(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
16 January 2020 (16.01.2020)

WIPO | PCT

(10) International Publication Number
WO 2020/014570 A1

(51) International Patent Classification:

C12N 1/04 (2006.01)	*C12N 15/09* (2006.01)
C12N 1/15 (2006.01)	*C12N 15/10* (2006.01)
C12N 1/16 (2006.01)	*C12N 15/11* (2006.01)
C12N 1/19 (2006.01)	*C12N 15/63* (2006.01)
C12N 5/04 (2006.01)	*C12N 15/79* (2006.01)
C12N 9/22 (2006.01)	*C12N 15/90* (2006.01)

(21) International Application Number:
PCT/US2019/041538

(22) International Filing Date:
12 July 2019 (12.07.2019)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/697,855 13 July 2018 (13.07.2018) US

(71) Applicant: **KANSAS STATE UNIVERSITY RESEARCH FOUNDATION** [US/US]; 2005 Research Park Circle, Suite 105, Manhattan, Kansas 66502 (US).

(72) Inventor: **FINNIGAN, Gregory**; 8900 Butterfly Lane, Manhattan, Kansas 66502 (US).

(74) Agent: **COOK, Crissa A.**; 10801 Mastin Blvd., Suite 1000, Hovey Williams LLP, Overland Park, Kansas 66210 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA,

(54) Title: MULTI-LOCUS GENE DRIVE SYSTEM

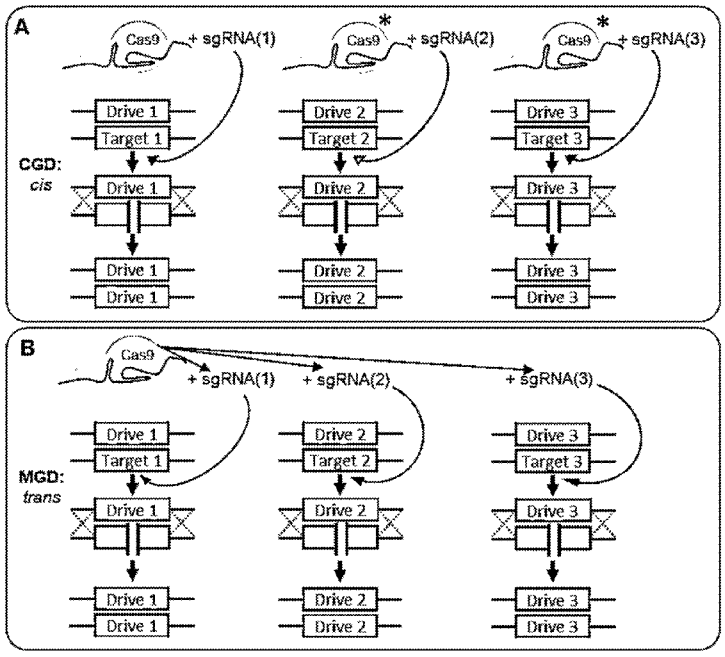


FIG. 1

(57) Abstract: Multi-locus CRISPR-Cas gene drive systems, methods, and host cells involving stable integration of a CRISPR-Cas gene drive system into a diploid eukaryotic cell genome at multiple loci. The gene drive system comprises a plurality of distinct gene drive constructs, wherein at least one gene drive construct (and preferably only one gene drive construct within the system) encodes for a functional CRISPR nuclease, along with two or more distinct gene drive constructs encoding respective sgRNA. Each of the gene drive constructs are configured for integration at separate distinct loci on the genome that are nonadjacent to one another, and preferably in remote positions of the genome, even on different chromosomes. As such, the functional CRISPR nuclease operates *in trans* with the sgRNA to drive stable integration of the drive elements at multiple loci in the genome simultaneously.

WO 2020/014570 A1

[Continued on next page]

SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

— *with international search report (Art. 21(3))*

— *with sequence listing part of description (Rule 5.2(a))*

MULTI-LOCUS GENE DRIVE SYSTEM

CROSS-REFERENCE TO RELATED APPLICATIONS

The present application claims the priority benefit of U.S. Provisional Patent Application Serial No. 62/697,855, filed July 13, 2018, entitled MULTI-LOCUS GENE DRIVE SYSTEM, incorporated by reference in its entirety herein.

FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

This invention was made with government support under Grant Number P20 GM103418 awarded by the National Institutes of Health, and Hatch Project 1013520 awarded by the United States Department of Agriculture. The government has certain rights in the invention.

SEQUENCE LISTING

The following application contains a sequence listing in computer readable format (CRF), submitted as a text file in ASCII format entitled "Sequence Listing," created on July 11, 2019, as 13KB. The content of the CRF is hereby incorporated by reference.

BACKGROUND

Field of the Invention

The present invention relates to multi-locus CRISPR-based gene drive systems.

Description of Related Art

The CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) gene editing system has emerged with the potential to revolutionize industries including basic research, biomedical application, and uses in agriculture. The discovery and rapid expansion of CRISPR-based gene editing across all fields of molecular biology has led to many unique applications of this biotechnology. These range from “traditional” editing (deletion, modification, or replacement of genomic loci), to genome-wide screens, to repurposing of dCas9 to modulate gene transcription, to imaging of chromosome dynamics in real-time.

Arguably one of the most powerful (and ethically concerning) uses of CRISPR is within the nuclease-based “gene drive.” CRISPR-associated protein-9 nuclease (Cas9) from *Streptococcus pyogenes* is a particularly well-studied arrangement of this system, and was pioneered by Esvelt et al. (WO 2015/105928, filed January 8, 2015, incorporated by reference

herein in its entirety). The gene drive consists of the Cas9 nuclease, a single-guide RNA, optional "cargo" DNA, and homology arms flanking the CRISPR expression cassette that match the two genomic sequences immediately adjacent to either side of the target cut site. The drive can be introduced via exogenous insertion and/or propagation after sexual reproduction. An expression vector comprising the gene drive is introduced into the cell, where the Cas9 and sgRNA gene products combined to induce a double strand break at the target site, and inducing the cell to repair the damage by copying the drive sequence onto the damaged chromosome, using the driving-containing expression vector as the template. Once the gene drive has been inserted, it will self-propagate. That is, the RNA-loaded Cas9 expressed from the mutated allele again combine targets the chromosomal DNA and creates a double stranded break (DSB) at the corresponding position on the wild-type allele, which is then repaired via (HDR) using the intact chromosome carrying the gene-drive element as a template.

Thus, within a diploid cell, the homologous chromosome pair serves as the DNA "donor" and, via homology directed repair (HDR), will copy the entire gene drive (Cas9 and its sgRNA) to the second chromosome, modifying the endogenous sequence in the process. As a result, the gene drive components and any "cargo" are copied into the homologous chromosome. Example modifications include inserting or installing the gene drive "cargo" via the DSB at the desired genomic location, or even deleting and replacing the entire endogenous copy of the gene. However, the action of cutting and copying of Cas9 and guide RNA remain the same for both versions. In any event, this process "forces" the heterozygous individual into the homozygous state, ensuring rapid propagation through a population if the transgenic organisms harboring the gene drive are released into the wild. The technology allows targeted bypass and editing of genetic material (in the form or one or more genes). The CRISPR-based gene drive bypasses traditional genetic actions and "forces/drives" a predetermined genetic element into a population at a very high rate of speed and penetrance. Indeed, early studies in the laboratory have demonstrated the unnatural power of a gene drive system to impose strong selection and over a 95% reduction in a population in only a few generations.

CRISPR-based systems have been proposed as a technique to control pest populations, and are currently being studied in model systems including mosquitos for their role as hosts for pathogens including malaria. Control of biological populations is an ongoing challenge in many fields including agriculture, biodiversity, ecological preservation, pest control, and the spread of disease. In some cases, such as insects that harbor human pathogens (e.g. malaria), elimination or reduction of a small number of species would have a dramatic impact across the globe. Destruction

of critical genes can cause a reduction in organism breeding, spread, and population size. This can cause a dramatic decrease in population levels of key insects, pests, or pathogens in wild populations. Given the recent discovery and development of the CRISPR/Cas9 gene editing technology, a unique arrangement of this system—a nuclease based "gene drive"—allows for the Super-Mendelian spread and forced propagation of a genetic element through a population. Recent studies have demonstrated the ability of a gene drive to rapidly spread within and nearly eliminate insect populations in a laboratory setting.

However, there are logistical, technical, and ethical considerations for utilizing this very new technology in real applications. Recent studies have demonstrated the ability of a gene drive to rapidly spread within and nearly eliminate insect populations in a laboratory setting. Further, given the incredibly strong selective pressure on a population, insect species have been shown to evolve *resistance* to the gene drive. In addition, wild populations may naturally have a diverse set of polymorphisms within a given gene target (and thus provide a "native" source to evade a gene drive). Moreover, there is ongoing competition between HR-based propagation of the drive and repair of the double-stranded break via non-homologous end-joining.

Before deployment of this powerful biological agent, we must consider mechanisms to control the unintentional, accidental, or malicious introduction of CRISPR-based genetic drive systems into native ecological systems, and potential unintended consequences of extinction-level population control of a given species, as well as naturally evolved resistance to gene drives that may render the CRISPR approach ineffective in the future.

While there are still ongoing technical challenges to design of a more optimal gene drive to be used in wild populations, there are also serious ecological and ethical concerns surrounding the nature of this powerful biological agent.

SUMMARY

Described herein are new methods for stable integration of a CRISPR-Cas gene drive system into a diploid eukaryotic cell genome at multiple loci (multi-locus CRISPR-Cas gene drive system). The method generally comprises introducing into the eukaryotic cell genome at a first locus, a first gene drive construct comprising a first nuclease nucleotide sequence encoding a functional CRISPR nuclease that induces a double-stranded break in or near a target site; and introducing into the eukaryotic cell genome at a plurality of loci remote from the first locus, a plurality of secondary gene drive constructs each comprising a guide nucleotide sequence encoding a single guide RNA (sgRNA) sequence complementary to respective endogenous target

sites in said cell genome, wherein the constructs are expressed in said cell to produce functional CRISPR nuclease and a plurality of sgRNA. Advantageously, the single CRISPR nuclease co-localizes with respective sgRNA in the cell at each respective target site, and induces a double-stranded break at each target site, wherein homology-directed repair integrates each of the first gene drive construct and plurality of secondary gene drive constructs into their respective target sites of the cell genome.

Also described herein are eukaryotic host cells comprising the inventive CRISPR-Cas gene drive system stably integrated into its genome at multiple loci. The modified cells comprise at least a first gene drive construct integrated at a first locus, which comprises a first nuclease nucleotide sequence encoding a functional CRISPR nuclease that induces a double-stranded break in or near a target site, and a plurality of secondary gene drive constructs integrated at a plurality of loci remote from the first locus, which each comprises a guide nucleotide sequence encoding a single guide RNA (sgRNA) sequence complementary to respective endogenous target sites in the cell genome. These constructs are each expressed in the cell to produce functional CRISPR nuclease and a plurality of sgRNA. Advantageously, the single CRISPR nuclease co-localizes with respective sgRNA in the cell at each respective target site, such that the CRISPR nuclease induces a double-stranded break at each target site, wherein homology-directed repair integrates each of gene drive constructs into their respective target sites of the cell genome. As such, the host cell becomes homozygous for the CRISPR-Cas gene drive system at each locus, which biases its propagation through the population.

Accordingly, the technology described herein is particularly suite for targeted inhibition, suppression, or extinction of a population, comprising releasing a CRISPR-Cas gene drive system into a targeted population, which is configured for stable integration into a eukaryotic diploid cell genome at multiple loci, according to various embodiments of the invention.

Advantages of the technology are exemplified by data demonstrating an *S. pyogenes* Cas9-based gene drive system installed at a single locus in the yeast genome (HIS3), as well as a second (SHS1) and third (DNL4) installed "minimal" drive system (sgRNA-expression cassette only). Further, the technology demonstrates application of this drive system to achieve either (a) full disruption of DNA Ligase IV (yeast Dnl4) or (b) inclusion of partial loss of function alleles of the enzyme as part of the multi-locus gene drive system. It will be appreciated that this same technique could be applied to target, delete, or partially replace a wide variety of endogenous sequences. Further, gene drives can be used to add exogenous DNA "cargo" from other species (native genes or artificially programmed DNA). Thus, one can not only simultaneously modify/delete the native

DNA, but also add useful exogenous genes or pathways that do not exist endogenously.

It will be appreciated that budding yeast is used herein as a safe and fully-contained model system to demonstrate mechanisms that allow for programmed regulation of gene drive activity. However, the technology is in no way limited to yeast and can be applied to any eukaryotic diploid or polyploid cell system. The data exemplifies development and design of a three-part gene drive system that can propagate three separate genetic elements at three separate chromosomal positions using a single copy of Cas9. The data and examples further demonstrate utility of such a multi-locus gene drive setup (rather than all installed at a single locus) by modifying an essential component of non-homologous end joining (NHEJ) repair systems, DNA Ligase IV.

These results demonstrate for the first time that a CRISPR gene drive can be split into separate genetic elements located on different chromosomes and that they can be simultaneously propagated with super-Mendelian inheritance, offering flexibility and safety while functioning as a full-GD.

BRIEF DESCRIPTION OF FIGURES

Fig. 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 (nuclease 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.

Fig. 2A: Design of a CRISPR/Cas9-based gene drive system in *S. cerevisiae* across three loci, and illustration of multi-locus artificial drive system.

Fig. 2B: Image of culture plates for haploid yeast harboring the triple drive mated to the triple target strain to form diploids.

Fig. 2C: A time course of galactose activation.

Fig. 2D: Table of presence of target.

Fig. 2E: Graph of gene drive activity.

Fig. 2F: PCR results of gene drive activity.

Fig. 3A: Phylogenic analysis of Ligase IV candidates across fungi and metazoans. Branch lengths correspond to the number of substitutions per site and the confidence of most branches is

illustrated as a decimal (red text).

Fig. 3B: 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), and multiple sequence alignment performed using Clustal Omega 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. The position of six alleles (K742, T744, L750, D800, G868, and G869) are also illustrated (red asterisk) that were identified from a previous study.

Fig. 3C: 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 and illustrated using Chimera.

Fig. 3D: Cas9-based genomic integration methodology for introduction of mutational substitutions to the native *DNL4*locus in yeast.

Fig. 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 KanR 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 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 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 Fig 11. (**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.

Fig. 5A: Methodology for construction of a CRISPR gene drive system across three loci.

Fig. 5B: Methodology for construction of a CRISPR gene drive system across three loci.

Fig. 6: Analysis of haploid genomes of the triple drive and triple target yeast strains by diagnostic PCRs.

Fig. 7: Re-activation of gene drive system in clonal isolates displaying incomplete initial drive activity.

Fig. 8: Conservation of the N-terminal domain of DNA Ligase IV.

Fig. 9: Further analysis of clonal isolates from the triple gene drive harboring dnl4 replacement alleles.

Fig. 10: 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.

Fig. 11: Original images of DNA agarose gels used for PCRs from Fig. 2F and Fig. 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. 2F. (**D**) PCRs from Fig. 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-GelTM Imager (ThermoFisher Scientific), cropped for clarity, but were not processed by any method.

Fig. 12: Original DNA gels. 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. 11.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

In one aspect, a multi-locus CRISPR-based gene drive system is described herein. The gene drive system is configured for integration into a diploid eukaryotic cell genome at two or more (and preferably three or more) target sites. Various techniques can be used to introduce

these foreign nucleic acid constructs into the target organisms, such as viral-based integration, transposon-based integration, bacterial-based integration (e.g., *Agrobacterium* in plants), or in fungi/yeast, HDR-based integration using added synthetic DNA. Further, it will be appreciated that CRISPR-based systems can themselves be used to create the CRISPR-based gene drive. This will vary based on the organism of interest. Moreover, although sequential insertion techniques, as exemplified in the working examples, can be used to introduce the foreign nucleic acid constructs, other approaches could be used, depending upon the organism. For example, depending on the exact location of where the drives (2, 3, or 4, etc.) are located, it could be possible to insert them into the genome in fewer steps (e.g., as combined steps). In theory, all the intended gene drive DNA could be inserted at once using CRISPR editing itself, or included as synthetic chromosome fragments, etc. Those skilled in the art will appreciate that there are many strategies to install the initial gene drive(s).

The system comprises a plurality of multi-locus drive constructs targeting more than one genetic locus (chromosomal location) acting as a "gene drive" in a single yeast diploid cell. This has never been accomplished before. The working examples illustrate three loci at the same time. This approach could theoretically extend to n-number of positions within a genome. Described herein are two general techniques for a multi-locus gene drive system across distinct chromosomal positions. In one or more embodiments, the multi-locus gene drive system is a "Complete" Gene Drive (CGD) comprising both a nuclease and corresponding guide RNA for each locus. Preferably, the multi-locus gene drive system is a multi-locus "Minimal" Gene Drive (MGD) comprising a single nuclease and sgRNA, and all other genetic loci would only contain additional guide RNA cassettes (without additional corresponding nucleases).

