCDC11::GFP::Bud3(1-1273)::ADH1(t)::Hyg^R*	This study

[a]Abbreviations of the fungal genus and species are included for the *CDC10* and *HIS5* genes: *S.c.*, *Saccharomyces cerevisiae; S.p. Schizosaccharomyces pombe.*

[b]The mCherry sequence used:
VSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLP
FAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSL
QDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLKDG
GHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERAEGRHSTGGMDE
LYK.

[c]The GFP sequence used:
MGRRIPGLINSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTG
KLPVPWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTR
AEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIR
HNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAA
GITHGMDELYK.

[d]The commonly used *prMX* and *MX(t)* sequences were included flanking all drug resistance cassettes (such as Hyg^R).

[e]The strategy for plasmid construction included use of a parental vector built on pRS315 that included pr*CDC11*-GFP-[*SpeI*]-*ADH1*(t)-Hyg^R. Linearization was achieved by digestion of the unique *SpeI* restriction site between the GFP and *ADH1*(t) sequences. Digested plasmid was then co-transformed with amplified *BUD3* PCR(s) with flanking homology to upstream GFP sequence and downstream *ADH1*(t) sequence into yeast for *in vivo* plasmid assembly.

[f]The *Bos taurus* lactadherin C2 (Lact-C2) domain sequence [3,4] was fused to each construct at the 3' end of the GFP gene. The numbering scheme for Lact-C2 refers to the specific domain and not the full-length lactadherin protein. The amino acid sequence of the 158 residue domain:
CTEPLGLKDNTIPNKQITASSYYKTWGLSAFSWFPYYARLDNQGKFNAWTAQTNSASE
WLQIDLGSQKRVTGIITQGARDFGHIQYVAAYRVAYGDDGVTWTEYKDPGASESKIFPG
NMDNNSHKKNIFETPFQARFVRIQPVAWHNRITLRVELLGC.

[g]For constructs tagged with Kan^R (rather than Hyg^R), a similar cloning strategy was employed. Rather than use of a parental construct already containing GFP and the drug cassette, a plasmid with the *CDC11* promoter was linearized and used for *in vivo* assembly. The included GFP, *BUD3* fragments, *ADH1*(t) and Kan^R were included on amplified PCR fragments.

Supplemental Figure G.S1. DNA Sequences used in this study.

The universal *S. pombe HIS5*, Kan^R, Hyg^R MX-based cassettes [5] were used following the *ADH1* terminator sequence for integrated or plasmid-born constructs.

prCDC10-CDC10-mCherry-ADH1(t)
(from strain GFY-42)

GTATTGGCTCTCTTATAATGGCCGCATGATGGGCACTTATTCAAATGGTGAATCATCTTCAATTGCTTCTTACCTGG
ACCGTAAAGTTTTTGCCTTTTTTTTTTGGTGTGATACTTTTTTTTTAGGAACTGCTAGTAGAATACCATTATTAGAGA
AAAAATCTTTATCTTCAGTGTCCTTATGATCTTGTTGTAGCTTTTCTCTCAACCAACGGTGGATAGATCCCAGTTGT
TTTTGGCCTGGAATCTGCAGTGAAGGATTTCCTAGCGGTACCAACCACCCAATTGCAGTGGTAGGCACAATTTTGTG
AAATGCTAATTGTTTACCAAAAATCAAAGAATTCATAGTTTGTTTGCTGATATTCCCTGGTGTTTTTGTCTCCAAGT
TAAGGTATTATTTGGGAACATTGTGCCTTATATTAAGATTTGCTTTCTGAATGCTTGCGGAAATAATCGTATGATAC
GATGAGATTTCCTCGAGAGACCACAAATAGATCAGAATCGCTGAATAGAATATATGTAAACATGTATCAGTAATACT
TAACTTTTTTCAGGCAAAGACAAGAAAATACAAGGCCAAGCCCCACGGTTACTACAAGCACTCTATAAATATATTAA
TGGATCCTCTCAGCTCAGTACAGCCTGCTTCTTATGTTGGTTTTGATACCATCACGAATCAGATCGAACATCGTCTG
TTGAAGAAAGGTTTTCAATTTAATATAATGGTTGTTGGCCAATCCGGATTGGGTAAAAGTACTCTAATAAATACGTT
ATTTGCCTCACATTTGATTGATTCTGCTACTGGTGATGATATTTCTGCCCTGCCTGTTACAAAAACAACTGAAATGA
AAATTTCTACTCATACTCTTGTGGAGGACCGCGTTCGCTTGAATATTAATGTTATAGATACACCTGGATTTGGTGAC
TTTATTGACAATTCTAAAGCTTGGGAGCCTATTGTGAAGTACATTAAGGAACAACATTCTCAATACTTACGTAAAGA
ATTGACAGCCCAACGTGAAAGGTTTATTACTGATACAAGAGTTCATGCAATTCTTTATTTCCTGCAACCAAATGGAA
AGGAGTTGAGCCGCCTTGACGTTGAAGCCTTGAAAAGATTGACAGAAATAGCAAATGTTATACCAGTTATTGGCAAG
TCGGATACATTGACTTTAGATGAAAGAACGGAGTTTAGGGAGCTTATTCAAAATGAATTCGAAAAATACAATTTCAA
GATTTATCCTTATGATTCGGAAGAACTAACTGACGAGGAATTAGAACTAAACAGAAGTGTTAGATCTATCATTCCGT
TTGCAGTGGTTGGTTCTGAGAATGAGATTGAAATAAACGGTGAAACCTTCAGGGGAAGAAAAACTCGTTGGAGCGCT
ATTAATGTTGAGGATATCAACCAGTGTGATTTTGTATATTTAAGGGAATTTTTGATTCGAACTCATCTCCAAGACTT
AATCGAAACAACTTCCTACATTCATTATGAAGGGTTCAGAGCAAGACAATTAATTGCCTTGAAAGAAAATGCGAATA
GTCGTTCCTCAGCTCATATGTCTAGCAACGCCATTCAACGTGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATC
AAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGA
GGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACA
TCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTG
TCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTC
CTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGC
AGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAG
CAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCA
GCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGT
ACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGTAAGGCGCGCCACTTCTAAATAAG
CGAATTTCTTATGATTTATGATTTTTATTATTAAATAAGTTATAAAAAAAATAAGTGTATACAAATTTTAAAGTGAC
TCTTAGGTTTTAAAACGAAAATTCTTATTCTTGAGTAACTCTTTCCTGTAGGTCAGGTTGCTTTCTCAGGTATAGTA
TGAGGTCGCTCTTATTGACCACACCTCTACCGGCAGATCCGCTAGGGATAACAGGGTAATAT

prCDC11-GFP-Bud3(1-1636)-ADH1(t):
(from plasmid pGF-IVL-1636)

