# The dynamics and diversity of copy number variants during adaptive evolution

by

Stephanie Lee Lauer

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Biology

New York University

January 2019

_____

David Gresham, PhD

# DEDICATION

to garrett

# ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. David Gresham, for his mentorship and his role in inspiring such a great project. I'd like to thank my committee members Dr. Edo Kussell, Dr. Gregory Lang, Dr. Matt Rockman, and especially Dr. Andreas Hochwagen for encouraging me to try Southern blotting and for welcoming me into his lab. I would like to thank all Gresham lab members past and present (Chris, Niki, Benjy, Naomi, Jungeui, and Gunjan) for their support throughout my graduate career -- I'll always remember the board game nights, lab hikes, holiday celebrations, and beer and ice cream outings. I'd especially like to thank Dr. Darach Miller for showing me how to be truly passionate about research and for giving me an amazing example of what it means to be a scientist. I'd like to thank Grace Avecilla and Dr. Pieter Spealman, my co-authors, for their fantastic contributions to our paper, which I am very proud of. Grace joined the lab during a difficult time for me, and she really made my P.h.D. better -- I am grateful to have spent the last couple years working with such a brilliant scientist and fun person. I'd also like to thank Farah Abdul-Rahman and Siyu Sun for being incredible lab mates -- we've been through a lot together, and I feel lucky to have had their support through it all. I'd like to thank Nathan Brandt for always being cool under pressure because I truly could not have done my P.h.D without his steady presence and guidance. I would like to thank the undergrads who worked with me, Saroj Gourkanti and Jill Trivedi. Working with Saroj for a few years and seeing his dedication and passion for research was inspiring, and he has (hopefully) taught me to be a better mentor. I'd like to thank Christine Vogel and the members of the Vogel lab for being awesome floormates and for their helpful suggestions at lab meetings. I'd like to thank Andres Mansisidor for encouraging me to help him organize a symposia, for helping me perform my first Southern blot, and for being such a cool person. I'd like to thank my entire cohort (especially Alex Grote), for all the happy hours, potluck dinners, and fun moments during graduate school. I'd like to thank the NYU Gencore for all their help with sequencing, flow cytometry, and FACS.

I would like to thank my fiancé, Garrett Dripchak, for his unwavering love and support. Doing a P.h.D. made me a little crazy sometimes, but I've stayed (relatively) sane and happy thanks to him. I'd also like to thank my incredible parents who nurtured my interest in science and encouraged me to pursue graduate school. Even though they had no idea what I was studying, they were always so proud of me and that means the world.

# ABSTRACT

For a quantitative understanding of adaptive evolution, it is critical to define the rate new genetic variants are introduced to an evolving population, how selection acts on these variants, and how they are maintained. Copy number variants (CNVs) are an important class of genetic variation that contributes to rapid adaptive evolution in scenarios ranging from local adaptation to tumor evolution. CNVs comprise duplications and deletions of genomic sequence that can result in large-scale changes in mRNA expression and protein abundance driving their selective advantage. Although CNVs contribute to inter- and intra-species phenotypic variability and have been directly linked with human disease, fundamental questions concerning their dynamics and mechanisms of generation remain unresolved. An impediment to addressing these questions has been the lack of accurate and sensitive methods to detect and analyze CNVs in complex evolving populations. Current technology requires that they exist at high population frequency before they can be identified. To address this challenge, I developed a novel fluorescent reporter assay to detect, track, and isolate single cells that undergo gene duplication or deletion. During nutrient limitation in chemostats, CNVs occur in genes encoding high-affinity nutrient transporters, including the general amino acid permease *GAP1*. Preliminary data suggest that *GAP1* duplications arise via multiple mechanisms, which makes *GAP1* an ideal candidate for studying both the temporal dynamics of CNVs and how different molecular processes contribute to their generation, selection, and ultimate fate. In chapter 1, I review the role of CNVs in driving rapid adaptive evolution across diverse systems and discuss the specific challenges inherent in studying CNVs. In chapter 2, I quantified the temporal dynamics with which CNVs are generated and selected at the *GAP1* locus, determined the molecular mechanisms underlying CNV diversity, and defined the role of clonal interference among competing CNV lineages. In chapter 3, I performed a comparative analysis of CNV dynamics by extending the use a CNV reporter to

other loci: the high affinity glucose transporters *HXT6/7* and the high-affinity ammonia transporter, *MEP2*. In chapter 4, I explored the use of a reporter for quantifying the rate of a specific molecular event underlying *GAP1* deletion. Together, these studies underscore the range of processes underlying CNV formation that undoubtedly contribute to the diverse array of CNV alleles segregating in evolving populations. The broad implications of these findings are presented in Chapter 5.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# Chapter 1: Copy number variants (CNVs): drivers of rapid evolutionary change

This chapter is based on the review paper “Copy number variants (CNVs): drivers of rapid evolutionary change” by Stephanie Lauer, Grace Avecilla, Zoe Lye, Michael Purugganan and David Gresham. I generated all of the figures and wrote the majority of the manuscript text with contributions from GA and ZL (excluding the sections “CNVs underlie domestication of plants and animals” and “Natural populations as emerging models for studying CNVs” which were written by ZL and GA but edited by me).

## 1.1: Abstract

Comparative genomics has revealed the pervasiveness of changes in gene copy number across the evolutionary history of genomes in different organisms. But increasingly, population surveys of intraspecific variation have illuminated the prevalence of copy number variants (CNVs) within species. As alleles of large effect, CNVs can drive rapid adaptive evolution over short evolutionary timescales in diverse scenarios including microbial and pathogen evolution, domestication, and human adaptation. In multicellular organisms, somatic CNVs can drive tumorigenesis, exhibiting parallels with evolutionary scenarios. We review recent findings that have revealed the role of CNVs in adaptive evolution. CNVs represent a unique class of genetic variation that differ from nucleotide variation in function and population genetics in several respects. We discuss the unique aspects of CNVs with respect to their dynamics and fate in populations, the role of antagonistic pleiotropy and the potential effects on cellular homeostasis. We define open questions, and proposed approaches, to studying the role of CNVs in short term evolutionary scenarios.

## 1.2: Introduction

Genomes can vary in both nucleotide sequence and in the configuration and structure of segments of DNA. Copy number variants (CNVs) are genomic variants consisting of DNA segments that are increased or decreased in copy number. When inherited through the germline, this variation can result in variable copy number, or polymorphisms, within populations. In somatic tissue, mitotic errors can result in somatic variation where cells containing CNVs are found within an individual. CNVs are ubiquitous in nature and comprise a significant portion of the human genome: de novo CNVs introduced each generation affect substantially more DNA than point mutations (Itsara et al. 2010). Among individuals, between 4.8-9.5% of the genome contains CNVs, which are typically less than 3,000 base pairs in length (Zarrei et al. 2015). However, CNVs can range in size from tens to hundreds of kilobases (Zarrei et al. 2015; Itsara et al. 2009). Individual CNVs on the order of megabases have also been identified (Itsara et al. 2009). One single-cell sequencing study demonstrated that at least one megabase-scale CNV is present in ~10% of human somatic cells (Knouse, Wu, and Amon 2016). Aneuploidy, the duplication or deletion of an entire chromosome, is an extreme case of copy number variation and has been previously discussed in relationship to cancer (C. J. Ye et al. 2018; Gordon, Resio, and Pellman 2012), disease (Oromendia and Amon 2014), adaptation in microbial populations (Pavelka et al. 2010; Mulla, Zhu, and Li 2014), and its effect on transcription in a diverse range of organisms (Sheltzer et al. 2012).

For the purpose of this review we will consider CNVs as any change in copy number that comprises less than a whole chromosome. CNVs have previously been defined based on the length of the sequence that is duplicated or deleted, including >50 (Zarrei et al. 2015), >100

(Zhang et al. 2009), and >1000 base pairs (Feuk, Carson, and Scherer 2006; Itsara et al. 2009). Historically, these definitions have been influenced by the resolution of available technologies used to detect CNVs, such as array-based comparative genomic hybridization (array CGH) and comparative genome sequencing (Iafrate et al. 2004; Sebat et al. 2004; Feuk, Carson, and Scherer 2006; Korbel et al. 2007). Because CNV detection continues to improve and definitions based on sequence length are arbitrary, a functional definition for CNVs is required (Xi et al. 2011; MacDonald et al. 2014; F. Zhang et al. 2009). For example, at what size should a deletion be considered a CNV as opposed to an insertion deletion polymorphism (indel)? Since CNVs and indels are generated by distinct mechanisms, determining the origin of deletion event could be used to distinguish between the two types of variants (Carvalho and Lupski 2016; Garcia-Diaz and Kunkel 2006). However, our understanding of the underlying mechanisms of CNV and indel formation is insufficiently detailed to enable accurate prediction of mechanisms solely on the basis of sequence. Therefore, we propose the following definition for CNVs: polymorphisms resulting in a change in copy number for a functional unit of DNA. Our definition includes duplication or deletion of a section of DNA containing at least one entire gene, intron, exon, promoter, enhancer or other regulatory region. As an upper bound we define CNVs as those events that include less than a chromosome arm. Large-scale duplications or deletions that encompass more than a chromosome arm are typically considered partial (i.e. segmental) aneuploidies (Natesuntorn et al. 2015).

CNVs have been identified as the causative agent of several diseases and disorders (F. Zhang et al. 2009). Because increases or decreases in gene copy number result in altered mRNA and protein abundance, CNVs can have dramatic effects on cell physiology. Germ-line CNVs underlie a range of human diseases including Crohn's disease, autism and several

developmental disorders (reviewed in (Girirajan et al. 2011; Marshall and Scherer 2012; Polley et al. 2016). Somatic CNVs contribute to neuronal diversity in both healthy and diseased individuals (McConnell et al. 2013; Cai et al. 2014; Bruder et al. 2008). Somatic CNVs also have a role in promoting tumorigenesis; nearly 40% of cancer-related genes are found in CNVs (Stratton, Campbell, and Futreal 2009; Shlien and Malkin 2009). While CNVs can have negative phenotypic consequences, they play important roles in evolution, including human adaptation.