Thus, preferred embodiments of the multi-locus gene drive system comprise a plurality of distinct gene drive constructs, wherein at least one gene drive construct (and preferably only one gene drive construct within the system) comprises a "nuclease nucleotide sequence" encoding for a functional CRISPR nuclease that induces a double-stranded break in or near the target sites, and a plurality of distinct gene drive constructs each comprising a "guide nucleotide sequence" encoding a single guide RNA sequence complementary to respective endogenous target sites. Each of the gene drive constructs are configured for integration at separate distinct loci on the genome that are nonadjacent to one another, and preferably in remote positions of the genome, even on different chromosomes. As such, the functional CRISPR nuclease operates *in trans* (i.e., at a distance) with the sgRNA to drive stable integration of the drive elements at multiple loci in the genome simultaneously.

In one or more embodiments, the multi-locus gene drive system comprises a first gene drive construct comprising a first guide nucleotide sequence encoding a single guide RNA sequence complementary to a first endogenous target site and located on the expression cassette between a first pair of endogenous flanking sequences homologous to sequences adjacent the first target site of integration. In one or more embodiments, the first gene drive construct further comprises a first nuclease nucleotide sequence encoding for a functional CRISPR nuclease that induces a double-stranded break in or near the first target site. In such embodiments, the sgRNA and CRISPR nuclease coding sequences are both located on the same expression cassette between the first pair of flanking sequences in the construct. The first nuclease nucleotide sequence encoding for a functional CRISPR nuclease may alternatively be introduced via a separate expression cassette. The multi-locus gene drive system further comprises a second, distinct gene drive construct comprising a second guide nucleotide sequence encoding a single guide RNA sequence complementary to a second endogenous target site, and a second pair of endogenous flanking sequences homologous to sequences adjacent a second target site of integration. In the second gene drive construct, the sgRNA is located between the pair of flanking sequences on the same expression cassette. In one or more embodiments, the second gene drive construct does not include a nuclease-encoding nucleotide sequence. Further, the second gene drive construct is preferably a distinct expression cassette from the first gene drive construct. The multi-locus gene drive system further comprises a third gene drive construct comprising a third guide nucleotide sequence encoding a single guide RNA sequence complementary to a third endogenous target site, and a third pair of endogenous flanking sequences homologous to sequences adjacent the third target site of integration. In the third gene drive construct, the sgRNA is located between the third pair of flanking sequences on the same expression cassette. In one or more embodiments, the third gene drive construct does not include a nuclease-encoding nucleotide sequence. Further, the third gene drive construct is a distinct expression cassette from each of the first and second gene drive constructs.

As noted, the system can extend to n-number of additional gene drive constructs comprising sgRNA and flanking sequences for n-number of endogenous target sites, each driven by the CRISPR enzyme from the first gene drive construct or from a separately introduced expression cassette. It will be appreciated that any one of the foregoing gene drive constructs may further comprise optional donor DNA for integration at any one of the target sites. It will also be appreciated that each guide nucleotide sequence may encode for more than one sgRNA, likewise, the nuclease nucleotide sequence may encode for more than one functional CRISPR nuclease.

Further, if desired, although not required, the second and/or third (and so one) gene drive constructs may include a nuclease-encoding nucleotide sequence if desired.

In the expression cassettes, the sgRNA and/or CRISPR enzyme sequences may be operably linked to one or more respective regulatory elements, e.g., promoters. In one or more embodiments, the constructs include only a single promoter driving nuclease expression. The CRISPR enzyme is preferably a nuclease, and preferably a type II CRISPR system enzyme, and more preferably a Cas9 enzyme. There are many orthologs of Cas9 that have been developed today, including SpCas9, SaCas9, NmCas9, and mutated versions of Cas9 ("enhanced" Cas9 that has several mutations added artificially), to artificially programmed "minimal" CasX. Exemplary Cas9 enzymes include *S. pyogenes* (SEQ ID NO:9), *S. aureus*, *S. pneumoniae*, *S. thermophilus*, or *N. meningitidis* Cas9 nucleases, as well as Cas9 orthologs and mutant/variant Cas9 derived from these organisms, so long as the mutant remains functional (i.e., can bind and cleave the target nucleotide sequence). As noted above, alternative CRISPR nucleases that could be used in the CRISPR system include type V (Cas12a, previously called Cpf1), as well as type I CRISPR/Cas3 "shredder," and functional mutants and derivatives of Cas9 that otherwise maintain DNA nuclease activity.

The sgRNA used in each construct can be any short nucleotide sequence (i.e. less than 100 nt, preferably less than 50 nts, more preferably less than 25 nts, even more preferably less than 20 nts) that is complementary to the target sequence and therefore capable of hybridizing to the target sequence of a DNA molecule in a eukaryotic cell for the desired genetic modification. Various sgRNA expression cassettes have been established and are known in the art for various target organisms.

The CRISPR enzyme (e.g., Cas9) expressed from one loci in the genome interacts with each sgRNA (via noncovalent binding) to form a respective ribonucleoprotein (RNP) complex and thereby directs sequence-specific binding of respective CRISPR complexes to each target sequence via base pairing of each sgRNA and its target genomic DNA sequence. The sgRNA preferably contains the pre-programmed CRISPR RNA (crRNA) and tracer RNA (tracrRNA) which is specific to the particular species of enzyme used. For S. *pyogenes* Cas9, a particular tracrRNA is used, but can include modified forms of tracrRNA (e.g., different size, position of stem-loop, chemical modifications, etc.). Generally, the tracrRNA is required either fused to the crRNA (a "single" guide), or expressed as its own piece of RNA. Thus, the "CRISPR complex" comprises the CRISPR enzyme complexed with the sgRNA that is hybridized to the target sequence. Examples of various CRISPR enzymes and sgRNA systems can be found in the art.

For example, the crRNA may change length, and the tracrRNA can also change length or include other modifications but, in general, it is the "corresponding paired RNA sequence" with the exact species of interest. Thus, for SpCas9, one would use the *S. pyogenes* guide RNA. The pioneering work in CRISPR coined the phrase "single guide" RNA, which engineered the two halves crRNA + tracrRNA into one "sgRNA" widely used now. By combining tracrRNA and crRNA into a single synthetic gRNA, this resulting one-component system not only simplifies the experimental design, it also yields equal or greater guiding efficiency. Different constructs of sgRNA have been designed and tested, each with different lengths of tracrRNA sequence at their 3' end, with the editing rate increasing as the proportion of the original tracerRNA sequence increases in the overall sgRNA design. Accordingly, even with *S. pyogenes*, there are numerous variations on SpCas9 guide RNAs that have been published. Those skilled in the art recognize the fundamental feature is that the selected nuclease pairs with the corresponding guide RNA (modified or not).

Once integrated, the sgRNA and CRISPR nuclease are then expressed from their respective integrated gene drive constructs at their respective loci, and the sgRNA targets the resulting CRISPR complex to the corresponding loci on the "wild-type" (or endogenous) copy of the target DNA on the homologous chromosome in a diploid cell, usually although not necessarily, at the same site of insertion of the original gene drive construct. In the context of the invention, the CRISPR enzyme expressed from the first gene drive construct (or from its own separate construct) is recruited to generate distinct, respective CRISPR complexes with each of the first, second, third (and so on) sgRNA expression products, thereby targeting each of the first, second, third (and so on) endogenous target sites.

Advantageously, the multiple endogenous target sites can be positioned in a variety of different locations within the genome, without departing from the scope of the invention. For example, endogenous target sites can be selected at different positions on the same chromosome arm, on different arms of the same chromosome, or even on different chromosomes in the genome. Further, it will be appreciated that the associated gene products are not necessarily limited to targeting their own loci, but instead could target each other, or have redundancy in their targeting. For example, the first gene drive construct could target the second endogenous site, the second gene drive construct could target the first endogenous site, and the third gene drive construct could target the third endogenous site. Alternatively, the second gene drive construct could target both the first and second endogenous sites, etc., and various combinations of the foregoing, provided that each of the gene drive constructs are copied over to subsequent chromosomes. That is, position and targeting of the guide RNAs can be fluid and can be re-arranged in many

combinations. The examples demonstrate guide 1 designed for target 1, guide 2 for target 2, guide 3 for target 3, etc. This could be altered so guide 1 is designed for target 3, guide 2 for target 4, guide 4 for target 2, target 3, *and* target 4, etc. There can also be more than one guide at each location, such as one drive with three guides, or one drive with ten guides. In this "cross-targeting" approach, the main requirement is that the system must contain one guide RNA to target each drive for the effect to act as a drive. Any single location can have n-number of guides. Any combinations after that are possible. Including multiplexing Drive 1 to Drive 1, but with 10 guide RNAs. Further, it is possible to utilize multiple nucleases across drives. The examples demonstrate one nuclease for simplicity, safety, and experimental "efficiency." However, one could add in a second or third copy of the same nuclease or different nuclease (as long as it can be paired with its proper guide RNA). This could add redundancy, or targeting or many separate gene targets, and/or allow for increased multiplexing or programmed self-regulation. It will be appreciated that the particular advantage of the invention is the ability to use a CRISPR enzyme expressed in a separate and distinct part of the genome to drive recombination of sgRNA (and optional donor DNA) sequences in multiple loci in the genome. That is, the basic principle for more than one drive is that each "drive" (position in genome) just needs a minimum of a single cut within a target (via one guide RNA) to allow for copying of the drive construct, and super-Mendelian inheritance (to work as a "drive."). This is true for all intended positions/drives. This can be accomplished by any number of loci and any number of guides and/or nucleases. There is no restriction (as far as our data shows) based on where these are, how many times you cut, and how many nucleases are possible.

In practice, the CRISPR enzyme in the bound CRISPR complex cleaves both strands in or near the target sequence in the eukaryotic cell, such that repair of the targeted gene is activated within the given cell. In addition to the sgRNA, site-specific Cas9 activity can be targeted by presence of a protospacer adjacent motif (PAM) sequence adjacent to the target DNA sequence. That is, targeting specificity of the CRISPR system is determined by the nucleotide sequence at the 5' end of the sgRNA. In the *S. pyogenes* CRISPR Cas9 system, the desired target sequence must immediately precede the PAM sequence, such that editing will not occur at any location other than the one at which the Cas9 in the complex recognizes the PAM. PAM sequences are selected to correspond to the particular Cas9 species selected for the complex. The conventional PAM the sequence for *S. pyogenes* is 5'-NGG-3' where "N" is any nucleotide, and in the Examples GGG was typically used. However, different PAMs are associated with different Cas9 proteins (e.g., 5'-NGA-3', 5'-YG-3'), and attempts have also been made to engineer Cas9 to recognize different

PAMs. In addition, other nucleases, such as Cpf1 have been shown to recognize other PAM sequences (e.g., 5'-YTN-3'). In some cases, the PAM sequence is present at the designated position in the genome prior to integration of the CRISPR complex, such that the gene drive is inserted next to the PAM. In some cases, the PAM sequence is part of the gene drive package, such that both the PAM sequence and gene drive sequence are inserted into the genome.

The foregoing approaches can be used in methods of altering expression of two or more gene products from a target nucleotide sequence in a diploid eukaryotic cell, *in vitro*, *in vivo*, or *ex vivo*. Embodiments of the invention can be used for general research and study related to a variety of gene editing technologies, including therapeutic approaches in humans and animals, agricultural technology (both plant and pest), and the like. The methods generally comprise introducing into the eukaryotic cell the described gene drive system according to the various embodiments described herein under conditions where each gene drive construct is expressed in the cell to produce functional CRISPR nuclease and multiple sgRNAs, which co-localize in the cell at the endogenous target sites, so that the CRISPR nuclease can induce a double-stranded break at each site, wherein homology-directed repair mediated by the flanking sequences integrates each gene drive construct into (or in replacement of) the target site. The CRISPR-Cas gene drive constructs or components can be introduced into the cell as a nucleic acid construct encoding the components for expression in the cell. Further, transgenic organisms harboring copies of the gene drive system can be introduced into the wild type population for mating and propagation of the gene drive system through the population. Methods of the invention may involve activating regulatory elements, such as inducible promoters to induce expression of the construct. The components can also be introduced as preassembled proteins or RNP complexes.

Thus, aspects of the invention concern novel design of a multi-locus CRISPR gene drive (e.g., triple system with three separate genetic loci), exemplified by *S. pyogenes* Cas9, which can operate *in trans* and allow for propagation of three (or more) genetic elements in a diploid cell. In one or more embodiments, one of the targets can include a highly conserved molecular component of non-homologous end joining DNA repair systems, DNA Ligase IV (yeast, "Dnl4") and we demonstrate the ability of a multi-locus gene drive system to disrupt or install modified alleles of the Ligase IV. Also described herein are methods of integrating the multi-gene drive into a eukaryotic cell genome at a target site. The methods generally comprise introducing into the eukaryotic cell a multi-CRISPR-Cas gene drive system according to the various embodiments described herein under conditions where the gene drive is expressed in the cell. Embodiments of the invention are also concerned with eukaryotic host cells comprising a multi-CRISPR-Cas gene

drive system or multi-gene editing system according to the various embodiments described herein. Organisms comprising such modified or altered (engineered) cells are also contemplated.

This technology could be applied to genetically modified insects, plants, animals, or fungi harboring a multi-locus gene drive. Further an "Anti-gene drive" drive-containing species (used to safeguard use of the initial drive) can be developed.

In comparison to previous approaches to modulating gene drive activity with four independent approaches, this "multi-drive" system presents a novel arrangement and the ability to install more than one gene drive in a single genome. The working example of modifying DNA Ligase IV also pilots the idea of suppressing NHEJ within a drive system.

The technology illustrates a "minimal" arrangement requiring only one copy of the Cas9 nuclease but three distinct guide RNA expression cassettes targeting three unique positions within the genome. Cas9 can function *in trans*—expressed and located at one position within the genome, but the enzyme could act/function at multiple locations and induce multiple breaks. This will apply to both the nuclease of choice (Cas9 or other), as well as the guide RNA (sgRNA). The working examples and data demonstrate this point with the nuclease; however, the same approach would apply for the placement (*in trans*) of the sgRNA cassette.

Further, it will be appreciated that the multi-gene drive system exemplified herein demonstrates total deletion of an endogenous gene. The method also illustrates adding a modified (native) "cargo" gene as part of the drive itself (rather than deleting a gene). This method could also incorporate non-native gene cargos as well.

In one or more embodiments, the described approach also pilots use of a loss-of-function allele(s) of the highly conserved DNA Ligase IV enzyme (termed "Dnl4" in yeast, "Lig4" in humans). This strategy could be employed to reduce (or eliminate) NHEJ repair in the organism of choice. NHEJ counteracts the action of the gene drive. Therefore, suppressing (using targeted alleles, or suppression of transcript via dCas9 or promoter-replacement methods, or total deletion (as in yeast)), would bias the action of the gene drive to HR-based repair, rather than NHEJ. It will be appreciated that the inventive technique would also aid in combating the formation of "resistant" alleles that can escape action of the gene drive. Therefore, it will be appreciated that the described technology encompasses many ways to "reduce," but not eliminate, DNA Ligase IV (or any other contributing protein member of the NHEJ pathway), via transcriptional regulation, direct enzymatic inhibition, or use of mutational substitutions (as shown in the examples).

Advantageously, the specific components utilized in the yeast system herein (Cas9 nuclease, guide RNA) would be nearly identical if used in any other organism for this type of drive

arrangement. Further, the mode of action would also be the same in other organisms. That is, the action of all "gene drives" is identical in that homologous recombination (HR)-based DNA repair is required to "copy" the drive to the second broken chromosome.

It will be appreciated that the targeting of DNA Ligase IV (or any component of the NHEJ repair pathway) would also be very well conserved across various organisms. That is, although the exact positioning of the targeted mutation(s) may vary from organism-to-organism, the enzymatic function of this protein is highly conserved across organisms. Accordingly, the work exemplified herein in the working examples can be applied beyond yeast.

Advantageously, a "multi-locus" strategy—installing a gene drive or multiple gene drives—across a single genome has wide application in the design and type of "drive" one is interested in. These can include "daisy-chain" drives, "over or under-dominance" drives, or "anti-drives." The inventive technique would aid in redundancy of the action of the drive (i.e., two of the same drive in two places in the genome as a "back-up"). Further, it provides pathway redundancy (i.e., destroying one gene from one pathway, and another gene from a secondary pathway). Moreover, this technique can be applied to exogenous cargo delivery systems, especially those that are "large", where entire sets of genes, pathways, or extremely large genetic material could be separated across a genome, rather than propagated in one position. The technique can also be applied to augment safety or security features—genetic forms of inhibitors (such as AcrIIA2 and AcrIIA4 anti-CRISPR genes from PCT/US2018/016231, filed January 31, 2018, incorporated by reference herein), or inducible anti-drives, or inducible guide RNAs (causing self-cleavage/self-destruction). It will also be appreciated that this method of redundancy aids in combatting drive resistance (natural or evolved), and further could allow for "intelligent" or pre-programmed titration of one or more drives triggered by various conditions (environmental or man-made).