CCATGTATTTACTGACTTTGCAAAAGTCCAAATATCTGTAAATTAACATGCTATTATAAATATATATATATATATAT
ATTTGAATTTTCTATTTTTTCATCTACAAAAGCAGGTTATAGCTCCGTTAAACCCTTTGCTCATTGAATTGTCCTTC
ACGGTTTTTTGGTGGAAAACACAACATGGAACATAACATTTAAACATCGTTCTCAATCCATTTCATCGCAGTAACAT
ATTTGGTCGGGCACATACGGACACGTATATGTTATACAGATAAATATAGCTTAAATATGGCTATATTTACACAACAC
AATGCTCGTGGGTCGTTACCCGACCATGCAGGATCGCCAGTATTCTTTATTTTTCCGCCTTCTAACATAAAGAAAAA
TAAACAAAAAAGTATTTGATCGAAAAGTAAAATAGGTAGACACCACGTATTGGCGACCCGATCTGGAGCCGTTTAGA
AAGTCAATCATCACAAGGCCTAAAGTTGCTAACCACCAGCCATGGGTCGACGGATCCCCGGGTTAATTAACAGTAAA
GGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGT
CAGTGGGGAGGGTGAAGGTGATGCAACATACGGAAAACTTACCCTTAAATTTATTTGCACTACTGGAAAACTACCTG
TTCCATGGCCAACACTTGTCACTACTTTGACTTATGGTGTTCAATGCTTTTCAAGATACCCAGATCATATGAAACGG
CATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAAAGAACTATATTTTTCAAAGATGACGGGAACTA
CAAGACACGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATAGAATCGAGTTAAAAGGTATTGATTTTAAAG

AAGATGGAAACATTCTTGGACACAAATTGGAATACAACTATAACTCACACAATGTATACATCATGGCAGACAAACAA
AAGAATGGAATCAAAGTTAACTTCAAAATTAGACACAACATTGAAGATGGAAGCGTTCAACTAGCAGACCATTATCA
ACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCCACACAATCTGCCCTTTCGA
AAGATCCCAACGAAAAGAGAGACCACATGGTCCTTCTTGAGTTTGTAACAGCTGCTGGGATTACACATGGCATGGAT
GAACTATACAAAATGGAGAAAGACCTGTCGTCTCTTTACTCTGAAAAGAAAGACAAAGAGAACGATGAAACCTTATT
TAACATCAAACTATCCAAATCTGTTGTCGAGACCACACCGCTAAATGGTCATTCATTGTTTGATGATGATAAATCAC
TTTCAGACTGGACGGATAATGTGTTCACTCAATCAGTATTCTATCACGGGTCAGATGACTTGATATGGGGGAAGTTC
TTTGTCTGCGTGTACAAGTCCCCCAACAGCAATAAGTTGAACGCTATAATATTCGACAAATTAGGAACATCATGCTT
CGAATCCGTCGATATATCTTCCAACTCGCAATACTATCCGGCCATTGAGAATTTGAGTCCAAGTGATCAGGAAAGCA
ATGTTAAGAAATGCATTGCTGTCATTCTGTTACAGCGCTATCCATTACTTTCACCATCAGACTTATCACAAATATTG
TCCAATAAATCGGAAAATTGCGACTATGACCCCCCTTATGCTGGAGATTTGGCTAGTAGTTGCCAGTTGATAACAGC
AGTTCCTCCAGAAGATCTGGGGAAGCGCTTCTTTACATCAGGACTTCTGCAAAATAGATTTGTCAGCTCTACCCTGT
TAGATGTTATTTATGAAAACAATGAATCCACCATCGAACTAAATAATAGGTTGGTATTCCATCTGGGTGAACAACTT
GAACAACTTTTTAACCCAGTCACAGAATACTCACCGGAACAGACAGAATATGGTTATAAGGCGCCAGAGGACGAATT
ACCCACAGAATCGGATGATGATCTTGTCAAGGCCATTTGCAACGAGTTATTACAACTACAAACAAATTTTACTTTCA
ATTTGGTAGAATTTTTGCAAAAATTCCTGATCGCCTTGAGAGTCAGAGTACTCAATGAAGAAATTAATGGGTTATCC
ACAACCAAATTAAATCGACTCTTCCCACCTACAATAGATGAAGTCACAAGAATCAATTGTATTTTTCTAGACTCGCT
AAAGACAGCAATCCCTTACGGTTCCCTCGAAGTACTGAAGGCATGCAGCATTACTATTCCTTATTTCTACAAAGCAT
ATACAAGACACGAGGCGGCCACAAAGAACTTCAGCAAAGATATTAAATTGTTTATTAGGCATTTCAGCAATGTAATT
CCAGAAAGAGAGGTCTACACGGAAATGAAAATCGAGAGTATAATTAAGGGACCTCAGGAAAAACTACTGAAGCTAAA
GTTAATTATAGAGAGATTGTGGAAGTCGAAAAAATGGAGACCGAAAAATCAAGAAATGGCAAAAAAATGCTACAACA
ATATCATTGATGTCATTGATTCGTTTGGAAAATTAGATTCCCCACTTCATTCTTATAGTACCAGAGTATTTACTCCA
TCGGGAAAAATCCTTACAGAATTAGCCAAATGCTGGCCCGTAGAACTGCAATACAAATGGCTGAAGAGAAGGGTAGT
CGGTGTGTATGATGTAGTGGATTTGAATGATGAAAATAAGAGAAATTTATTAGTCATATTCAGTGATTATGTGGTTT
TCATCAATATACTGGAGGCAGAAAGTTACTACACTTCAGATGGATCAAACAGGCCCTTAATCTCAGATATTTTAATG
AACTCATTGATCAACGAAGTTCCGTTGCCCTCCAAGATCCCTAAGTTGAAAGTGGAGCGTCATTGCTATATAGATGA
GGTTCTAGTTTCTATATTAGACAAAAGCACTCTACGTTTTGATCGATTGAAGGGAAAAGATTCTTTCTCAATGGTAT
GTAAATTATCCTCTGCATTTATCTCTTCTTCGTCAGTTGCTGACTTGATTACGAAGGCTAGAATTTTGGAAAAAGAC
ACTGCATTTCATTTATTTAAAGCTAGTAGAAGCCATTTTACATTATATTCTACTGCTCACGAGCTTTGCGCTTATGA
TTCCGAAAAAATAAAATCAAAATTTGCCTTATTCCTGAACATACCACCATCCAAGGAGATATTGGAGGTCAACAACC
TTCATTTGGCTTTTTTTGCAAGATTTTGCAGTAACGATGGTAGAGATAACATCGTAATCTTAGACGTCTTAACCAAA
CATGACGATAAACATATAGAAGTTACATCCGATAACATTGTTTTCACCATAATTAATCAATTGGCCATTGAAATACC
GATATGCTTTTCTTCCTTAAACTCATCGATGGCCAAAGATTTACTCTGTGTAAATGAGAATTTGATAAAAAACTTAG
AACATCAATTGGAAGAGGTCAAGCACCCTTCAACAGACGAACATAGGGCTGTTAATAGCAAACTTTCCGGTGCATCC
GATTTCGATGCTACTCACGAGAAGAAAAGATCATACGGTACCATAACAACATTTAGAAGCTATACAAGCGACTTGAA
GGACAGTCCATCAGGCGATAATAGTAATGTCACCAAGGAAACTAAGGAAATTTTACCAGTGAAACCTACGAAAAAGT
CTTCAAAAAAACCAAGAGAAATTCAAAAGAAGACCAAGACAAACGCCTCTAAAGCAGAGCACATAGAAAAGAAGAAG
CCTAACAAAGGCAAAGGGTTTTTTGGCGTGTTAAAAAATGTTTTTGGAAGTAAAAGCAAGAGCAAGCCTTCACCAGT
TCAAAGAGTGCCTAAAAAAATATCGCAGAGGCATCCTAAGTCTCCAGTGAAGAAGCCAATGACCTCAGAAAAGAAAT
CCTCCCCTAAAAGGGCAGTCGTTTCATCTCCCAAAATTAAAAAGAAAAGTACTTCTTTTTCCACAAAAGAATCACAA
ACTGCTAAATCTTCTCTTCGAGCAGTTGAATTCAAATCTGATGACTTGATCGGAAAACCACCTGATGTTGGAAATGG
CGCACATCCTCAAGAAAATACCAGAATATCTTCAGTAGTAAGGGATACAAAATATGTCTCCTACAATCCCTCTCAGC
CTGTGACAGAAAATACCAGTAACGAAAAAAATGTCGAACCAAAAGCGGATCAATCCACAAAGCAGGATAACATTTCC
AATTTTGCAGATGTAGAGGTATCTGCGTCTTCTTATCCTGAAAAACTTGATGCAGAAACAGATGATCAAATAATTGG
GAAGGCGACGAATTCGTCATCAGTTCATGGAAATAAAGAGCTGCCAGACCTTGCTGAGGTGACTACAGCAAATAGGG
TTTCTACAACATCGGCTGGGGACCAACGTATTGATACCCAAAGCGAATTTTTACGTGCAGCTGATGTTGAAAACTTA
AGTGATGACGATGAACACAGACAGAATGAAAGTAGAGTTTTTAACGATGACCTCTTTGGTGATTTTATTCCTAAGCA
TTACCGTAATAAACAGGAGAACATTAACAGCTCGAGTAATTTGTTTCCAGAGGGAAAGGTGCCCCAAGAAAAGGGCG
TATCAAATGAAAACACTAACATATCTCTCAAAACTAATGAAGATGCATCTACATTGACGCAGAAACTCTCTCCACAA
GCGAGTAAAGTGCTGACAGAAAATTCTAATGAATTAAAAGATACCAACAATGAAGGGAAGGACGCAAAGGACATAAA
ATTAGGAGATGATTACAGTGATAAAGAAACAGCGAAAGAAATAACTAAACCAAAAAATTTTGTTGAAGGAATAACTG
AACGGAAAGAAATATTCCCCACTATTCCTAGGTTAGCGCCGCCAGCTTCAAAAATTAACTTTCAAAGGTCACCATCC
TATATTGAGCTCTTTCAAGGAATGAGGGTGGTTTTAGATAAGCATGATGCCCATTATAACTGGAAACGCTTGGCTAG
TCAAGTCTCCTTAAGTGAGGGACTAAAAGTCAATACTGAGGAAGATGCGGCAATTATAAATAAAAGTCAGGATGATG
CCAAGGCGGAAAGAATGACTCAAATTTCTGAAGTGATTGAGTATGAAATGCAGCAACCTATCCCAACTTATTTGCCT
AAGGCGCATCTAGATGACTCGGGTATTGAAAAAAGTGATGACAAATTCTTCGAAATTGAAGAAGAACTTAAGGAAGA
ATTGAAGGGCAGCAAAACCGGTAATGAAGATGTCGGTAATAATAATCCATCCAATTCTATTCCAAAAATCGAGAAGC