CNVs drive adaptation in diverse scenarios ranging from niche adaptation to speciation (Zuellig and Sweigart 2018; Dhami, Hartwig, and Fukami 2016; K. M. Turner et al. 2017; Geiger, Cox, and Mann 2010; Stratton, Campbell, and Futreal 2009). CNVs are common in human populations (Barreiro et al. 2008; Iskow et al. 2012; Zarrei et al. 2015), but are also pervasive among domesticated and wild populations of animals and plants, (Ramirez et al. 2014; Clop, Vidal, and Amills 2012; Żmieńko et al. 2014), pathogenic and non-pathogenic microbes (Greenblum, Carr, and Borenstein 2015; Nair et al. 2008; Iantorno et al. 2017; Dulmage et al. 2018), and viruses (Gao et al. 2017; Rezelj, Levi, and Vignuzzi 2018; Elde et al. 2012). While CNVs play an important role across long evolutionary timescales by providing the substrate for innovation and gene family expansion, they also drive rapid adaptation in response to stress and changes in the environment. As alleles of large effect, CNVs can have dramatic effects on phenotype and organismal fitness. The goal of this review is to discuss our emerging view of CNVs and their role in evolutionary processes, while simultaneously highlighting the unique aspects of CNVs as a source of genetic variation and the challenges of studying them.

## 1.3: Functional effects of CNVs

Susumo Ohno and others (Ohno 1970; Austin L. Hughes 1994; R. P. Anderson and Roth 1977) highlighted the role of gene duplication in generating evolutionary novelty and diversity over long time scales (Conant and Wolfe 2008; J. B. Walsh 1995; B. Walsh 2003; M. Lynch and Force 2000; Michael Lynch and Conery 2000). Importantly, CNV formation can also have immediate consequences for organismal fitness (**Figure 1.1**). CNVs are alleles of large effect with a single CNV affecting $10^2$-$10^6$ base pairs of DNA, compared with 1 base pair for single nucleotide variants (SNVs). Since CNVs typically encompass large regions of the genome, they can affect multiple protein-coding genes and regulatory regions simultaneously. Large duplications and deletions leading to increases or decreases in gene dosage can subsequently result in widespread protein abundance changes (reviewed in (A. M. Rice and McLysaght 2017; Tang and Amon 2013). CNVs can affect neighboring loci, leading to concomitant increases or decreases in expression for genes outside the CNV boundary (Molina et al. 2008; Merla et al. 2006). CNVs can also have effects in trans, by changing the expression of distal transcripts (Gamazon, Nicolae, and Cox 2011), by affecting global levels of transcription (Henrichsen et al. 2009), and by changing the topology of chromatin organization (Lupiáñez, Spielmann, and Mundlos 2016; Lupiáñez et al. 2015; Franke et al. 2016).

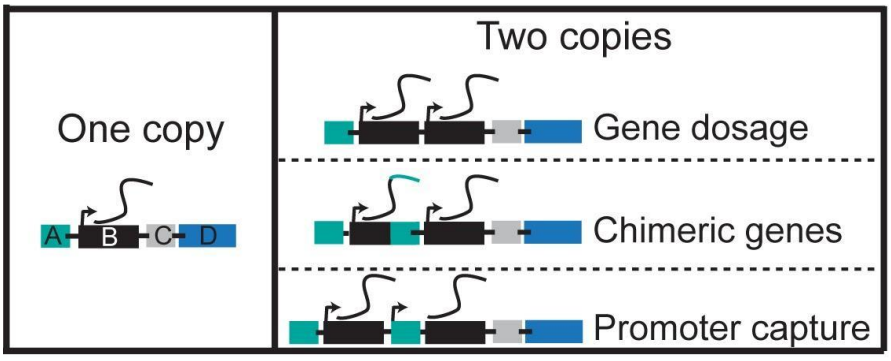

**Figure 1.1. Functional consequences of copy number variation.** Four different genes are represented by the letters A, B, C, and D and an active promoter is represented by arrows. CNVs can cause phenotypic consequences through several different mechanisms, a few of which are depicted here. CNVs can affect gene dosage by directly altering transcript and protein abundance. During formation of CNVs, juxtaposition of two sequences can result in the formation of a chimeric transcript (for example, a combination of gene A and B). Position effects can result in promoter capture events, where the duplicated copy acquires altered regulation by a non-native promoter (here, the promoter from gene A regulates expression of gene B).

In addition to the gene expression changes described above, CNVs can affect the fitness of an organism through other mechanisms (**Figure 1.1**). A recent study proposed that gene duplications can increase fitness by buffering fluctuations in gene expression (Rodrigo and Fares 2018). Similarly, gene duplications can mask deleterious mutations (Gu et al. 2003) or promote heterozygote advantage (Sellis et al. 2016). It has also been proposed that gene duplication can resolve genetic conflicts arising due to sexual antagonism, where sex-biased fitness effects are constrained by genes shared between both sexes (Connallon and Clark 2011). While CNVs are often thought of as a substrate for future innovation through duplication and subsequent divergence, de novo CNVs can immediately provide new functionality. For example, gene duplications, deletions, and unbalanced translocations can lead to the formation of chimeric genes (Rippey et al. 2013; Mayo et al. 2017; Arguello et al. 2006; Aigner et al. 2013; Schrider et al. 2013). Therefore, CNVs can drive important adaptive innovations during short-term evolutionary scenarios.

We typically think of CNVs as protein-coding gene deletions or duplications, however, adaptive copy number changes in intergenic sequences have also been identified. CNV formation can result in position effects that disrupt or modify regulatory elements (Koszul et al. 2004; Chan et al. 2010). Promoter capture, where spatial re-arrangement of an amplified DNA segment leads to its regulation by a different promoter (**Figure 1.1**), has been observed repeatedly in diverse systems (Usakin et al. 2005; Adam, Dimitrijevic, and Schartl 1993;

Whoriskey et al. 1987). During Richard Lenski's long term evolution experiment, promoter capture enabled E. coli cells to grow aerobically on citrate (present as a chelating agent, not a nutrient source), which greatly increased the effective population size (Blount, Borland, and Lenski 2008). This adaptive innovation involved tandem duplication and subsequent co-option of the rnk promoter, allowing expression of a citrate transporter gene under aerobic conditions where it is normally repressed (Blount et al. 2012). Similarly, a 500 base pair deletion in a key regulatory region leads to adaptive morphological phenotypes in freshwater populations of stickleback fish (Bell 1987; Chan et al. 2010). Intron 1 amplifications for the gene encoding SOX5, an important developmental transcription factor, lead to the cold-adapted Pea-comb phenotype in chickens (Wright et al. 2009). In humans, intronic CNVs have been associated with decreases in lifespan (Iakoubov et al. 2013) and disease (H.-S. Lee et al. 2014; Tsai et al. 2016). In a recent comparative study between exons and introns, intronic deletions were the most common CNV, with 20% of ancient human genes (pre-tetrapod lineages) showing variation in intron length, and with many of these intronic deletions associated with expression-QTLs (Rigau et al. 2018). Collectively, these findings demonstrate that a single class of mutation can provide a range of functional effects and adaptive phenotypes.

Determining the functional consequences of a CNV remains challenging. Empirical tests such as competition assays can be used to directly measure the fitness effect of CNVs (Payen et al. 2014; Gresham et al. 2010). However, determining which individual gene duplications or deletions within a large CNV allele contribute directly to phenotypic effects requires testing of individual candidate loci (Payen et al. 2014; A. Selmecki et al. 2008). As a result, functional studies typically rely on well-characterized CNVs that have been directly implicated in adaptation through multiple lines of experimental evidence. As alleles of large effect, CNVs are

also more likely to result in pleiotropy, where one mutation simultaneously influences multiple phenotypic characteristics (Gamazon and Stranger 2015). Thus, resolving the genetic basis of complex traits becomes increasingly difficult, especially in natural populations. For example, causal associations can be obscured by SNPs that are in linkage disequilibrium with CNVs (Parkes et al. 2007; McCarroll et al. 2008). CNVs can affect gene expression in cis and in trans, as described above, which further complicates direct associations to genes within copy number variable regions. Dosage compensation mechanisms can prevent correlated increases in gene expression across the entire CNV allele, further altering the expression landscape (Stenberg and Larsson 2011; Torres et al. 2010; Stingele et al. 2012). Profiling gene expression may therefore be integral to determining all the functional effects of a CNV. In addition, CNVs themselves can be polymorphic within populations (i.e., the same gene is routinely duplicated, but different variants include variable lengths of the chromosome and adjacent genes (**Figure 1.2**). Because of this, CNVs segregating within populations can have different fitness benefits and trade-offs in alternative environments (Valdivia-Anistro et al. 2015; Lauer et al. 2018).

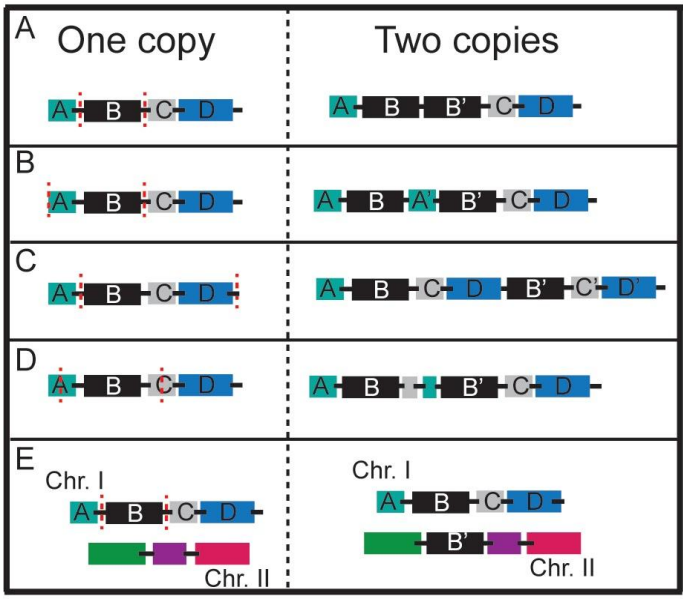

**Figure 1.2. Copy number variant alleles segregating within populations can be polymorphic.** This schematic demonstrates several different ways that gene B can be duplicated, leading to the formation of distinct CNV alleles. Four different genes are represented by the letters A, B, C, and D. Red dashed lines represent left and right breakpoints for each CNV event. (A) A simple CNV event where only gene B is duplicated to produce a second copy B'. (B) A CNV event where genes A and B are both duplicated. (C) A more complex event that includes genes B, C, and D, leading to duplicate copies B', C', and D'. (D) A CNV event with breakpoints inside genes A and B, interrupting the function of both duplicate copies. Only B' is functional. (E) A non-reciprocal translocation event in which gene B is copied to chromosome II.