The examples demonstrate a further example, which is to preserve a "native" gene at its own native position within the genome (DNL4), where it maintains its own promoter sequence, own precise position on precise chromosome, and epigenetic marks. The guide RNA can be "hidden" downstream and pair with this native gene. Further, there may be a "safe" position where one can position the nuclease and then extend guide RNAs to many, many other positions within the genome to accomplish the tasks at hand (modifying many genes in a pathway, or destroying many genes, adding many mutations, etc.). This allows a combination of low-risk (reliable drive with nuclease) and higher-risk loci that may or may not recombine at optimal rates at poorly-studied positions in a genome of interest. As demonstrated and discussed in the examples,

biosafety and containment is another useful factor of the inventive system. The ability to have an n-number gene drive with many targets but can be shut-off "easily" by only targeting one locus (that has Cas9) or one anti-CRISPR, allows best of both worlds without sacrificing safety. The inventive technique also helps address the size limitation of a single "drive" position. For example, if one wanted to introduce 20 new genes into a species as part of a drive, there are inherent limitations based on size of the single drive. No one has explore the upper limits of HDR-based repair and whether there is a limitation of loss of efficiency when trying to copy huge amounts of DNA at one position. One alternative approach would be to separate the 20 genes over 4 positions (e.g., 5 genes at each position) for copying using the inventive gene drive system, reducing the overall cargo load per position.

Further, in the context of anti-CRISPR drives taht immunize a population against an invading drive, or to re-write over an existing drive, a multi-drive system allows a convenient way to modulate “tuning” of an original drive. That is, if a drive has positions 1-4, and we wish to remove “part 2” we could create a system to easily counter position 2, and still allow the original drive to hit targets 1, 3 and 4 as intended. This could be much more difficult if this was only a single drive with many cargo and would not allow a “modular” approach to refining or tuning a drive’s original intended targets. In the same way, one could envision adding targets 5 and 6 to an original drive 1-4. One could piggyback the success and programming of the original drive 1-4 and simply have a new population of 5+6 that would eventually converge and become drives 1-6 seamlessly. This could not be easily done with only a single drive. These approaches further tie into one form of “anti-drive” the “defensive” drive. It uses the same mechanism-that, at a distant locus (in fact, any position in the genome), one of two things could be easily delivered into the cell: 1) expression of an anti-CRISPR (to directly block the nuclease and simply stop the drive) or 2) expression of a self-targeting guide RNA. It would target the nuclease gene itself--causing it to cut and cleave itself and, in theory, kill the cell or destroy the drive movement forward. It will be appreciated that the underlying principle of “action at a distance,” such as with multiple guide RNAs scattered across the genome for use *in trans*, applies to these other techniques as well.

Additional advantages of the various embodiments of the invention will be apparent to those skilled in the art upon review of the disclosure herein and the working examples below. It will be appreciated that the various embodiments described herein are not necessarily mutually exclusive unless otherwise indicated herein. For example, a feature described or depicted in one embodiment may also be included in other embodiments, but is not necessarily included. Thus, the present invention encompasses a variety of combinations and/or integrations of the specific

embodiments described herein.

As used herein, the phrase "and/or," when used in a list of two or more items, means that any one of the listed items can be employed by itself or any combination of two or more of the listed items can be employed. For example, if a composition is described as containing or excluding components A, B, and/or C, the composition can contain or exclude A alone; B alone; C alone; A and B in combination; A and C in combination; B and C in combination; or A, B, and C in combination.

The term “gene drive” refers to a nucleic acid construct that is capable of copying itself into the genome of the cell into which it is introduced, along with any optional cargo. The present application is particularly concerned with endonuclease gene drives in diploid (or polyploid) organisms, wherein the gene drive induces a double-stranded break in the chromosome and induces the cell to copy the gene drive sequence to repair the damaged sequence via homologous recombination using the gene drive construct as the template. As a result, this process will re-occur in each organism that inherits one copy of the modification (and one wild-type copy). In this manner, gene drives are self-propagating, in comparison to the self-limited nature of traditional gene editing techniques.

The phrase, “operably linked” refers to the linkage of nucleic acid sequences in such a manner that a nucleic acid molecule (regulatory element) is capable of directing the transcription of a given gene and/or the synthesis of a desired protein molecule is produced. The term also refers to the linkage of amino acid sequences in such a manner so that a functional protein is produced. “Regulatory element” refers to promoters, enhancers, and other expression control signals that direct constitutive expression of a nucleotide sequence. Such elements may be host-specific or may drive expression broadly across various host cell types. Such elements may also be inducible and direct expression only under certain conditions (e.g., active or “on” only under an external stimulus, tissue-specific, or developmentally determined parameter).

A “host cell” or “target cell” as used herein, refers to the eukaryotic cell into which the CRISPR vector system has been introduced, include the progeny of the original transformed cell. A “host” or “subject” as used herein refers to an individual organism targeted for altered gene expression via CRISPR-based gene editing. Likewise, a “host” or “target” population refers to a plurality of individual host organisms which may be targeted for altered gene expression through CRISPR-based gene editing, such as a population of mosquitoes or other pests.

In general, and throughout this specification, the term “expression cassette” refers to a nucleic acid molecule capable of transporting nucleic acids into a host cell to thereby produce

transcripts, proteins, or peptides encoded by the nucleic acids in the cell. The term includes recombinant DNA molecules containing a desired coding sequence(s) and appropriate nucleic acid sequences (e.g., promoters) necessary for the expression of the operably linked coding sequence in a particular host organism. It will be appreciated that the design of the expression cassettes will depend on various factors, including the organism species, host cell type, and type of editing desired.

In the context of the present invention, the term "wild type" is not limited to truly naturally-occurring sequences, but includes previously mutated (whether naturally or artificially) sequences that are nonetheless different from the modified sequence being sought through CRISPR gene editing or incorporation of the gene drive.

The present description also uses numerical ranges to quantify certain parameters relating to various embodiments of the invention. It should be understood that when numerical ranges are provided, such ranges are to be construed as providing literal support for claim limitations that only recite the lower value of the range as well as claim limitations that only recite the upper value of the range. For example, a disclosed numerical range of about 10 to about 100 provides literal support for a claim reciting "greater than about 10" (with no upper bounds) and a claim reciting "less than about 100" (with no lower bounds).

EXAMPLES

The following examples set forth methods in accordance with the invention. It is to be understood, however, that these examples are provided by way of illustration and nothing therein should be taken as a limitation upon the overall scope of the invention.

EXAMPLE 1

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

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. Priming of the nuclease with a pre-programmed guide RNA fragment targets a specific genomic sequence for a double strand 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, alter epigenetic landscapes, or introduce mutational substitutions *sans* any DNA cleavage event.

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. Some of the possible benefits of this technology include eradication of invasive species, agricultural pest management, and elimination of insect-borne diseases such as malaria. A number of recent studies have demonstrated the potency and success of CRISPR-based gene drives in fungi, and metazoans.

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. There is an immediate need for further study (*in silico* and *in vivo*) of gene drive systems that focus on issues of safety, control and reversal, and optimal design.

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. 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.

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. 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, mechanisms for fortifying drive systems are still being elucidated. The proposal to increase the number of targeted double strand breaks (and corresponding sgRNAs) by the single nuclease of choice (e.g. *S. pyogenes* Cas9) would greatly aid in combatting resistance. 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. 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, by induced self-excision, or by removal of the Cas9-containing drive guide RNA. 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.

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. 2A). As shown in Fig. 2, the left panel illustrates the artificial gene drive 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. The right panel shows artificial (u1) and (u2) sites used flanking the gene drive at the *HIS3* locus (Chromosome XV) and the *S.p.HIS5* selectable marker. The use of the artificial programmed sites (see WO/2017/189336, filed April 20, 2017, incorporated by reference herein) was included in this study, but is not required for use of this invention. The artificial sites are used for biosafety and for convenience, but the invention can be applied without use of these (u1) and (u2) sites. 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. Thus, in this design, 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), and novel "self-editing" integration events (Fig. 5). 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).

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, the *DNL4* locus was deleted and replaced with the Kan^R drug cassette (Fig. 2A).

As shown in Fig. 2B, the triple MGD strain (Haploid yeast harboring the triple drive (GFY-3675)) was first mated with the triple target strain (GFY-3596) to form diploids, and Cas9 was activated by culturing in medium containing 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. In the absence of nuclease expression (Fig. 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. 2B, *bottom*).

Fig. 2C illustrates a time course of galactose activation using the [GFY-3675 x GFY-3596] diploid in triplicate (Error, SD), showing 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. 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), Seven haploid strains (GFY-3206, 3593, 3264b, 3578, 3594, 3623, and 3596) were tested as in (**B**) against the triple drive strain (GFY-3675). Thus, 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* and the results are illustrated in Fig. 2D. Then, each of the diploids from (D) were cultured for 5 hr and quantified for drive success. Error, SD. 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. 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. 2B) and confirmed both the growth profile and ploidy status of random samples (Fig. 2F, *bottom*). In Fig. 2F, 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 of all six distinct loci to assay for the presence or absence of each engineered drive and target (Fig. 2F, Fig. 6). 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 the tables below. 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. See also Fig. 10. All PCR reactions were performed using identical conditions.

Oligonucleotides 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). In Fig. 2, oligonucleotide 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. See also Fig. 11.

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. 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. 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 (Fig. 7). 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 described1. 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). 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.

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.

The NHEJ pathway is highly conserved from yeast to humans and functions to directly fuse exposed DNA ends. DNA Ligase IV (Dnl4 in yeast, Lig4 in humans) is required for the final step of DNA ligation along with other conserved binding partners. 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. 3A). Note, the branching of *Z. nevadensis* (termite) was poorly supported and has been previously shown to be included within the *Insecta* class. 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. 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. 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. 3C, Table 1).

Table 1. 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).

[b]The TM-score is a metric for structural similarity and scores > 0.5 indicate high confidence.

[c]Root-mean-square deviation of atomic positions.

The N-terminal region also displayed strong structural homology using the human Lig4 crystal structure as a template (Fig. 8). 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 and aligned using MatchMaker in Chimera. Each I-TASSER model was individually aligned against the human Lig4 N-terminal structure.

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. 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. 3D). 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. In a strain harboring integrated Cas9 at the *HIS3* locus, 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.

We utilized these eight haploid strains to quantify the level of NHEJ repair (Fig. 4). Our system of DSB formation followed by DNA repair utilized the dual programmed (u2) sites flanking the Cas9 expression cassette (Fig. 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. 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.

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. 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. 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. 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. 4E). Finally, each of the *dnl4* alleles was tested within our MGD system as a native cargo-based delivery system (Fig. 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. 4F, *bottom*). Moreover, PCRs from clonal isolates confirmed the presence or absence of each drive and target locus (Fig. 9). 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 *dnl4Δ::Kan*^{R} target cassette. 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 Fig. 12. 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).

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. 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, self-excised by our flanking (u2) sites, 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* and *in vivo* 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. This would provide greater than 99% confidence in eliminating an *A. gambiae* population on a continent-wide scale. 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 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 and flies, it is embryonic lethal in mouse and not tolerated in mosquito. 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*. 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.

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 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.

Materials and Methods

Yeast Strains and Plasmids

Standard molecular biology protocols were used to engineer all *S. cerevisiae* strains (Table 2) used in this study. 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 (Fig. 5). Briefly, DNA constructs were first assembled onto *CEN*-based plasmids (typically pRS315) using *in vivo* assembly in yeast. If necessary, point mutations were introduced using PCR mutagenesis. 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, and integrated at the desired genomic locus.

Table 2. Yeast Strains used in Study

Strain	Genotype
BY4741	*MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*
BY4742	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*
GFY-3675[a]	BY4741; *his3Δ::prHIS3:(u2)::prGAL1/10::SpCas9::NLS::ADH1(t):(u2)::HIS3(t) shs1Δ::prSHS1::sgRNA(GFP)::SHS1(t) dnl4Δ::prDNL4::sgRNA(Kan)::DNL4(t)*
GFY-3206[b]	BY4742; *his3Δ::prHIS3::(u1)::prCDC12::mCherry::NLS::SHS1(t)::prCCW12::SpHIS5:: MX(t)::(u1)::HIS3(t)*
GFY-3593[c]	BY4742; *shs1Δ::prSHS1::SHS1::GFP::CDC10(t)::prMX::CaURA3::SHS1(t)*

GFY-3264b[d]	BY4742; *dnl4Δ::prDNL4::prMX::Kan^R::MX(t)::DNL4(t)*
GFY-3578	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)*
GFY-3594	BY4742; *his3Δ::prHIS3::(u1)::prCDC12::mCherry::NLS::SHS1(t)::prCCW12::SpHIS5::MX(t)::(u1)::HIS3(t)* *dnl4Δ::prDNL4::prMX::Kan^R::MX(t)::DNL4(t)*
GFY-3623	BY4742; *shs1Δ::prSHS1::SHS1::GFP::CDC10(t)::prMX::CaURA3::SHS1(t)* *dnl4Δ::prDNL4::prMX::Kan^R::MX(t)::DNL4(t)*
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)*
GFY-3611	BY4741; *his3Δ::prHIS3:(u2)::prGAL1/10::SpCas9::NLS::ADH1(t):(u2)::HIS3(t)* *shs1Δ::prSHS1::sgRNA(GFP)::SHS1(t) DNL4*
GFY-2383	BY4741; *his3Δ::prHIS3:(u2)::prGAL1/10::SpCas9::NLS::ADH1(t):prMX::Kan^R::MX(t)::(u2)::HIS3(t)*
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)*
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)*

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)*
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)*
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)*
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)*
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)*
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)*
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)*
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)*
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)*
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)
GFY-3872	BY4741; *his3Δ::prHIS3:(u2)::prGAL1/10::SpCas9::NLS::ADH1(t):(u2)::HIS3(t) shs1Δ::prSHS1::sgRNA(GFP)::SHS1(t) dnl4Δ::prDNL4::dnl4(G868A G869A)::CDC11(t)::sgRNA(Kan)::DNL4(t)*
GFY-3875	BY4741; *his3Δ::prHIS3:(u2)::prGAL1/10::SpCas9::NLS::ADH1(t):(u2)::HIS3(t) shs1Δ::prSHS1::sgRNA(GFP)::SHS1(t) dnl4Δ::prDNL4::DNL4(WT)::CDC11(t)::sgRNA(Kan)::DNL4(t)*

[a]Strain GFY-3675 was derived from GFY-2383. First, *S. pyogenes* Cas9 expression was activated as previously described 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'; SEQ ID NO:8) where the PAM is in bold, 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'; SEQ ID NO:1) includes a PAM (bold) and the 20 bp target site for *Sp*Cas9. 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 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'; SEQ ID NO:4) and (5' **CCT**GAGGAGGATTTCCCCGTAGT 3'; SEQ ID NO:6) where the PAM (bold) and target sequences 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 3. Plasmids used in study.

Plasmid	Description

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

[a]The sgRNA(u1) sequence is 5' CGGTGGACTTCGGCTACGTA 3' (residues 5-24 SEQ ID NO:1). All sgRNA constructs include 269 bp of the *SNR52* promoter, the 79 bp tracrRNA, and the 20 bp *SUP4* terminator, as modeled from published studies.

[b]The sgRNA(u2) sequence is 5' GCTGTTCGTGTGCGCGTCCT 3' (SEQ ID NO:8 with GGG PAM sequence).

[c]The sgRNA(DNL4-A) sequence is 5' GACTATGTCACGGAGGACAC 3' (SEQ ID NO:5)

[d]The sgRNA(DNL4-B) sequence is 5' ACTACGGGGAAGTCTTCTTC 3' (SEQ ID NO:7). The target is present on the non-coding strand at the 3' end of the *DNL4* gene.

PCR was used to diagnose proper chromosomal position for each integration event followed by DNA sequencing. Expression cassettes for sgRNA were based on a previous study, 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. 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. 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). 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. 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 Figs. 11-12 for comparison.

Data analysis (Fig. 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. Homologous sequences to the yeast DNA Ligase IV (Dnl4) protein were obtained using multiple BLAST (NCBI) searches within either the fungal or metazoan clade (Table 4).

Table 4. Species used for alignments and phylogeny of DNA Ligase IV.

Species Name (Strain) [Protein Name]	Amino Acids	Reference ID[x]
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

[X] Reference sequences incorporated by reference herein.

The phylogenetic tree of DNA Ligase IV was created using the Phylogeny.fr software. Multiple sequence alignments were performed using Clustal Omega. The predicted structures of the human, yeast, and mosquito Ligase IV enzyme were generated using I-TASSER. The template structures included the human Lig4 N-terminus (PDB:3W1B) and the yeast Dnl4 C-terminus (PDB:1Z56). Predicted models were individually aligned against the crystal structures using MatchMaker in Chimera.