CCCCAGCATTCAAAGTTATTAGAACATCGCCTGTGAGAATTATCGGAAGGACTTTTGAAGACACTAGAAAATATGAA
AATGGCTCTCCATCTGATATTTCGTTCACTTATGATACTCACAACAATGATGAACCTGACAAAAGGCTGATGGAATT
AAAATTTCCATCCCAAGATGAAATTCCGGATGACAGATTCTATACTCCAGCAGAGGAACCCACTGCTGAATTTCCGG
TGGAAGAACTTCCAAATACTCCGCGAAGTATTAACGTTACAACTTCAAATAACAAGAGCACAGACGATAAGTTGAGT
AGCGGTAATATTGATCAAAAACCTACCGAACTGTTAGATGATTTAGAATTCAGTTCATTTAATATAGCATTTGGAAA
TACCTCCATGAGTACTGACAATATGAAAATATCATCCGACTTAAGTTCGAATAAAACCGTGTTAGGAAATGCTCAGA
AAGTTCAAGAGTCTCCTAGTGGACCATTAATCTATGTTTTGCCTCAGAGTAGCACAAAGCATGAGAAAGAGGGGTTC
CTTCGAAAGAAACAAAAAGACGAGCCCATTTGGGTTTCCCCTAGCAAAATTGACTTTGCTGATCTAAGTAGGAGAAC
TAAAGCATTGACGCCAGAGCGTAATACTGTTCCTTTGAAAAACAACGACAGTAGAAAATACAAATATACTGGAGAGG
GATCTATCGGTAATATGACAAATATGCTGTTAACTAAAGATGCTTCGTACGCGTACTTAAAAGATTTTGTTGCGTTG
AGTGACGATGAAGATGAAGATGGGAAACAGAACTGCGCTGTTGGTGGCCCAGAGAAACTGAAATTTTATTGAGGCGC
GCCACTTCTAAATAAGCGAATTTCTTATGATTTATGATTTTTATTATTAAATAAGTTATAAAAAAAAAATAAGTGTAT
ACAAATTTTAAAGTGACTCTTAGGTTTTAAAACGAAAATTCTTATTCTTGAGTAACTCTTTCCTGTAGGTCAGGTTG
CTTTCTCAGGTATAGTATGAGGTCGCTCTTATTGACCACACCTCTACCGGCAGATCCGCTAGGGATAACAGGGTAAT
AT

Supplemental Figure G.S2. Fungal proteins similar to yeast Bud3 used for domain alignment.

The full-length Bud3 protein sequence from *S. cerevisiae* was used as a query to search for similar proteins within the fungal kingdom using BLAST (NCBI). The following proteins were used to perform a CLUSTAL-W alignment:

NP_009914.2 Bud3p [*Saccharomyces cerevisiae* S288C]

MEKDLSSLYSEKKDKENDETLFNIKLSKSVVETTPLNGHSLFDDDKSLSDWTDNVFTQSVFYHGSDDLIWGKFFVCV
YKSPNSNKLNAIIFDKLGTSCFESVDISSNSQYYPAIENLSPSDQESNVKKCIAVILLQRYPLLSPSDLSQILSNKS
ENCDYDPPYAGDLASSCQLITAVPPEDLGKRFFTSGLLQNRFVSSTLLDVIYENNESTIELNNRLVFHLGEQLEQLF
NPVTEYSPEQTEYGYKAPEDELPTESDDDLVKAICNELLQLQTNFTFNLVEFLQKFLIALRVRVLNEEINGLSTTKL
NRLFPPTIDEVTRINCIFLDSLKTAIPYGSLEVLKACSITIPYFYKAYTRHEAATKNFSKDIKLFIRHFSNVIPERE
VYTEMKIESIIKGPQEKLLKLKLIIERLWKSKKWRPKNQEMAKKCYNNIIDVIDSFGKLDSPLHSYSTRVFTPSGKI
LTELAKCWPVELQYKWLKRRVVGVYDVVDLNDENKRNLLVIFSDYVVFINILEAESYYTSDGSNRPLISDILMNSLI
NEVPLPSKIPKLKVERHCYIDEVLVSILDKSTLRFDRLKGKDSFSMVCKLSSAFISSSSVADLITKARILEKDTAFH
LFKASRSHFTLYSTAHELCAYDSEKIKSKFALFLNIPPSKEILEVNNLHLAFFARFCSNDGRDNIVILDVLTKHDDK
HIEVTSDNIVFTIINQLAIEIPICFSSLNSSMAKDLLCVNENLIKNLEHQLEEVKHPSTDEHRAVNSKLSGASDFDA
THEKKRSYGTITTFRSYTSDLKDSPSGDNSNVTKETKEILPVKPTKKSSKKPREIQKKTKTNASKAEHIEKKKPNKG
KGFFGVLKNVFGSKSKSKPSPVQRVPKKISQRHPKSPVKKPMTSEKKSSPKRAVVSSPKIKKKSTSFSTKESQTAKS
SLRAVEFKSDDLIGKPPDVGNGAHPQENTRISSVVRDTKYVSYNPSQPVTENTSNEKNVEPKADQSTKQDNISNFAD
VEVSASSYPEKLDAETDDQIIGKATNSSSVHGNKELPDLAEVTTANRVSTTSAGDQRIDTQSEFLRAADVENLSDDD
EHRQNESRVFNDDLFGDFIPKHYRNKQENINSSSNLFPEGKVPQEKGVSNENTNISLKTNEDASTLTQKLSPQASKV
LTENSNELKDTNNEGKDAKDIKLGDDYSDKETAKEITKPKNFVEGITERKEIFPTIPRLAPPASKINFQRSPSYIEL
FQGMRVVLDKHDAHYNWKRLASQVSLSEGLKVNTEEDAAIINKSQDDAKAERMTQISEVIEYEMQQPIPTYLPKAHL
DDSGIEKSDDKFFEIEEELKEELKGSKTGNEDVGNNNPSNSIPKIEKPPAFKVIRTSPVRIIGRTFEDTRKYENGSP
SDISFTYDTHNNDEPDKRLMELKFPSQDEIPDDRFYTPAEEPTAEFPVEELPNTPRSINVTTSNNKSTDDKLSSGNI
DQKPTELLDDLEFSSFNIAFGNTSMSTDNMKISSDLSSNKTVLGNAQKVQESPSGPLIYVLPQSSTKHEKEGFLRKK
QKDEPIWVSPSKIDFADLSRRTKALTPERNTVPLKNNDSRKYKYTGEGSIGNMTNMLLTKDASYAYLKDFVALSDDE
DEDGKQNCAVGGPEKLKFY

