UNIVERSITY OF CALIFORNIA SAN DIEGO

Evolutionary consequences of a CRISPR/Cas9-based gene drive in *Saccharomyces cerevisiae*

A Thesis submitted in partial satisfaction of the requirements
for the degree Master of Science

in

Biology

by

Sean Edward Gomez Guy

Committee in charge:
Professor Sergey A. Kryazhimskiy
Professor Ethan Bier
Professor Justin Meyer
Professor Scott Rifkin

2019

The Thesis of Sean Edward Gomez Guy is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____
Chair

University of California San Diego

2019

EPIGRAPH

7 Wisdom is the principal thing; therefore get wisdom: and with all thy getting get understanding.
8 Exalt her, and she shall promote thee: she shall bring thee to honour, when thou dost embrace her. 9 She shall give to thine head an ornament of grace: a crown of glory shall she deliver to thee.

Proverbs 4: 7-9. King James Version.

TABLE OF CONTENTS

Signature Page ……. iii

Epigraph …….. iv

Table of Contents ……. v

List of Supplemental Files ……....……...…. vi

List of Figures …….....……. vii

List of Tables ..……. viii

Abstract of the Thesis …….. ix

Introduction …….. 1

1. Results …….. 6
    1.1. Asexual Evolution Experiment …… 6
    1.2. Mutation Accumulation Assay …… 10

2. Discussion ……..... 10

3. Materials and Methods ……. 17
    3.1. Asexual Yeast Evolution …….. 17
        3.1.1. Evolution of Replicate Populations …….. 17
        3.1.2. Sample Preparation for Population-wide Genome Sequencing ……. 18
        3.1.3. Calling Small-Nucleotide Variants …….. 18
        3.1.4. Filtering Variant Calls and Allele Trajectories …… 20
    3.2. Mutation Accumulation Assay …… 22
        3.2.1. Strain Construction …….. 22
        3.2.2. Chromosomal Integration of CRISPR/Cas9 Cassettes into Haploid Parent Strains …… 23
        3.2.3. Mating Haploid Parents ……... 23
        3.2.4. Cell Culture and Single-cell Bottlenecks ……. 24

Appendix: Supplemental Figures and Tables …….. 26

Works Cited ……..…… 31

LIST OF SUPPLEMENTAL FILES

File 1. combine_raw_reads.py - The Python 2.7 script (van Rossum, 1997) used to sort and combine FASTQ files from the Illumina sequencer.

File 2. trimmomatic_fastq_040119.sh - The bash script for Nextera adapter removal using Trimmomatic.

File 3. bowtie_samtools_040119.sbatch - The bash script for read alignment using Bowtie 2.

File 4. picard_sam_042219.sh - The bash script for indexing alignments and marking duplicate reads.

File 5. submit_depth_table_042119.sh - The bash script for calculating read coverage in each sample by chromosomal position.

File 6. run_GATK_041819.sh - The bash script for calling variants with GATK UnifiedGenotyper.

File 7. combine_vcfs_042019.sh - The bash script for compiling variant call data across all replicates and timepoints.

File 8. add_depth2vcf.v0.1.py - The Python 2.7 script for adding the actual read depth to the variant calls.

File 9. format_filter_allele_freq.v0.2.py - The Python 2.7 script for filtering allele trajectories.

File 10. plot_allele_freq.v0.4.py - The Python 2.7 script for plotting allele trajectories.

File 11. summarize_dynamics.v0.2.py - The Python 2.7 script for comparing allele trajectories across strains and replicate populations.

LIST OF FIGURES

Figure 1. Asexual evolution of replicate populations of yeast strains with the following constructs: Cas9(-), gRNA(-); Cas9(+), gRNA(+); and Cas9(+), gRNA(-)…. 7

Figure 2. De novo mutations rise in frequency and fix at different rates depending on the presence of Cas9 nuclease. …….. 9

Figure 3. Allele trajectories of gene conversions at loci that were heterozygous in the ancestors vary slightly depending on the presence of Cas9. ……. 11

Figure 4. An overview of the mutation accumulation assay to capture the mutagenic effects of a CRISPR/Cas9-based gene drive and Cas9 nuclease in S. cerevisiae. …… 13

Figure 5. Cas9(-), gRNA(-) populations' allele trajectories from the asexual evolution experiment. …… 28

Figure 6. Cas9(+), gRNA(+) populations' allele trajectories from the asexual evolution experiment. …… 29

Figure 7. Cas9(+), gRNA(-) populations' allele trajectories from the asexual evolution experiment. …… 30

LIST OF TABLES

Table 1. Primers used for gene drive cassette construction with homology arms for RM11-1a strains. …… 26

Table 2. Selection media for plasmid maintenance, transformant selection, and mating selection. …… 27

ABSTRACT OF THE THESIS

Evolutionary consequences of a CRISPR/Cas9-based gene drive in *Saccharomyces cerevisiae*

by

Sean Edward Gomez Guy

Master of Science in Biology

University of California San Diego, 2019

Professor Sergey Kryazhimskiy, Chair

Vector-borne diseases pose a significant health threat to millions. Genetic modification through CRISPR/Cas9-based gene drive technologies may be able to solve this problem by rapidly spreading disease resistance throughout vector populations. However, the long-term effects of active gene drive elements at evolutionary scales has yet to be extensively explored. To address this issue, we constructed *S. cerevisiae* strains that express both Cas9 and gRNA, Cas9 and no gRNA, or neither Cas9 nor gRNA. We founded replicate populations and propagated these over 1000 generations. From whole-population, whole-genome sequences, we identified changes in allele frequency in each population. By comparing the numbers of alleles to appear or fix in each population, we find evidence of higher mutation rates in strains expressing Cas9

nuclease. In parallel, we directly measure the mutation rate differences due to gene drive or Cas9 alone. Our work underscores the need to control CRISPR/Cas9 activity to limit unintended effects in natural populations.

Introduction

The spread of tropical, vector-borne diseases such as malaria, chikungunya, dengue fever, lyme disease, Chagas disease, and Zika virus poses a serious risk to the health of millions across the globe. Malaria in particular was attributed to approximately 219 million cases worldwide in 2017 (WHO, 2018). Given that *Plasmodium*, the parasite group that cause malaria, infects humans via transmission by mosquitoes of the genus *Anopheles* (WHO, 2018), governments and public health programs have reduced disease burden by attempting to reduce human contact with the vector (WHO, 2018). Despite such measures, the number of malaria cases is increasing across regions of Africa and Eastern Asia due in part to climate change and limited resources for vector population control (WHO, 2018; Martens et al, 1995; Tanser, Sharp, & le Sueur, 2003; Tulu, 1996; Caminade et al, 2014). It should be noted, however, that the extent of climate change's influence on malaria transmission is still up for debate (Hay et al, 2002; Shanks et al, 2002; Pascual et al, 2006; Gething et al, 2010; Caminade et al, 2014). Concern for the communities threatened by debilitating, vector-borne diseases drives a global effort to improve on current current vector control methods to eradicate such diseases.

To guide the research of newer disease control methods, the strengths and limitations of measures already in place or in development must be considered. Pesticides and pesticide-treated nets have been applied in many regions (WHO, 2018), but these may also pose health risks for people (Bouwman & Kylin, 2009; International Agency for Research on Cancer, 1997; Chen et al, 2019) while also being difficult to apply over large areas (WHO, 2018). Worryingly, mosquito resistance to four main classes of insecticides have been reported in multiple countries (WHO, 2018). Protein-mediated DNA recognition gene editing systems, such as zinc finger nucleases (ZFNs; Bibikova et al, 2002; Porteus & Carroll, 2005; Doyon et al, 2008) and Transcription-

Activator Like Effector Nucleases (TALENs; Christian et al, 2010; Li et al, 2010; Mahfouz et al, 2011; Miller et al, 2011; Smidler et al, 2013; Basu et al, 2015) may be able to replace naturally occurring mosquito populations with ones resistant to the malaria parasite, but these approaches require continually releasing large numbers of modified organisms to outcompete natural strains (Smidler et al, 2013), a resource- and time-intensive endeavor. Another system involves introducing *Wolbachia*-infected, sterile males into a population (Laven, 1967; Curtis & Adak, 1974; Xi et al, 2005; Xi et al, 2006). Uninfected females who mate with the sterile males produce sterile eggs, so continuously releasing sterile males reduces the total population size over time (Laven, 1967; Curtis & Adak, 1974; Xi et al, 2005; Xi et al, 2006). *Wolbachia*-mediated population reduction, however, is liable to accidentally fail: infected, female mosquitoes can breed normally with *Wolbachia*-infected males and replenish the population (Curtis & Adak, 1974; Stouthamer et al, 1999). To improve on current population control systems, an ideal system would minimize risks to the surrounding communities and the environment while having high enough efficacy to make implementation feasible.

One approach with minimal risk to the environment would be to displace or outcompete extant populations with a new strain of disease-resistant vectors. Genetic engineering with ZFNs or TALENs can generate the desired strains, but testing different constructs and loci for incision would involve labor-intensive protein construction (Boch et al, 2009; Christian et al, 2010; Li et al, 2010; Miller et al, 2011; Reyon et al, 2012) and an additional system to spread the modifications throughout a population (Smidler et al, 2013). In 2012, the publication of a CRISPR/Cas9-based gene editing system (Jinek et al, 2012) improved the precision and throughput of genetic engineering, opening more avenues to vector population control (Sternberg & Doudna, 2015). The CRISPR/Cas9 gene editing system derives its sequence-specificity from base-pair complementation of a short-guiding RNA (gRNA) to a target DNA sequence upstream

of a protospacer adjacent motif (PAM) site, canonically a pair of guanines preceded by any other nucleotide or 5'-NGG-3' (Mojica et al, 2009; Jinek et al, 2012). Bound to and positioned by the gRNA, the Cas9 endonuclease then cleaves the double-stranded DNA (dsDNA) three to four base pairs upstream of the PAM sequence (Jinek et al, 2012; Jinek et al, 2013). In response to the dsDNA break, the host's own DNA repair mechanisms complete the gene editing process, either by introducing small insertions and deletions via non-homologous end-joining (NHEJ) and micro-homology directed repair (MHDR) pathways or by inserting new sequences copied from a homologous chromosome or any available template flanked by homologous sequences via homology-directed repair (HDR). As a genetic editing tool, CRISPR/Cas9 provides unprecedented power to manipulate virtually any sequenced genome or locus (Wright, Nuñez, & Doudna, 2016).

Soon after the discovery of CRISPR/Cas9, scientists designed a system that could rapidly distribute the tool throughout a population. CRISPR/Cas9-based gene drives were soon capable of copying themselves from one chromosome to its homologous partner (Gantz & Bier, 2015a; Burt, 2003; Esvelt et al, 2014). In these gene drive systems, the process starts by inserting DNA- or plasmid-encoded CRISPR/Cas9 gene drive elements into a cell (Gantz & Bier, 2015a). Once expressed, the Cas9 nuclease encoded in the cassette cleaves dsDNA at the target locus. When the cell repairs the break via HDR, the gene drive sequence can then be used as a template for repair. This process occurs at both homologous target loci, turning the cell homozygous for the gene drive. If germline cells receive the gene drive, the gene drive be inherited at super-Mendelian rates because the heterozygous gene drive allele will insert itself into the wildtype allele (Gantz & Bier, 2015a). By avoiding the pressures of natural selection, CRISPR-Cas9-based gene drives have the potential to rapidly spread itself throughout a population without the need of repeated treatments as with other systems.

CRISPR/Cas9-based gene drives provide two main approaches to minimizing vector-borne disease transmission: suppressing the vector's population (Deredec, Godfray, & Burt, 2011; Galizi et al, 2014; Hammond et al, 2016; Kyrou et al, 2018) or spreading a parasite-resistance phenotype throughout a population (Burt, 2003; Windbichler et al, 2011; Gantz et al, 2015c). The former approach may result in unexpected ecological effects for the populations competing with and preying upon the vector (Collins et al, 2018). For the latter approach, a small inoculum of gene-drive-carrying individuals should quickly modify most of, if not the entire, vector population to prevent the disease from transmitting without the drawbacks of removing the vector from its ecological niche. Due to our limited ability to replicate and control insect ecology in the lab, we decided to focus on the latter method of population-wide genetic modification.