CLAIMS:

1. A method for stable integration of a CRISPR-Cas gene drive system into a diploid eukaryotic cell genome at multiple loci, said method comprising:

introducing into said eukaryotic cell genome at a first locus, a first gene drive construct comprising a first nuclease nucleotide sequence encoding a functional CRISPR nuclease that induces a double-stranded break in or near a target site;

introducing into said eukaryotic cell genome at a plurality of loci remote from said first locus, a plurality of secondary gene drive constructs each comprising a guide nucleotide sequence encoding a single guide RNA (sgRNA) sequence complementary to respective endogenous target sites in said cell genome,

wherein said constructs are expressed in said cell to produce said functional CRISPR nuclease and plurality of sgRNA, wherein said CRISPR nuclease co-localizes with respective sgRNA in said cell at each respective target site, said CRISPR nuclease inducing a double-stranded break at each target site, wherein homology-directed repair integrates each of said first gene drive construct and said plurality of secondary gene drive constructs into their respective target sites of said cell genome.

2. The method of claim 1, wherein said first gene drive construct further comprises a first guide nucleotide sequence encoding a sgRNA sequence complementary to a first endogenous target site and located on the expression cassette adjacent said nuclease nucleotide sequence encoding and between a first pair of endogenous flanking sequences homologous to sequences adjacent the first target site.

3. The method of claim 1, wherein said plurality of secondary gene drive constructs comprise at least:

a second gene drive construct comprising a second guide nucleotide sequence encoding a sgRNA sequence complementary to a second endogenous target site, and a second pair of endogenous flanking sequences homologous to sequences adjacent the second target site; and

a third gene drive construct comprising a third guide nucleotide sequence encoding a sgRNA sequence complementary to a third endogenous target site, and a third pair

of endogenous flanking sequences homologous to sequences adjacent the third target site.

4. The method of claim 1, said plurality of secondary gene drive constructs being free of nuclease-encoding nucleotide sequences.

5. The method of claim 1, wherein any one of said gene drive constructs further comprises donor DNA for integration into the cell genome at any one of said target sites.

6. The method of claim 1, wherein said first locus is on a different arm of the same chromosome as said plurality of loci, or wherein said first locus is on a different chromosome from said plurality of loci.

7. The method of claim 1, wherein one of said target sites is adjacent an endogenous gene in said genome, wherein said endogenous gene is deleted or partially replaced via said integration of one or more said plurality of secondary gene drives into said genome via said homology-directed repair.

8. The method of claim 1, wherein said cell is fungal cell, a plant cell, an insect cell or a non-human mammalian cell.

9. The method of claim 1, wherein said functional CRISPR nuclease is Cas9.

10. The method of claim 9, wherein said Cas9 is selected from the group consisting of *S. pyogenes*, *S. aureus*, *S. pneumoniae*, *S. thermophilus*, *N. meningitidis* Cas9 nucleases, and functional Cas9 orthologs and mutant Cas9 derived from these organisms.

11. The method of claim 1, wherein said homology directed repair is mediated by said flanking sequences in said gene drive constructs to integrates said gene drive construct into said respective target site.

12. The method of claim 1, said method further comprising growing said diploid eukaryotic cell into an organism.

13. The method of claim 12, said method further comprising mating said organism with a wild-type organism of the same species to produce offspring, wherein said offspring has a genome that is at least heterologous for said CRISPR-Cas gene drive system.

14. A eukaryotic host cell produced by the method of any one of claims 1-13.

15. A eukaryotic host cell comprising a CRISPR-Cas gene drive system stably integrated into its genome at multiple loci, comprising:

a first gene drive construct integrated at a first locus, said first gene drive construct comprising a first nuclease nucleotide sequence encoding a functional CRISPR nuclease that induces a double-stranded break in or near a target site;

a plurality of secondary gene drive constructs integrated at a plurality of loci remote from said first locus, said plurality of secondary gene drive constructs each comprising a guide nucleotide sequence encoding a single guide RNA (sgRNA) sequence complementary to respective endogenous target sites in said cell genome,

wherein said constructs are expressed in said cell to produce said functional CRISPR nuclease and plurality of sgRNA, wherein said CRISPR nuclease co-localizes with respective sgRNA in said cell at each respective target site, said CRISPR nuclease inducing a double-stranded break at each target site, wherein homology-directed repair integrates each of said first gene drive construct and said plurality of secondary gene drive constructs into their respective target sites of said cell genome, such that said host cell is homozygous for said CRISPR-Cas gene drive system at each of said loci.

16. The eukaryotic host cell of claim 15, wherein said first gene drive construct further comprises a first guide nucleotide sequence encoding a sgRNA sequence complementary to a first endogenous target site and located on the expression cassette adjacent said nuclease nucleotide sequence encoding and between a first pair of endogenous flanking sequences homologous to sequences adjacent the first target site.

17. The eukaryotic host cell of claim 15, wherein said plurality of secondary gene drive constructs comprise at least:

a second gene drive construct comprising a second guide nucleotide sequence encoding a

sgRNA sequence complementary to a second endogenous target site, and a second pair of endogenous flanking sequences homologous to sequences adjacent the second target site; and

a third gene drive construct comprising a third guide nucleotide sequence encoding a sgRNA sequence complementary to a third endogenous target site, and a third pair of endogenous flanking sequences homologous to sequences adjacent the third target site.

18. The eukaryotic host cell of claim 15, wherein said host cell is free of any additional nucleic acid constructs encoding for functional CRISPR nuclease.

19. An organism comprising the eukaryotic host cell of claim 15.

20. A method for targeted inhibition, suppression, or extinction of a population, comprising releasing a CRISPR-Cas gene drive system into a targeted population, wherein said CRISPR-Cas gene drive system is configured for stable integration into a eukaryotic diploid cell genome at multiple loci, said CRISPR-Cas gene drive system comprising:

a first gene drive construct configured for integration at a first locus, said first gene drive construct comprising a first nuclease nucleotide sequence encoding a functional CRISPR nuclease that induces a double-stranded break in or near a target site;

a plurality of secondary gene drive constructs for configured integration at a plurality of loci remote from said first locus, said plurality of secondary gene drive constructs each comprising a guide nucleotide sequence encoding a single guide RNA (sgRNA) sequence complementary to respective endogenous target sites in said cell genome,

wherein said constructs are expressed in said cell to produce said functional CRISPR nuclease and plurality of sgRNA, wherein said CRISPR nuclease co-localizes with respective sgRNA in said cell at each respective target site, said CRISPR nuclease inducing a double-stranded break at each target site, wherein homology-directed repair integrates each of said first gene drive construct and said plurality of secondary gene drive constructs into their respective target sites of said cell genome, such that said eukaryotic diploid cell genome becomes homozygous for said CRISPR-Cas gene drive system at each of said loci.

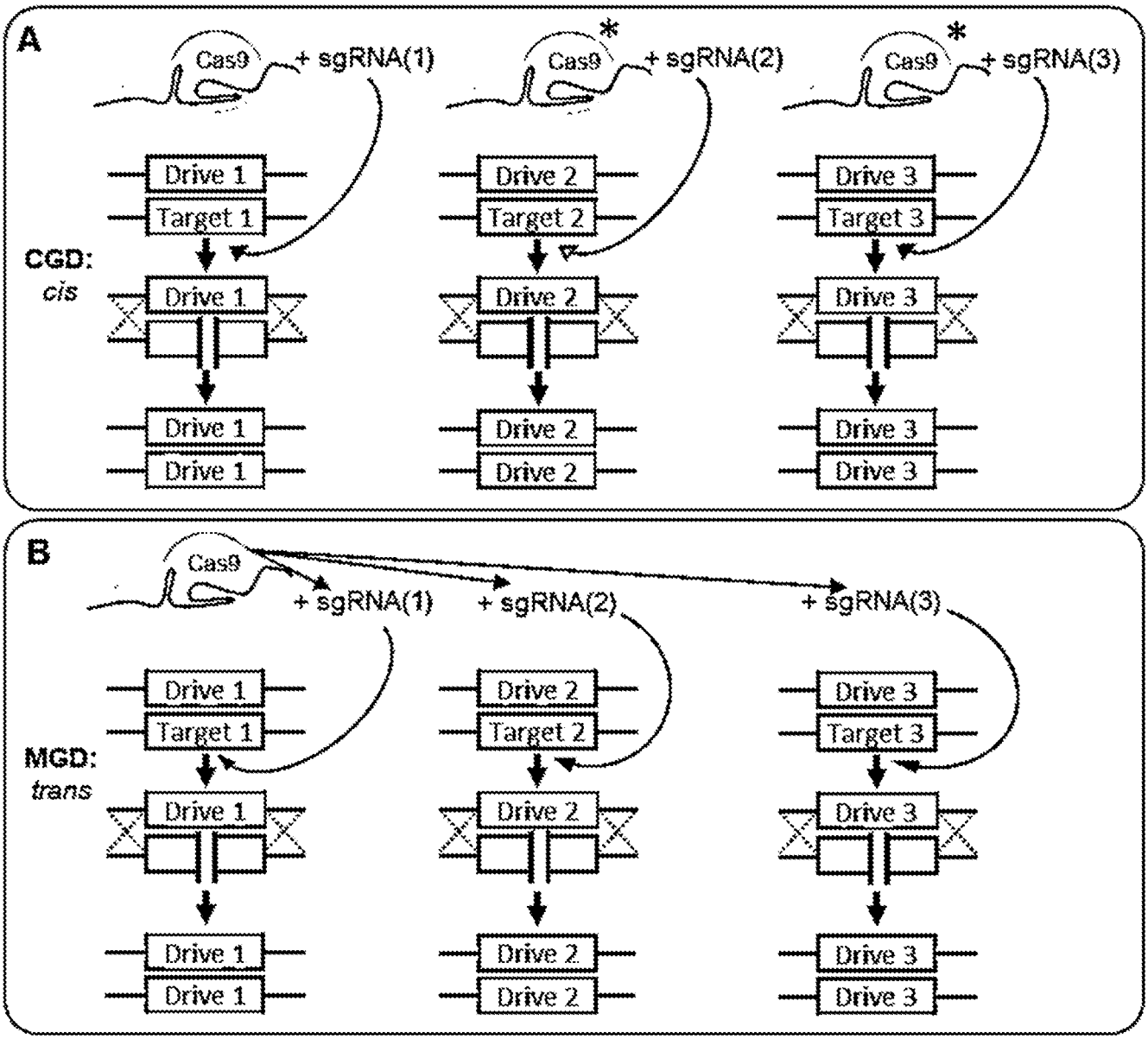
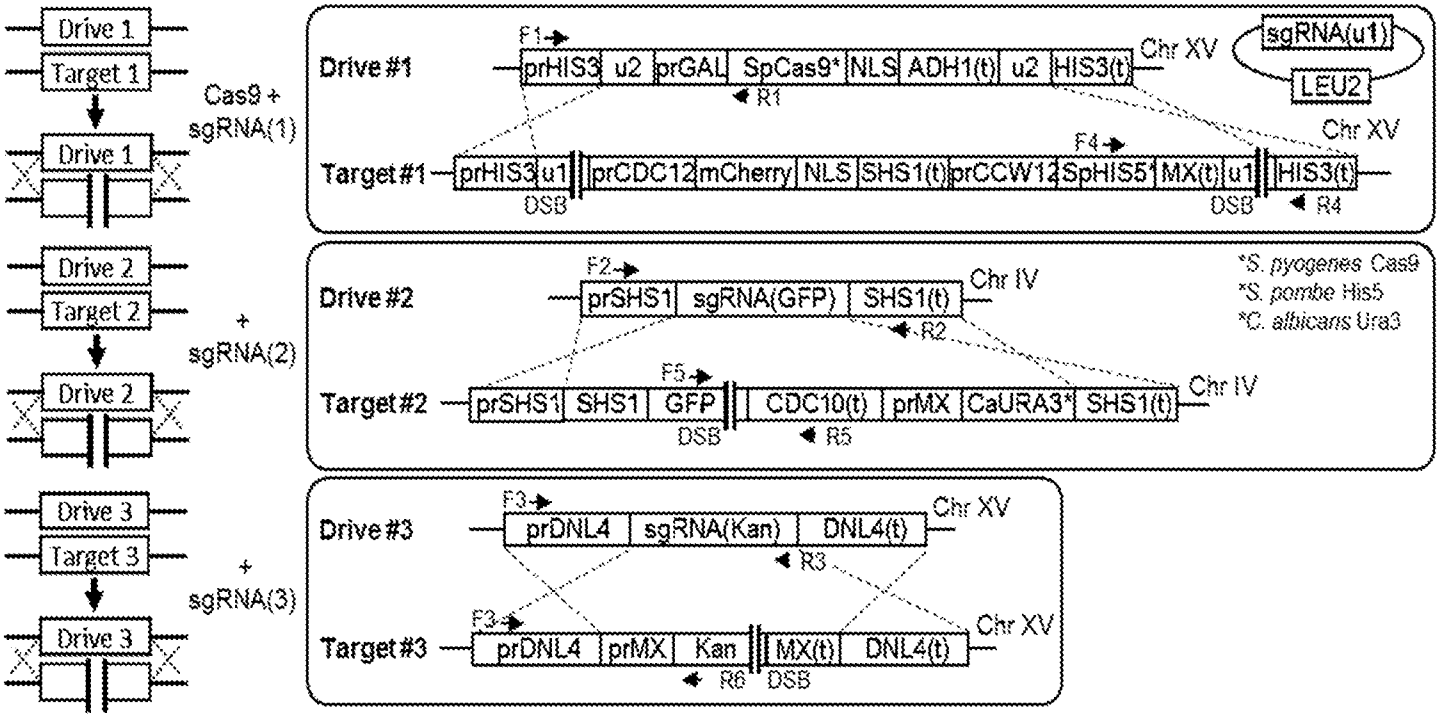

A
Cas9
+ sgRNA(1)
Cas9
*
+ sgRNA(2)
Cas9
*
+ sgRNA(3)
Drive 1
Target 1
Drive 2
Target 2
Drive 3
Target 3
CGD:
cis
Drive 1
Drive 2
Drive 3
Drive 1
Drive 1
Drive 2
Drive 2
Drive 3
Drive 3
B
Cas9
+ sgRNA(1)
+ sgRNA(2)
+ sgRNA(3)
Drive 1
Target 1
Drive 2
Target 2
Drive 3
Target 3
MGD:
trans
Drive 1
Drive 2
Drive 3
Drive 1
Drive 1
Drive 2
Drive 2
Drive 3
Drive 3

FIG. 1



u1 Target PAM
CGGTGGACTTCGGCTACGTAGGG (SEQ ID NO:1)

GFP Target PAM
AAAGGAGAAGAACTTTTCACTGG (SEQ ID NO:2)

Kan Target PAM
GCCATCCTATGGAACTGCCTCGG (SEQ ID NO:3)