## 1.4: A variety of mechanisms generate CNVs

CNVs are formed through complex and diverse processes, but the molecular basis of these events is not well understood (see (Hastings, Lupski, et al. 2009; Reams and Roth 2015) for excellent reviews). CNVs can be generated through replication errors (Koszul et al. 2004; Cardoso-Moreira, Arguello, and Clark 2012; Chen et al. 2015) that involve template switching (Slack et al. 2006), sequence microhomology (Hastings, Ira, et al. 2009), and/or the generation of extrachromosomal circles or circular intermediates (Gresham et al. 2010; Brewer et al. 2011, 2015; Møller et al. 2015; Cohen and Segal 2009; K. M. Turner et al. 2017). Replication stress has been directly linked to increased generation of CNVs in human cells, including variants associated with disease and tumorigenesis (S. G. Durkin et al. 2008; Arlt et al. 2009). The extent of replication-mediated CNV formation may have been previously underestimated, since many early studies focused on recurrent disease-related variants formed by non-allelic homologous recombination, which can be easier to detect and characterize.

Non-allelic (or "illegitimate") homologous recombination between repetitive sequences is a major driver of CNV formation. In prokaryotes, small insertion sequence (IS) elements are flanked by long terminal repeats (LTRs) (Siguier, Gourbeyre, and Chandler 2014; Brügger et al. 2002), which are also frequently found in the yeast genome (Carr, Bensasson, and Bergman 2012). Many eukaryotes have longer repetitive elements including segmental duplications,

which are >1 kilobase in size and are dispersed throughout the genome (Eichler 2001). Extensive homology between repetitive sequences enables recombination and can lead to increases or decreases in copy number (Peng et al. 2015). Recurrent CNVs, which repeatedly occur in specific regions of the genome, typically underlie re-occurring germline mutations and human disease (Itsara et al. 2009; Girirajan, Campbell, and Eichler 2011). Current evidence suggests that CNVs are enriched in pericentromeric and subtelomeric chromatin (Zarrei et al. 2015), and that recurrent CNVs arise due to specific features of the neighboring genomic sequence including: repetitive elements (Farslow et al. 2015), tRNA genes (Bermudez-Santana et al. 2010), origins of replication (Di Rienzi et al. 2009), and replication fork barriers (Labib et al. 2007). Stress can lead to increases in genome-wide mutation rates in both bacteria and yeast (Foster 2007; Galhardo, Hastings, and Rosenberg 2007; Shor, Fox, and Broach 2013), and active transcription units may play a role in elevating mutation rates and generating these hotspots (Thomas and Rothstein 1989; Skourti-Stathaki and Proudfoot 2014; Wilson et al. 2015). Increases in the rate of transcription lead directly to amplification of the rDNA and other loci (Jack et al. 2015; Hull et al. 2017). Together, these findings demonstrate that certain regions of the genome are more susceptible to mutation. Understanding the full repertoire of mechanisms that underlie CNV formation and whether these processes can be directly stimulated by the environment are important open questions in the field.

## 1.5: The role of CNVs in driving rapid adaptive evolution

CNVs were first characterized early in the 1900s, with much of the research focusing on Drosophila (Sturtevant 1925) and the theoretical, long-term consequences of gene duplication (reviewed in (Taylor and Raes 2004). Later work in the 1970s and 1980s revealed the extent of

gene family expansions, whole genome duplications, and polyploidy in natural populations (Stuber and Goodman 1983; Ferris and Whitt 1979; Avise and Kitto 1973; Hopkinson, Edwards, and Harris 1976; Schughart, Kappen, and Ruddle 1989). Here, we discuss adaptive CNVs, which explicitly provide fitness benefits to competing individuals. Below we review the identification of adaptive CNVs and cases in which these variants drive short-term, or rapid, adaptation.

### 1.5.1: Experimental evolution reveals CNVs as a major source of adaptation

The highly controlled and replicated selective conditions of laboratory evolution provides an efficient means of gaining insight into evolutionary processes (Lenski et al. 1991; Good et al. 2017). CNVs are an important mechanism of adaptation in these regimes where organisms are subjected to a strong, well-defined selective pressure. A subset of these experiments have focused on the adaptive response to nutrient limitation using chemostats. Depending on the specific limiting nutrient, CNVs including the gene encoding for that nutrient transporter were routinely identified and directly linked to increases in fitness. This includes Saccharomyces cerevisiae limited for carbon, phosphorus, nitrogen, and sulfur (Brown, Todd, and Rosenzweig 1998; Hansche 1975; Gresham et al. 2008; Hong and Gresham 2014), as well as Salmonella typhimurium under different carbon limitations (Sonti and Roth 1989), and Escherichia coli limited for lactose (Horiuchi, Horiuchi, and Novick 1963). CNVs in genes encoding nutrient transporters have also been identified in natural populations (Dhami, Hartwig, and Fukami 2016; Kettler et al. 2007), including species of the human gut microbiome (Greenblum, Carr, and Borenstein 2015). These CNVs are adaptive through direct increases in gene dosage that

increase the effective number of transporters on the cell membrane. Importantly, CNVs are not typically reported during experimental evolution in batch culture. Batch culture requires daily dilution, which leads to cycles of boom and bust. These fluctuations in population size and nutrient content result in a complex and dynamic environment where the selective pressure is inconsistent and poorly defined (Gresham and Hong 2014).

Evolution experiments have been performed with other model organisms including the worm Caenorhabditis elegans and the fruit fly Drosophila melanogaster. In contrast to microbial evolution experiments where asexual propagation is routinely used, these organisms can reproduce sexually, leading to important insights for evolutionary processes. A study using C. elegans imposed 200 generations of selection for recovered fecundity after reduced productivity due to mutation accumulation and inbreeding. Duplications and deletions increased in frequency over time, and CNVs were enriched for genes related to reproduction and development (Farslow et al. 2015). CNVs spanning the same region were observed in many replicate populations, suggesting that CNVs containing specific genes were under strong positive selection. One duplication and one deletion with the exact same breakpoints were found in several experimental populations and a control population. This high rate of recurrent CNV formation suggests that mutation at this locus can have as large of an impact as selection (Farslow et al. 2015; Lipinski et al. 2011). Researchers also noted that the median size of duplications was larger (191.5 kb) than duplications found in a prior mutation accumulation experiment (2 kb), a trend that was recapitulated for deletions (Lipinski et al. 2011). Genomic analysis of a fly line reared in darkness for 1400 generations found about 150 putative CNVs. One verified CNV contained a 500 base pair deletion within CG459, a gene of unknown function whose mammalian homologues are involved in fatty acid metabolism in the mitochondria (Izutsu

et al. 2012). Despite their importance and prevalence in Drosophila populations (Zichner et al. 2013), CNVs are not as frequently addressed as SNVs in many evolution experiments (Burke et al. 2010; T. L. Turner et al. 2011; Zhou et al. 2011; Remolina et al. 2012; Jalvingh et al. 2014; Kang et al. 2016), and should be considered an active area of research for further study.

### 1.5.2: CNVs drive adaptation in response to extreme environments

CNVs have played an important role in adaptation to stressful environments across the domains of life. In microbial systems, CNVs promote heat tolerance in E. coli (Riehle, Bennett, and Long 2001; Christ and Chin 2008) and metal tolerance in various wild species: the heavy metal resistant bacterium Cupriavidus metallidurans (von Rozycki and Nies 2009), Ralstonia pickettii in copper-contaminated lakes (Yang et al. 2010), and strain-specific arsenite resistance in the fungal pathogen, Cryptococcus neoformans (Chow et al. 2012). Tandem duplications of the CUP1 locus in yeast are important for copper tolerance in both laboratory-evolved and natural populations of S. cerevisiae (Fogel et al. 1983; Y. Zhao et al. 2014; Adamo et al. 2012), and may be directly stimulated by copper exposure (Hull et al. 2017). CNVs are responsible for frost tolerance in certain varieties of wheat (Zhu et al. 2014) and duplication of the Bot1 gene, a boron efflux transporter, confers tolerance of high boron soils in barley (Sutton et al. 2007). Cultured human cells generate de novo copy number variants after exposure to toxic chemicals and ionizing radiation (Arlt et al. 2011, 2014). Archaea have specifically adapted to extreme environments (for example, high salt tolerance), and one potential mechanism underlying this adaptation is through large-scale genome amplifications that lead to coordinated up-regulation of gene expression (Dulmage et al. 2018). While we do not discuss it here, increases in ploidy in

archaea, bacteria, and S. cerevisiae may be generally adaptive during stress (A. M. Selmecki et al. 2015; Scott et al. 2017; Zerulla and Soppa 2014; Pecoraro et al. 2011).

### 1.5.3: CNVs underlie domestication of plants and animals

CNVs have been recognized as the underlying causal variants for many domestication traits. Domestication is a well-studied example of recent adaptive evolution, as humans controlling plant and animal reproduction (as a means of improving their own fitness) create a strong selective pressure (Meyer and Purugganan 2013). Domestication traits include loss of seed shattering in many grain crops, and development of docile behaviors in animals. These traits can improve local adaptation or arise as a result of cultural preference. Selection for culturally preferred traits, such as coat color, are characterized by higher selection coefficients in the later stages of domestication (Meyer and Purugganan 2013). For example, in both cows and pigs, the KIT gene is duplicated causing white coat color (K. Durkin et al. 2012; Giuffra et al. 2002). Similarly, variation in coat color is caused by duplication of Agouti signaling protein gene in both goats and sheep (Norris and Whan 2008; Fontanesi et al. 2009). Domestication variants caused by CNVs in this category represent a clear example of how CNVs can contribute to rapid adaptation.

In addition to directed selection based on human preference, there are climatic and dietary changes associated with domesticated species. Control of flowering time across different climates is important for the reproductive success of plants. The geographical distribution of CNV genes in winter wheat populations demonstrate the role of CNVs in this process. Increased copy number of the Photoperiod-B1 (Pbd-B1) gene causes early flowering, allowing plants in warmer climates to escape summer drought (Würschum et al. 2015). Conversely, increased

copy number of Vernalization-1 (Vrn-1) lengthens the required period of cold exposure before flowering and is more common in temperate environments (Würschum et al. 2015). Similarly, the copy number of amylase genes among different dog populations has been associated with the dietary starch content of regional human diets (Reiter, Jagoda, and Capellini 2016). Duplication of the salivary amylase gene (AMY1) improves starch digestion, and in humans, provided one of the first examples of CNVs under positive selection (Perry et al. 2007). AMY1 duplications are beneficial for human populations living in agricultural societies and facilitated an adaptation to this change in human dietary history. During domestication, amplification of the amylase gene similarly underwent positive selection when dogs began to share in a starch rich diet with humans (Axelsson et al. 2013).