CRISPR/Cas9-based gene drives have been shown to drive itself and cargo sequences throughout mosquito populations on short-term time-scales (Gantz et al, 2015c; Champer et al, 2019; Hammond et al, 2016; Hammond et al, 2017). Over evolutionary time scales, the consequences of maintaining CRISPR/Cas9-based gene drive elements in eukaryotic organisms have not been well-explored. To this end, we model a gene drive's effects in populations of baker's yeast, *Saccharomyces cerevisiae*. Relative to multicellular insects, yeasts replicate rapidly, allowing us to observe dynamics over hundreds of generations with a matter of months. Maintaining large populations in liquid culture reduces the effects of genetic drift while emphasizing deterministic evolutionary pressures, namely natural selection (Jain & Krug, 2007). On a practical level, extensive genetic databases and protocols make drug and nutrient selection, chromosomal or plasmid integration, controlled mating, and genetic variant analysis much more feasible in yeast than in less well-characterized organisms like mosquitoes. By leveraging yeast as our model organism, we are able to conduct controlled experiments at evolutionarily relevant timescales and population sizes.

By incorporating CRISPR/Cas9-based gene drives into yeast strains, we aim to explore the long-term consequences of the drive elements maintained within a population. Many researchers have noted that Cas9 nuclease activity produces unintended changes to the genome both around the cut site (Allen et al, 2019) and at off-target sequences *in vivo* (Wienert et al, 2019; Zuo et al, 2019; Jin et al, 2019). At the target loci, unintended edits arise from the host's DNA repair mechanisms and how the cell decides to repair the site (Allen et al, 2019). In the context of a gene drive, small indels from NHEJ and MHDR at the target site may disrupt the gRNA's ability to recognize the target sequence, creating a drive-resistance allele (Noble et al, 2017). Off-target DNA damage may arise from promiscuous gRNA complementation (Wu et al, 2014; Singh et al, 2016; Boyle et al, 2017), but this mechanism alone does not explain all off-target effects (Tsai et al, 2017; Newton et al, 2019). How off-target DNA damage affects insect populations remains unclear due to complex interactions among genotype, phenotype, and environment. In short, the unpredictability of DNA damage and repair are cause for reservations concerning the long-term efficacy and safety of CRISPR/Cas9-based gene drive technologies in natural populations.

To model the long-term effects of a CRISPR/Cas9-based gene drive in *S. cerevisiae*, we took two approaches. In the first approach, replicate populations of yeast with the gene drive, without the drive, and with an incomplete drive were propagated for 1000 generations. The aim of this experiment was to observe the patterns in adaptation that result from gene drive. In the second approach, we run a mutation accumulation assay. From this assay, we will profile the mutagenic effects of gene drive elements. Taken together, these experiments set the basis for whether CRISPR/Cas9-based gene drive technology can have a significant effect on the evolutionary fate of a population. We hope that our results will help guide research into

minimizing risks to the public and the environment as we work toward reducing the transmission of infectious disease with gene drive technology.

1. Results

1.1. Asexual Evolution Experiment

This experiment was conducted to observe how population genetics might be affected by the presence of CRISPR/Cas9 gene drive elements. The original *S. cerevisiae*, W303-derived, haploid strains yPH3 and yPH4 were provided by the Desai Lab. We incorporated different CRISPR/Cas9 gene drive elements and mated the haploids to generate three strains. The test strain expressed a "split-drive" system (Esvelt et al, 2014), denoted here as "Cas9(+), gRNA(+)", with its Cas9 nuclease gene on a drug-selectable plasmid and the gRNA in the *ade2* locus. Separating Cas9 and gRNA prevents cells from maintaining the full drive system without drug selection, decreasing the overall chance that the gene drive would successfully enter outside populations (Esvelt et al, 2014). As another cautionary measure, the gRNA targets only an altered gene, *ADE2*, to prevent our construct from driving in wildtype yeast strains (Esvelt et al, 2014). To test Cas9 nuclease's activity independent of the gRNA, we constructed a strain with the full Cas9 plasmid but no gRNA. We also constructed a strain for a baseline for mutation rate under drug selection. This strain had no gRNA and a truncated Cas9 plasmid. From each of the three strains, eight replicate populations were propagated for roughly 1000 generations at bottleneck sizes about 10^5 cells per population (Fig. 1).

We extracted genomic DNA from these populations every 100 generations to obtain whole-genome, whole-population DNA sequences from throughout the 1000 generations of the propagation (Fig. 1). Using a variant calling pipeline from McDonald, Rice, & Desai, 2016, with

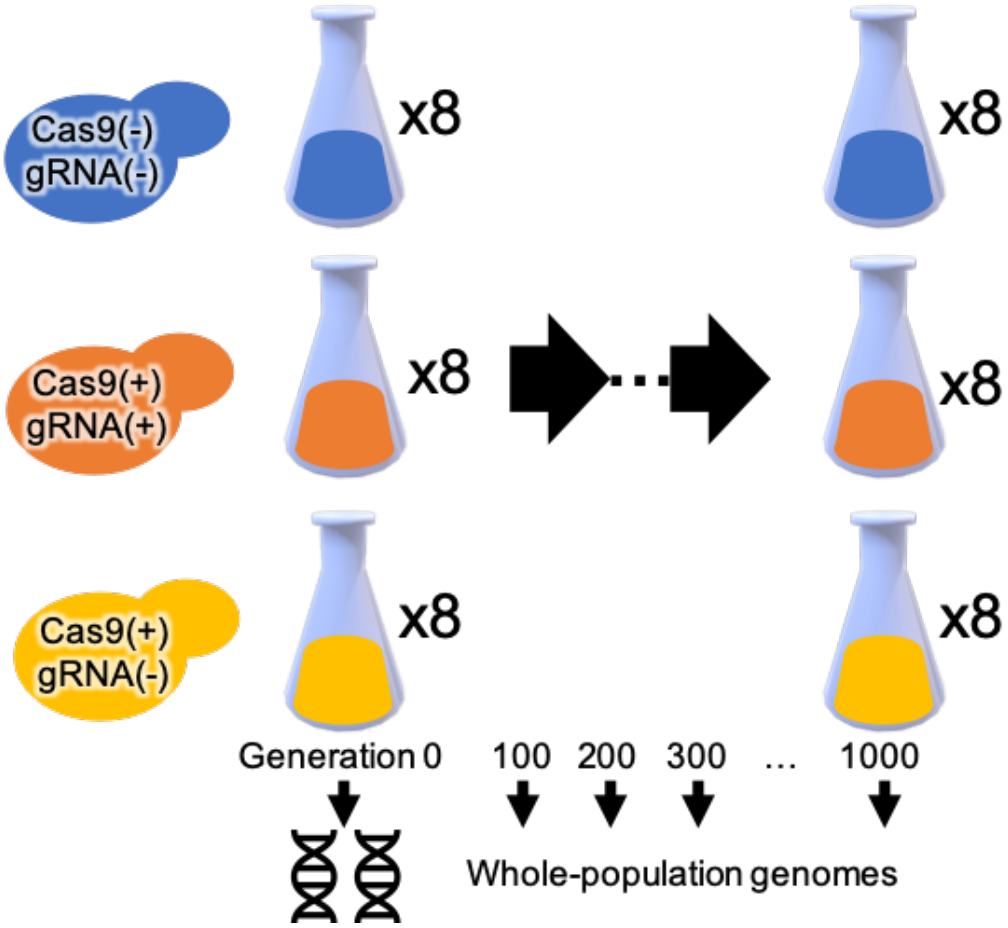

Figure 1. Asexual evolution of replicate populations of yeast strains with the following constructs: Cas9(-), gRNA(-); Cas9(+), gRNA(+); and Cas9(+), gRNA(-).

filters based on Lang et al, 2013, we identified *de novo* mutations and tracked their frequency over time (Fig. 2A-C). Such mutations were expected to start at a frequency of 0.0 and then rise in frequency when part of an adaptive sweep (Atwood, Scheider, & Ryan, 1951). Once a *de novo* mutation rises in frequency, its trajectory will have one of three fates: go to extinction due to clonal interference, or competition with another adaptive lineage (Fig. 2E-G, horizontal shading); go to fixation as a heterozygous allele, which would maintain an allele frequency of 0.5 for the remainder of the experiment (Fig. 2E-G, vertical shading); or go to fixation as a homozygous allele, which would stay at a frequency of 1.0 (Fig. 2E-G, dotted shading). Clonal interference and fixation are normal features of a population adapting under natural selection (Lang et al, 2013). Interestingly, we observed that the Cas9(+), gRNA(+) populations produced more segregant *de novo* mutations than the Cas9(-), gRNA(-) populations (Fig. 2D; $p = 0.233$, Welch's t-test). This difference was even more exaggerated in the Cas9(+), gRNA(-) populations (Fig. 2D; $p = 0.110$, Welch's t-test). However, the proportion of mutations to reach fixation as either heterozygous or homozygous alleles was significantly lower when directly comparing the Cas9(-), gRNA(-) results to the combined results of the Cas9(+) strain (Fig. 2E-G; $p = 0.028$, Welch's t-test). These trends among the *de novo* mutations indicate that Cas9-carrying strains may have a higher mutation rate due to Cas9 nuclease acting independently of its gRNA.

We observe that a significant proportion of *de novo* mutations reach fixation as homozygous alleles across all three strains (Fig. 2E-G, dotted shading). Assuming that all *de novo* mutations entered the population as heterozygous alleles, their conversion into homozygous alleles indicated that these alleles were undergoing gene conversion. This phenomenon was interesting because gene conversions can be the result of dsDNA cleavage followed by HDR (Szostak et al, 1983). To better quantify the occurrence of gene conversions, we identified a total of 111 pre-existing, heterozygous alleles in the three strains and tracked the trajectories of those