FIG. 2A

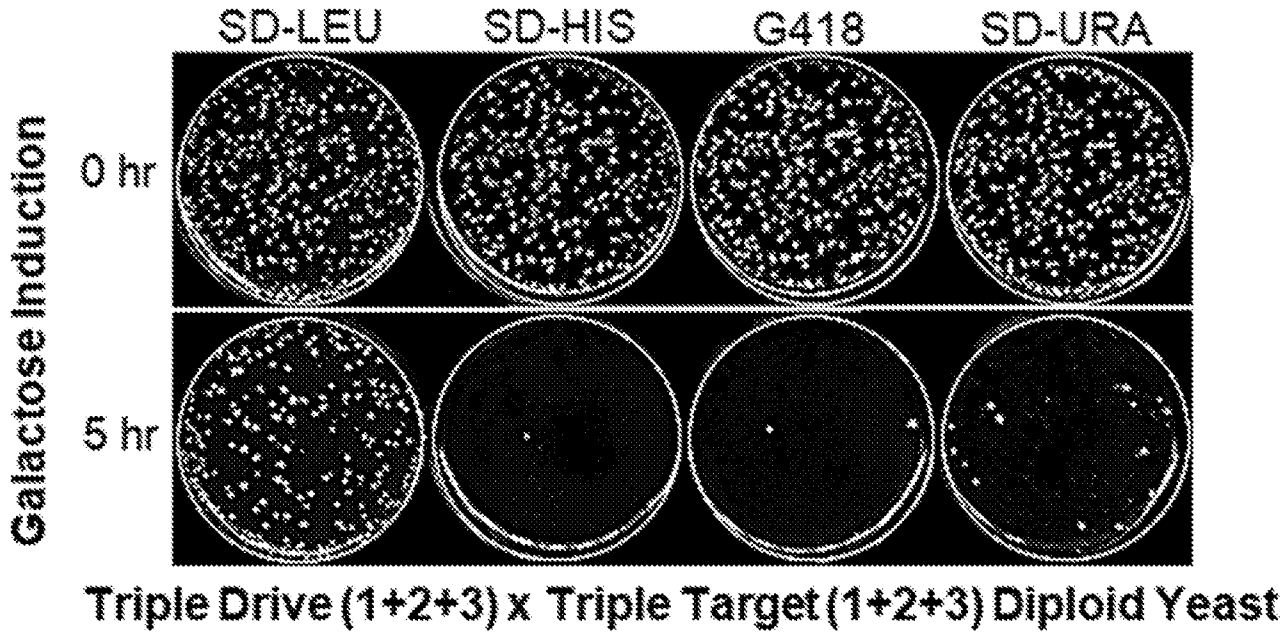


FIG. 2B

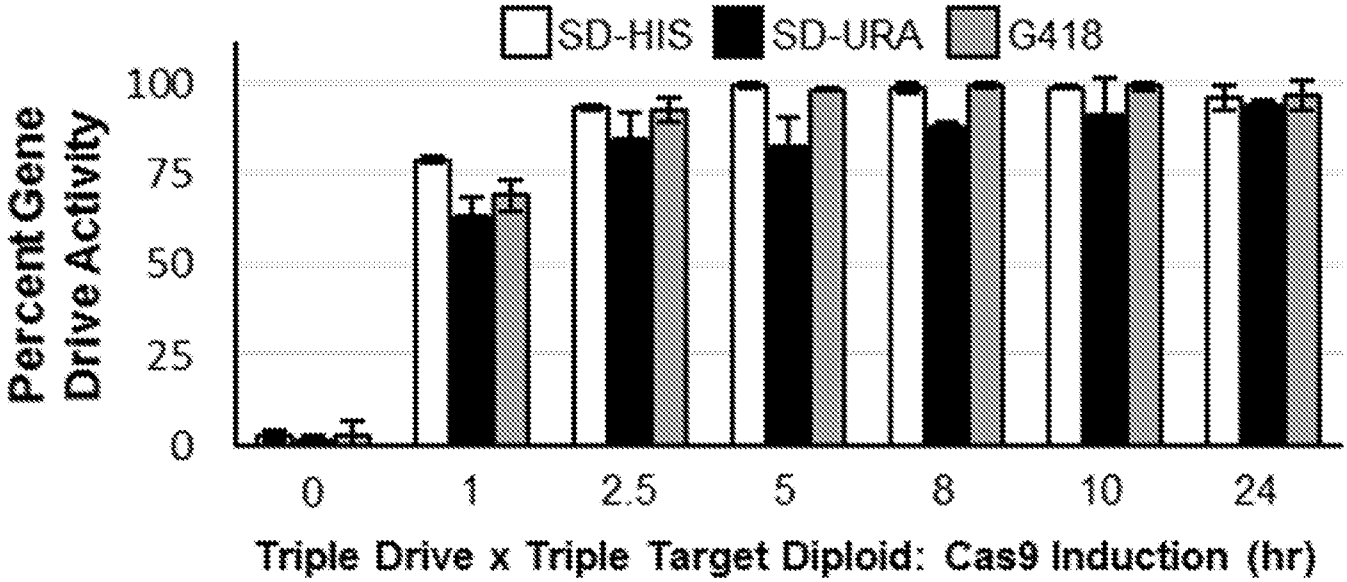


FIG. 2C

Diploid Yeast Strains with Various Targets Present (1, 2 or 3)

Presence of Drive & Target Loci

	1	2	3	4	5	6	7
Drive 1	√	√	√	√	√	√	√
Drive 2	√	√	√	√	√	√	√
Drive 3	√	√	√	√	√	√	√
Target 1	√			√	√		√
Target 2		√		√		√	√
Target 3			√		√	√	√

FIG. 2D

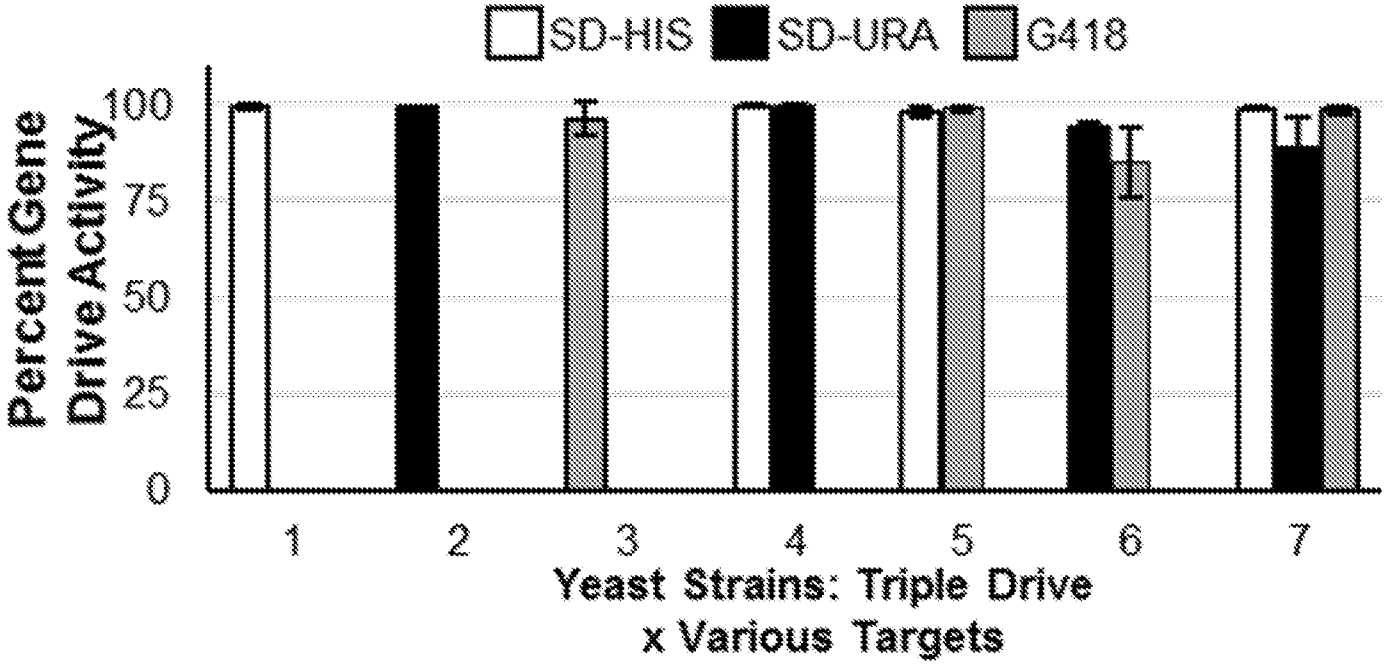
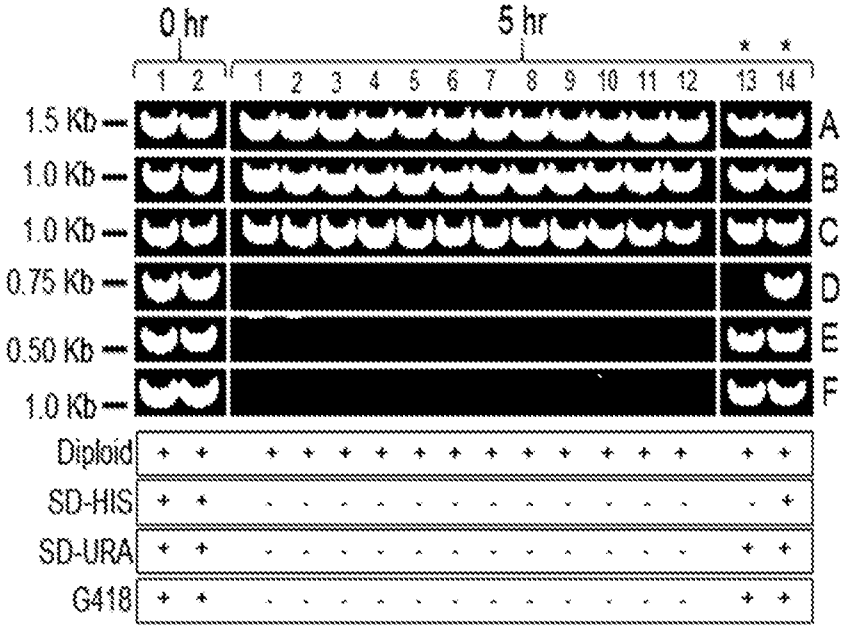


FIG. 2E



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

FIG. 2F

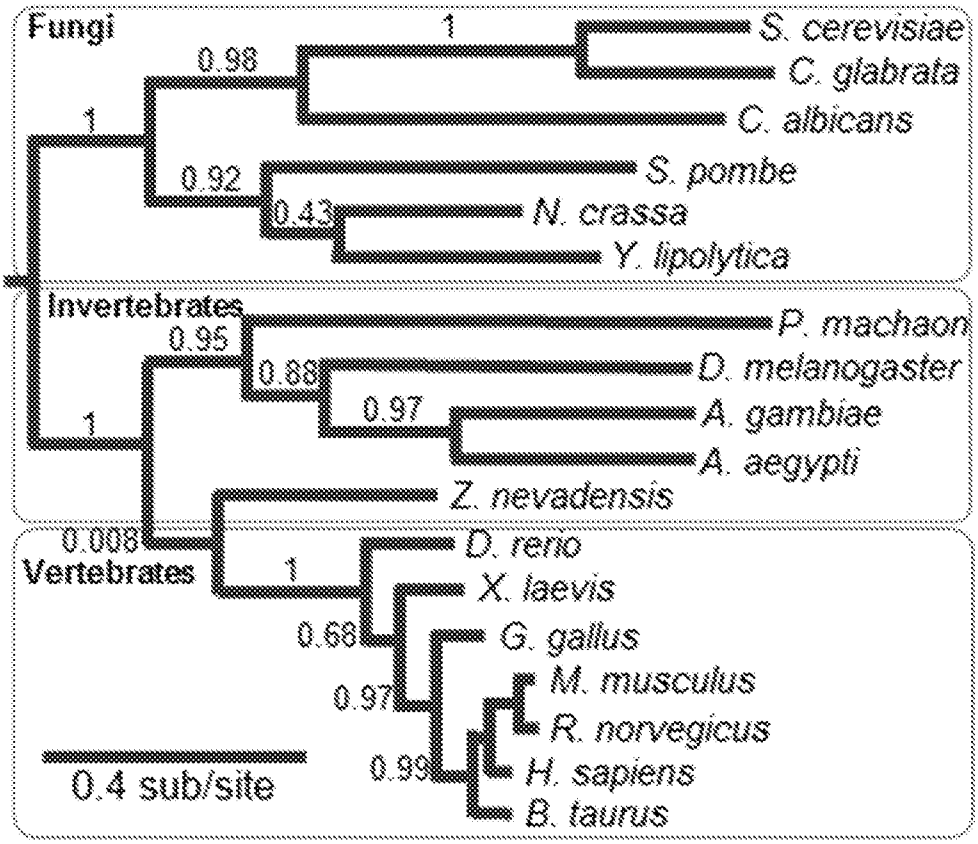

Fungi
1
0.98
1
0.92
0.43
S. cerevisiae
C. glabrata
C. albicans
S. pombe
N. crassa
Y. lipolytica
Invertebrates
0.95
0.88
0.97
1
P. machaon
D. melanogaster
A. gambiae
A. aegypti
Z. nevadensis
0.008
Vertebrates
1
0.68
0.97
0.99
D. rerio
X. laevis
G. gallus
M. musculus
R. norvegicus
H. sapiens
B. taurus
0.4 sub/site

FIG. 3A

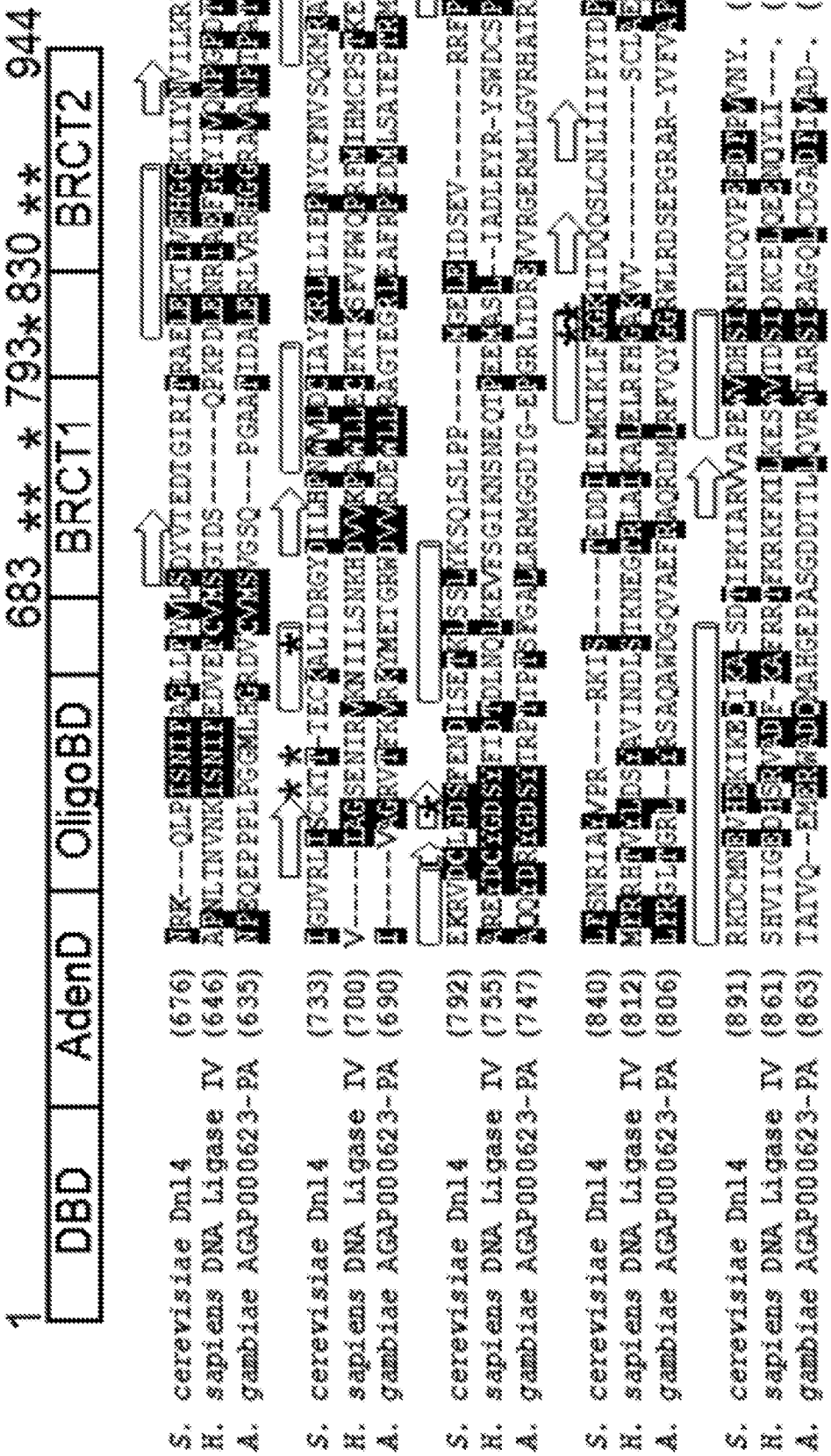

1
683 ** * 793* 830 ** 944
DBD
AdenD
OligoBD
BRCT1
BRCT2
S. cerevisiae Dnl4 (676)
H. sapiens DNA Ligase IV (646)
A. gambiae AGAP000623-PA (635)
S. cerevisiae Dnl4 (733)
H. sapiens DNA Ligase IV (700)
A. gambiae AGAP000623-PA (690)
S. cerevisiae Dnl4 (792)
H. sapiens DNA Ligase IV (755)
A. gambiae AGAP000623-PA (747)
S. cerevisiae Dnl4 (840)
H. sapiens DNA Ligase IV (812)
A. gambiae AGAP000623-PA (806)
S. cerevisiae Dnl4 (891)
H. sapiens DNA Ligase IV (861)
A. gambiae AGAP000623-PA (863)
(732)
(699)
(691)
(791)
(754)
(746)
(839)
(811)
(805)
(890)
(860)
(862)
(944)
(911)
(914)