### 1.5.4: CNVs arm pests and parasites in the coevolutionary arms race

Pathogenic microbes often have CNVs containing genes that increase virulence or promote antibiotic resistance. The first discoveries of CNVs conferring antibiotic resistance were identified as tandem amplifications on plasmids in the bacteria Proteus mirabilis (Rownd and Mickel 1971; Perlman and Rownd 1975) and Enterococcus faecalis (Clewell, Yagi, and Bauer 1975; Yagi and Clewell 1976), and as chromosomal amplifications of β-lactamase in E. coli (Normark et al. 1977). Other examples of CNVs conferring antibiotic resistance in bacteria are discussed in (Sandegren and Andersson 2009). CNVs also provide resistance to antifungals in Candida glabrata (Marichal et al. 1997) and Candida albicans (A. Selmecki et al. 2008). In the latter study, increases in drug resistance were associated with the formation of an isochromosome comprised of two copies of the left arm of chromosome 5 (A. Selmecki, Forche, and Berman 2006), indicating that CNVs involved in large, structural genomic alterations can

have important adaptive effects. The bacteria Renibacterium salmoninarum is a pathogen found in salmonid fish that has detrimental effects on both salmon farming and conservation. CNVs encoding multiple virulence factors have been identified (Rhodes, Coady, and Deinhard 2004; O'Farrell and Strom 1999), and the authors of one study hypothesized that strong selective pressures have led to multiple, independent events that generated these CNVs in R. salmoninarum (Brynildsrud et al. 2016). Similarly, adaptive CNVs that promote virulence have been identified in double-stranded DNA viruses including the poxvirus, vaccinia (Elde et al. 2012; Cone et al. 2017) and may represent a generally applicable and frequent mechanism of dsDNA viral evolution (Gao et al. 2017; Bayer, Brennan, and Geballe 2018).

CNVs promote virulence and antibiotic resistance through diverse mechanisms. CNVs in membrane-bound transporters important for drug resistance have been identified in the unicellular eukaryotic parasites Leishmania (Iantorno et al. 2017; Mary et al. 2010; Monte-Neto et al. 2015), Trypanosoma (Vincent et al. 2010) and Plasmodium (Nair et al. 2007; Sidhu et al. 2006; Barnes et al. 1992; Suwanarusk et al. 2008). One study looking at clinical isolates of Leishmania found that copy number at the DNA level accounted for more than 85% of gene expression variation among differentially expressed genes, of which membrane-bound transporters were among the most highly differentially expressed (Iantorno et al. 2017). While CNVs are important for up-regulation of membrane bound transporters such as efflux pumps, CNVs can confer drug resistance through additional mechanisms including the over-production of metabolic enzymes in targeted pathways (Nair et al. 2008; Marichal et al. 1997; Brochet et al. 2008) and the degradation or sequestration of antibiotic (Dumetz et al. 2018; Leprohon et al. 2009). One proposed mechanism of antibiotic resistance for Streptococcus pneumoniae and other bacteria involves replication fork stalling such that early-replicating, origin-proximal genes

are transiently up-regulated. Up-regulation of origin-proximal genes increases competence, which allows uptake of exogenous beneficial DNA such as drug-resistant plasmids (Slager et al. 2014). Secondary, compensatory CNVs that reduce the fitness burden of a primary antibiotic resistance mutation have also been identified in Salmonella enterica (Nilsson et al. 2006).

CNVs have been shown to aid in endowing insects with pesticide resistance. Mosquitos are a major pest and act as a vector of numerous diseases that affect millions of people each year (Benelli 2015). Insecticides are important for controlling these pests and the spread of vector-borne disease, but many mosquito populations have developed insecticide resistance. In at least two mosquito species that are disease vectors, amplification of detoxification genes including cytochrome P450s and esterases has been associated with resistance (Riveron et al. 2013, 2014; Itokawa et al. 2010; Mouches et al. 1986). Arthropod pests such as Lucilia cuprina, a blowfly, pose additional problems for agriculture. L. cuprina is a major pest of sheep populations in Australia and New Zealand, and there have been extensive attempts to eradicate using different organophosphates since the early 20th century (Mackerras 1936), with resistance developing rapidly thereafter. A study investigating multiple resistance in 41 strains of blowflies found that a duplication in the gene LcαE7, which encodes an esterase, had arisen in 7 strains. Within these, six shared at least one haplotype, and one strain had two unique haplotypes, suggesting that the duplication arose at least twice independently (Newcomb et al. 2005).

Conversely, polymorphic CNVs segregating within human populations can protect from parasites and infectious disease. Deletion alleles of the alpha-globin genes HBA1 and HBA2 protect against malaria and can be found at 20% population frequency in areas where malaria is endemic (Williams et al. 2005; Mockenhaupt et al. 2004; May et al. 2007). Additional CNVs are

likely to be involved in susceptibility to malaria infection (Faik, de Carvalho, and Kun 2009). Erythrocyte receptor genes CR1 and the glycophorins GYPA, GYPB, and GYPE are copy number variable among individuals and might affect the ability of Plasmodium falciparum to invade host cells (reviewed in (Hollox and Hoh 2014). CNVs have been implicated in susceptibility to a variety of other diseases including hepatitis C, tuberculosis, and HIV (Hollox and Hoh 2014).

### 1.5.5: Natural populations as emerging models for studying CNVs

While empirical studies performed in laboratory settings are important for determining the role of CNVs in adaptive evolution, there are a few caveats. Natural environments are complex, and can fluctuate (for example, in temperature, predation rates, or nutrient content). Even subtle variations in the environment can cause selective pressures to vary, or can increase the consequences of antagonistic pleiotropy. This is in direct contrast to adaptive laboratory evolution and even domestication, where a single, strong selective pressure is applied. In a recent study using Leishmania, the authors detected whole-chromosomal aneuploidies as major drivers of adaptation during in vitro culture, but identified smaller CNVs from clinical isolates adapting in the field (Bussotti et al. 2018). Field studies that examine the role of CNVs in situ are integral to a comprehensive understanding of the role of CNVs in adaptation.

The three-spined stickleback fish, Gasterosteus aculeatus, is a model for molecular divergence and adaptive evolution in natural environments. This fish occupies both marine and freshwater habitats throughout the northern hemisphere. Since the last Ice Age (~11,700 years ago), marine sticklebacks have repeatedly colonized freshwater lakes and streams in the northern hemisphere in formerly glaciated regions, resulting in many independent parallel

evolutions of adaptive phenotypes (Bell and Foster 1994). Hybridization between marine and freshwater populations allows genetic mapping of phenotypes (Chan et al. 2010; Peichel and Kingsley 2006; Colosimo et al. 2005; C. T. Miller et al. 2007), and the generation of a reference genome allows for comparative genomic analyses (Jones et al. 2012; Chain et al. 2014; Hirase, Ozaki, and Iwasaki 2014).

The first adaptive CNV in sticklebacks to be described was the deletion of a tissue specific regulatory enhancer of the Pituitary homeobox transcription factor 1 (Pitx1) gene (Chan et al. 2010). In marine species (the ancestral phenotype), the pelvic girdle is articulated with prominent serrated spines. Over two dozen freshwater populations have lost this structure (Bell 1987) because of a deletion in a 500 base pair regulatory region that drove expression of Pitx1 in the pelvis (Bell 1987; Chan et al. 2010). Multiple independent freshwater populations have Pitx1 enhancer deletions ranging in size from 488-1868 base pairs, and a reduction in heterozygosity around this locus which is a signature of selection (Bell 1987; Chan et al. 2010). Subsequent studies discovered significant standing CNVs in the marine (ancestral) populations (Feulner et al. 2012), and that these CNVs were adaptive and underwent positive selection during the transition to freshwater environments (Hirase, Ozaki, and Iwasaki 2014; Chain et al. 2014). In one comparative study between 20 stickleback genomes from marine environments across the northern hemisphere and 20 from freshwater environments, the authors identified 24 genes with significant differences in copy number (Hirase, Ozaki, and Iwasaki 2014). Among these were genes with roles in pathogen immunity and migration/brooding behavior, which could represent adaptations to novel life histories in freshwater lakes (Hirase, Ozaki, and Iwasaki 2014). Further studies will illuminate the role of CNVs in freshwater adaptation.

Arabidopsis lyrata is a small flowering plant closely related to the model organism A. thaliana which can be found in varied habitats throughout temperate and subarctic regions of the northern hemisphere (Hu et al. 2011). A study that compared two A. lyrata populations in serpentine soil, which has high heavy metal content and low mineral nutrients, to two populations from non-serpentine soils found that there were many CNVs in the serpentine populations: 94 duplications and 373 deletions. These CNVs included several genes involved in toxic compound extrusion and genes shown to protect roots from inhibitory compounds in A. thaliana (T. L. Turner et al. 2010). However, because the researchers only looked at four populations, they could not draw broad conclusions about the population genetics of CNVs in adaptation to serpentine soils. A. lyrata populations have independently colonized serpentine soils many times, and are often found near non-serpentine populations that can be used for comparison. Because A. lyrata is an outcrossing species, populations will show stronger local adaptation and these signals will be easier to detect than in the selfing model A. thaliana (Yant and Bomblies 2017). However, the abundance of resources and knowledge about A. thaliana will allow comparative and experimental elucidation of the mechanisms of adaptation in A. lyrata. Comparisons of CNVs segregating in both A. thaliana and A. lyrata may also allow inferences of which CNVs were segregating in ancestral populations and which CNVs arose de novo.

The species discussed above are particularly tractable for further studying the adaptive role of CNVs in natural populations. They have high quality reference genomes available, are comprised of recently diverged populations that can still interbreed, and natural selection has produced distinct phenotypes for well-defined ecological niches. These features enable genetic mapping of traits under selection and provides an opportunity to dissect fitness contributions by

measuring the variance explained by QTL. Parallel divergences between many different population pairs allows for the study of parallel adaptations at phenotypic and genetic levels, as well as the population dynamics underlying generation and selection of CNV alleles. By studying CNVs in these species and others like them, we may begin to understand generalizable characteristics of CNVs in natural evolving populations.