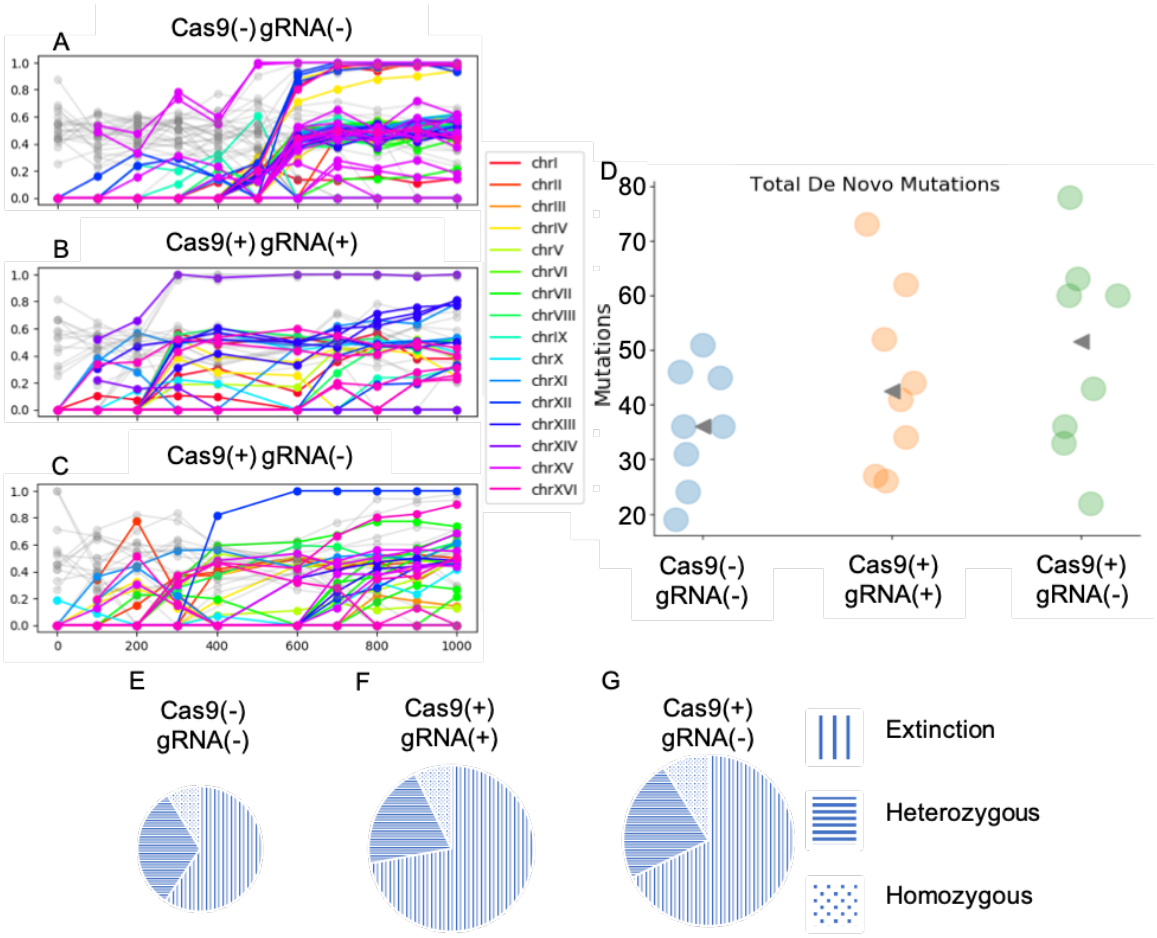

Figure 2. *De novo* mutations rise in frequency and fix at different rates depending on the presence of Cas9 nuclease. (A-C) For the first replicates of the Cas9(-), gRNA(-) strain (A), the Cas9(+), gRNA(+) strain (B), and the Cas9(+), gRNA(-) strain (C), segregant *de novo* mutations are plotted on an axis of allele frequency over time in generations. Individual trajectories are colored by the chromosome on which they appear. Grayed-out trajectories are segregating, initially heterozygous mutations. (D) The total number of segregating, *de novo* allele trajectories detected is compared across the three strains. Each dot represents a single replicate population. Gray arrows designate the median for each strain. Horizontal positions of each point were randomly scattered to reduce the amount of overlapping points. (E-F) The fates of pooled allele trajectories together are compared across strains. Areas shaded with vertical lines denote the fraction of alleles to go extinct. Areas shaded with horizontal lines denote the fraction of alleles to fix at around 0.50 frequency. Areas shaded with dots denote the fraction of alleles to fix at around 1.0 frequency. The relative area of the circles is proportional to the total number of *de novo* mutations found for the corresponding strain.

that deviated from their expected frequency of 0.50 (Fig. 3A-C). Analogous to trends for *de novo* mutations, we observe that both the Cas9(+), gRNA(+) ($p = 0.958$, Welch's t-test) and Cas9(+), gRNA(-) ($p = 0.329$, Welch's t-test) strains encounter slightly more gene conversions sweeping to detectable frequencies (Fig. 3D), while the overall fraction of heterozygous loci to fix in the population after gene conversion appeared to be consistent across all strains (Fig. 3E; $p = 0.592$, Welch's t-test between Cas9(-) and Cas9(+) groups; $p = 0.939$, Kruskal-Wallis test across all three strains). These trends for gene conversions are also an indication that dsDNA damage may occur more frequently in the presence of Cas9 nuclease.

From the asexual evolution experiment, we found evidence suggesting that the overall mutation rate may be elevated in the presence of Cas9 nuclease, especially when it is expressed without a gRNA. We observed similar trends across both segregating *de novo* and initially heterozygous alleles; more *de novo* alleles or gene conversions appeared at detectable frequencies in both the Cas9(+), gRNA(+) and Cas9(+), gRNA(-) populations than in the Cas9(-), gRNA(-) populations (Figs. 2D and 3D). These trends suggest that there may be differences in how these strains are adapting under the same conditions, so further investigation into the underlying processes is warranted.

1.2. Mutation Accumulation Assay

Given that each strain in the asexual evolution experiment was represented by only eight replicate populations, we anticipated that the evolution experiment may have been too underpowered to yield statistically significant results. To determine whether CRISPR/Cas9 was mutagenic to the yeast genome, I designed a mutation accumulation assay wherein replicate lines accumulated new mutations under minimal selective pressure (Zhu et al, 2014; Kondrashov & Kondrashov, 2010). By comparing the ancestral genomes to their descendents from the end of the assay, we will construct a mutation rate profiles for both the full gene drive and Cas9 alone.

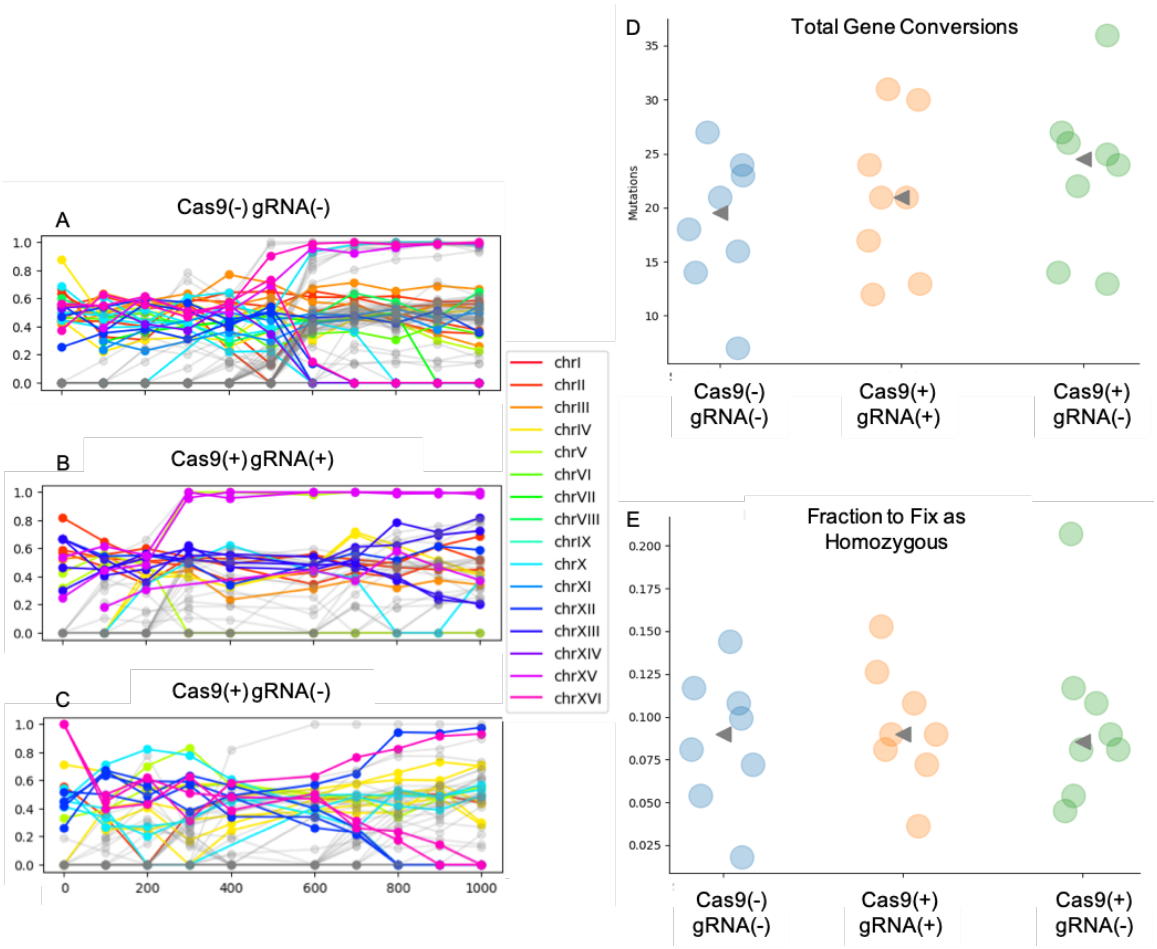

Figure 3. Allele trajectories of gene conversions at loci that were heterozygous in the ancestors vary slightly depending on the presence of Cas9. (A-C) For the first replicates of the Cas9(-), gRNA(-) strain (A), the Cas9(+), gRNA(+) strain (B), and the Cas9(+), gRNA(-) strain (C), segregating, initially heterozygous alleles are plotted on an axis of allele frequency over time in generations. Individual trajectories are colored by the chromosome on which they appear. Grayed-out trajectories are *de novo* mutations from the corresponding population. (D) The total number of segregating, initially heterozygous allele trajectories and (E) the fraction of the total 111 detected loci to fix as homozygous alleles are compared across the three strains. Each dot represents a single replicate population. Gray arrows designate the median for each strain. Horizontal positions of each point were randomly scattered to reduce the amount of overlapping points.

In the asexual evolution assay, gene conversions were particularly difficult to reliably measure given that we could only observe gene conversions at 111 heterozygous loci across the ancestral genome. To address this, I constructed three new *S. cerevisiae* strains analogous to those in the asexual evolution experiment. For this set of strains, we mated haploids from two highly diverged backgrounds, YAN501 derived from RM11-1a (Brem et al, 2002) and BY4741 (Brachmann et al, 1997) from the Andrews Lab and Desai Lab, respectively. These strains differ by approximately $4.6x10^4$ small-nucleotide polymorphisms (SNPs; Qi et al, 2009), so a cross between the two strains should yield that many heterozygous loci to monitor for gene conversion rates. Relative to the strains used in the asexual evolution experiment, the new strains are capable of yielding more precise measurements of the rate of gene conversion and better inform us of Cas9 nuclease's activity throughout the genome.

We passaged 95 replicate lines per strain through over 40 bottlenecks, or approximately 500 generations. DNA samples from the ancestral strains and their 285 total descendants are in the process of preparation for high-throughput sequencing. We will compare the genomes of our ancestral strains to the descendents to observe the numbers and type of mutations over time. By incorporating read depth data, we may be able to measure frequencies of large chromosomal rearrangements. By the end of the data analysis, we will have mutation profiles for each of the three strains that will inform us of whether Cas9 actually introduces significant numbers of off-site mutations.

2. Discussion

In the asexual evolution experiment, we observed elevated rates of the *de novo*, segregant mutations appearing in both Cas9(+), gRNA(+) and Cas9(+), gRNA(-) yeast populations (Fig. 2D), and a similar trend was observed for gene conversion rates (Fig. 3D). These results suggest that Cas9 nuclease increases mutation rates in yeast, which in turn affects the allele dynamics

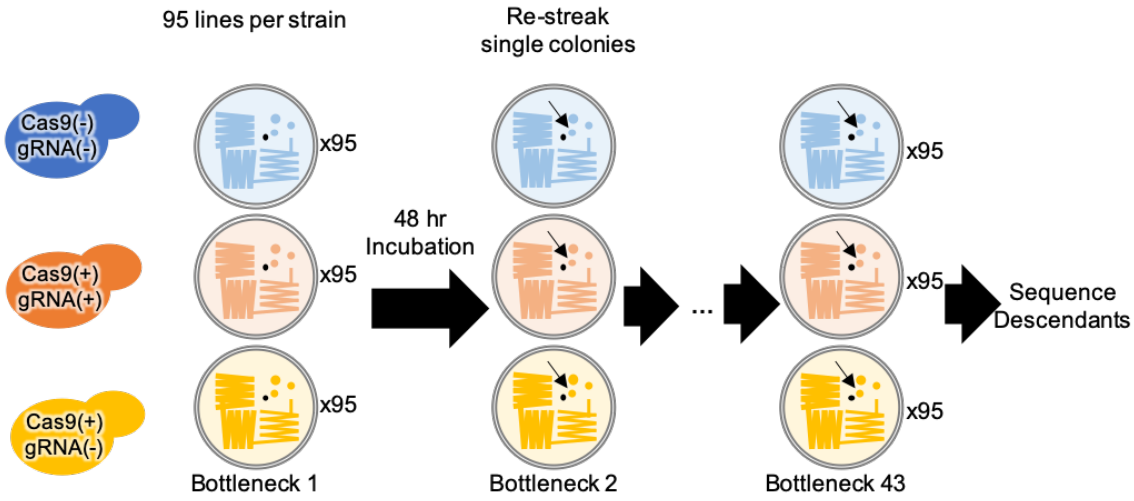

Figure 4. An overview of the mutation accumulation assay to capture the mutagenic effects of a CRISPR/Cas9-based gene drive and Cas9 nuclease in *S. cerevisiae*.

within populations. Further analysis of the evolution experiment's allele trajectory data may clarify the underlying processes by which Cas9 causes these changes. Additionally, the mutation accumulation assay will inform us of Cas9-specific mutagenic effects in the absence of selection. Once we arrive at the full data set, we will determine whether CRISPR/Cas9-based gene drive elements actually affected the evolutionary dynamics of our populations.