FIG. 3B

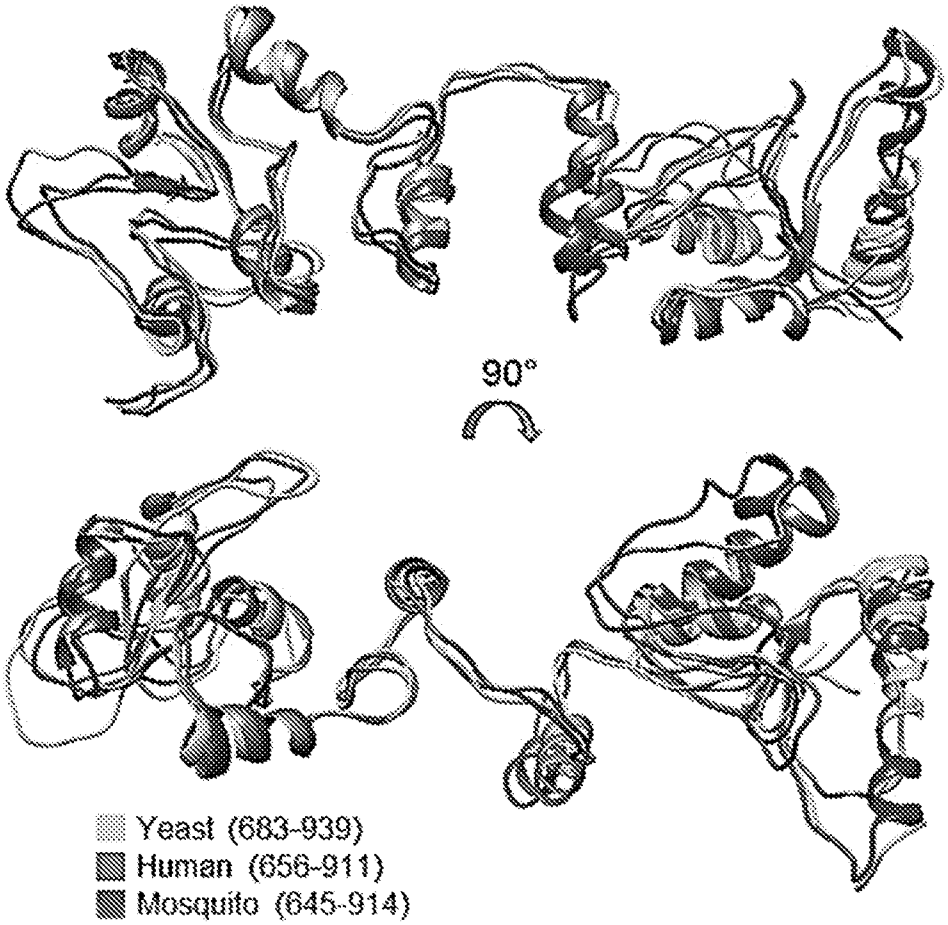


FIG. 3C

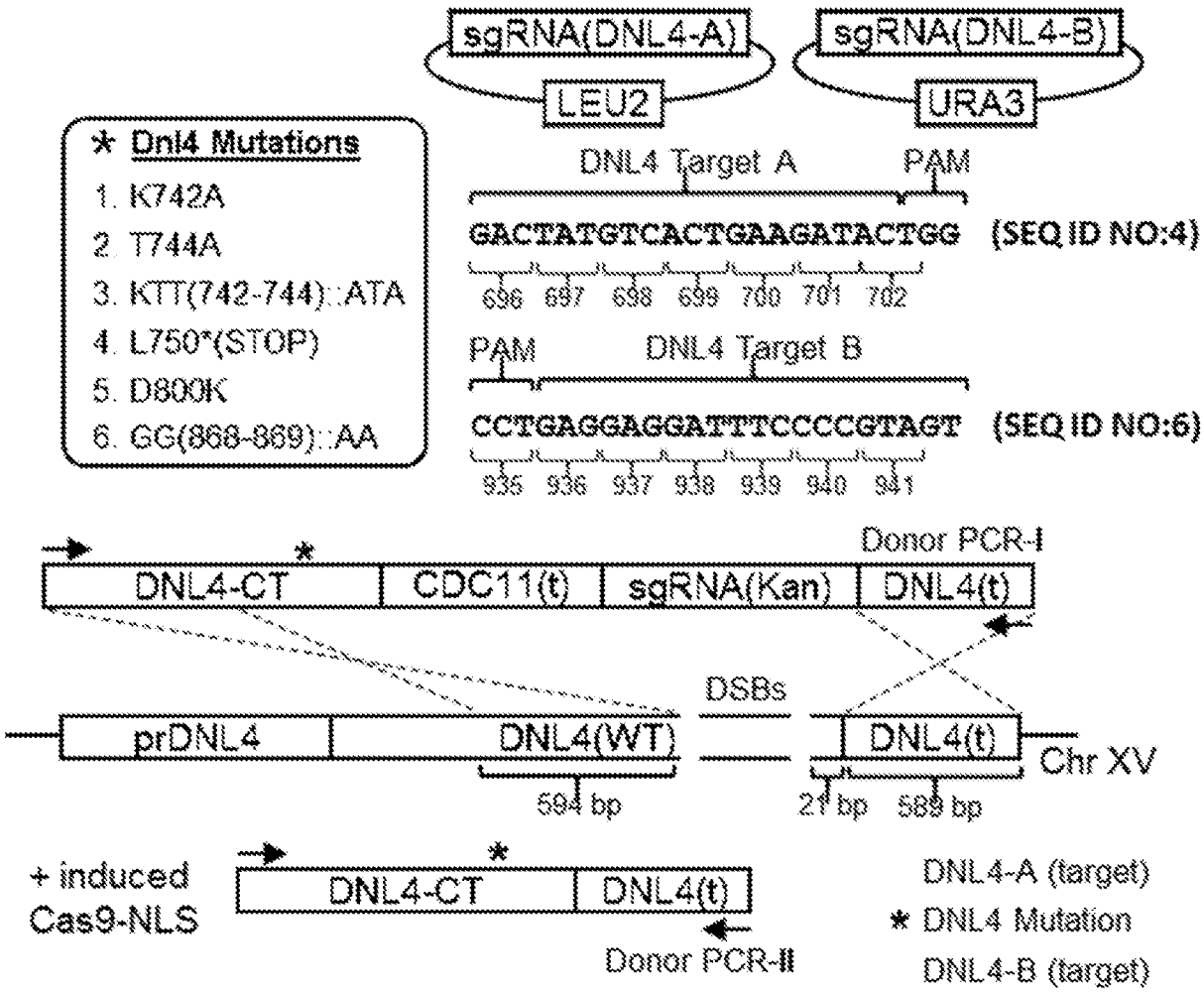

sgRNA(DNL4-A)
LEU2
sgRNA(DNL4-B)
URA3
* Dnl4 Mutations
1. K742A
2. T744A
3. KTT(742-744)::ATA
4. L750*(STOP)
5. D800K
6. GG(868-869)::AA
DNL4 Target A
PAM
GACTATGTCACTGAAGATACTGG (SEQ ID NO:4)
696 697 698 699 700 701 702
PAM
DNL4 Target B
CCTGAGGAGGATTTCCCCGTAGT (SEQ ID NO:6)
935 936 937 938 939 940 941
Donor PCR-I
DNL4-CT
CDC11(t)
sgRNA(Kan)
DNL4(t)
DSBs
prDNL4
DNL4(WT)
DNL4(t)
Chr XV
594 bp
21 bp
589 bp
+ induced
Cas9-NLS
DNL4-CT
DNL4(t)
Donor PCR-II
DNL4-A (target)
* DNL4 Mutation
DNL4-B (target)

FIG. 3D

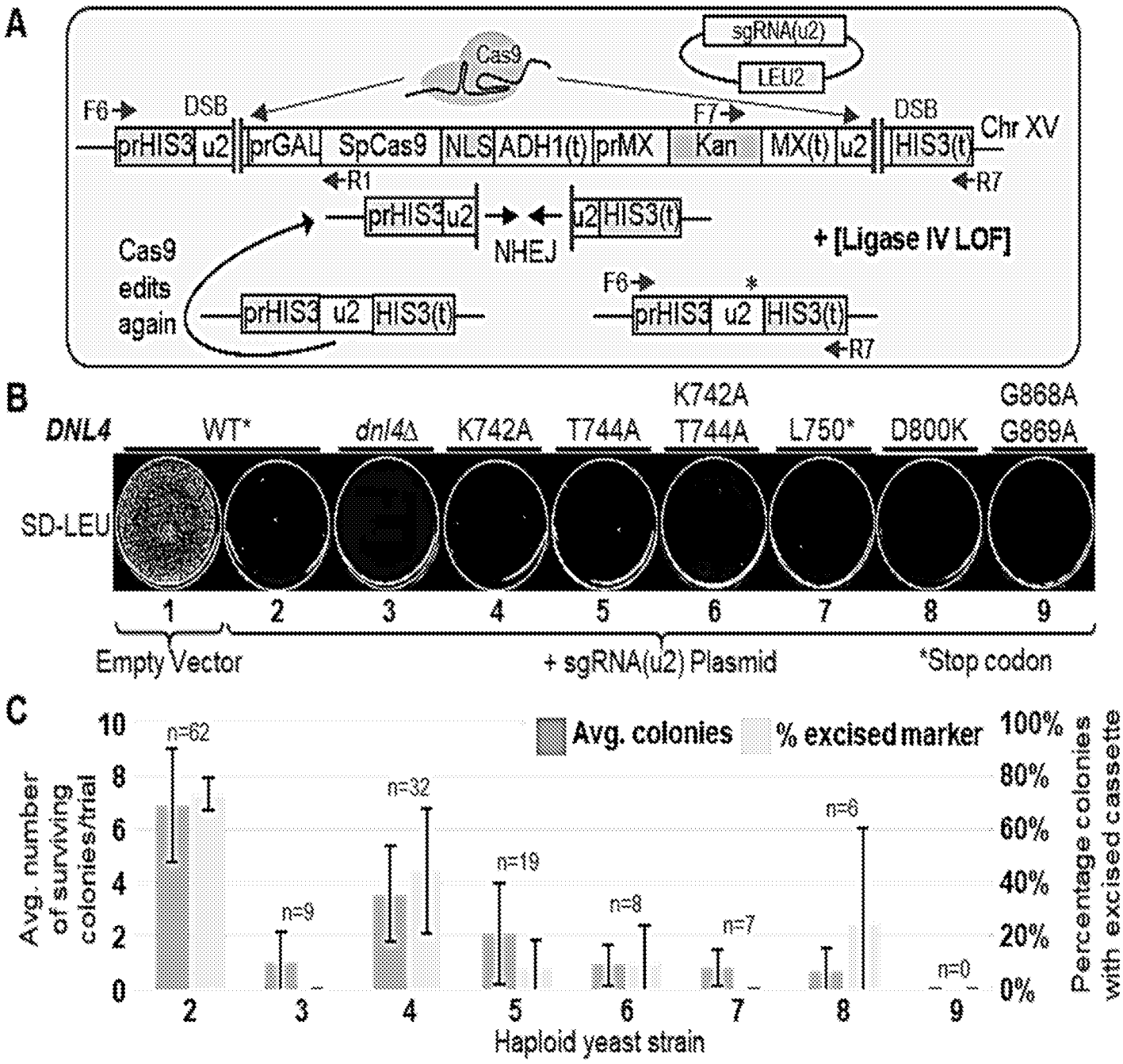

A
Cas9
sgRNA(u2)
LEU2
F6
DSB
F7
DSB
prHIS3 u2
prGAL SpCas9 NLS ADH1(t) prMX Kan MX(t) u2
HIS3(t)
Chr XV
R1
R7
prHIS3 u2
u2 HIS3(t)
NHEJ
+ [Ligase IV LOF]
Cas9 edits again
prHIS3 u2 HIS3(t)
F6
*
prHIS3 u2 HIS3(t)
R7
B
DNL4
WT*
dnl4Δ
K742A
T744A
K742A T744A
L750*
D800K
G868A G869A
SD-LEU
1 2 3 4 5 6 7 8 9
Empty Vector
+ sgRNA(u2) Plasmid
*Stop codon
C
Avg. colonies
% excised marker
Avg. number of surviving colonies/trial
Percentage colonies with excised cassette
n=62
n=9
n=32
n=19
n=8
n=7
n=6
n=0
Haploid yeast strain

FIG. 4

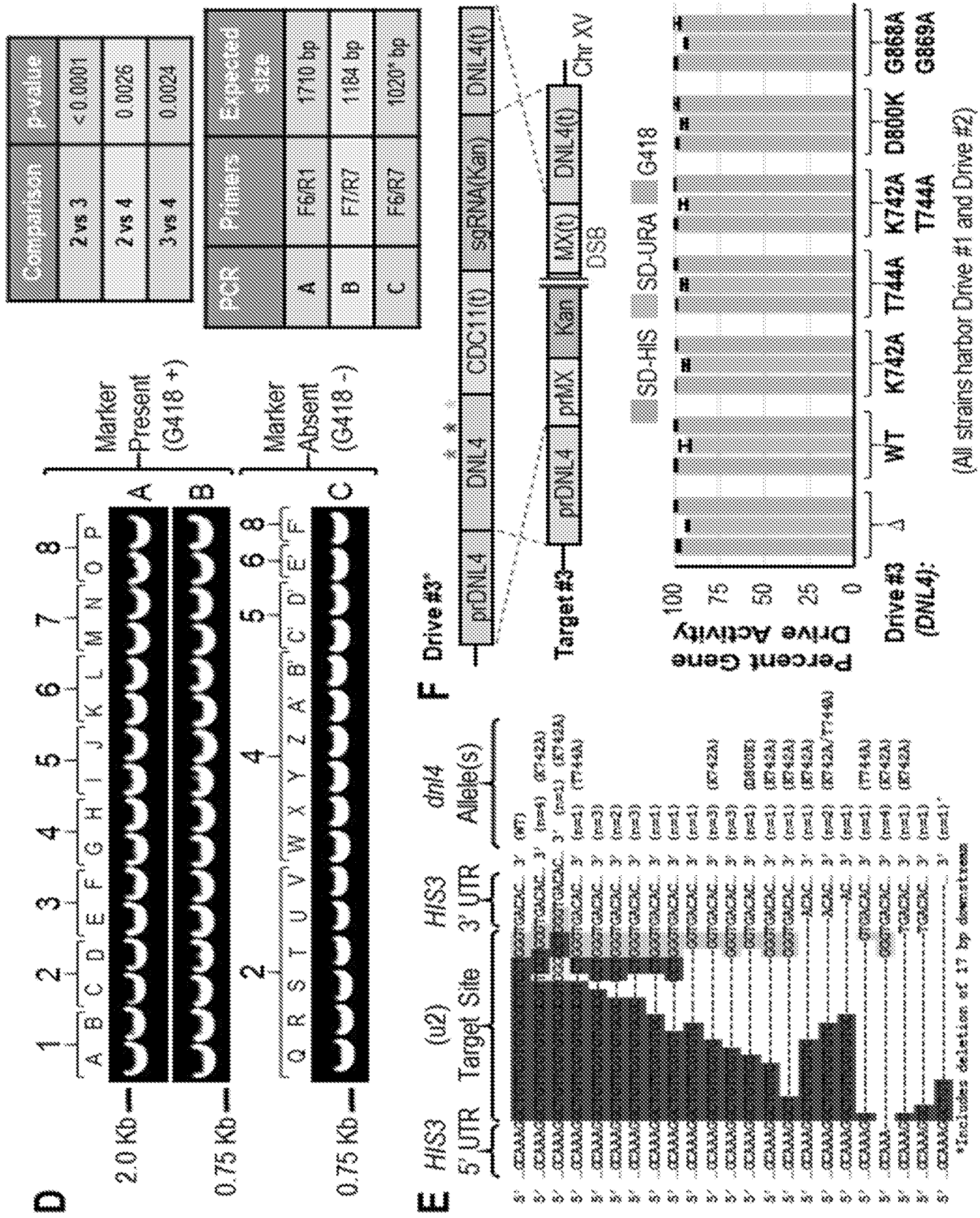

D
2.0 Kb
0.75 Kb
0.75 Kb
Marker Present (G418 +)
Marker Absent (G418 -)
A
B
C
PCR Primers Expected size
A F6/R1 1710 bp
B F7/R7 1184 bp
C F6/R7 1020* bp
Comparison p-value
2 vs 3 <0.0001
2 vs 4 0.0026
3 vs 4 0.0024
E
HIS3 5' UTR
(u2) Target Site
HIS3 3' UTR
dnl4 Allele(s)
*Includes deletion of 17 bp downstream
F
Drive #3*
prDNL4
DNL4
CDC11(0)
sgRNA(Kan)
DNL4(t)
prDNL4
prMX
Kan
MX(t)
DNL4(t)
Target #3
DSB
Chr XV
SD-HIS
SD-URA
G418
Percent Gene Drive Activity
100
75
50
25
0
Drive #3 (DNL4):
WT
K742A
T744A
K742A T744A
D800K
G868A
G869A
(All strains harbor Drive #1 and Drive #2)

FIG. 4 (cont'd)