## 1.6: Population genetics of CNVs

### 1.6.1: Challenges for identifying CNVs under selection in evolving populations

Successful and accurate identification of CNVs is a serious challenge in the field. CNVs can be difficult to detect, especially if they are present at low population frequency or if they have been segregating in wild populations for many generations. Mutations can accumulate on copies of the duplicated gene and have various effects. A dominant negative mutation can promote selection for reduction back to a single copy (Cooke et al. 1997). SNVs and indels can change the function of copies, leading to processes such as pseudogenization, subfunctionalization or neofunctionalization (Ohno 1970; R. A. Jensen 1976; A. L. Hughes 1994; Force et al. 1999). These processes can promote the maintenance of amplification by requiring all copies for full gene function (subfunctionalization) or by acquiring new functions that have a selective benefit (neofunctionalization). However, mutations in divergent gene copies can be homogenized by interlocus gene conversion (J. B. Walsh 1987), which drives concerted evolution of gene duplicates (Dumont 2015; Hartasánchez et al. 2014) and affects our ability to accurately date the origin of a duplication (Pan and Zhang 2007). The interplay between

different types of mutations and their effects on divergent gene copies requires further study, especially with respect to the ultimate fate of a CNV in evolving populations.

Issues of CNV detection are compounded when determining whether or not a CNV is under selection. While there are many established metrics to identify signatures of selection for SNVs (Helyar et al. 2011; Leinonen et al. 2013; Stapley et al. 2010), there is less consensus on how to determine if selection is acting on CNVs. The relative roles of selection vs. neutral evolution (often referred to as “drift”) remain a hot topic of debate (Kern and Hahn 2018; Yoder et al. 2018; de Koning and De Sanctis 2018). Understanding the relative roles of these processes, as well as conducting further studies on the dynamics of selection (e.g. determining whether mutations are undergoing hard vs. soft sweeps), are often dependent on identifying when selection has acted on the genome. Studies of natural populations have employed a variety of techniques to infer selection on CNVs including: comparing the level of nucleotide change in CNVs compared to non-CNV genes (Chain et al. 2014), comparing maps of regions under selection as determined by SNVs (Roudnitzky et al. 2016), and identifying loss of heterozygosity or linkage disequilibrium around a CNV (Lowe et al. 2018; Conrad et al. 2010; Cardoso-Moreira et al. 2016). Other studies have identified CNVs under selection through associations with traits known to be under selection (Perry et al. 2007) or ecological variables (Lowe et al. 2018). Another method incorporates distribution of allele frequencies in order to determine the role of selection (Emerson et al. 2008). We know that at least some of these methods are problematic, as was illustrated by a study in which researchers found that fewer than 20% of their CNV associations had a corresponding SNV association (Stranger et al. 2007). We believe that a major area of research now and in the future will be determining the proper metrics for identifying CNVs under selection and their phenotypic associations,

especially in a high-throughput manner and in a way that does not just consider protein coding regions.

### 1.6.2: The population dynamics of CNVs

Gene duplications and deletions occur at a higher rate than single nucleotide variants (SNVs). Early on, researchers studying gene duplications, including the bar mutation, noticed this phenomenon (Sturtevant 1925). A clever genetic screen in E. coli revealed that mutation rates were high and that cells with amplifications quickly rose to a high frequency (Cairns and Foster 1991; Hastings et al. 2000). Reported frequencies of duplications per locus per generation range from $10^{-2}$ to $10^{-6}$ in E. coli and Salmonella (R. P. Anderson and Roth 1977; Horiuchi, Horiuchi, and Novick 1963; Reams et al. 2010; P. Anderson and Roth 1981; Starlinger 1977; Langridge 1969), $10^{-6}$ in yeast (Michael Lynch et al. 2008), $10^{-4}$ to $10^{-6}$ in Drosophila (Gelbart and Chovnick 1979; Shapira and Finnerty 1986), and $10^{-5}$ to $10^{-7}$ in human sperm (Lam and Jeffreys 2006; D. J. Turner et al. 2008). Further discussion of CNV formation rates are in (Katju and Bergthorsson 2013).

While these early estimates indicated that CNVs occurred at high rates, the dynamics with which CNVs undergo selection in evolving populations remain relatively unknown. Recent studies in S. cerevisiae demonstrate that CNVs arise early and predictably, then rise to high population frequency during many independently replicated evolution experiments (Gresham et al. 2008; Payen et al. 2014; Lauer et al. 2018). This striking degree of parallelism has also been seen in other systems: the algae Chloralla variablis co-evolving with a virus (Frickel et al. 2018) and Caenorhabditis elegans undergoing experimental evolution for increased fecundity (Farslow et al. 2015). CNVs have been identified early at high frequency during other evolution

experiments: after just 5 generations of heat stress at 28°C, researchers observed CNVs affecting hundreds of genes in A. thaliana, 52% of which were genes duplicated in tandem (DeBolt 2010). In one study with Salmonella, the steady-state population frequency of cells with duplications reached 20% by generation 48 of chemostat culture (Sun et al. 2012).

Segregating CNVs can undergo positive selection and become fixed in a population (Cardoso-Moreira et al. 2016; Kondrashov et al. 2002; Kondrashov 2012). However, the relative importance of “hard” vs. “soft” selective sweeps during rapid evolution is a major question in the field of population genetics (Messer and Petrov 2013; J. D. Jensen 2014; Harris, Sackman, and Jensen 2018). Hard sweeps occur when a single de novo variant arises and increases to fixation in a population (Smith and Haigh 1974), while soft sweeps occur when multiple adaptive alleles increase in frequency in a population at the same time (soft sweeps could refer to selection on existing variation and/or recurrent de novo mutation, but usually refers to the former) (Hermisson and Pennings 2005; Nair et al. 2007). Population genetics models have historically assumed mutation limited scenarios and therefore a prevalence of hard sweeps, but it is becoming apparent that rapid adaptation is actually more often facilitated by soft sweeps (Messer and Petrov 2013). SNVs have been the focus of the majority of population genetic studies investigating selective sweeps while the role of CNVs in selective sweeps is rarely considered.

Since CNVs tend to have both a higher mutation rate and larger effect size than SNVs, the relative frequency of hard vs. soft sweeps in CNVs may not be the same. Evidence from sticklebacks (Chain et al. 2014), humans (Conrad et al. 2010), and microbial populations (Nair et al. 2007; Payen et al. 2014; Lauer et al. 2018) suggests that soft sweeps are also a dominant mode of adaptation for CNVs, but not at the exclusion of hard sweeps. Further studies in both

laboratory and natural populations are needed to address this question. Additionally, it is of interest to discover how much of rapid adaptation is facilitated by selection on standing variation versus de novo variation, that is, the relative impact of “survival of the fittest” and “arrival of the fittest” (Gruber et al. 2012). Studies in bacteria (R. P. Anderson and Roth 1977; Sonti and Roth 1989) and yeast (Lauer et al. 2018) demonstrate that both forms of variation are likely to contribute. Additional studies show that there is a significant amount of apparently neutral CNVs in populations (Zarrei et al. 2015), and that when environments change, selection acts on this variation (Chain et al. 2014d). This may be because more CNVs tend to be neutral or nearly neutral more often than previously believed, or it may be that the high mutation rate of CNVs leads to a large supply.

Non-additive, or epistatic, interactions can constrain the outcome and order of mutational events during adaptive evolution (D.-H. Lee and Palsson 2010; Chou et al. 2011; Khan et al. 2011). However, interactions between beneficial SNVs and CNVs remain largely uncharacterized and it is unclear whether the principles governing epistasis among SNVs are maintained for these interactions. In one laboratory evolution experiment where S. cerevisiae was grown in glucose-limited conditions, two independent adaptive mutations were identified: a loss-of-function mutation in MTH1, which negatively regulates glucose-sensing signaling pathways, and duplication of the high-affinity hexose transporters HXT6/7 (Kvitek and Sherlock 2011). When combined, these two mutations produced a strain less fit than the ancestor, resulting in negative epistasis (Kvitek and Sherlock 2011). The authors hypothesized that both mutations lead to an over-production of glucose transporters on the cell membrane. A neutral large-scale duplication potentiated an adaptive deletion in E. coli (Maharjan et al. 2013), demonstrating that there can also be epistatic relationships among two CNV alleles.

Co-occurence of two CNV alleles has been observed in other systems, including CNVs in two virulence factors for the fish pathogen, R. salmoninarum (Brynildsrud et al. 2016). Further studies are required to disentangle the potential effects of positive and negative epistasis among CNVs and between CNVs and SNVs.

## 1.7: Consequences and costs of CNVs

CNVs can have dramatic consequences for organismal fitness because they are very rarely duplications or deletions encompassing a single gene. While tandem duplications or gene amplifications may be targets of selection due to a single adaptive locus, additional protein-coding genes and intergenic sequences across the entire CNV can affect fitness. In some systems, the co-duplication of adjacent genes specifically provides a fitness benefit (Reams and Neidle 2004). In one study where the authors imposed a selection for increasing copies of one particular gene, they instead found extremely large spontaneous duplications (Koszul et al. 2004). This phenomenon has been repeatedly seen in yeast adapting to nutrient limitation, where neighboring genes are amplified in addition to the gene encoding a specific nutrient transporter (Lauer et al. 2018). In these studies, CNVs are over 100 kilobases long and can encompass an entire chromosome arm (Lauer et al. 2018). Such large CNVs are likely to be associated with fitness costs, which can be attributed to disruption of cellular homeostasis at multiple levels: inherent costs due to increases in genome size (Elde et al. 2012), changes to local and global gene expression (Sheltzer et al. 2012), increased translational capacity and changes to protein stoichiometry (Torres et al. 2007), or increased burden on protein degradation machinery (Torres et al. 2010; Stingele et al. 2012).

In one study, there was a 0.15% reduction in fitness for every kilobase pair amplified in E. coli (Adler et al. 2014). Selection of large-scale deletions that remove unnecessary "accessory genes" (i.e. genes that are beneficial only in specific environments) has been observed in other systems (M.-C. Lee and Marx 2012). In another study, there was no correlation between the size of the duplicated region and fitness reduction for the organism (Pettersson et al. 2009). In order to determine the precise fitness consequences of any given CNV, exhaustive experiments must be performed in a variety of environments and conditions, as trade-offs between these and the original selective condition are likely to exist.