QNG12518.1 uncharacterized protein GWK60_B04697 [*Candida glabrata*]

MVSEQSSNYSKESIEEKPILLHNNLAHTSPDVQVSLTEIKDTLTIFRGSDELIWGDFLVAIGMDNNTNFNSIMVNKF
GATTFNNVNISKNSKYYPAIENLDPKYKDSNARKCLAVTLLKIYPLFEKHITTVSDLPFEYDQTHAGELASTATLIP
NVDPEVFLLKLKNYGMLNDTSRVITSTMVDVVYENNEDVIDYNNQLVYYLGEQLEQLFNPVTEYSPEQTEYAYKAPD
EEPSLFNSDTSLLKSVCNELLQLQTKFTYDLVEFLQEFLIILRVNVLNEEIEGLSTLKLNRLFPPTIDEVTRINCIF
LDSLKAAVPFGSFEVLKACNITIPYFYKAYTRHEAATKNFSKDIKLFLENFGDVIPKKEEYTEMKMEALMKGPQEKL
LKIKLIIERLWESTPDWGNQEKEATSFYNNIIDIIDSFGKLESPLHSYTTRVFTPSGKILTELAKGWPIELQYKWLK
RRVVGVYDIIDSSDISKRKLLVIFSDYVVFLDIQNSESYYKNPNRPMLSDILTNSLINEVPLPSKIPRMKVQKYSYI
DDVQVTIVDGNILRFDCIREVDPFSISCRLKSKSTTEKRIADLITKAKILEKDTAFHLFKAEVNGVTLYSTAHEYQA
YRSERQKSKVALFLNLSPSPSYITENSLYAGLFMKFEDTTRLDKIMITTVLYDGTKSNYVVRPEQMITFVVRQLSQL
LPNCYSSTRSPLATSLLNLQAQLISELVKPKIAASSKSNDSAKLMDMNNFIAEKADKSDKYDAKHEKKRSYGTITTF
RSYKSDLKDVESSGEIYSREHNTSTKISKNNVQKISKNVRTKVTESPGRKQRVSKTQADKSKKTSGLSNFFKSIFKG
SGKKTKRPTNEKIQLKRIGSNKNISHVHISSKPRNKAVIKDTTNTAELEEPLISRAVTSVEISAEKVDRKDEDERVL
SVVHNKQFEETKANSEAPADNERMGQDIHHDAGNTSDLLIKEIADEVVEQKEFLEQLSKNMSIQKENKTVQSQVTQK
PTSIKNQASALYLYDDDLFGDFKQKPNDHDNISVSLENELTAEGSENEDEDPKKVSHQYQEETIEYQENEPKNDEPD
NRSKEADEQADSKENVELLPTKSEVSPGKPLVFPTIQKVVKKPQQIQRSDSFYELYKGMRMVLDDTDVKYNWKRLPS
LVSLSVQNAVNSDKSKHAFEKIAHARETPLLVENLMEKKELPMQATETQVREARKLKDAAVIKPVENNFKPVSDTLK
SPFKAYPELTKSTEYVNNDSINIFSELEKAFEVPQAKVSVPTFKVVRTSPTRYVQLESGSIRTDRMVPSPELSQSLS
YSSFVEELQNKKPQRLVEIADSITDMQSNLTKDDETIASSNAPSHVSSEFEQQKDISQQSDSSLEEEHYSAKKPLSQ
EINSSGILESLEFSSFTMGFEDTIDVVNSSHIPSGDLLPETGSRNMVLTRPKRNDAHPVYLLPRYSISTTKSLNQTR
ERMQYDEDAIWVSPTKLSFSEKDSSIRLVKEGTTPTKTLTHQGNKTKVDIPREESSFGYLSSLLAVDSNNNNILEFV

XP_003959137.1 hypothetical protein KAFR_0I02230 [*Kazachstania africana* CBS 2517]