One confounding variable hampering our ability to differentiate among strains in the asexual evolution experiement was the use of zeocin to maintain the Cas9 plasmids. Zeocin selection was required in these strains because other drug resistance and auxotrophic markers were already in use for the controlled mating in a related experiment. At 50ng/µl, the concentration of zeocin during evolution likely did not kill all cells that lacked a functioning plasmid and prevent them from spontaneously gaining resistance (Gatignol, Baron, & Tiraby, 1987). For reference, an earlier study had estimated that spontaneous resistance in *S. cerevisiae* strains OL1 and GRF 18 occurs at a rate of $3x10^{-6}$ at 1ug/mL phleomycin on YPD plates (Gatignol et al, 1987). Under the assumption that our strains gain resistance at the same rate, we expect resistance to spontaneously arise once every three to four generations in our populations of 10^5 cells. Resistance mutations to drug selection may be responsible for driving most adaptive sweeps even in strains with functional ShBleoR genes. In addition to adding selective pressure in the evolution experiment, zeocin itself generates dsDNA breaks (Povirk et al, 1977), so the basal mutation rate is expected to be elevated relative to evolution in non-selective media (i.e., YPD broth alone). Despite the added selective pressure and mutagenic properties from zeocin selection, we still trust that our experiment captures the effects of Cas9 since all populations were grown in the same conditions. The relative differences among the strains should be agnostic to the particular selection pressures and basal mutation rates due to external conditions. In light of

zeocin's properties, the strains for the mutation accumulation assay were designed to not require drug selection since the gene cassettes were incorporated into the genome.

Another confounding factor arise from the mutagenic effects of strain construction. As stated before, we identified 111 total initially heterozygous, chromosomal loci across the three strains in the evolution experiment. Ideally, we expected these strains to be completely homozygous, owing to their parents yPH3 and yPH4 coming from the same evolved strain H02.6 from the Desai Lab. The initially heterozygous loci may have come from original strain H02.6 or from transformation with gRNA and Cas9 cassettes. If mutations were introduced by strain construction on the haploids, the three strains may have differed at loci outside of the Cas9 plasmid and gRNA insertion. As a consequence, one strain might have started with a known mutator phenotype (Smith et al, 2004; Serero et al, 2014; Stirling et al, 2014) or a significantly different population-wide fitness. In either case, evolutionary dynamics of the populations may have been biased due to alleles outside of the gene drive elements. In consideration of these heterozygous mutations, we plan to identify which heterozygous mutations are from which strain and cross-check them with the known set of mutator alleles. We also will conduct growth curve assays to compare population-wide fitness across strains before and after evolution. Ultimately, we will be able to identify whether unintended alleles from before the evolution likely skewed our results and adjust our calculations accordingly.

Moving forward, we would like to determine how gene drive elements might affect selective pressures in a population. CRISPR/Cas9 constructs may impose a fitness cost by using up significant resources for gene expression (Lang, Murray, & Botstein, 2009) or causing cytotoxicity via DNA damage (Aguirre et al, 2016; Morgens et al, 2017). By starting at a lower fitness cost, populations with active Cas9 may find that some otherwise neutral mutations confer a fitness advantage by specifically counteracting Cas9. With more beneficial mutations initially

available, Cas9(+) strains might have access to more evolutionary paths than without Cas9. In the evolution experiment, we observe that more *de novo* mutations appear in populations with active Cas9 elements, yet similar numbers of alleles fix across all strains. Connecting these trends to underlying adaptive processes will require additional information. The key data currently being analysed include the cellular functions that were included in adaptive sweeps and the mutations likely to occur in each strain. These data will require identifying the genomic context of segregant alleles and building the mutagenic profiles from the results of the mutation accumulation assay. Taken together, all of this information should help us piece together whether the presence or absence of Cas9 had any effect of the fate of these populations.

Through further analysis of our *S. cerevisiae* CRISPR/Cas9 gene drive system, we hope to better understand the evolutionary consequences of gene drive for populations. From the asexual evolution experiment, we found that CRISPR/Cas9 elements may introduce more mutations population-wide level over time. and that Cas9 that has lost a specific gRNA may introduce even more mutations as a result. In the context of modifying mosquito vectors to resist malaria, changing the evolutionary path of vector populations may end up accelerating their adaptation toward pesticide or gene drive resistance. Our results underscore the importance of researching more controlled, more precise systems for CRISPR/Cas9-based gene drive (e.g. Cox et al, 2017; Gaudelli et al, 2017; Noble et al, 2019; Gangopadhyay et al, 2019) before modifying the genetic makeup of a natural population. With a mutagenic profile from the mutation accumulation assay, we may be able to improve assays to identify chromosomal artifacts directly caused by unintended Cas9 activity. Alternatively, if the mutation accumulation assay finds that the mutation rate of Cas9-carrying lines were not higher than the control, we would try to better understand the cause of cytotoxicity of Cas9 expression with and without its gRNA. Ultimately, this work aims to help assess the safety of CRISPR/Cas9-based gene drives. As with almost any

nascent technology, there is still much to learn about the short- and long-term impacts on the environment and people whom scientists aim to help with gene drive technologies. With the potential risks involved in genetic engineering, it is the ethical responsibility of the scientific community to mitigate the negative impacts of our collective work.

3. Materials and Methods

3.1. Asexual Yeast Evolution

3.1.1. Evolution of Replicate Populations

Cas9(-), gRNA(-); Cas9(+), gRNA(+); and Cas9(+), gRNA(-) split-drive strains were constructed by K. Carolino. These were designed to also be suitable for a controlled mating protocol (McDonald et al, 2016), though this part of the project is not discussed here. To begin the evolution experiment, the three strains were streaked to single colonies on YPD agar with 50ng/µl zeocin and incubated for about 48 hours. From a single colony of each strain, eight replicate populations were started by inoculating 10ml of YPD broth with 50ng/µl zeocin in non-baffled, non-airtight, 50ml erlenmeyer flasks. The selection by zeocin maintained the Cas9 and truncated Cas9 plasmids in these populations. The cultures were incubated at 30°C while shaking upright at 110 rpm. Every 24 hours, 10ul from each culture was transferred to a new flask with 10ml of YPD zeocin broth.

Samples were collected and frozen at about every 50 generations, or every five dilutions. Each sample was stored separately in 2ml centrifuge tubes. In each well, 500ul of autoclave-sterilised 75% glycerol solution was vortexed with 1ml of cell culture. The tubes were then stored in -80°C until the DNA extraction step.

3.1.2. Sample Preparation for Population-wide Genome Sequencing

DNA from each replicate and 100-generation time-point was extracted via the column-based, chloroform extraction protocol in the YeaStar DNA Extraction kit (Zymo Research). Whole-population, genomic DNA (gDNA) samples were prepared for sequencing on the Illumina HiSeq platform using a modified version of Illumina's Nextera protocol (Kryazhimskiy et al, 2014). Samples were stored at -80°C until ready to submit for sequencing. Generations 100 to 400 of replicate populations 1, 2, and 3 for each strain were sequenced separately before the other samples. The remaining samples were sequenced twice to increase read coverage to over 100X.

3.1.3. Calling Small-Nucleotide Variants

Our DNA sequence alignment and variant calling pipeline follows the one in Lang et al, 2013. In summary, this data analysis pipeline took the raw reads from multiple Illumina Hi-Seq lanes, filtered and aligned the reads to reference genomes, called putative mutations from the alignments, and mapped allele frequency trajectories over time to explore the evolutionary dynamics among our replicate populations.

As mentioned above, sequencing was completed in three separate batches. First we combined raw read sequence files by DNA sample (Supplemental Files 1). Next, Illumina adapter sequences were removed using Trimmomatic V0.32 (Bolger, Lohse, & Usadel, 2014; Supplemental Files 2). In the process, Trimmomatic also sorted reads into paired and unpaired categories depending on whether each read had a reverse-complementary read in both or one of the paired FASTQ files.

After trimming the reads, we used Bowtie2 (Langmead & Salzberg, 2012) to align the reads to strain-specific reference genomes constructed by J. P. Shaffer (Supplemental Files 3). We then indexed the alignment files with samtools (Li et al, 2009) before marking duplicate reads with GATK MarkDuplicates (McKenna et al, 2010; Van der Auwera et al, 2013; Supplemental

Files 4). These reads likely came from PCR amplification and therefore should not have been considered as separate reads when scanning for putative mutants. I must note that some BAM files would crash MarkDuplicates; these files were excluded from further analyses, so some time points were unavailable in some replicate populations. From these alignments, I used the samtools depth function (Li et al, 2009) to generate tables that mapped coverage across the genome for each of the samples (Supplemental Files 5).

Next, we used GATK HaplotypeCaller (McKenna et al, 2010; Van der Auwera et al, 2013) to identify mutations in each of the sample alignment files. Unfortunately, GATK HaplotypeCaller heavily subsampled reads according to the samtools depth (Li et al, 2009) coverage measurements. Unsatisfied with the apparent subsampling and its effects on allele frequencies estimates, I reverted back to legacy software, UnifiedGenotyper from GATK 2.6 (McKenna et al, 2010; Van der Auwera et al, 2013; Supplemental Files 6). Overall, UnifiedGenotyper appeared to subsample less than HaplotypeCaller, so we used UnifiedGenotyper's output to calculate allele frequencies instead of HaplotypeCaller. Since UnifiedGenotyper was legacy software, all GATK programs further down the pipeline were also from GATK version 2.6.

At this point, all putative variant calls were stored in multiple VCF files (Danecek et al, 2011), one file per replicate population per time point. The next step was merging all files with GATK CombineVariants (Supplemental Files 7). One minor bug was that CombineVariants assumed that the reference sequence was the same for all variant calls. We had to remove a variant call at chr XVI:683022 due to this limitation of CombineVariants. This variant only appeared at a single time point, so it would have been ignored anyway when creating allele trajectories. After the files were merged, another script combined the UnifiedGenotyper data with the read depths measured by samtools depth (Supplemental Files 8).

3.1.4. Filtering Variant Calls and Allele Trajectories

With variant calls and coverage across all replicates and timepoints stored on a single, table-delimited table, I built a script to handle the organization and filtering of the allele frequency information (Supplemental Files 9) for plotting in Matplotlib (Hunter, 2007). Filters were modeled after those used in Lang et al, 2013 and McDonald et al, 2016. All of the following table manipulations and filtering steps were completed using the *pandas* Python library (Oliphant, 2007; Millman & Aivazis, 2011; McKinney, 2010), the NumPy package (Oliphant, 2006), and the statsmodels package (Seabold & Perktold, 2010). From several iterations of plotting allele frequency time courses post-filtering for whole populations, I noted time points at which allele frequencies for all or most putative mutations dramatically dropped to below significance simultaneously only to reappear again in the following time point at roughly the same frequency as before the drop. Such time points were excluded from analysis by the formatting and filtering script. Next, individual allele frequency calls were filtered on the basis of how many reads were output by UnifiedGenotyper to produce the overall allele frequency estimation; those with below 20 total reads were likely to yield imprecise data and were thus excluded. Variant calls were excluded from all time points if they had appeared at above 95% frequency by generation 100 in three or more replicate populations. Such variants were already features of the strains prior the asexual evolution. Mutations were also excluded if they were only at non-zero frequency no more than once or twice in any time course. After filtering this set of mutations out of the data set, values of zero frequencies were assigned to replicate-time-points where no variant was called despite having significant coverage at above 20 bp in the original BAM file alignments. Flags placed by the preceding coverage, fixation, and time point filters were used to prevent overwriting allele frequencies that were previously removed.