Since CNVs can confer substantial fitness costs, it has been proposed that they occur transiently and are not effective long-term solutions for organisms adapting to stressful conditions. Early experiments in bacteria demonstrated that CNVs were lost spontaneously at very high rates after removing the selective pressure (R. P. Anderson and Roth 1977; Kugelberg et al. 2006; Winfield and Falkinham 1981). Gene duplications can be rapidly eliminated by recombination (Adler et al. 2014) or purged by the cell if they are on extrachromosomal circles (Gresham et al. 2010). One proposed mechanism of alleviating fitness costs is the use of "genomic accordions," which involve expansions and contractions of genic arrays (Roth and Andersson 2012). Gene duplications occur at a high rate, and incremental increases in gene dosage improve cell growth such that cells with the duplication rise to high population frequency. Multiple gene copies, as well as many individuals with multiple copies, increase the likelihood of generating beneficial SNVs in the gene under selection (Sun et al. 2009). If these SNVs provide significant fitness benefits, selection on maintenance of multiple gene copies is relaxed, and the alternative copies can be subsequently lost. This phenomenon has been observed in a variety of systems: viruses adapting to host defenses (Elde et al. 2012), bacteria growing in

lactose-limiting environments (Slechta et al. 2003), the evolution of antibiotic resistance (Pränting and Andersson 2011; Paulander, Andersson, and Maisnier-Patin 2010), and the evolution of metabolic enzymes (reviewed in (Copley 2012)).

While CNVs have been identified as drivers of rapid adaptation across the domains of life, determining their ultimate fate in complex, evolving populations remains difficult. While empirical evidence is rare, the authors of one laboratory evolution experiment with S. enterica found that duplicate genes can undergo specialization (Näsvall et al. 2012). To perform this study, the authors generated mutations in the histidine biosynthesis enzyme, HisA, that enabled the promiscuous catalyzation of a step in tryptophan biosynthesis. After 3,000 generations of selection for histidine and tryptophan synthesis, the modified HisA gene was amplified and subsequently diverged to generate two copies with distinct enzyme specificities. Evolution experiments such as these provide us with an opportunity to witness long-term evolutionary processes in short-term laboratory scenarios, and additional experiments should help elucidate the ultimate fate of CNVs.

## 1.8: Open Questions

In this review, we discussed the emerging understanding of CNVs as drivers of rapid adaptation. Across the domains of life, adaptive CNVs are important for sensing and responding to changes in the environment under a broad range of conditions including: the presence of new ecological niches, the availability of nutrient content, extreme temperatures, host defenses, and the presence of heavy metals, toxins or antibiotics. CNVs are distinct from SNVs in several respects, making them more difficult to characterize. As large-effect alleles, they are more likely to result in pleiotropic effects, dramatic loss-of-function phenotypes, and trade-offs in alternative

environments. There are many aspects of CNVs that are not well understood, and we identify several open questions that are outstanding in the field.

CNVs provide fitness benefits through diverse processes: increasing gene dosage, masking deleterious mutations, promoting heterozygote advantage, resolving genetic conflicts, generating positional effects or chimeric genes, and globally altering expression landscapes. Determining the functional effects of any given CNV remains challenging and usually requires multiple, independent empirical tests. As a result, identifying causal variants in natural populations or performing genome-wide analysis is difficult. Researchers have begun to solve this problem by combining genome sequence analysis with RNA sequencing to determine if there are any direct effects of CNV formation on transcription levels. Recent evidence indicates that CNVs may even alter the chromatin landscape, indicating that further integration of sequencing analysis (including chromosome conformation capture and identification of open chromatin by ATAC-seq and other methods) may be required to fully address the functional effects of CNVs.

CNVs are a complex class of structural variant that can be generated by a diverse range of processes and mechanisms. Repetitive sequences including segmental duplications are known to generate CNVs through non-allelic homologous recombination. Recent advances in the field have highlighted the role of replication-based errors in CNV formation. However, the extent to which replication, transcription or environmental condition affect the generation, selection, or maintenance of CNVs in evolving populations remains unresolved. While the environment has been shown to stimulate CNV formation at two specific loci (the rDNA and the CUP1 locus in yeast), further studies are needed to address these questions. To determine the

precise effect of transcription on CNV mutation rate, empirical experiments that decouple this process from any potential environmental stimulation are also required.

The population dynamics of CNVs have been difficult to resolve, especially due to limitations in accurately detecting CNVs in complex, heterogeneous evolving populations if they are present at low frequency. This is compounded when determining whether a given CNV is under selection. While some CNVs may be generated at a high rate and quickly rise to high frequency within populations, we have yet to resolve the respective contributions of standing genetic variation and de novo variation, hard and soft selective sweeps, and epistatic interactions. Recent advances, including the use of barcode lineage tracking (Levy et al. 2015; Lauer et al. 2018), are poised to further our understanding of these population genetics questions and controversies. While adaptive laboratory evolution experiments have begun to dissect the roles of CNV in driving these processes, we propose that the use of non-model natural populations will provide additional insights.

While CNVs can provide fitness benefits, they can also underlie maladaptive phenotypes and disease. Importantly, the adaptive potential of any given CNV is context- and environment-dependent, meaning that any beneficial effects observed are likely to be specific. Changes to the environment and/or the selective pressure (using fluctuating temperatures rather than a static high temperature for example) may therefore decrease or eliminate any fitness benefit conferred by the CNV. Moreover, the larger the CNV allele, the more likely it is to have negative consequences in alternative environments. In addition, the adaptive potential and fitness benefits of CNVs are complex, which makes it difficult to predict the long-term fate of CNVs segregating within populations. To address these concerns, future studies should focus on the fitness effects of a CNV in multiple environmental conditions, under fluctuating selective

pressures, in different subpopulations or strain backgrounds, and across longer evolutionary timescales.

## 1.9: Conclusions

We have proposed several open questions concerning the functional effects, mechanisms, population genetics, and trade-offs associated with CNV alleles. Bridging these gaps in knowledge will require further empirical and observational studies. A synthesis among the fields of population genetics, genomics, and evolutionary biology and the combined use of laboratory evolution experiments and studies of natural populations will be useful for answering these questions. For example, an in vitro evolution experiment in the malarial parasite P. falciparum revealed previously known and unknown candidate CNVs involved in drug resistance (Cowell et al. 2018), which can be used to inform future clinical field studies. We have also discussed the utility of integrating multiple genomics approaches and using understudied, natural populations as models for future CNV studies. As drivers of rapid adaptation in diverse scenarios, CNVs are an important source of genetic diversity and evolutionary potential, and new technical advancements in the field are key to understanding this understudied class of mutations.

# Chapter 2: Single-cell copy number variant detection reveals the dynamics and diversity of adaptation

This chapter is based on the research paper “Single-cell copy number variant detection reveals the dynamics and diversity of adaptation” by Stephanie Lauer, Grace Avecilla, Pieter Spealman, Gunjan Sethia, Nathan Brandt, Sasha F. Levy and David Gresham, with an anticipated publication in PLoS Biology December 2018. I generated all of the data for Figures 2.1-3 and Table 2.1, contributed to generating the data for Figures 2.4-5 and Tables 2.2-4, wrote

the manuscript text, and generated most of the supplementary figures presented here (2.S11 and 2.S12 being the exceptions).

## 2.1: Abstract

Copy number variants (CNVs) are a pervasive source of genetic variation and evolutionary potential, but the dynamics and diversity of CNVs within evolving populations remains unclear. Long-term evolution experiments in chemostats provide an ideal system for studying the molecular processes underlying CNV formation and the temporal dynamics with which they are generated, selected, and maintained. Here, we developed a fluorescent CNV reporter to detect de novo gene amplifications and deletions in individual cells. We used the CNV reporter in Saccharomyces cerevisiae to study CNV formation at the GAP1 locus, which encodes the general amino acid permease, in different nutrient-limited chemostat conditions. We find that under strong selection, GAP1 CNVs are repeatedly generated and selected during the early stages of adaptive evolution resulting in predictable dynamics. Molecular characterization of CNV-containing lineages shows that the CNV reporter detects different classes of CNVs including aneuploidies, non-reciprocal translocations, tandem duplications, and complex copy number variants. Despite GAP1's proximity to repeat sequences that facilitate intrachromosomal recombination, breakpoint analysis revealed that short inverted repeat sequences mediate formation of at least 50% of GAP1 CNVs. Inverted repeat sequences are also found at breakpoints at the DUR3 locus, where CNVs are selected in urea-limited chemostats. Analysis of 28 CNV breakpoints indicates that inverted repeats are typically 8 nucleotides in length and separated by 40 bases. The features of these CNVs are consistent with origin dependent inverted repeat amplification (ODIRA) suggesting that replication-based mechanisms of CNV formation may be a common source of gene amplification. We combined the CNV reporter with

barcode lineage tracking and found that $10^2$-$10^4$ independent CNV-containing lineages initially compete within populations resulting in extreme clonal interference. However, only a small number (18-21) of CNV lineages ever comprise more than 1% of the CNV subpopulation, and as selection progresses, the diversity of CNV lineages declines. Our study introduces a novel means of studying CNVs in heterogeneous cell populations and provides insight into their dynamics, diversity, and formation mechanisms in the context of adaptive evolution.

## 2.2: Introduction

Copy number variants (CNVs) drive rapid adaptive evolution in diverse scenarios ranging from niche specialization to speciation and tumor evolution (Conant and Wolfe 2008; Zuellig and Sweigart 2018; Shlien and Malkin 2009; Stratton, Campbell, and Futreal 2009). CNVs, which include duplications and deletions of genomic segments, underlie phenotypic diversity in natural populations (Barreiro et al. 2008; Iskow et al. 2012; Clop, Vidal, and Amills 2012; Żmieńko et al. 2014; Greenblum, Carr, and Borenstein 2015; Zarrei et al. 2015), and provide a substrate for evolutionary novelty through modification of existing heritable material (Ohno 1970; Michael Lynch and Conery 2000; Austin L. Hughes 1994; R. P. Anderson and Roth 1977). Beneficial CNVs are associated with defense against disease in plants, increased nutrient transport in microbes, and drug resistant phenotypes in parasites and viruses (Iantorno et al. 2017; Cowell et al. 2018; Dolatabadian et al. 2017; Elde et al. 2012; Greenblum, Carr, and Borenstein 2015). Despite the importance of CNVs for phenotypic variation, evolution and disease, the dynamics with which these alleles are generated and selected in evolving populations are not well understood.