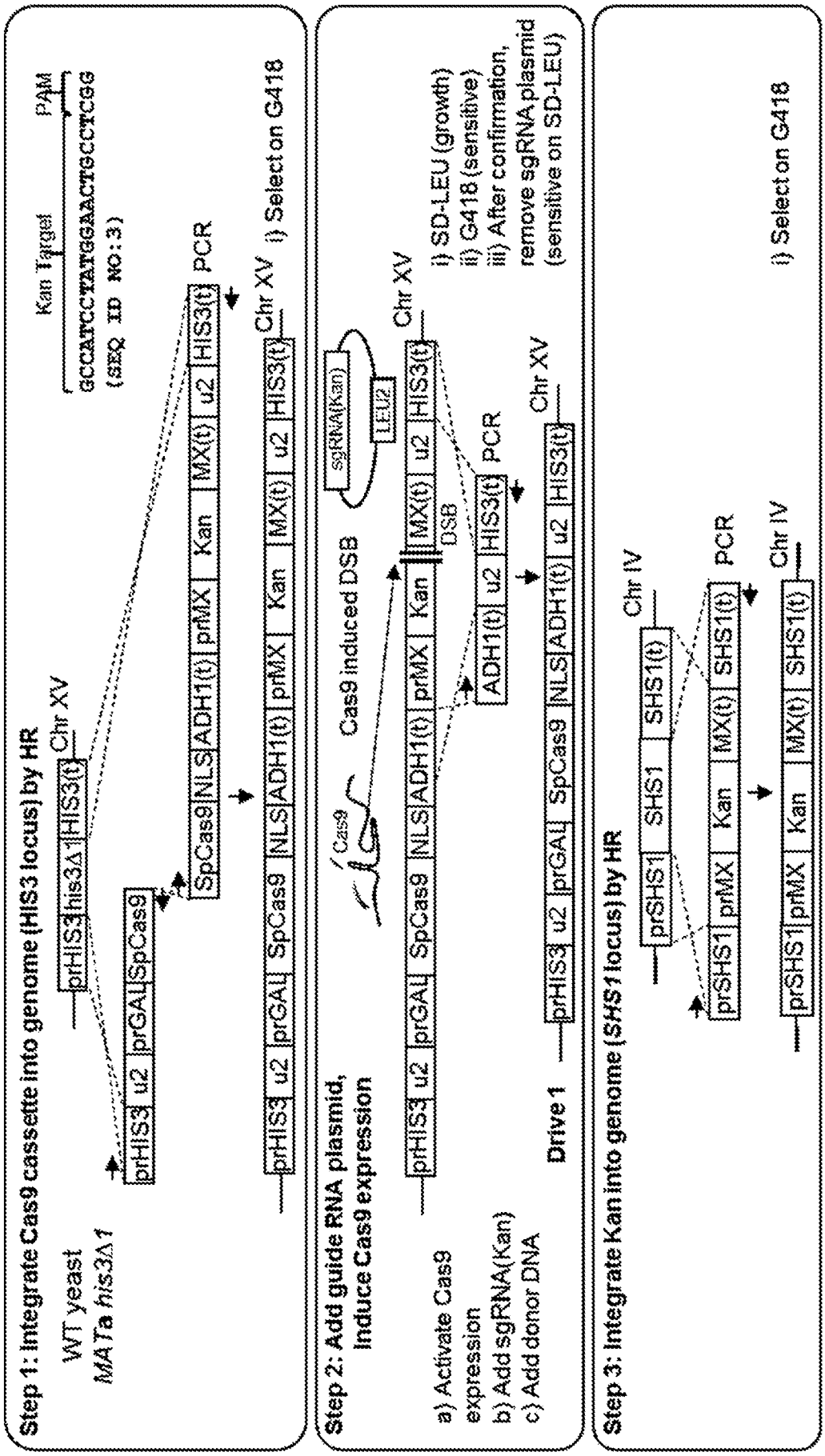

Step 1: Integrate Cas9 cassette into genome (HIS3 locus) by HR
WT yeast
MATa his3Δ1
Kan Target
PAM
GCCATCCTATGGAACTGCCTCGG
(SEQ ID NO:3)
prHIS3 his3Δ HIS3(t)
Chr XV
prHIS3 u2 prGAL SpCas9
SpCas9 NLS ADH1(t) prMX Kan MX(t) u2 HIS3(t)
PCR
prHIS3 u2 prGAL SpCas9 NLS ADH1(t) prMX Kan MX(t) u2 HIS3(t)
Chr XV
i) Select on G418
Step 2: Add guide RNA plasmid,
Induce Cas9 expression
Cas9
Cas9 induced DSB
sgRNA(Kan)
LEU2
Chr XV
prHIS3 u2 prGAL SpCas9 NLS ADH1(t) prMX Kan
MX(t) u2 HIS3(t)
DSB
i) SD-LEU (growth)
ii) G418 (sensitive)
iii) After confirmation,
remove sgRNA plasmid
(sensitive on SD-LEU)
a) Activate Cas9
expression
b) Add sgRNA(Kan)
c) Add donor DNA
ADH1(t) u2 HIS3(t)
PCR
Chr XV
Drive 1
prHIS3 u2 prGAL SpCas9 NLS ADH1(t) u2 HIS3(t)
Step 3: Integrate Kan into genome (SHS1 locus) by HR
Chr IV
prSHS1 SHS1 SHS1(t)
prSHS1 prMX Kan MX(t) SHS1(t)
PCR
Chr IV
prSHS1 prMX Kan MX(t) SHS1(t)
i) Select on G418

FIG. 5A

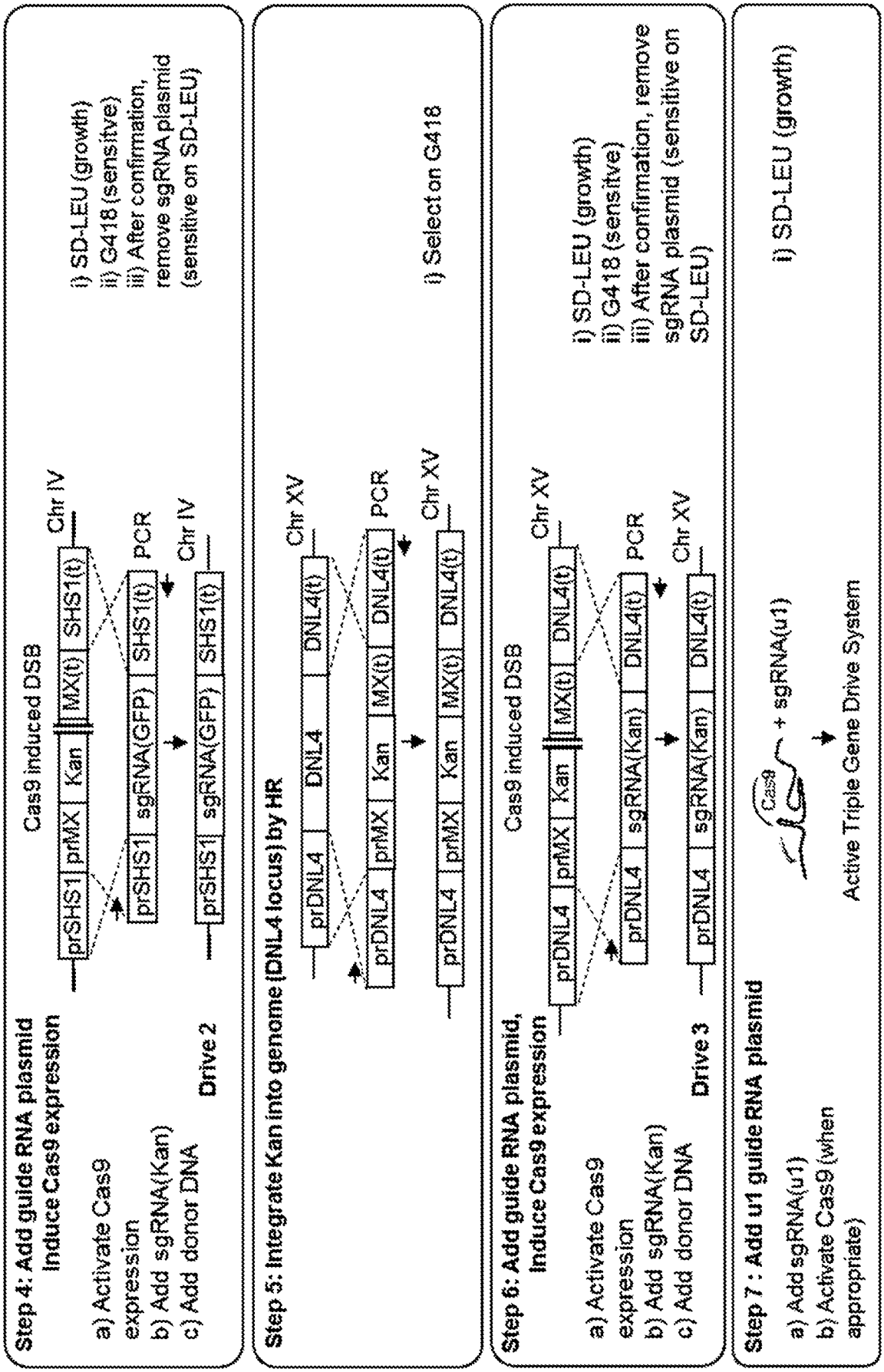
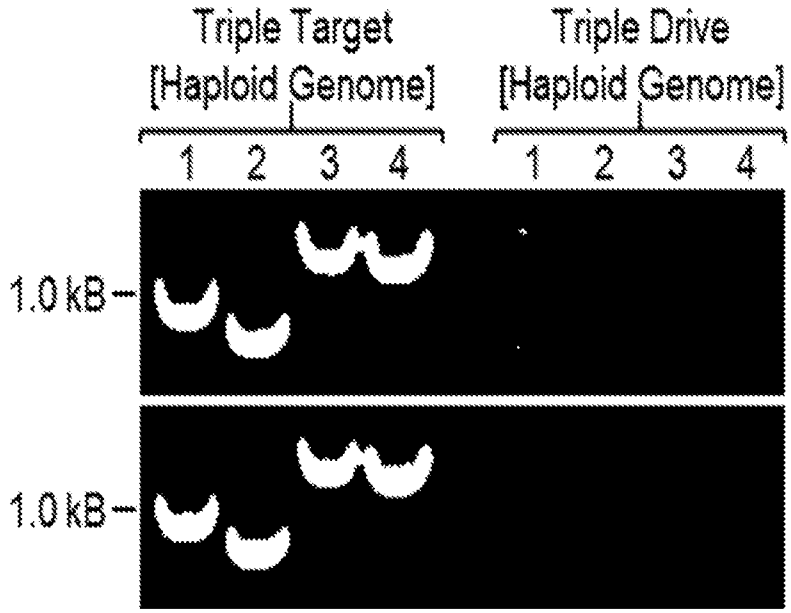

Step 4: Add guide RNA plasmid
Induce Cas9 expression
a) Activate Cas9 expression
b) Add sgRNA(Kan)
c) Add donor DNA
Cas9 induced DSB
prSHS1 prMX Kan MX(t) SHS1(t)
Chr IV
prSHS1 sgRNA(GFP) SHS1(t)
PCR
prSHS1 sgRNA(GFP) SHS1(t)
Chr IV
Drive 2
i) SD-LEU (growth)
ii) G418 (sensitive)
iii) After confirmation, remove sgRNA plasmid (sensitive on SD-LEU)
Step 5: Integrate Kan into genome (DNL4 locus) by HR
prDNL4 DNL4 DNL4(t)
Chr XV
prDNL4 prMX Kan MX(t) DNL4(t)
PCR
prDNL4 prMX Kan MX(t) DNL4(t)
Chr XV
i) Select on G418
Step 6: Add guide RNA plasmid,
Induce Cas9 expression
a) Activate Cas9 expression
b) Add sgRNA(Kan)
c) Add donor DNA
Cas9 induced DSB
prDNL4 prMX Kan MX(t) DNL4(t)
Chr XV
prDNL4 sgRNA(Kan) DNL4(t)
PCR
prDNL4 sgRNA(Kan) DNL4(t)
Chr XV
Drive 3
i) SD-LEU (growth)
ii) G418 (sensitive)
iii) After confimation, remove sgRNA plasmid (sensitive on SD-LEU)
Step 7 : Add u1 guide RNA plasmid
a) Add sgRNA(u1)
b) Activate Cas9 (when appropriate)
Cas9
+ sgRNA(u1)
Active Triple Gene Drive System
i) SD-LEU (growth)