MNDLSSIYSQELELTERSVVISLFNNLLASKNTHQQTLDWIHERSEKEWLDDIFINSVLYHNVDRVIWGPFFICIYK
DPKTDKFGSLTLDRFGITHINSIDLSRKSAYYPAIENLHDNDKNSNVKKCIAVSLLQKFSNISMQHLKYLTEDKINY
DPVHAGDLSSGCKLVTSISPELFGKRLISAGLLTGRLINSTLMDVIYENNESTIDSNNRLVFHLGEQLEQLFDPVTE
YSPEQTEYGYKPPEDDKPTETDSELVQAICNELLQLQSNFTFTLVEFLQKVLIALRVKVLNGEIDALSTVKLNRMFP
PTIDEVTRINCIFLDSLKSSMPYGSLEVLKACSITIPYFYKAYTRHEAATKLFSRDIKQFLKHFRNSIPECEEYSEM
KLEAIIKGPQEKLLKLKLIIDRLYHSKEWANEENKIIGKKNYDNIIDVIDSFGRLNEPVSSYSTRVFTPSGKILTEL
AKGWPVELQYKWLKRRVVGVYDIIDQTFSNKRALLVIFSDYIVFLSITDYELYYTDDGSNKPLISDILMNSLINEVA
LPSKIPKLNVQNYCYIGDVLVSIFDNDTIRFDALRPDAFSISCKLATMKEPIDAKKVAALVTKAKILEKDTAFHLFK
ACRDDISLYSTAHELEAYNNEKLKSKFALFLNIKPASQFLSLYNLHFAVFASFVGTDDTNKVRLSVITRSNKDTVKT
LEIFPDNIVDSIIGQLSTEYPLCYSSIQSSLIKELFAVTTYLANSIGKVGEKEHTLREAAAIVKDSNKSSKTNTLVN
SDAEKKHSKAKDIDAKASKKIKKKTDNSTQKRISKIPESKNVMNLKPQEVKKKSLIEKLKSIFKSKRRSKKDISGPI
VVNSKSYSSKSSLPHKKNNSPLPISPNKNILANTRSEKNLNAKNSSGKNPGVKSNDDAENLRISSVVRDTTYDASGA
RFQYTPRIGDLSKEKTEDIAEPHTPDEYTQLSQIPSPHIAEKLFLPKESKIHEPVDEVSEHEDIAGDISALTQSTSQ
ENVEVLSAGQSGKEKDVSKRAHNQSQLYNSDLFEDFVPPGKECGHNIEIKQASEELSDDNLPTEKSSTQNQLHNVEK
NNAVTLNSGEDNTKEKKNGTFHAGKQDPFVDIANPPIVQENIGVLPKKLNIFPTIPKLAPISRMEFTRSPSLIELFE
GMRVVLDETDAHYNWKRLSSEVSLNQKYITNSGNTPENNNNFRNFAHAAAFNLAIPKNVETIVIPDEKAELESSHST
NGTLKTGDDHSKTTVPSVEEVSEEAVGKSPLKHEENPSLAANDTNRDRALDSERKFKVINTSPTRYSNIGSSEKQLA
STTSETDFSKASFENPEDLQLSIARDISSQKLSVDTTTKMVPDFSFLTEMNREVTTRLFELNLASQEDVNDEEYYTP
TTNAPKENMAASTSDQVETENETSIGLDVATGLENDLAKIRETSETHPEGDKSQNVATEESQNFLEDLEFSSFAMTF
DNSLFNQYSSGLQEATLTHNSYETTKILDNIPELPAEKEGPVVYTLTNDLFSSNDIRIGINQKAGFMDNSNDAEDPI
WVSPSKLDFDDITRTHVRTLGENIKDHAITRKNTKPDPPSNKLDDSELKHDLSYAFLANLVQTSEFDEIEEENYNDD
KPTRLQFKS

KAG0668095.1 hypothetical protein C6P45_004993 [*Kazachstania exigua*]

MVMADIEINEFSSIYSHDSERNKHIDHLVQTTDIQGSSTPEINQSEIPKNENDIVLFNNYLSGADDSTTYNDIQSFS
KIKSYWTQDVFDNSKIYTSYDDLIWGKFVLCIHHSKDTYNCLMMDDFGIRELLNIRTSLESLYYGAVEGLNSDDKNS
EIKKCIAILLLRRYSELSSIQKIELKSLYPTRSIEQYDPAHAGQVASGCIEVKEVTASKFGERLISQGFLDKHIVKS
TLIDVIYESKESTIELNNRLVFHLGEQLEQLFNPITEYSPEQTEYGYKAPDNDIATETDTQLVEAICNELLEVQSNF
TFNLVEFLQKFLITLRVQVLNNEIEGLSTVKLNRLFPPTIDEVTRINCIFLDSLKSATPFGSFEVLKACNLTIPYFY
KAYTRHEAATKNFSKDIKLFIRNFSDIMPNKDIYTEMKLETIIKGPQEKMLKLKLIIDRLYKKKVWSAENSNNAKKY
YDNIIDIINSFGHLEAPMSSYNTRVFTPSGKILTELAKGWPVELQYKWLKRRIVGVFDIIDSNNPSKRNLLVIFSDY
IVFLNILNYKKYYTSDKSNKKPLISDILMNSLINELPLPPKIPKLEVEKYCYINDIHTSVTNGDILRFDAIRGKGKQ
PFSMVCQLASKTNSADDIADLITKAKILEKDTAFHLFKSVIKESTLYFTAHELESYNTERIKSKFGLFLNMEPSNEL
LISNDLHSAAFAKFTDPEKSNMVKLTVIRRHDRLSSKTDTIPADDIVCTLMQQLSSDIPICYSSILSEDAMKLVCVN
NIVTNKVIEEELFENITDDISTQNIASQVNEENKKSYGTITTFRSDVSDLVDVPHMKEDKRRDADGTSANKASETKK
ESTKTKAPKSAKKSSRTDKTNQKKTRATNKKTDDKKSTPKERKNKGFFGAVKSIFSNSSRKEKRTIGKPVIVKQKTT
NRDLEVKQHIKSKSHPLSPVKVTKPSDSKTSDEKSKLEEVKVEKNDRKVDKTEKDDNSEEDDKVEERKCSEIYNVRI
SSVVHNIEYTGSSVNTQQGSNKTEITNSIPKVAPAFKIMSPTKELPDIPVKVENDDLVISDVSRRTDSFDETEGNDT
DIKLANPITHQSKLFDNDLFGDFIPEKVHSKQQEIEKASPEKKDKIEVHPESSEFKLTLNTPVDIPAEHNIELPTLD
DINLEPNDDINEGLLTDIHEPRIKKKVQIFPEIANVEPPRSRIKFEKSPSFIELFRDMRIVLDDSDAKYNWKRLSTE
VSLNEKYLVNNEVPSSENVSNKHGLKTIVESNYNLPQKQSPMMRLSNVDPVSPARQKESHSETSPFRALAHNKLSDL
TVNDLHKFNSEAKNENITLRTPLKNPVNNASPNIGSPSARSSPLKTGSIFKVINKSPTRITNSPSKEATMELSHNLK
ENRKFLEQQALSTGMNLQPPPLMNSKSNNMNERNIASDFSFASDFNGDSNRRWVKLNVNSKEDLLDDTFHTPLEEPS
ETFSDALFGDINVTTNSAVPQLPQIPAKQTNSQIAAVVDINKNSSQGNHSKEKADILDDLEFSSFDMTFNTSTTTND
MEPTVPTDMSNSKNILVGLNNNDIPKMDPPVVYKYSKEQSFNTRKNLYNDTKYDDDSNEPFWISPSKVDFTSLTKSI
TKPTRNGTTLNPNIYTADAHTETPMKRNFKSTNREINLTQDMSFAFLGSLVDIDGTGDNLDDQAQKMYLHVYNLSNR
DIYKN

SCW02591.1 LAFE_0F09868g1_1 [*Lachancea fermentati*]

MENGAKEGLIDSSNATVDKMDTGCRIKLRPNLALNGKPKQWIIDRSTEEWVSQIFPTAAIFKGYDNLIWGNFFIVVY
REILTEKFSSIIIDKLGLTHFNSIDISRQSRFYPAVENLTDVNQKSNVRKLIAISLLRKYADIDSKLIKKISPEMKY
DYDPTIAGDLANSCELFGQCTPEEFGESLSQLGVLRDRFINSILLDVVYENKAKIIDSNNELVFHLGEQLEQLFNPL
SEYSPEQTEYVYKSPDDGYVVESEDPLIVSICNELLQLQTNFTLSLVEFLQNFLIPLRIEVANDEIDGLSIPKLNRL
FPPTIDEVTRINCIFLDALKSATPYGSAEVLKACSVTIPYFYKAYTRHEAATKNFSKDIKLFLSKFGHLTPNRDVYT
ELKIETLIKGPQEKIMKLKLIIERLWSNENLVHDDAAKKRYDDIIETIHSFGTLEKPMSAYSTRVFTPSGKILTELA