Long term experimental evolution provides an efficient means of gaining insights into

evolutionary processes using controlled and replicated selective conditions (Lenski et al. 1991; Good et al. 2017). Chemostats are devices that maintain cells in a constant nutrient-poor growth state using continuous culturing (Gresham and Dunham 2014). Nutrient limitation in chemostats provides a defined and strong selective pressure in which CNVs have been repeatedly identified as major drivers of adaptation. CNVs containing the gene responsible for transporting the limiting nutrient are repeatedly selected in a variety of organisms and conditions including Escherichia coli limited for lactose (Horiuchi, Horiuchi, and Novick 1963), Salmonella typhimurium in different carbon source limitations (Sonti and Roth 1989), and Saccharomyces cerevisiae in glucose-, phosphate-, sulfur- and nitrogen-limited chemostats (Hong and Gresham 2014; Gresham et al. 2010; Payen et al. 2014; Gresham et al. 2008; Brown, Todd, and Rosenzweig 1998; Kao and Sherlock 2008; Hansche 1975). CNVs confer large selective advantages and multiple, independent CNV alleles have been identified within experimental evolution populations (Payen et al. 2014; Gresham et al. 2008; Kvitek and Sherlock 2011; Gresham et al. 2010). These findings suggest that CNVs are generated at a high rate, but estimates differ greatly, ranging from $1 \times 10^{-10}$ to $3.4 \times 10^{-6}$ duplications per cell per division, with variation in CNV formation rates potentially differing between loci and/or condition (Dorsey et al. 1992; Michael Lynch et al. 2008). A high rate of CNV formation suggests that multiple, independent CNV-containing lineages may compete during adaptive evolution resulting in clonal interference, which is characteristic of large, evolving populations (Lang et al. 2013; J. M. Hughes et al. 2012; Maddamsetti, Lenski, and Barrick 2015; Kao and Sherlock 2008). However, the extent to which clonal interference among CNV-containing lineages influences the dynamics of adaptation is unknown.

The general amino acid permease gene, GAP1, is well suited to studying the role of

CNVs in adaptive evolution. GAP1 encodes a high-affinity transporter for all naturally occurring amino acids, and it is highly expressed in nitrogen-poor conditions (Grenson, Hou, and Crabeel 1970; Stanbrough and Magasanik 1995). We have previously shown that two classes of CNVs are selected at the GAP1 locus in S. cerevisiae when a sole nitrogen source is provided: GAP1 amplification alleles are selected in glutamine and glutamate-limited chemostats and GAP1 deletion alleles are selected in urea- and allantoin-limited chemostats (Gresham et al. 2010; Hong and Gresham 2014). GAP1 CNVs are also found in natural populations. In the nectar yeast, Metschnikowia reukaufii, multiple tandem copies of GAP1 result in a competitive advantage over other microbes when amino acids are scarce (Dhami, Hartwig, and Fukami 2016). As a target of selection in adverse environments in both experimental and natural populations, GAP1 is a model locus for studying the dynamics and mechanisms underlying both gene amplification and deletion in evolving populations.

CNVs are generated by two primary classes of mechanisms: homologous recombination and DNA replication (Hastings, Lupski, et al. 2009; Reams and Roth 2015; Carvalho and Lupski 2016). DNA double strand breaks (DSBs) are typically repaired by homologous recombination and do not result in CNV formation. However, non-allelic homologous recombination (NAHR) can generate CNVs when the incorrect repair template is used, which occurs more often with repetitive DNA sequences such as transposable elements and long terminal repeats (LTRs) (Stankiewicz and Lupski 2002). During DNA replication, stalled and broken replication forks can re-initiate DNA replication through processes including break-induced replication (BIR), microhomology-mediated break-induced replication (MMBIR), and fork stalling and template switching (FoSTes) (J. A. Lee, Carvalho, and Lupski 2007; Hastings, Ira, et al. 2009; Payen et al. 2008). BIR is driven by homologous sequences, whereas MMBIR relies on shorter stretches

of sequence homology. Recently, origin-dependent inverted-repeat amplification (ODIRA) has been identified as a novel mechanism underlying amplification of the SUL1 locus in yeast (Brewer et al. 2011, 2015). ODIRA is mediated by short inverted repeat sequences that facilitate ligation of the the leading and lagging strands following regression of the replication fork during DNA synthesis. ODIRA is hypothesized to involve the formation of an extrachromosomal circular intermediate that replicates independently and therefore requires an origin of replication within the amplified region. Subsequent integration of the circle into the original locus via homologous recombination results in an inverted triplication. Extrachromosomal circular DNA is common in yeast (Møller et al. 2015), can drive tumorigenesis (K. M. Turner et al. 2017), and may represent a rapid and reversible mechanism of generating adaptive CNVs (Møller, Andersen, and Regenberg 2013; Cohen and Segal 2009). Previously, we found that some GAP1 amplifications are extrachromosomal circular elements. We hypothesized that $GAP1^{circle}$ alleles are generated as a result of NAHR between flanking LTRs resulting in their excision from the chromosome (Gresham et al. 2010). Identifying the mechanisms underlying CNV formation is required for understanding the roles of CNVs in evolutionary processes and human disease.

A key limitation to the study of CNVs in evolving populations is the challenge of identifying them at low frequencies in heterogeneous populations. CNVs are typically detected using molecular methods including qPCR, Southern blotting, DNA microarrays and sequencing (Gresham et al. 2010; Payen et al. 2014; Hong and Gresham 2014). However, using any of these methods, de novo CNVs are undetectable in a heterogeneous population until present at high frequency (e.g. >50%). This precludes analysis of the early dynamics with which CNVs emerge and compete in evolving populations. As CNVs usually comprise genomic regions that include multiple neighboring genes (Hong and Gresham 2014), we hypothesized that CNVs

could be identified on the basis of increased expression of a constitutively expressed fluorescent reporter gene inserted adjacent to a target gene of interest. A major benefit of this approach is that it detects CNVs independently of whole genome sequencing, enabling a high-resolution and efficient assay of CNV dynamics with single-cell resolution in evolving populations.

In this study, we constructed strains containing a fluorescent CNV reporter adjacent to GAP1 in S. cerevisiae and performed evolution experiments in different selective environments using chemostats. The CNV reporter allowed us to visualize selection of CNVs at the GAP1 locus in real time with unprecedented temporal resolution. We find that CNV dynamics occur in two distinct phases: CNVs are selected early during adaptive evolution and quickly rise to high frequencies, but the subsequent dynamics are complex. We find that GAP1 CNVs are diverse in size and copy number, and can be generated by a range of processes including aneuploidy, non-reciprocal translocations and tandem duplication by NAHR. Nucleotide resolution analysis of GAP1 CNV breakpoints revealed that CNV formation is mediated by short, interrupted inverted repeats for half of the resolvable cases, suggesting that replication-based mechanisms also underlie gene amplification at the GAP1 locus. The presence of inverted repeats, in combination with a replication origin and inverted triplication, is consistent with GAP1 CNV formation through ODIRA. ODIRA may be a major source of de novo CNVs in yeast, as these breakpoint features also characterize CNVs at an additional locus identified in our study, DUR3. To determine the underlying structure of the CNV subpopulation, we generated a lineage-tracking library using random DNA barcodes. FACS-based fractionation of CNV lineages and barcode sequencing identified hundreds to thousands of individual CNV lineages within populations, consistent with a high CNV supply rate and extreme clonal interference.

Together, our results show that CNVs are generated repeatedly by diverse processes, resulting in predictable dynamics, but that the long term fate of CNV-containing lineages in evolving populations is shaped by clonal interference and additional variation.

## 2.3: Results

### 2.3.1: Protein fluorescence increases proportionally with gene copy number

We sought to construct a reporter for CNVs that occur at a given locus of interest. Based on previous studies (Suzuki et al. 2011; Gruber et al. 2012; Kafri et al. 2016; Steinrueck and Guet 2017), we hypothesized that CNVs that alter the number of copies of a constitutively expressed fluorescent protein gene would facilitate single cell detection of de novo copy number variation. To test the feasibility of this approach, we constructed haploid S. cerevisiae strains isogenic to the reference strain (S288c) with one or two copies of a constitutively expressed GFP variant mCitrine (Griesbeck et al. 2001), and diploid strains with 1-4 copies of mCitrine, integrated into the genome.

Flow cytometry analysis confirmed that additional copies of mCitrine produce quantitatively distinct distributions of protein fluorescence (**Figure 2.1A**). Haploid cells with two copies of mCitrine have higher fluorescence than those with a single copy and there is minimal overlap between the distributions of fluorescent signal in the two strains. Normalization of the fluorescent signal by forward scatter, which is correlated with cell size, shows that the concentration of fluorescent protein is proportional to the ploidy normalized copy number of the mCitrine gene (i.e. one copy in a haploid results in a signal equivalent to two copies in a diploid and two copies in a haploid results in a signal similar to four copies in a diploid). Thus, the cell size-normalized fluorescent signal, or concentration, accurately reports on the number of copies

of the fluorescent gene in single cells. Therefore, integrating a constitutively expressed fluorescent protein gene proximate to an anticipated target of selection functions as a CNV reporter for tracking gene amplifications and deletions in evolving populations (**Figure 2.1B**).