FIG. 5B



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

FIG. 6

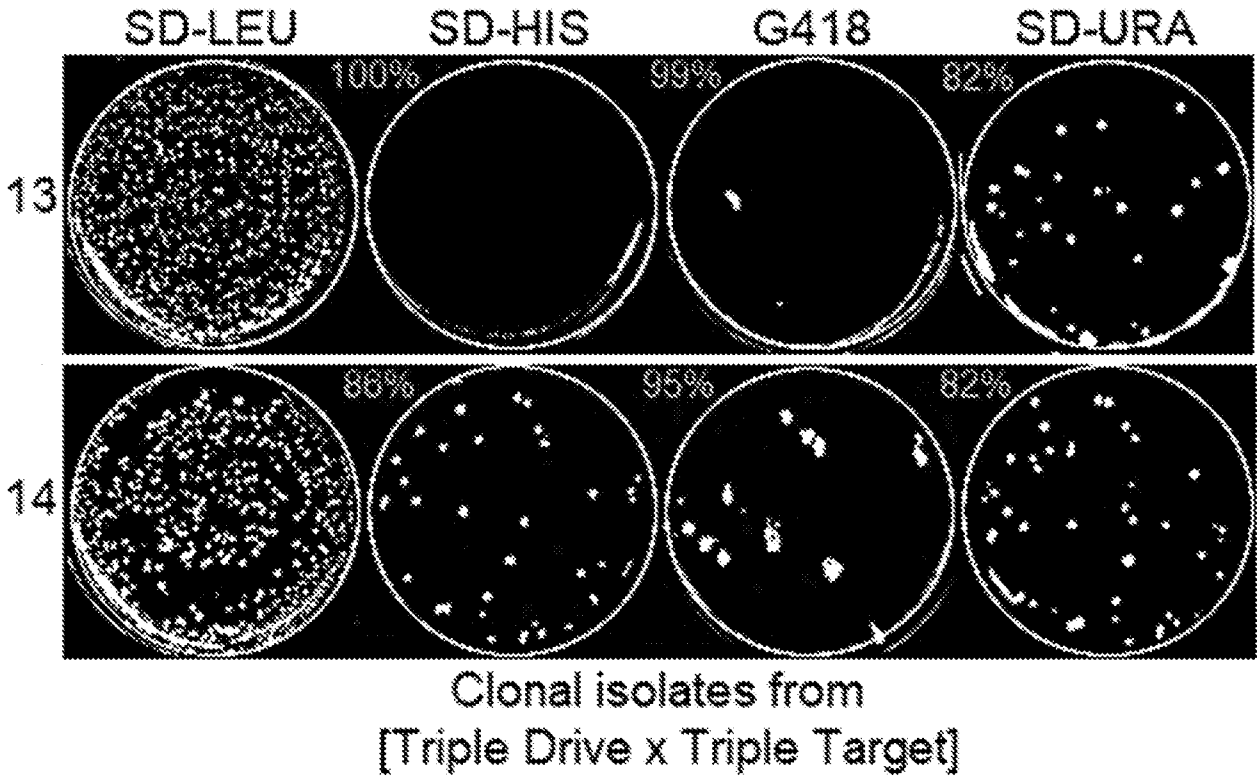


FIG. 7

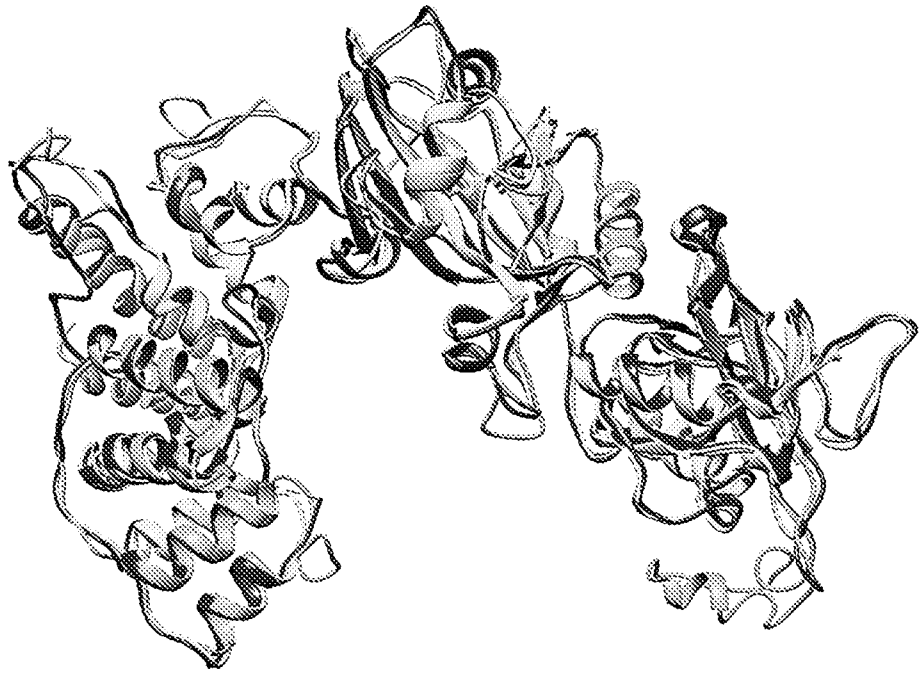
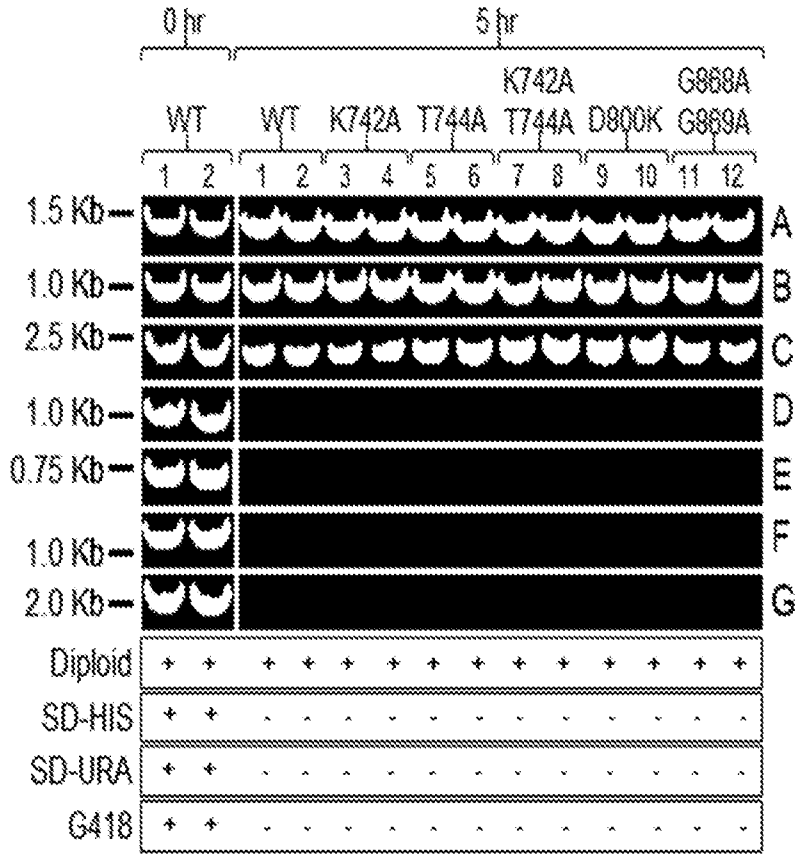


FIG. 8



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

FIG. 9

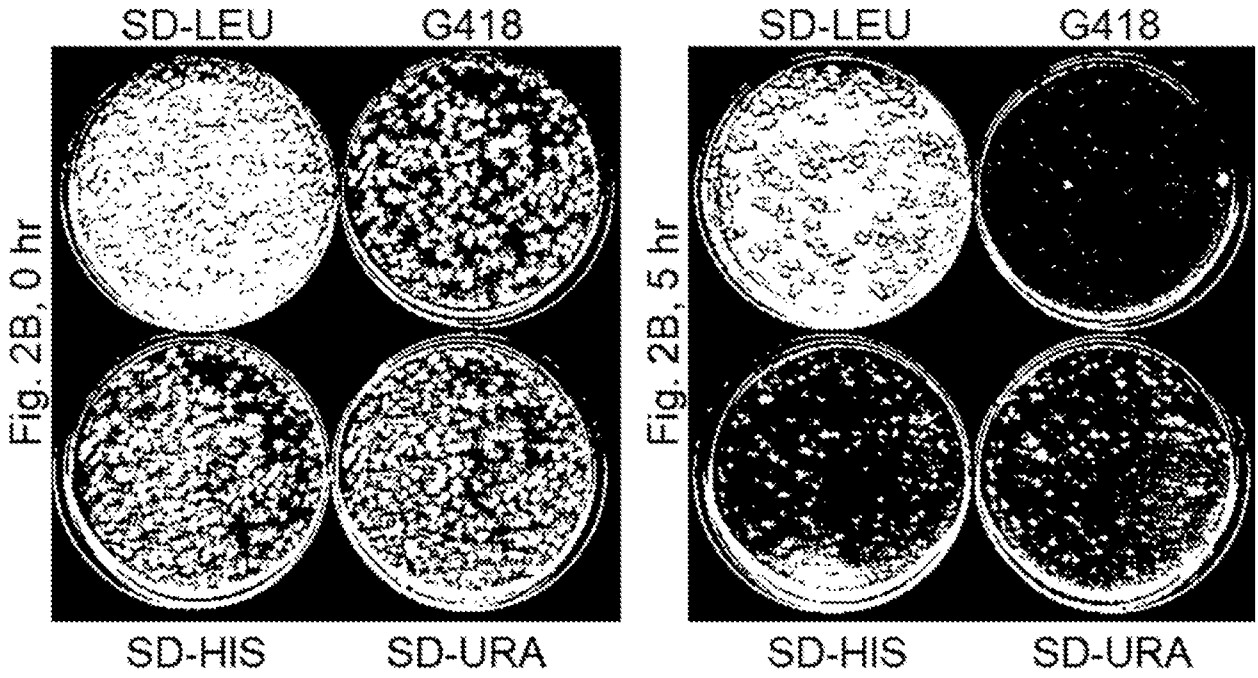

SD-LEU
G418
Fig. 2B, 0 hr
SD-HIS
SD-URA
SD-LEU
G418
Fig. 2B, 5 hr
SD-HIS
SD-URA

FIG. 10

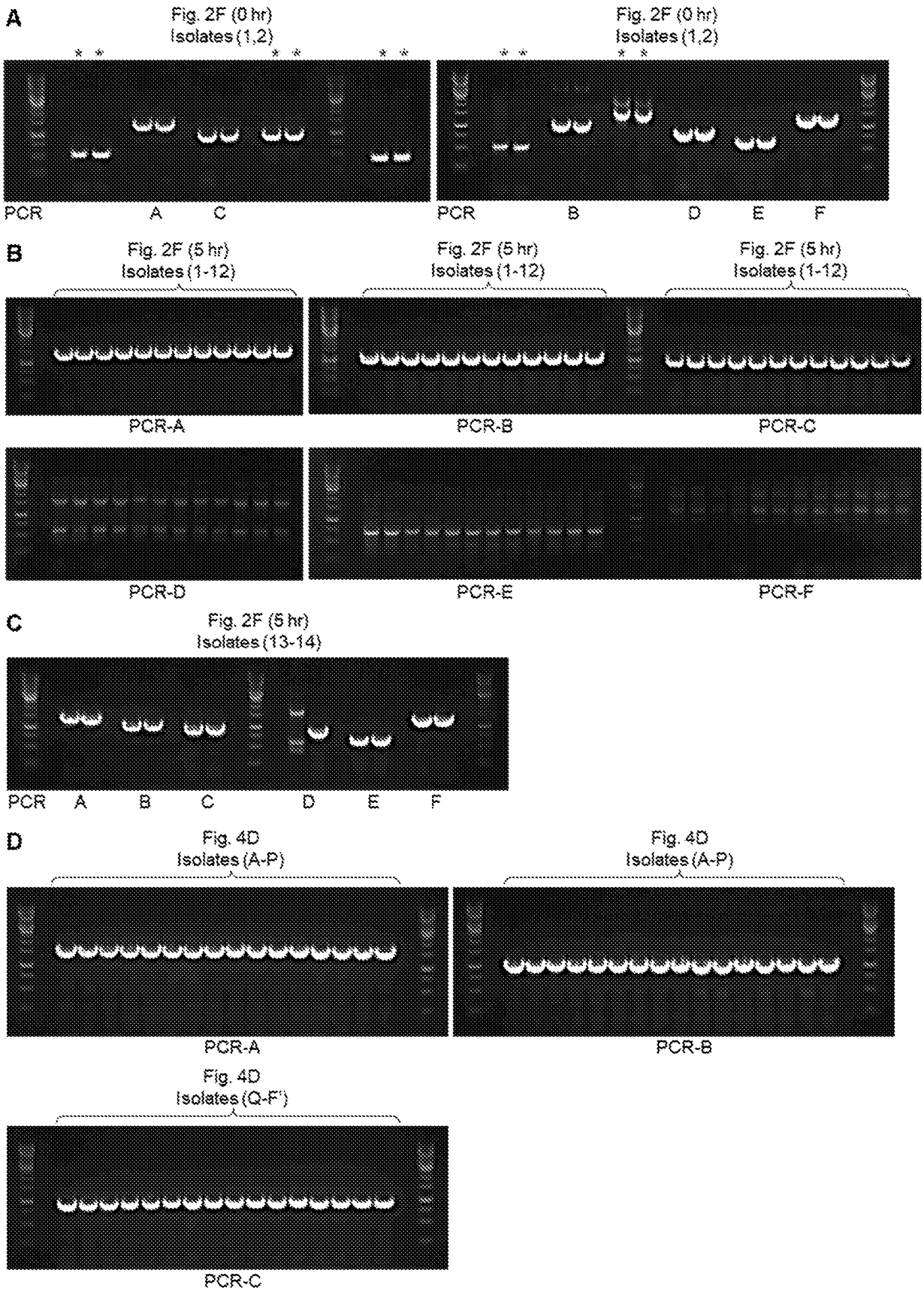

A
Fig. 2F (0 hr)
Isolates (1,2)
PCR
A
C
Fig. 2F (0 hr)
Isolates (1,2)
PCR
B
D
E
F
B
Fig. 2F (5 hr)
Isolates (1-12)
Fig. 2F (5 hr)
Isolates (1-12)
Fig. 2F (5 hr)
Isolates (1-12)
PCR-A
PCR-B
PCR-C
PCR-D
PCR-E
PCR-F
C
Fig. 2F (5 hr)
Isolates (13-14)
PCR
A
B
C
D
E
F
D
Fig. 4D
Isolates (A-P)
Fig. 4D
Isolates (A-P)
PCR-A
PCR-B
Fig. 4D
Isolates (Q-F')
PCR-C

FIG. 11

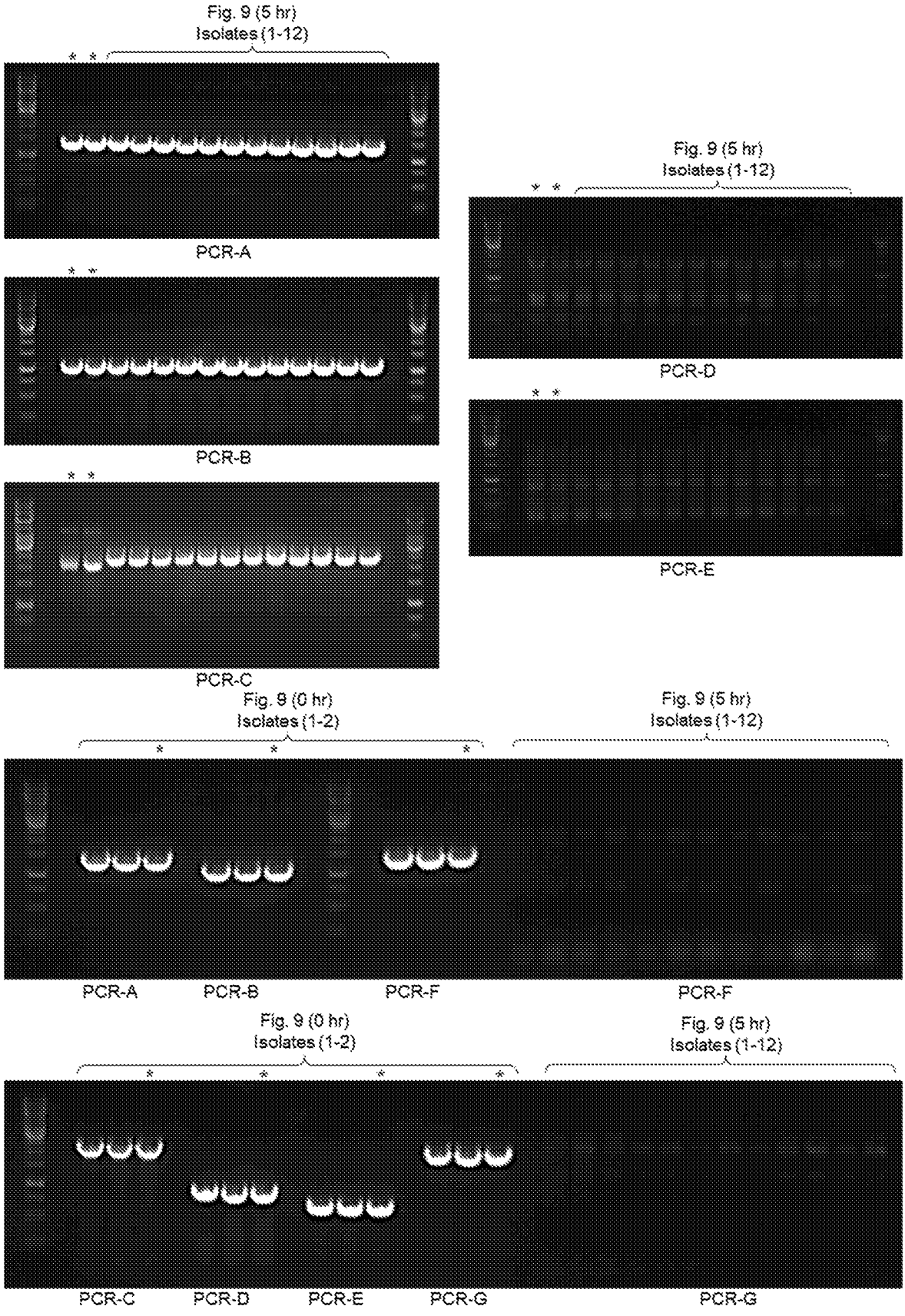

Fig. 9 (5 hr)
Isolates (1-12)
PCR-A
PCR-B
PCR-C
Fig. 9 (5 hr)
Isolates (1-12)
PCR-D
PCR-E
Fig. 9 (0 hr)
Isolates (1-2)
Fig. 9 (5 hr)
Isolates (1-12)
PCR-A
PCR-B
PCR-F
PCR-F
Fig. 9 (0 hr)
Isolates (1-2)
Fig. 9 (5 hr)
Isolates (1-12)
PCR-C
PCR-D
PCR-E
PCR-G
PCR-G

FIG. 12

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US19/41538

A. CLASSIFICATION OF SUBJECT MATTER

IPC - C12N 1/04, 1/15, 1/16, 1/19, 5/04, 9/22, 15/09, 5/10, 15/11, 15/63, 15/79, 15/90 (2019.01)

CPC - C12N 1/04, 1/14, 1/16, 5/04, 9/22, 15/09, 5/10, 15/11, 15/63, 15/79, 15/90, 15/102, 15/113, 15/111

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
See Search History document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2017/058839 A1 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE) 6 April 2017; page 1, 3rd paragraph; page 5, 1st paragraph; page 6, 2nd paragraph; page 6, 3rd paragraph; page 8, 1st paragraph; page 9, 5th paragraph; page 10, 5th paragraph; page 11, 3rd paragraph; page 12, 2nd paragraph; page 17, 1st paragraph; page 18, 3rd paragraph; page 19, 1st paragraph; page 19, 3rd paragraph; page 26, 3rd paragraph; page 28, 4th paragraph; page 35, 3rd paragraph; page 35, 4th paragraph; page 41, 4th paragraph; page 59, 2nd paragraph; page 63, 1st paragraph; page 66, 3rd paragraph; page 68, 5th paragraph; page 72, 1st paragraph; page 74, 3rd paragraph; page 77, 1st paragraph; page 78, 1st paragraph; Figures 2A, 14A	1-13, 14/1-13, 15-20
A	WO 2018/049287 A2 (MASSACHUSETTS INSTITUTE OF TECHNOLOGY et al.) 15 March 2018; entire document	1-13, 14/1-13, 15-20
A	WO 2017/189336 A1 (THE REGENTS OF THE UNIVERSITY OF CALIFORNIA) 2 November 2017; entire document	1-13, 14/1-13, 15-20
A	WO 2018/035388 A1 (THE BROAD INSTITUTE, INC. et al.) 22 February 2018; entire document	1-13, 14/1-13, 15-20
P,X	(YAN, Y et al.) Development of a multi-locus CRISPR gene drive system in budding yeast. Nature. 22 November 2018, Vol. 8; pages 1-12; DOI: 10.1038/s41598-018-34909-3	1-13, 14/1-13, 15-20

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"D" document cited by the applicant in the international application
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
27 September 2019 (27.09.2019)	18 OCT 2019
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Shane Thomas Telephone No. PCT Helpdesk: 571-272-4300