KGWPVELQYKWLKRRVVGVFDMVDADDKAKRDILVIFSDYVVFLRVIGGEDYYNMDGNKPLISDILMNSLINEVPLP
SKIPKLEVLAHTYIDKVLVSTYGSSFLRFDLRDKNFCTPLTYELASNSMTTSEVADLVTKAQILEKDTAFHLFKYSN
DYLQIYSTAHELSAYSTEKIKSQFALFLNLEPSIELIEKYGLCMALFASFRKDHQVNLTRITVDNARTELTVPLHDL
AETLANELVDAIPIYFSSLSSPFYEELLQLNAKLVQRIGKSFRKDELTIQENKNDKAPEKTSSAEIIYHHQKNKSFG
TITTFRSFPSDMKELSDNNTINEKIKDKSKHGGSKARKTVNRDIANHPPKQMGLIKTFKNLFGRRKKSRKENERLTI
SSPHITSPKTESKANSSITRVKKTIGDPKSHVRESQEVEHLATQPAESSRVSSVIHTPSRNITDCVDISKESMSSMN
RKAQAGVEKNSTENSSSTQLQDYEINHSSLSVHSEKEVLPDSVKATGTPYKESLRQSKVFNDDPFASVISSDENESI
GPAKDNLCGESLTGETSVSVKDSKLVGKEADEPTIHKEDNHFAKTLAEKVELKPSIPEVHDTAKSFEGPIDSRRLSE
GRCGQEFENAQIFPKVQGLKTKTIDFCRSPSFVELFDGMRAILDITDESTNWRRLSSEGSLTIQTTPHAEDVKALDK
NCIKGNSSAIMLDKSNHSNMPDARKSSPDHRDENDDARIFASSKVSEPEYGQLNLDFSKSEGTLTPVFPTFKVVNTS
PAKIVNFATSINNVHPDLTNNSTRESSSPVKNSYSSDLNAENLRLVELSFNSQDDVYSYDSRQKPNDILETLPPSHS
IHSPRLSNSEVNSKPLADQLSATTSVEVANLQNELVIEPLESNILGDLDFSSFNMTFDISNELNDSTQESIMDEPKN
PFVRKPHLSPRDPVFYRLPNSTRSDETFFSCVDDQRNKGQKRKSYINSLPLECEDEPMWVSPSKIDMFDLSKQPDSV
FQELKLRSKQTREAKFQLPGKANRSSTEEQLLLPDSSYAYLGSLLTDDDMQLAEISSDDRPVRLKFNP

XP_003980081.1 hypothetical protein NDAI_0G04200 [*Naumovozyma dairenensis* CBS 421]

MATTQDLSSLYSNHNNDITGKENANTLTIPLYRNLIESYLPESISIGNHFLQNEKQLNACFDDSSHEASRYAGIFKG
FDELLWGQFIVCLYKETASKKYNAIMLDKLGPTTFNSIDVSKDSQFYAAIENLNTTNRKSTLRKAIAVILLQKFVEI
PKEGVHKIIQPNKTLNFEYDPQYVGELTNSCEEISSSNHNNSSNSYQVISFMKKLYTMGSFQRKYINSTLLDVVYEN
NESTIEINNRLVFHLGEQLEQLFNPVIEYSPEHTGYVYKPPDEEESTSNISNNGNDNSNGKNDQEDNALIKAICNEL
LELQTNFTLSLVEFLQNFLIVLRVKVLNDEIDRLSTRKLNRLFPPTIDEVTRINCIFLDSLKAAIPYGSLEVLKACS
MTIPYFYKAYTRHEAATKNFSKDIKSFLRNFDDAIPKNQSLTEMKISTIINGPQEKLLKVKLIIDRLYTSKKNWNIP
NKTKADKYYNNIVDVVDSFGKLSNTSLSSSYNTRVFTPSGKILTELAKGWPVELQYNWLKRRVVGVFDIVDETYGGR
KKLLVVFSDYIIFLNIINPDSYYSIETDNQDNSLIKPHLSDILMNSLINEVPLPAKIPKLEVDNYCYIDDIFASTFE
ENYLRFDAFKEDKSFSVSYRLSTESTPVSTVADLIMKAKILEKDTAFHLFKASFNNITLYSTAHELAAYNTEKIKAK
YALFLNIEPSKELLIQNDLHMAIFAKFANENEGKFINLDILTSTEVHRKDTFEAAKIIPELIGQLLLEIPIYRSSIV
SSQAIKLMAINEGLVKQCISCLEKETPSIPITDSYNKPAQTDKQAKEKPYETTLTSKNIVNNSRALRGKSVNTKTAV
EKDNTVKKIEKNTNKKLPSAKNLPRKQTGKLPTNVPSQLLAKKEKRRSGLITAFRGIFGSKKNKDAEKGRNQLDDKS
KNYQKTTPRQKLVPKVSQKDSIRKENTTPQPIKNESVTSFARVSKNQKIEPPMLSEEQRITSVIHNRQYTEDIQGNK
VMNNLDSESKENVLSETATTTPLTSKEISVSSDKQLRTALESPPGSIGNSEKVLPDPNRSLSNKNTTSRMPQIQIFT
DDLFGDFVKPPSLDKQTAKQEETECDNGDEDAEHYQKHLQEDALGEGVSSKDNSPLPTTANATSENRGFIAPSENPS
FRDGEIQHEKVKIFPDIPKMQLSKIQFDRSTSFVELFKGMRLVLDDSDAKYNWKRIPSEQGLNINNLVNNATTIKNT
DRNMVGNEIEKVIPKEKCYSEAAQEQRPSTPESSSNITEAEEEPREEKEGGHITSPQPVLSTIALLPDQTRRRLPTF
KVTKTSPVRIINKTLNSPQKVVGKVENDSCFSSDVTIGNIENNKRLHKLTFRSQEDVFDAKICTASRENIDTIEATV
QADRSKNNEEQSSKLHDRLARSGSNENLLEAFEFSSFNMSFQDLPENNEECVTSIHPVTIHDTVVPPLQNLEPAPMV
YRLSNEHSSNKNKLEADDDPIWVSPSKLDLLSSNKLRDRVLVNKTENEALSNQKTPIASIKNKSIPSQKLLDTSLIR
NASYVYLTDFVDFHDDEKYEDKPTRLQFQ

XP_003679944.1 hypothetical protein TDEL_0B06040 [*Torulaspora delbrueckii*]

MTEVDLSSLCSATPKDAVKLHGESMKVRLFENLLERHIKNRQWLKDRTGEEWVEKVFKEAAFYEGVDELIWGQFFIC
VYKDPQSTKLSALIIDKFGLSHFNPVDISPKSQFYPAIENLDPKERKSNVRKCIAVSLLQKYAQLPDGCIHRIQPNG
LKFDYDPTHAGDLASACKLVDFCTPENAGKSIKSLGILGRRHVQSSLLDVVYESNSNNSEENNRMVFNLGEQLEQLF
DPLSEYSPEQTEYIYKAPENDESLYQDNQLVKSILNELLELQMAFTLSLVEFLQGFLISLRVKVLNNEIEGLSTVKL
NRLFPPTIDEVTRINCIFLDSLKSATPYGASEVLKACSVTIPYFYKAYTRHEAATKNFSKDIKLFLRRFQDLIPEKD
VYTEMKIETIIKGPQEKLLKIKLIIDRLYEEKEWPEELQEEAQRNYNSVVEVIDSFGKLEVPLSSYNTRVFTPSGKI
LTELAKGWPVELQYKWLKRRIVGIFDVVDANDSSKRKLLVIFSDYVVFLDIARAELYYAADGKNRPQISDILMNSLI
NEVPLPPKIPKLTVNSYCFIDDVLVSLVEGNSIRIDALREGESFSTTFRLASNSTTASTVAELVVKAKILEKETAFH
LFRAMQDNMTIYSTAHELEAYQNEKIKSRFSVFLNMKPSKDILHKNRLHLAIFLKFVTVGQNEQIQLEALTSDERTR
KATFPPEELVPALIEQLSIEIPVCYSSIYSPLYSILIEINDLLVKKIGHHFNPPIEEKDLANSDQASSFIREHEKKK
SFGTITTYRSHVSDFKDATNEQTNSGESHSTHKIPARTTDKTTKAAANHQKLHSSRKIDNAKKRRSIVGLVKGIFSG
KRSKDDVIKNKSKSTTKQSPKMRSKNMRADALVISKPTISKPNPPKNETEIYEKSQRISSVIRKTEFSPVEDPLINA
IEEPKIGSGATTKPALQNEAVSQNPLDVDISIHSSTIEPNLVERVEREEASSNHSTIRIAESAKQFYRQAGRQSQLF
NDDLFGELGSEPATEANQKQVDSAKIESKPKRKEMITRQDSSTLDSHTEGSGGSDFDEAPHTKNTDEGEMSTAHCFD
TETGRDSSAKTQQEVEKLVQDDIVNQKKIQIFPSISPPKLSRINIQRSSSFIELFEGMRLILDESDAQYNWKSLSND
GFLREVQESDEPATIPHVFRPIAIANTKLIPDETENLLNGKPSLSLNQCEDMFENQESQEISSPALKTNNVSPDASK