The next set of filters relied on time course analyses for each of the variant trajectories. First, variants were sorted two categories: *de novo* or pre-existing heterozygous. The criteria for being heterozygous was that the variant had to appear in at least two replicate populations of one strain and maintain an average frequency of 0.50 ± 0.10 and a standard deviation below 0.10 across all time points in those populations. Mutations that failed these requirements were considered to be *de novo* mutations. *De novo* mutations were further categorized into being present in under three populations or over three populations. Then each variant checked by two filters. The first filter identified noisy allele trajectories by setting a minimum threshold for autocorrelation (Seabold & Perktold, 2010; Parzen, 1963). The second filter set a minimum threshold for the difference between each trajectory's initial frequency, excluding time points filtered due to low coverage, and its maximum or minimum frequency. Thresholds for autocorrelation and frequency change measurements were set differently among the heterozygous, single- or double-population, and multi-population variants. We set the strictest autocorrelation and frequency change thresholds for multi-population variants, a minimum of 0.5 and 0.2 respectively. Heterozygous alleles were given a lower autocorrelation threshold at 0.2. Single- and double-population alleles were held to a minimum frequency change of 0.1 and an autocorrelation of at least 0.2. The results of both filters were combined, and any variants for which all trajectories were flagged by either the autocorrelation or frequency change filters were removed from the dataset. The data structure was then exported bitwise through cPickle for visualization with Matplotlib via a separate Python script (Supplemental Files 10).

To categorize the resulting allele trajectories after filtration, I relied again on the *pandas* package to set different thresholds for different changes in frequency (Seabold & Perktold, 2010; Supplemental Files 11). For heterozygous alleles, I determined which alleles had a maximum or minimum frequency at least 0.2 away from 0.5, the presumed frequency of a heterozygous allele

prior to any gene conversion between homologous alleles. Those trajectories that did pass the 0.2 threshold were counted as gene conversion events for their particular replicate population. I also set a separate threshold that determined whether a particular allele ever fixed within a population by setting a minimum maximum frequency of at least 0.95 or less than 0.10 in the case of the homozygous wildtype genotype outcompeting the heterozygous one. Note that no attempt to sort alleles into distinct genotypes was made, though I did create an unfinished portion of the statistical summary script that would do so by grouping trajectories with high correlation among one another. Alternatively, *de novo* mutations simply had thresholds going in one direction, since in theory they should have started from a frequency of around 0. An increase of at least 0.2 was deemed an adaptive event; reaching at least 0.4 and never going above 0.6 meant that a heterozygous allele had fixed; trajectories that pass 0.7 were likely due to a gene conversion; and those that reached at least 0.90 were considered fixed in their population. After each allele frequency had been sorted by their frequency change threshold, the data tables for both heterozygous and *de novo* alleles were merged in order to count the total number of putative selective sweeps and the total number of gene conversions observed in each population.

3.2. Mutation Accumulation Assay

3.2.1. Strain Construction

To control for potential mutations from the strain construction process, we generated multiple replicate strains per construct. Should a known mutator genotype arise in any particular strain at or before generation 0, we would expect that any mutagenic effect of our construct would be confounded by the resulting elevated mutation rate. First, we knocked-in a functional *HIS3* gene into the BY4741 strain to make the diploid selection easier later on (Table 1). Linear DNA cassettes to be integrated into the *ADE2* locus of the haploid strains were amplified off of a

plasmid encoding a gRNA, Cas9 nuclease, GFP fluorescence marker, and zeocin-resistance gene (Table 1). For the first set of strains, the primers included 40bp homology arms to regions upstream and downstream of *ADE2* in BY4741 and RM-11a (Table 1). In subsequent rounds of PCR, I amplified the gene drive constructs from genomic DNA of strains which already contained integrated cassettes, including 800bp homology arm regions to improve transformation efficiency (Table 1).

3.2.2. Chromosomal Integration of CRISPR/Cas9 Cassettes into Haploid Parent Strains

In the first round of transformations, we inserted a functional *HIS3* into BY4741 at its native locus, providing one of two diploid markers for the mating protocol. Having generated the gene drive constructs, we next integrated the constructs into parent haploid strains, following a standard lithium acetate yeast transformation protocol. Transformants were selected by plating onto YPD agar plates with 100μg/mL nourseothricin (Table 2). The transformation protocol was repeated until several replicate transformant strains were generated for each of the three cassettes, those being the full gene drive, the gene drive without its gRNA, and the drug marker without gene drive elements, in both parent backgrounds, YAN501 and BY4741. These replicates served to control for the potential mutagenic effects of the chromosomal integration step itself. For confirmation and storage, a single colony of each strain was individually grown in 2mL of YPD broth shaking at 220 rpm in 30°C for about 48 hours. 1mL of each strain's culture was vortexed with 333uL of 75% glycerol in 2mL microcentrifuge tubes to be stored at -80°C.

3.2.3. Mating Haploid Parents

A key design element of the mutation accumulation assay was to use diploid strains whose parents were sufficiently diverged to form numerous heterozygous loci, which would assist us in monitoring the overall gene conversion rates of each strain. As described above, the gene drive and control cassettes were inserted into each haploid separately. Following the

transformation, I paired transformants from the YAN501 parent strain with those of BY4741 with matching gene drive elements.

Prior to mating, I started cultures in 2mL YPD from frozen stock of each transformant haploid. After 48 hours of shaking at 220 rpm in 30°C, the haploids from either parent were crossed by combining 0.5mL of liquid culture from either parent and 8mL of YPD broth. These combined cultures were gently shaken by hand in loosely capped 50mL erlenmeyer flasks and incubated without agitation at 30°C overnight. To select a single from each replicate mating pair, I cross-streaked the mated cultures onto post-mating, double-selection CSM -his +G418 agar plates (Table 2). After incubating the plates for approximately 48 hours, single colonies were selected from each post-mating, double-selection plate and cross-streaked again onto double-selection plates, ensuring that each replicate strain had a single diploid ancestor. Another 48-hour, 30°C incubation later, single colonies from each mating culture were picked to start 2mL YPD broth cultures, which were then shaken at 220 rpm and 30°C overnight. 1mL of the culture was vortexed with 333uL of 80% glycerol to store at -80°C.

3.2.4. Cell Culture and Single-cell Bottlenecks

For the duration of the assay, replicate lines were grown on YPD agar media (Cold Spring Harbor Protocols). To begin the assay, multiple parent strains per treatment, each replicate of the chromosomal integration process, were streaked onto YPD clonNAT agar from frozen 20% glycerol stock stored at -80°C. A single colony of each respective strain was selected by nearest proximity to a premarked position on the agar plate, and those colonies were used to streak seven to twenty-four replicate lines for the mutation accumulation. At the same time, 2ml liquid YPD cultures were started from the same colonies, shaken in culture tubes for two overnights at 30°C, then frozen at -80°C in 15% glycerol for later DNA extraction as the ancestral references. A total of 285 parallel lines underwent about 43 cycles of plating onto YPD agar and incubation at 30°C

for 48 hours. The petri dishes used during the mutation accumulation step were 50 mm in diameter and made of polypropylene, which could withstand repeated bleaching and autoclaving. Roughly 300 sterile dishes were filled with approximately 5ml of YPD agar at least one day prior to each day of cell plating. Prior to plating, each plate was marked with a unique strain number and a single dot. At 2:30 PM, one to three lab members would begin the streaking process. We selected the colonies closest to the marked dot. Then we picked the colonies using autoclave-sterilised, flat, wooden toothpicks and cross-streaked them onto a clean YPD agar plate three to four times. The newly streaked plates were incubated at 30°C for 48 hours.

Appendix: Supplemental Figures and Tables

Table 1. Primers used for gene drive cassette construction with homology arms for RM11-1a strains.

Construct	Primer Homology Arms	5' Annealing Site	5' Primer Sequence	3' Annealing Site	3' Primer Sequence
Cas9(-), gRNA(-) for YAN501	40 bp in around ADE2 in RM11-1a	5' end of GFP promoter	GTTGCAGGGCTACGA ACCGGGTAATATTAA GTGATTGACTACTAGA ACTGAAAGGCCATT	3' end of NatMX terminator	TATATTATTTGCTGTG CAAGTATATCAATAAA CTTATATACAGTATAG CGACCAGCATTC
Cas9(+), gRNA(+) for YAN501	40 bp in around ADE2 in RM11-1a	5' end of the Cas9 promoter	GTTGCAGGGCTACGA ACCGGGTAATATTAA GTGATTGACTTCTTTG AAAAGATAATGTAT	3' end of NatMX terminator	TATATTATTTGCTGTG CAAGTATATCAATAAA CTTATATACAGTATAG CGACCAGCATTC
Cas9(+), gRNA(-) for YAN501	40 bp in around ADE2 in RM11-1a	5' end of gRNA promoter	GTTGCAGGGCTACGA ACCGGGTAATATTAA GTGATTGACTCATAGC TTCAAAATGTTTCT	3' end of NatMX terminator	TATATTATTTGCTGTG CAAGTATATCAATAAA CTTATATACAGTATAG CGACCAGCATTC
Cas9(-), gRNA(-) for BY4741	40 bp in around ADE2 in BY4741	5' end of GFP promoter	GTTGCATGGCTACGA ACCGGGTAATACTAA GTGATTGACTACTAGA ACTGAAAGGCCATT	3' end of NatMX terminator	ATAATTATTTGCTGTA CAAGTATATCAATAAA CTTATATACAGTATAG CGACCAGCATTC
Cas9(+), gRNA(+) for BY4741	40 bp in around ADE2 in BY4741	5' end of the Cas9 promoter	GTTGCATGGCTACGA ACCGGGTAATACTAA GTGATTGACTTCTTTG AAAAGATAATGTAT	3' end of NatMX terminator	ATAATTATTTGCTGTA CAAGTATATCAATAAA CTTATATACAGTATAG CGACCAGCATTC
Cas9(+), gRNA(-) for BY4741	40 bp in around ADE2 in BY4741	5' end of gRNA promoter	GTTGCATGGCTACGA ACCGGGTAATACTAA GTGATTGACTCATAGC TTCAAAATGTTTCT	3' end of NatMX terminator	ATAATTATTTGCTGTA CAAGTATATCAATAAA CTTATATACAGTATAG CGACCAGCATTC
Any cassette from gDNA in *ade2*	N/A	843bp upstream of *ADE2*	AGGGAGATTCGACCA ACACT	927bp downstream of *ADE2*	AATATGAGATCGGAG GGCAT

Table 2. Selection media for plasmid maintenance, transformant selection, and mating selection.