AVIKESSVDSRSSDPSKVKSPRKNGFKVVKSSPTRIIKKPFQQINIEVPDQNMTYNFSISSDLNQGADKRWFELKLP
SQEDLNSEIFHTPHEEPSSEFPEEQHNSETPSPFESDPNVIETSQDTITTDKPLQKGQDALLEDLEFSSFHMTFDTA
EGNSEQSPDTSSSDAAGATSSPKANSLVQNNKKDGPLLYRLPMTFSTKLSQSERAPTGDVGEDDDPIWVSPSKLDFY
DLSNTADSGATNYDLKVSARRDERKEPSIGDENMGTNRYSLRELSYAYLASLVSPTETSFEVDDKPQRLQFQ

XP_001646494.1 hypothetical protein Kpol_1048p67 [*Vanderwaltozyma polyspora* DSM 70294]

MRQVSYTSLYSDTVNVAAPFNNVNKYLDPTKGEIDAQQGSVSNIELFKNLIQYKLKDDKYQIEKTDKNWLDQVFEGS
KIYYGLDEILWGKFMVCVQCDKLNRKQTVLILDKLGVTFSDVIDVSNESQFYPAIENLDPKHNNSLERKCIAVALLQ
KYSLLSEEEISKIKSTVVKNNYDQIHAGDLASRCEILETKDVKKFGETILNTGLLQDRTVKSTLMDVIYKNESTSIE
LNNRLVYHLGEQLEQLFNPLTEYSPEQTEYTYKPPEEEKIDNTDSELTKAICNELLQFQTKFALNLVDFLQNFLIAL
RVEVLNEKINGLSTVKLNRLFPPTIDEVTRINCIFLDSLKAAIPFGAFEVLKACSVTIPYFYKAYTRHEAATKKFNK
DIKLFLANFGKQIPSPQVYTEMKIETIIKGPQEKLMKLKLIMDRLWESQKWSDEEKDKAKEYYDNIIEIIHSFGTLK
EPISSYSTRVFTPSGKILTELAKKWPVELQYKWLKRRVVGVFDVVLGPEFNDRGLLVIFSDYIVFLDIMDAHDYYND
KESNKPWLANVLMNSLINEVPLPSKIPTLNVSNYCYIDKVIVSVYERDIIRFDSFYDSDPFSISCKVISKTTSTNLV
ADLITKAKILEKDTAFHLFRANENGILLYSTAHEIDAYEKEKIKSNFSLFLNIDPSAELLDNYNIHVAFFAKLIDRP
NGETIQLDILTRDNRKYQVNISPDKLIPEILHQLSYEIPICYSSLRSNNFSDLLNINKVIISNLITGSVDKKDIVDE
PNGKTSADSFTMIHSKHRSYGSITTFRSYKSDMKDGSAEPIKEKVDATTPIENKRDTIRTQPQTILKDSGKTVPKTK
TVHSQKQAVKTEVSAAKEIERPRKSNVKDNKKKEKRKSFVGAFIGLFGSKEKKNKSSERNNNVKSKEGSKSFLKRKQ
KQKASKKTISKPQAIDKSGESNKIADNNLNIAERNNKSSERHEDKIEDSEQQRVSSVIRNKDYYSSKDNQPATVEKA
EMVNKRTDNKVIELKKSEILTKEVDGIQNNSKISTKDIFDKSFNDDLFGNYKVGNDTTVDKVKEQNNQDSNPNTENL
KQNTSSKHVPIQTTIPWAKDIDTLTFALNITKEDNNSKLVDTDETSKNEIHTEKAIEQTSKSDSKIGTIVDKDNEDA
ILVEKNIRMKDQSIFPVVPKLIQKNIDFSRSASYKELFEITRLVLDELDAQYNWKRLKQEASLSENHVISASFGSTE
DLVETEIESKKDIDNVTPEVVKSSDSNKILKISDIPILEPDSSLDNSEITKDFLNDLENIVDDVLKSNDFKKQSPSI
QNQHKVFKVIKTTSPTMVNKNFDKKSSTDSNSDARSNEIPKKLLPSVDLVPGIKLEDEEEEEDQYLTPSAEPSAVFE
QDMENETSKDSSDESTVNDTNKSVRAEILEDLNFSSFHMSFDDSKENEVENSVNLGANSELLNQLTKKNDTFQSPFA
GQAPLIYRLPKSSPIKNTANNISGAKKLNRDDNDAIWVSPSKIPFYDFNGKPESNVQGKKSKAAVQEQLTNLQISGN
SNMNKDISYTYLAGFIDSPDLGESNSNRLKFKE

XP_037143118.1 uncharacterized protein HG535_0B04320 [*Zygotorulaspora mrakii*]

MLERRENSLEMTAADLSSAYLHEVRHEMGHADPDAATDIILFGNLLEHEVERRQWVKDRTPKEWIQQVFRDSAVFHG
IDELIWGEFLVCIYKDPITNMISALFIDKLGVTHFNPVDITNKSQFYPAIENLTDNDKDSNVKKSIAVSLLRKYAQL
SSKDVSRLQPDGLRFNYDPTHAGDLASRCKLVESCAPETMGTRILSQGSLQRRCNKSILLDVIYERKESLTEINNKL
VFHLGEQLEQLFNPLTEYSPEQTEYTYKPPDYDPMSHRTSSKVVSIIDELLQLQTNFTFSLVEFLQGFLISLRIKVL
NEQIKDLSTVKLNRLFPPTIDEVTRINCIFLDSLKAATPFGALEILKACSATIPYFYKAYTRHEAATKNFVKNFKLF
LKSFEYCIPRDNGYSNMKIESIINGPQEKLVKLKLIIDRLYRTEEWPPNEQEEAKSNFDNIALAIDSFGKMKSPLSS
YNTRVFTPSGKILTELAKGWPEELQYKWLKRRVVGVFDVVGNDENKERQVLVIFSDYIVFLEVVNGSQYYSGTNNNK
PKIADILMNSLINEVPLPTRIPKLKVHNYCYIDDIIVTELENQSLRFDAYRGVNSFSVTCRLSTKSASASTVAGLVT
KAKILEKETAFHLFKANLEENEIYSTAHELEAYKNETLKSKFGLFLNMNPSRQVLAGSRSHVAIFASFISDMDASDV
QLTILTSDGRENASVVSPEDMILRIFEVLKHEVPICYSSLYSPLLPALLKINEQIVHKVLYSADPSDKFTNNTEGDA
LTFTQDHEKKKSFGTITTFRSYTSDYKDVESNQSRVSRSGTAKNETLGVEKSHKIEVKAPKAINSKNKGRINSKKRK
SIIDVFKGLFEIKRSGKKEADIAGTVKISKPIVKRGTNERLPKKKAASQEVKDPLKQREHKTGGDRITSVIKHKEFA
SNEGIDIASQNTSKNEPNEHIRNNILKEQKSFESISRDPAEKSSKDRSIATAQDEDSTAQGPSGVAQATSLRESVHS
ILSQTLKLPETPESRKVGDCAAPNGVQIHSAGQNGFYRGQSRQSQLFNDDLYGELAVEAAKEKSSDSSAETVSGHNT
KEVDAQKNHPDTTAPIQIIESPTIKQITPSEQIGSEKRISTVENFEVNISQYDGKETTEKNRIFPIIPTLDPPKLGF
ERSTSFVELFRGMRLILDENDAQYNWKRLSSESSLNEKQAFIAPIQNAHIFKTLAQDPDRRLGPPQEGNQIEMVKIP
SSPTGKLSTEDSSGIPDLANDTSARKLNNSNQSVVLQNIPGLPLDDSESKLNTQPKPSFLAAPNFKVLRTSPTRIVN
KLTKLHEVNRDSLNYDISLPLDLDMVSNKRWVELELPSQEDLQGDKFYTPAEQHNSEFPKEALEIDDHSTASEFAKE
ANETSQETSIEDSSKLIDAEDLLEDMEFSSFHMTFDASVSKDDISNYTSSPVQHDNNVNPQCPTTRHVRQEPVLYRL
PKDIFSTSRTISQLNDGGIGERIGHGRNPSLDDDPIWVSPSKIDFYNLSKDSEESTMANRKAAALLEYPYDNKTYEH
HNEDTNSLRELSYAYLASFVGSDQESERFDDKPVRLHFCD

References

1. Finnigan GC, Takagi J, Cho C, Thorner J. Comprehensive genetic analysis of paralogous terminal septin subunits Shs1 and Cdc11 in *Saccharomyces cerevisiae*. *Genetics* 2015; 200(3): 821-841 DOI: 10.1534/genetics.115.176495
2. Sikorski RS, Hieter P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* 1989; 122(1): 19-27 DOI: 10.1093/genetics/122.1.19
3. Andersen MH, Graversen H, Fedosov SN, Petersen TE, Rasmussen JT. Functional analyses of two cellular binding domains of bovine lactadherin. *Biochemistry* 2000; 39(20): 6200-6206 DOI: 10.1021/bi992221r
4. Shao C, Novakovic VA, Head JF, Seaton BA, Gilbert GE. Crystal structure of lactadherin C2 domain at 1.7A resolution with mutational and computational analyses of its membrane-binding motif. *J Biol Chem* 2008; 283(11): 7230-7241 DOI: 10.1074/jbc.M705195200
5. Goldstein AL, McCusker JH. Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae*. *Yeast* 1999; 15(14): 1541-1553 DOI: 10.1002/(SICI)1097-0061(199910)15:14<1541::AID-YEA476>3.0.CO;2-K

Appendix H - Reconstructed evolutionary history of the yeast septins Cdc11 and Shs1

This chapter has been published in *Genes, Genomes, Genetics (Bethesda)*.

Julie Takagi[1,2]*, Christina Cho[1,3]*, Angela Duvalyan[1,4]*, Yao Yan[5], Megan Halloran[5,6], Victor Hanson-Smith[7], Jeremy Thorner[1], and Gregory C. Finnigan[5†]

[1]Department of Molecular and Cell Biology,
University of California, Berkeley, CA 94720-3202; and,
[5]Department of Biochemistry and Molecular Biophysics,
Kansas State University, Manhattan, KS 66506
Running Title: Septin gene duplication and divergence

[2]Current Address: Dept. of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139 USA
[3]Current Address: Endodontics Residency Program, College of Dental Medicine, Columbia University, New York, NY 10032 USA
[4]Current Address: Dept. of Internal Medicine, Univ. of Texas Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas TX 75390 USA
[6]Current Address: Dept. of Psychology, University of Kentucky, Lexington, KY 40506 USA
[7]Current Address: Verge Genomics, Inc., South San Francisco, CA 94080 USA

*These authors contributed equally to this work.

[†]Correspondence to:
Gregory C. Finnigan
Dept. of Biochemistry & Molecular Biophysics
Kansas State University
141 Chalmers Hall, 1711 Claflin Rd.
Manhattan, KS 66506 USA
Phone: (785) 532-6939; FAX; (785) 532-7278;
E-mail: gfinnigan@ksu.edu

Abstract

Septins are GTP-binding proteins conserved across metazoans. They can polymerize into extended filaments and, hence, are considered a component of the cytoskeleton. The number of individual septins varies across the tree of life—yeast (*Saccharomyces cerevisiae*) has seven distinct subunits, a nematode (*C. elegans*) two, and humans thirteen. However, the overall geometric unit (an apolar hetero-octameric protomer and filaments assembled there from) has been conserved. To understand septin evolutionary variation, we focused on a related pair of yeast subunits (Cdc11 and Shs1) that appear to have arisen from gene duplication within the fungal clade. Either Cdc11 or Shs1 occupy the terminal position within a hetero-octamer, yet Cdc11 is essential for septin function and cell viability, whereas Shs1 is not. To discern the molecular basis of this divergence, we utilized ancestral gene reconstruction (AGR) to predict, synthesize, and experimentally examine the most recent common ancestor ("Anc.11-S") of Cdc11 and Shs1. Anc.11-S was able to occupy the terminal position within an octamer, just like the modern subunits. Although Anc.11-S supplied many of the known functions of Cdc11, it was unable to replace the distinct function(s) of Shs1. To further evaluate the history of Shs1, additional intermediates along a proposed trajectory from Anc.11-S to yeast Shs1 were generated and tested. We demonstrate that multiple events contributed to the current properties of Shs1: (i) loss of Shs1-Shs1 self-association early after duplication; (ii) co-evolution of heterotypic Cdc11-Shs1 interaction between neighboring hetero-octamers; and, (iii) eventual repurposing and acquisition of novel function(s) for its C-terminal extension domain (CTE). Thus, a pair of duplicated proteins, despite constraints imposed by assembly into a highly conserved multi-subunit structure, could evolve new functionality via a complex evolutionary pathway.

Abbreviations

AA, amino acids; AGR, ancestral gene reconstruction; Anc, ancestral subunit; CC, coiled coil; CTE, C-terminal extension; NTE, N-terminal extension; DIC, differential interference contrast; 5-FOA, 5-fluoroorotic acid; GAL, galactose; GFP, green fluorescent protein; mCherry, monomeric red fluorescent protein derivative; PCR, polymerase chain reaction; PP, posterior probability; RAF/SUC, raffinose and sucrose; SD, synthetic drop-out medium with dextrose; YPD, yeast extract peptone dextrose; yeast species [*C.g.*, *Candida glabrata*; *C.a.*, *Candida albicans*; *A.g.*, *Ashbya gossypii*/*Eremothecium gossypii*; and, *S.c.*, *Saccharomyces cerevisiae*]