Media	Recipe
YPD + zeocin	• Yeast Extract, 20g • Peptone, 40g • 20% dextrose solution, 200mL • Ultrapure water, 1.8L • Zeocin, 100mg in stock solution
YPD + nourseothricin agar	• Yeast Extract, 10g • Peptone, 20g • Bacto-agar, 20g • 40% dextrose solution, 50mL • Ultrapure water, fill to 1L • 100mg/ml clonNAT (nourseothricin), 1mL
CSM –his +G418	• Complete Supplement Mixture –his (Formedium), 0.385g • Yeast Nitrogen Base, 0.85g • Monosodium glutamate, 0.5g • Bacto-agar, 10g • 40% dextrose solution, 25mL • Ultrapure water, fill to 500mL • G418, 100µg

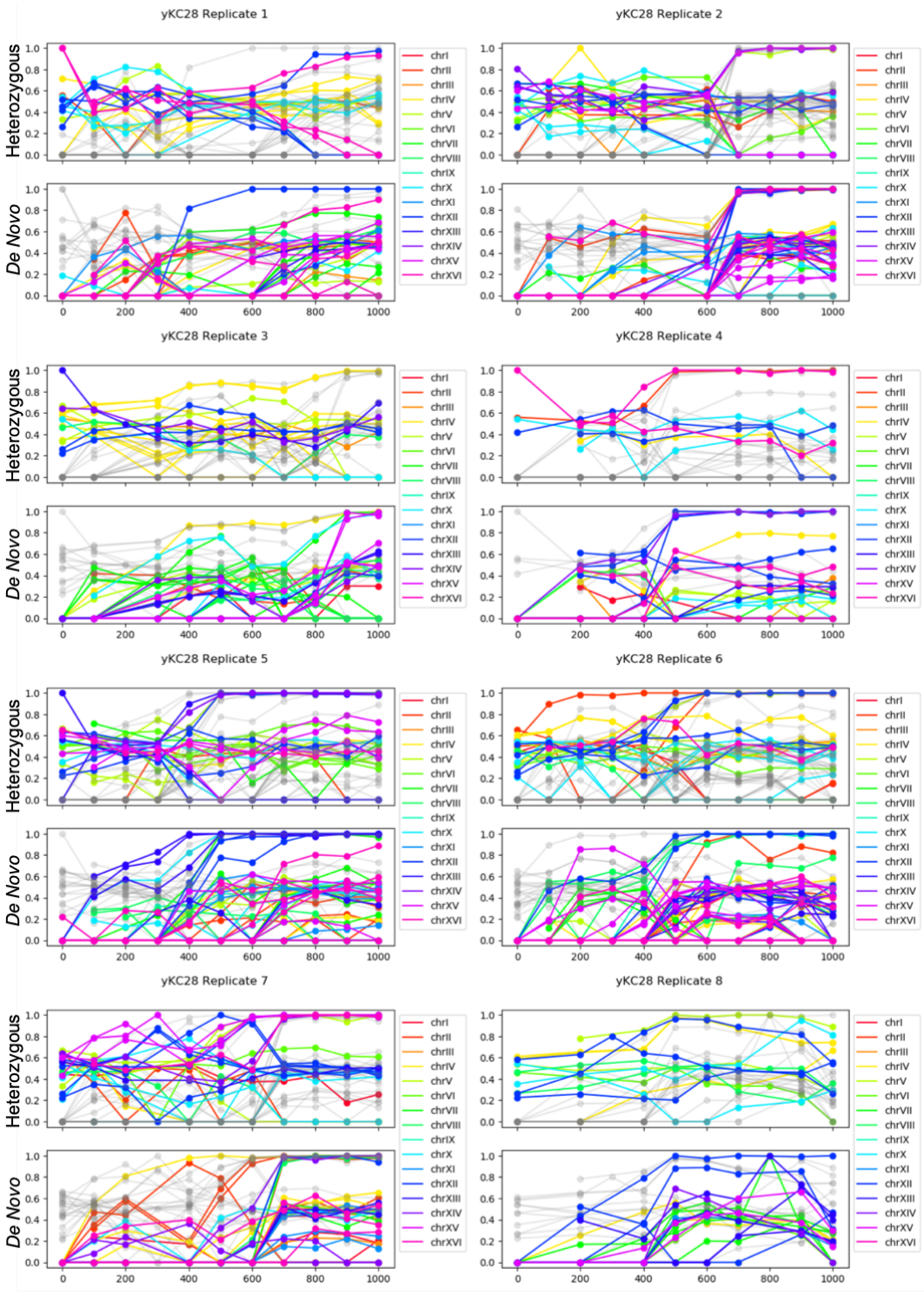

Figure 5. Cas9(-), gRNA(-) populations' allele trajectories from the asexual evolution experiment.

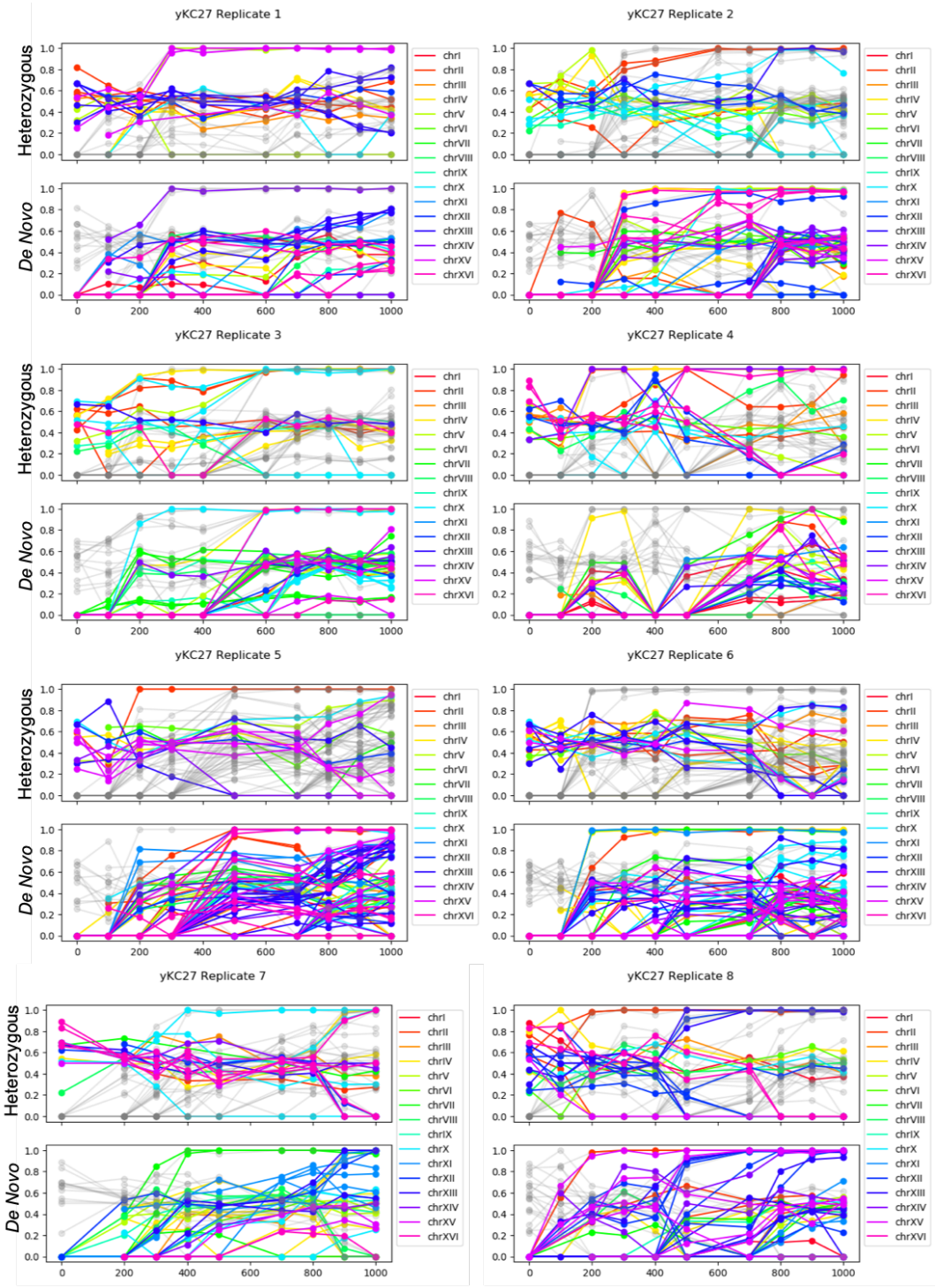

Figure 6. Cas9(+), gRNA(+) populations' allele trajectories from the asexual evolution experiment.

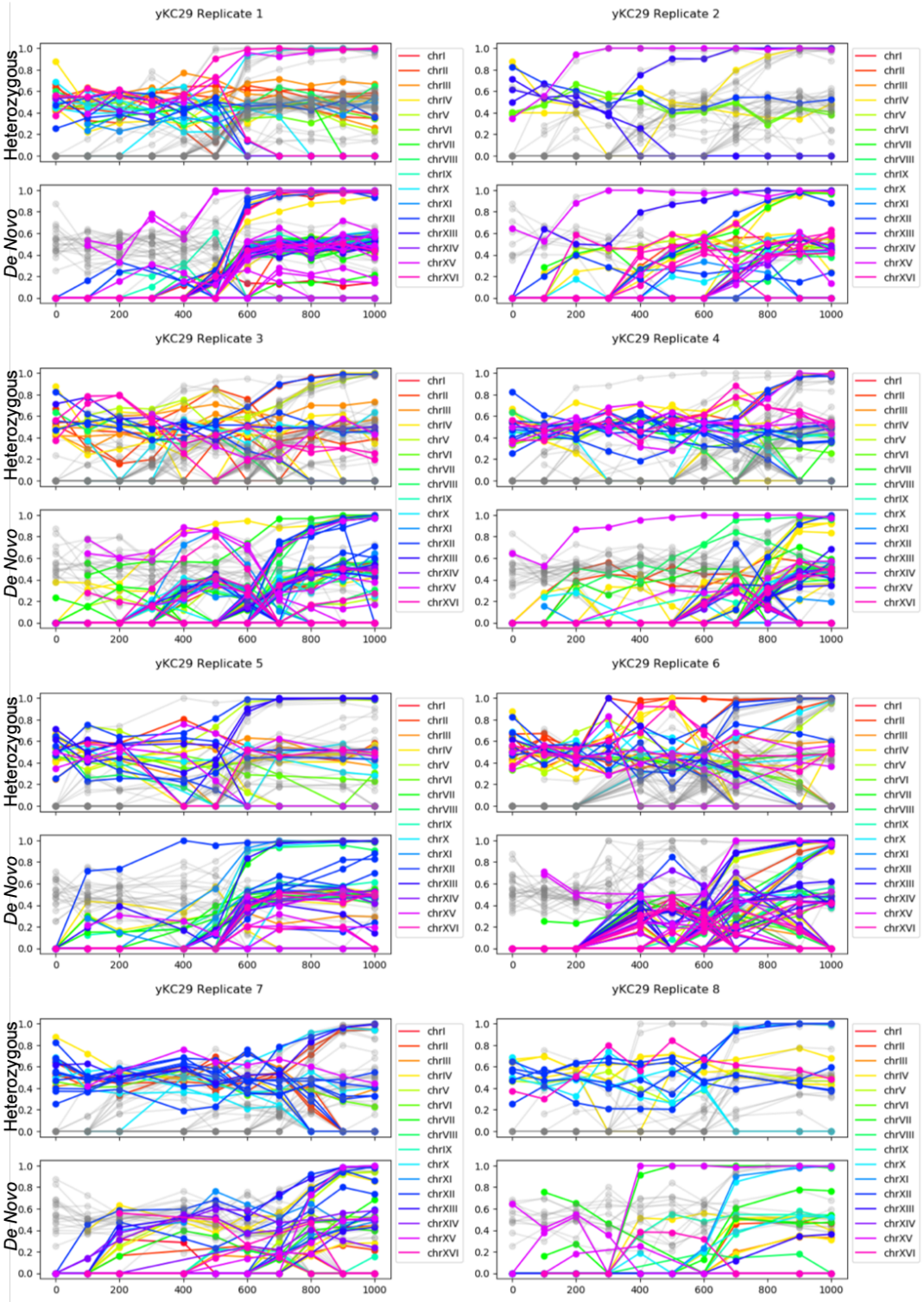

Figure 7. Cas9(+), gRNA(-) populations' allele trajectories from the asexual evolution experiment.

Works Cited

Aguirre, A.J., Meyers, R.M., Weir, B.A., Vazquez, F., Zhang, C.-Z., Ben-David, U., Cook, A., Ha, G., Harrington, W.F., Doshi, M.B., Kost-Alimova, M., Gill, S., Xu, H., Ali, L.D., Jiang, G., Pantel, S., Lee, Y., Goodale, A., Cherniack, A.D., Oh, C., Kryukov, G., Cowley, G.S., Garraway, L.A., Stegmaier, K., Roberts, C.W., Golub, T.R., Meyerson, M., Root, D.E., Tsherniak, A., & Hahn, W.C. (2016). Genomic Copy Number Dictates a Gene-Independent Cell Response to CRISPR/Cas9 Targeting. *Cancer Discovery. 6(8)*: 914-929. doi: 10.1158/2159-8290.CD-16-0154

Allen, F., Crepaldi, L., Alsinet, C., Strong, A.J., Kleshchevnikov, V., De Angeli, P., Páleníková, P., Khodak, A., Kiselev, V., Kosicki, M., Bassett, A.R., Harding, H., Galanty, Y., Muñoz-Martínez, F., Metzakopian, E., Jackson, S.P., & Parts, L. (2019). Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nat. Biotech. 37(1)*: 64-72. doi: 10.1038/nbt.4317

Atwood, K.C., Schneider, L.K., & Ryan, F.J. (1951). Periodic Selection in *Escherichia Coli*. *Proc. Natl. Acad. Sci. U.S.A. 37(3)*: 146-155. doi: 10.1073/pnas.37.3.146

Basu, S., Aryan, A., Overcash, J.M., Samuel, G.H., Anderson, M.A.E., Dahlem, T.J., Myles, K.M, & Adelman, Z.N. (2015). Silencing of end-joining repair for efficient site-specific gene insertion after TALEN/CRISPR mutagenesis in *Aedes aegypti*. *Proc. Natl. Acad. Sci. U.S.A. 112(13)*: 4038-4043. doi: 10.1073/pnas.1502370112

Bibikova, M., Golic, M., Golic, K.G., & Carroll, D. (2002). Targeted Chromosomal Cleavage and Mutagenesis in Drosophila Using Zinc-Finger Nucleases. *Genetics. 161(3)*: 1169-1175. Retrieved from https://www.genetics.org/content/161/3/1169

Brachmann, C.B., Davies, A., Cost, G.J., Caputo, E., Li, J., Hieter, P., & Boeke, J.D. (1997). Designer deletion strains derived from Saccharomyces cerevisiae S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast. 14(2)*: 115-132. doi: 10.1002/(SICI)1097-0061(19980130)14:2<115::AID-YEA204>3.0.CO;2-2

Brem, R.B., Yvert, G., Clinton, R., & Kruglyak, L. (2002). Genetic Dissection of Transcriptional Regulation in Budding Yeast. *Science. 296(5568)*: 752-755. doi: 10.1126/science.1069516

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., & Bonas, U. (2009). Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science. 326(5959)*: 1509-1512. doi: 10.1126/science.1178811

Bolger, A.M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics. 30(15)*: 2114-2120. doi: 10.1093/bioinformatics/btu170

Bouwman, H., & Kylin, H. (2009). Malaria Control Insecticide Residues in Breast Milk: The Need to Consider Infant Health Risks. *Environmental Health Perspectives, 117(10)*: 1477-1480. Retrieved from http://www.jstor.org/stable/40382917

Boyle, E.A., Andreasson, J.O.L., Chircus, L.M., Sternberg, S.H., Wu, M.J., Guegler, C.K., Doudna, J.A., & Greenleaf, W.J. (2017). High-throughput biochemical profiling reveals sequence determinants of dCas9 off-target binding and unbinding. *Proc. Natl. Acad. Sci. U.S.A. 114(21)*: 5461-5466. doi: 10.1073/pnas.1700557114

Burt, A. (2003). Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proc. R. Soc. Lond. B. 270*: 921–928. doi: 10.1098/rspb.2002.2319

Caminade, C., Kovats, S., Rocklov, J., Tompkins, A.M., Morse, A.P., Colón-González, F.J., Stenlund, H., Martens, P., & Lloyd, S.J. (2014). Impact of climate change on global malaria distribution. *Proc. Natl. Acad. Sci. U.S.A. 111(9)*: 3286-3291. doi: 10.1073/pnas.1302089111

Champer, J., Chung, J., Lee, Y.L., Liu, C., Yang, E., Wen, Z., Clark, A.G., & Messer, P.W. (2019). Molecular safeguarding of CRISPR gene drive experiments. *eLife. 8*: e41439. doi: 10.7554/eLife.41439

Chen, K.-X., Zhou, X.-H., Sun, C.-A., & Yan, P.-X. (2019). Manifestations of and risk factors for acute myocardial injury after acute organophosphorus pesticide poisoning. *Medicine (Baltimore). 98(6)*: e14371. doi: 10.1097/MD.0000000000014371

Christian, M., Cermak, T., Doyle, E.L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A.J., & Voytas, D.F. (2010). Targeting DNA Double-Strand Breaks with TAL Effector Nucleases. *Genetics. 186(2)*: 757-761. doi: 10.1534/genetics.110.120717

Cold Spring Harbor Protocols. (2010). *YPD media: Recipe*. Huntington, NY: Cold Spring Harbor Laboratory Press. doi: 10.1101/pdb.rec12315

Collins, C.M., Bonds, J.A.S., Quinlan, M.M., & Mumford, J.D. (2018). Effects of the removal or reduction in density of the malaria mosquito, Anopheles gambiae s.l., on interacting predators and competitors in local ecosystems. *Medical and Veterinary Entomology 33(1)*: 1-15. doi: 10.1111/mve.12327

Cox, D.B.T., Gootenberg, J.S., Abudayyeh, O.O., Franklin, B., Kellner, M.J., Joung, J., & Zhang, F. (2017). RNA editing with CRISPR-Cas13. *Science. 358(6366)*: 1019-1027. doi: 10.1126/science.aaq0180

Curtis, C.F. & Adak, T. (1974). Population replacement in *Culex fatigans* by means of cytoplasmic incompatibility. *Bull. World Health Organ. 51*: 249-255. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2366287/

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., & Durbin, R.; 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics. 27(15)*: 2156-2158. doi: 10.1093/bioinformatics/btr330

Deredec, A., Godfray, H.C.J., & Burt, A. (2011). Requirements for effective malaria control with homing endonuclease genes. *Proc. Natl. Acad. Sci. U.S.A. 108(43)*: E874-E880. doi: 10.1073/pnas.1110717108

Doyon, Y., McCammon, J.M., Miller, J.C., Faraji, F., Ngo, C., Katibah, G.E., Amora, R., Hocking, T.D., Zhang, L., Gregory, P.D., Urnov, F.D., & Amacher, S.L. (2008). Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nat. Biotech. 26(8)*: 702-708. doi: 10.1038/nbt1409

Esvelt, K.M., Smidler, A.L., Catteruccia, F., & Church, G.M. (2014). Emerging Technology: Concerning RNA-guided gene drives for the alteration of wild populations. *eLife. 3*: e03401. doi: 10.7554/eLife.03401

Galizi, R., Doyle, L.A., Menichelli, M., Bernardini, F., Deredec, A., Burt, A., Stoddard, B.L., Windbichler, N., & Crisanti, A. (2014). A synthetic sex ratio distortion system for the control of the human malaria mosquito. *Nat. Comm. 5*: Article 4977. doi: 10.1038/ncomms4977

Gangopadhyay, S.A., Cox, K.J., Manna, D., Lim, D., Maji, B., Zhou, Q., & Choudhary, A. (2019). Precision Control of CRISPR-Cas9 Using Small Molecules and Light. *Biochemistry. 58(4)*: 234-244. doi: 10.1021/acs.biochem.8b01202

Gantz, V.M. & Bier, E. (2015a). The mutagenic chain reaction: A method for converting heterozygous to homozygous mutations. *Science. 348(6233)*: 442-444. doi: 10.1126/science.aaa5945

Gantz, V.M. & Bier, E. (2015b). The dawn of active genetics. *Bioessays. 38*: 50-63. doi: 10.1002/bies.201500102

Gantz, V.M., Jasinskiene, N., Tatarenkova, O., Fazekas, A., Macias, V.M., Bier, E., & James, A.A. (2015c). Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito *Anopheles stephensi*. *Proc. Natl. Acad. Sci. U.S.A. 112(49)*: E6736-E6743. doi: 10.1073/pnas.1521077112

Gatignol, A., Baron, M., & Tiraby, G. (1987). Phleomycin resistance encoded by the ble gene from transposon Tn 5 as a dominant selectable marker in *Saccharomyces cerevisiae*. *Molecular and General Genetics. 207(2–3)*: 342–348. doi: 10.1007/BF00331599

Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H., Bryson, D.I., & Liu, D.R. (2017). Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature. 551*: 464-471. doi: 10.1038/nature24644

Gething, P.W., Smith, D.L., Patil, A.P., Tatem, A.J., Snow, R.W., & Hay, S.I. (2010). Climate change and the global malaria recession. *Nature 465*: 342–345. doi: 10.1038/nature09098

Hammond, A., Galizi, R., Kyrou, K., Simoni, A., Siniscalchi, C., Katsanos, D., Gribble, M., Baker, D., Marois, E., Russell, S., Burt, A., Windbichler, N., Crisanti, A., & Nolan, T. (2016). A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector *Anopheles gambiae*. *Nat. Biotech. 34(1)*: 78-83. doi: 10.1038/nbt.3439

Hammond, A.M., Kyrou, K., Bruttini, M., North, A., Galizi, R., Karlsson, X., Kranjc, N., Carpi, F.M., D'Aurizio, R., Crisanti, A., & Nolan, T. (2017). The creation and selection of mutations resistant to a gene drive over multiple generations in the malaria mosquito. *PLoS Genet. 13(10)*: e1007039. doi: 10.1371/journal.pgen.1007039

Hay, S.I., Cox, J., Rogers, D.J., Randolph, S.E., Stern, D.I., Shanks, G.D., Myers, M.F., & Snow, R.W. (2002). Climate change and the resurgence of malaria in the East African highlands. *Nature. 415*: 905-909. doi: 10.1038/415905a

Hunter, JD (2007). Matplotlib: A 2D Graphics Environment. *Computing in Sci. & Eng. 9(3)*: 90-95. doi: 10.1109/MCSE.2007.55

International Agency for Research on Cancer (1997). *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans: Polychlorinated Dibenzo- para-dioxins and Polychlorinated Dibenzofurans. 69.* Retrieved from https://monographs.iarc.fr/wp-content/uploads/2018/06/mono69.pdf

Jain, K., & Krug, J. (2007). Deterministic and Stochastic Regimes of Asexual Evolution on Rugged Fitness Landscapes. *Genetics. 175(3)*: 1275-1288. doi: 10.1534/genetics.106.067165

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., & Doudna, J.A. (2012). A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*. *337(6096)*: 816-821. doi: 10.1126/science.1225829

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., & Doudna, J. (2013). RNA-programmed genome editing in human cells. *eLife. 2*: e00471. doi: 10.7554/eLife.00471

Jin, S., Zong, Y., Gao, Q. Zhu, Z., Wang, Y., Qin, P., Liang, C., Wang, D., Qiu, J.-L., Zhang, F., & Gao, C. (2019). Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science. 364(6437)*: 292-295. doi: 10.1126/science.aaw7166

Kondrashov, F.A., & Kondrashov, A.S. (2010). Measurements of spontaneous rates of mutations in the recent past and the near future. *Philos. Trans. R. Soc. Lond. B Biol. Sci. 365(1544)*: 1169–1176. doi: 10.1098/rstb.2009.0286

Kryazhimskiy, S., Rice, D.P., Jerison, E.R., & Desai, M.M. (2014). Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science. 344(6191)*: 1519-1522. doi: 10.1126/science.1250939

Kyrou, K., Hammond, A.M., Galizi, R., Kranjc, N., Burt, A., Beaghton, A.K., Nolan, T., & Crisanti, A. (2018). A CRISPR–Cas9 gene drive targeting doublesex causes complete population suppression in caged Anopheles gambiae mosquitoes. *Nat. Biotech. 36(11)*: 1062-1066. doi: 10.1038/nbt.4245

Lang, G.I., Murray, A.W., & Botstein, D. (2009). The cost of gene expression underlies a fitness trade-off in yeast. *Proc. Natl. Acad. Sci. U.S.A. 106(14)*: 5755-5760. doi: 10.1073/pnas.0901620106

Lang, G.I., Rice, D.P., Hickman, M.J., Sodergren, E., Weinstock, G.M., Botstein, D., & Desai, M.M. (2013). Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature.* 500: 571–574. doi: doi:10.1038/nature12344

Langmead, B. & Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods. 9(4)*: 357-9. doi: 10.1038/nmeth.1923.

Laven, H. (1967). Eradication of *Culex pipiens fatigans* through Cytoplasmic Incompatibility. *Nature. 216*: 383-384. doi: 10.1038/216383a0

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R.; 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics. 25(16)*: 2078-2079. doi: 10.1093/bioinformatics/btp352

Li, T., Huang, S., Jiang, W.Z., Wright, D., Spalding, M.H., Weeks, D.P., & Yang, B. (2010). TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Research. 39(1)*: 359–372. doi: 10.1093/nar/gkq704

Mahfouz, M.M., Li, L., Shamimuzzaman, M., Wibowo, A., Fang, X., & Zhu, J-K. (2011). De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. *Proc. Natl. Acad. Sci. U.S.A. 108(6)*: 2623-2628. doi: 10.1073/pnas.1019533108

Martens, W.J., Niessen, L.W., Rotmans, J., Jetten, T.H., & McMichael, A.J. (1995). Potential impact of global climate change on malaria risk. *Env. Health Persp. 103(5)*: 458-464. doi: 10.1289/ehp.95103458

McDonald, M.J., Rice, D.P., & Desai, M.M. (2016) Sex speeds adaptation by altering the dynamics of molecular evolution. *Nature. 531*: 233–236. doi: 10.1038/nature17143

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res. 20(9)*: 1297-1303. doi: 10.1101/gr.107524.110

McKinney, W (2010). Data Structures for Statistical Computing in Python. *Proc. of the 9th Python in Science Conf*: 51-56. Retrieved from http://conference.scipy.org/proceedings/scipy2010/mckinney.html

Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J., Dulay, G.P., Hua, K.L., Ankoudinova, I., Cost, G.J., Urnov, F.D., Zhang, H.S., Holmes, M.C., Zhang, L., Gregory, P.D., & Rebar E.J. (2011). A TALE nuclease architecture for efficient genome editing. *Nat. Biotech. 29(2)*: 143–148. doi: 10.1038/nbt.1755

Millman, K.J. & Aivazis, M. (2011). Python for Scientists and Engineers. *Computing in Sci. & Eng. 13(22)*: 9-12. doi: 10.1109/MCSE.2011.36

Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J., & Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology. 155*: 733-740. doi: 10.1099/mic.0.023960-0

Morgens, D.W., Wainberg, M., Boyle, E.A., Ursu, O., Araya, C.L., Tsui, C.K., Haney, M.S., Hess, G.T., Han, K., Jeng, E.E., Li, A., Snyder, M.P., Greenleaf, W.J., Kundaje, A., & Bassik, M.C. (2017). Genome-scale measurement of off-target activity using Cas9 toxicity in high-throughput screens. *Nat. Comm. 8*: 15178. doi: 10.1038/ncomms15178

Newton, M.D., Taylor, B.J., Driessen, R.P.C., Roos, L., Cvetesic, N., Allyjaun, S., Lenhard, B., Cuomo, M.E., & Rueda, D.S. (2019). DNA stretching induces Cas9 off-target activity. *Nat. Str. & Mol. Biol. 26*: 185–192. doi: 10.1038/s41594-019-0188-z

Noble, C., Olejarz, J., Esvelt, K.M., Church, G.M., & Nowak, M.A. (2017). Evolutionary dynamics of CRISPR gene drives. *Sci. Adv. 3(4)*: e1601964. doi: 10.1126/sciadv.1601964

Noble, C., Min, J., Olejarz, J., Buchthal, J., Chavez, A., Smidler, A.L., DeBenedictis, E.A., Church, G.M., Nowak, M.A, & Esvelt, K.M. (2019). Daisy-chain gene drives for the alteration of local populations. *Proc. Natl. Acad. Sci. U.S.A. 116(17)*: 8275-8282. doi: 10.1073/pnas.1716358116

Oliphant, T.E. (2006). *A Guide to NumPy*. Trelgol Publishing. Retrieved from https://ftp.tw.freebsd.org/distfiles/numpybook.pdf

Oliphant, T.E. (2007). Python for Scientific Computing. *Computing in Sci. & Eng. 9(3)*: 10 - 20. doi: 10.1109/MCSE.2007.58

Parzen, E. (1963). On spectral analysis with missing observations and amplitude modulation. *Sankhya: The Indian Journal of Statistics, Series A*: 383-392.

Pascual, M., Ahumada, J.A., Chaves, L.F., Rodó, X., & Bouma, M. (2006). Malaria resurgence in the East African highlands: Temperature trends revisited. *Proceedings of the National Academy of Sciences. 103(15)*: 5829-5834. doi: 10.1073/pnas.0508929103

Porteus, M.H. & Carroll, D. (2005). Gene targeting using zinc finger nucleases. *Nat. Biotech. 23(8)*: 967-973. doi: 10.1038/nbt1125

Povirk, L.F., Wübker, W., Köhnlein, W., & Hutchinson, F. (1977). DNA double-strand breaks and alkali-labile bonds produced by bleomycin. *Nucleic Acids Research. 4(10)*: 3573–3580. doi: 10.1093/nar/4.10.3573

Qi, J., Wijeratne, A.J., Tomsho, L.P., Hu, Y., Schuster, S.C., & Ma, H. (2009). Characterization of meiotic crossovers and gene conversion by whole-genome sequencing in Saccharomyces cerevisiae. *BMC genomics. 10*: 475. doi:10.1186/1471-2164-10-475

Reyon, D., Tsai, S.Q., Khayter, C., Foden, J.A., Sander, J.D. & Joung, J.K. (2012). FLASH assembly of TALENs for high-throughput genome editing. *Nat. Biotech. 30*: 460-465.

Seabold, S & Perktold, J (2010). Statsmodels: Econometric and statistical modeling with python. *Proc. of the 9th Python in Science Conf*: 57-61. Retrieved from http://conference.scipy.org/proceedings/scipy2010/seabold.html

Serero, A., Jubin, C., Loeillet, S., Legoix-Né, P., & Nicolas, A.G. (2014). Mutational landscape of yeast mutator strains. *Proc. Natl. Acad. Sci. U.S.A. 111(5)*: 1897-1902. doi: 10.1073/pnas.1314423111

Shanks, G.D., Hay, S.I., Stern, D.I., Biomndo, K., & Snow, R.W. (2002). Meteorologic Influences on Plasmodium falciparum Malaria in the Highland Tea Estates of Kericho,

Western Kenya. *Emerging Infectious Diseases. 8(12)*: 1404-1408. doi: 10.3201/eid0812.020077

Singh, D., Sternberg, S.H., Fei, J., Doudna, J.A., Ha, T. (2016). Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. *Nat. Comm. 7*: Article number 12778. doi: 10.1038/ncomms12778

Smidler, A.L., Terenzi, O., Soichot, J., Levashina, E.A., & Marois, E. (2013). Targeted Mutagenesis in the Malaria Mosquito Using TALE Nucleases. *PLoS ONE. 8(8)*: e74511. doi:10.1371/journal.pone.0074511

Smith, S., Hwang, J.-Y., Banerjee, S., Majeed, A., Gupta, A., & Myung, K. (2004). Mutator genes for suppression of gross chromosomal rearrangements identified by a genome-wide screening in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U.S.A. 101(24)*: 9039-9044. doi: 10.1073/pnas.0403093101

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

Stirling, P.C., Shen, Y., Corbett, R., Jones, S.J.M., & Hieter, P. (2014). Genome Destabilizing Mutator Alleles Drive Specific Mutational Trajectories in Saccharomyces cerevisiae. *Genetics. 196(2)*: 403-412. doi: 10.1534/genetics.113.159806

Stouthamer, R., Breeuwer, J.A.J., & Hurst, G.D.D. (1999). *Wolbachia Pipientis*: Microbial Manipulator of Arthropod Reproduction. *Annu. Rev. Micro. 53*: 71-102. doi: 10.1146/annurev.micro.53.1.71

Szostak, J.W., Orr-Weaver, T.L., Rothstein, R.J., & Stahl, F.W. (1983). The double-strand-break repair model for recombination. *Cell. 33(1)*: 25-35. doi: 10.1016/0092-8674(83)90331-8

Tanser, F.C., Sharp, B., & le Sueur, D. (2003). Potential effect of climate change on malaria transmission in Africa. *The Lancet. 362(9398)*:1792-1798. doi: 10.1016/S0140-6736(03)14898-2

Tsai, S.Q., Nguyen, N.T., Malagon-Lopez, J., Topkar, V.V., Aryee, M.J., & Joung, J.K. (2017). CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR–Cas9 nuclease off-targets. *Nat. Methods. 14(6)*: 607-614. doi: 10.1038/nmeth.4278

Tulu, A.N. (1996). Determinants of malaria transmission in the highlands of Ethiopia: the impact of global warming on morbidity and mortality ascribed to malaria. *London School of Hygiene & Tropical Medicine*: PhD Thesis. doi: 10.17037/PUBS.00682286

Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K.V., Altshuler, D., Gabriel, S., & DePristo, M.A. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics. 43(1)*: 11.10.1-11.10.33. doi: 10.1002/0471250953.bi1110s43

Van Rossum, G (1997). Scripting the Web with Python. In "Scripting Languages: Automating the Web". *World Wide Web Journal. 2(2)*.

Wienert, B., Wyman, S.K., Richardson, C.D., Yeh, C.D., Akcakaya, P., Porritt, M.J., Morlock, M., Vu, J.T., Kazane, K.R., Watry, H.L., Judge, L.M., Conklin, B.R., Maresca, M., & Corn, J.E. (2019). Unbiased detection of CRISPR off-targets in vivo using DISCOVER-Seq. *Science. 364(6437)*: 286-289. doi: 10.1126/science.aav9023

Windbichler, N., Menichelli, M., Papathanos, P.A., Thyme, S.B., Li, H., Ulge, U.Y., Hovde, B.T., Baker, D., Monnat, R.J., Burt, A., & Crisanti, A. (2011). A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. *Nature. 473*: 212-215. doi: 10.1038/nature09937

World Health Organization (2018). *World Malaria Report 2018*. Retrieved from https://apps.who.int/iris/bitstream/handle/10665/275867/9789241565653-eng.pdf

Wright, A.V., Nuñez, J.K., & Doudna, J.A. (2016). Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell. 164(1)*: 29-44. doi: 10.1016/j.cell.2015.12.035

Wu, X., Scott, D.A., Kriz, A.J., Chiu, A.C., Hsu, P.D., Dadon, D.B., Cheng, A.W., Trevino, A.E., Konermann, S., Chen, S., Jaenisch, R., Zhang, F., & Sharp, P.A. (2014). Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotech. 32(7)*: 670-676. doi: 10.1038/nbt.2889

Xi, Z., Khoo, C.C.H., & Dobson, S.L. (2005). *Wolbachia* Establishment and Invasion in an *Aedes aegypti* Laboratory Population. *Science. 310(5746)*: 326-328. doi: 10.1126/science.1117607

Xi, Z., Khoo, C.C.H., & Dobson, S.L. (2006). Interspecific transfer of *Wolbachia* into the mosquito disease vector *Aedes albopictus*. *Proc. R. Soc. B. 273*: 1317-1322. doi: 10.1098/rspb.2005.3405

Zhu, Y.O., Siegal, M.L., Hall, D.W., & Petrov, D.A. (2014). Precise estimates of mutation rate and spectrum in yeast. *Proc. Natl. Acad. Sci. U.S.A. 111(22)*: E2310-8. doi: 10.1073/pnas.1323011111.

Zuo, E., Sun, Y., Wei, W., Yuan, T., Ying, W., Sun, H., Yuan, L., Steinmetz, L.M., Li, Y., & Yang, H. (2019). Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science. 364(6437)*: 289-292. doi: 10.1126/science.aav9973

Zymo Research. *YeaStar Genomic DNA Kit™, Catalog No. D2002: Instructions*. Irvine, CA. Retrieved from https://files.zymoresearch.com/protocols/_d2002__yeastar_genomic_dna_kit.pdf