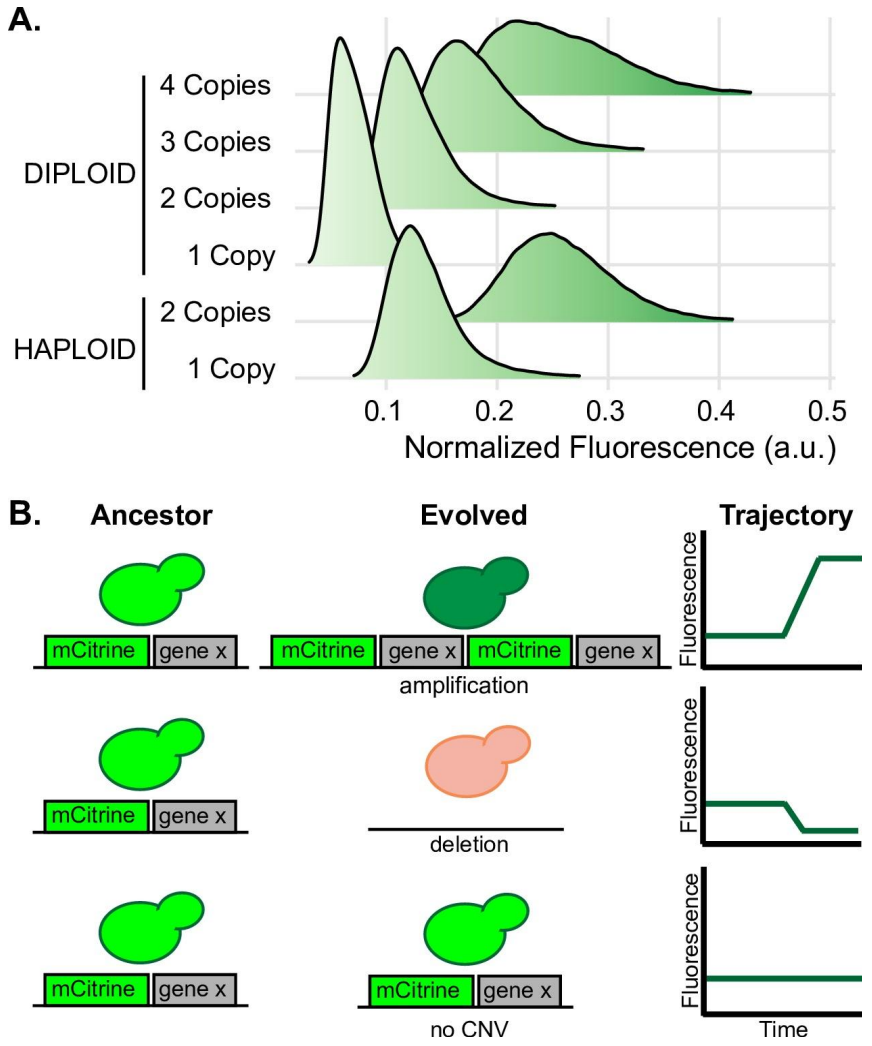


**Figure 2.1. Fluorescent protein signal is proportional to gene copy number.** (**A**) Protein fluorescence increases with increasing copies of the mCitrine gene. We determined the fluorescence of haploid and diploid cells containing variable numbers of a constitutively expressed mCitrine gene integrated at either the HO locus and/or the dubious ORF, YLR123C. The two copy diploid is heterozygous at both loci. Each distribution was estimated using 100,000 single cell measurements normalized by forward scatter. (**B**) Schematic representation of how the fluorescent reporter enables CNV detection in heterogeneous evolving populations through quantitative changes in protein fluorescence.

### 2.3.2: A CNV reporter tracks the dynamics of GAP1 CNVs in real time

Previous work has shown that spontaneous GAP1 amplifications are positively selected when glutamine is the sole limiting nitrogen source during evolution experiments in chemostats (Gresham et al. 2010). GAP1 copy number amplifications result in increased amino-acid transporters on the plasma membrane, providing cells with a selective advantage when nitrogen is scarce (Gresham et al. 2010; Hong and Gresham 2014). Conversely, GAP1 deletions provide a fitness benefit and are selected in urea-limited conditions (Gresham et al. 2010), which may be due to two non-exclusive reasons: either 1) because GAP1 is highly expressed regardless of the type of limiting nitrogen source (Airoldi et al. 2016) but unable to transport urea, it confers a gene expression burden, or 2) when the extracellular concentration of amino acids is low compared to the intracellular concentration, the electrochemical gradient drives their export through the GAP1 permease. Thus, the use of different nitrogen sources in nitrogen-limited chemostats enables the study of both GAP1 amplification and deletion, making it an ideal system for studying the dynamics of CNV selection in evolving populations.

We constructed a haploid strain containing a mCitrine CNV reporter located 1,118 bases upstream of the GAP1 start codon to ensure that the native regulation of GAP1 was unaffected (Stanbrough and Magasanik 1996). We inoculated the GAP1 CNV reporter strain into 9 glutamine-, 9 urea- and 8 glucose-limited chemostats, for a total of 26 populations. For each of the three selection conditions, we included two control populations: one containing a single copy of the mCitrine CNV reporter at a neutral locus (one copy control) and one containing two copies of the mCitrine CNV reporter at two neutral loci (two copy control). All populations were maintained in continuous mode (dilution rate = 0.12 culture volumes/hr; population doubling time

= 5.8 hours) for 267 generations over 65 days. We sampled each of the 32 populations every 8 generations and used flow cytometry to measure fluorescence of 100,000 cells per sample.

Experimental evolution in a glutamine-limited chemostat resulted in clear increases in fluorescence in individual cells containing the GAP1 CNV reporter by generation 79 (**Figure 2.2A**). By contrast, populations containing one or two copies of mCitrine at neutral loci exhibited stable fluorescence for the duration of the experiment (**Figure 2.2A**). Maintenance of protein fluorescence in one and two copy control populations is consistent with the absence of a detectable fitness cost associated with one or two copies of the CNV reporter in glutamine-limited chemostats, which we confirmed using competition assays (**Figure 2.S1**). Analysis of eight additional independent populations evolving in glutamine-limited chemostats showed qualitatively similar dynamics of single-cell fluorescence over time (**Figure 2.S2**). To summarize the dynamics of CNVs in evolving populations, we determined the median normalized fluorescence in each population at each time point. The fluorescent signal of the GAP1 CNV reporter increases during selection in all populations evolving in glutamine-limited chemostats (**Figure 2.2B**), consistent with the de novo generation and selection of CNVs at the GAP1 locus in all 9 populations.

Populations evolving in urea-limited and glucose-limited chemostats do not show substantial changes in fluorescence with one exception **(Figure 2.2B)**. In a single urea-limited population (ure_05), we detected a complete loss of fluorescent signal by generation 125, indicating the occurrence of a GAP1 deletion that subsequently swept to fixation. Thus, the GAP1 CNV reporter detects both amplification and deletion alleles at the GAP1 locus in evolving populations. The absence of increases or decreases in fluorescence in all

glucose-limited populations is consistent with the absence of selection for GAP1 CNVs in conditions that are irrelevant for GAP1 function.

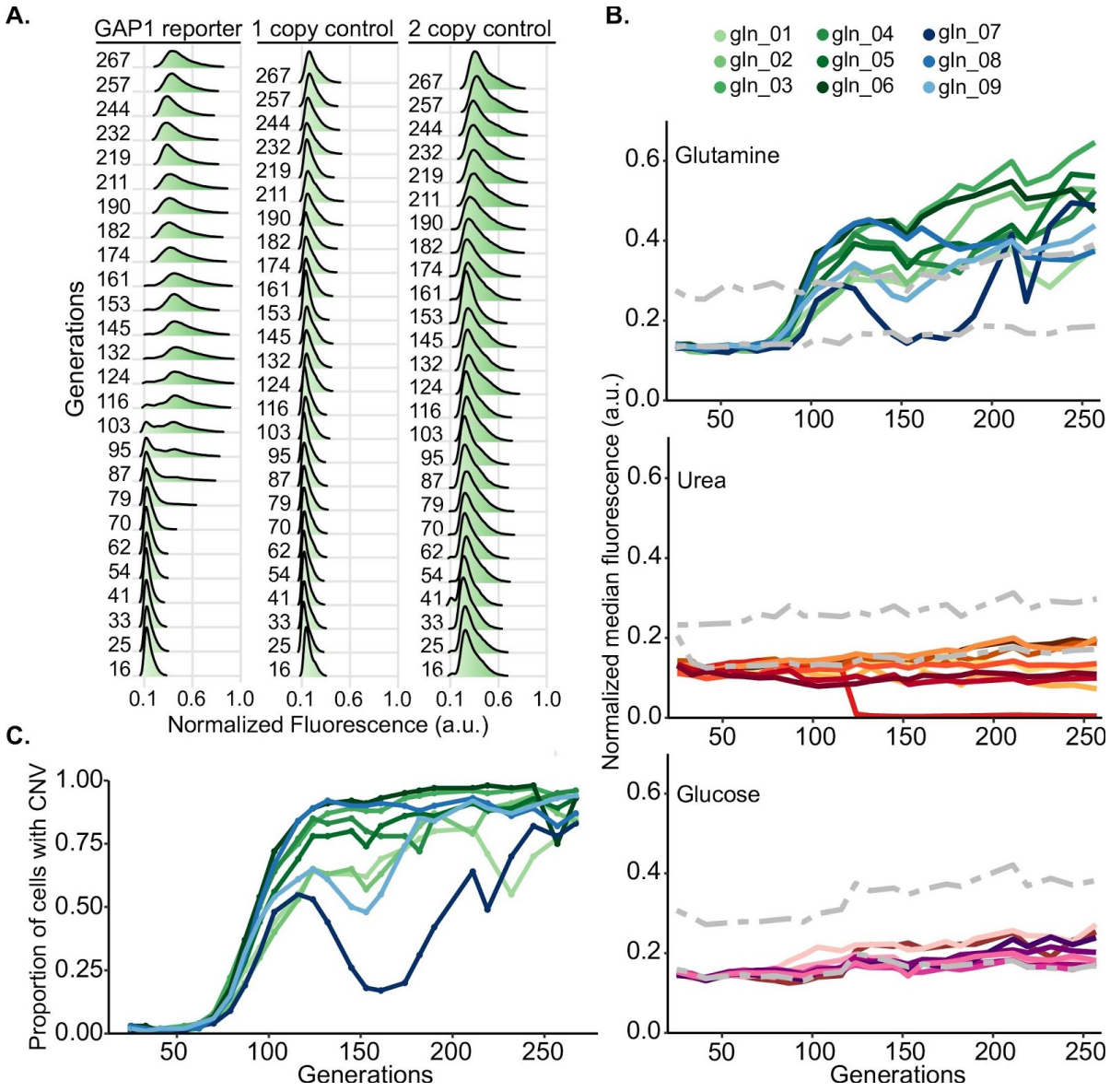


**Figure 2.2. Dynamics of** GAP1 **CNVs in evolving populations.** (A) Normalized distributions of single-cell fluorescence over time for a representative GAP1 CNV reporter strain and one and two copy control strains evolving in glutamine-limited chemostats. Single cell fluorescence is normalized by the forward scatter measurement of the cell. (B) Normalized median fluorescence for each population evolving in glutamine- (n = 9), urea- (n= 9) and glucose- (n = 8) limited chemostats. The fluorescence of the one and two copy control strains is plotted for reference (grey dotted lines). (C) Estimates of the proportion of cells with GAP1 amplifications over time for nine glutamine-limited populations containing the GAP1 CNV reporter.

To quantify the proportion of cells containing a GAP1 duplication, we used one and two copy control strains to define flow cytometry gates. We found that the fluorescence of control strains varied slightly (**Figure 2.S3A**), which may be indicative of either instrument variation or changes in cell physiology and morphology during the experiment as suggested by systematic changes in forward scatter with time (**Figure 2.S3B**). Using a conservative method to classify individual cells containing GAP1 amplifications (**methods**), we find that GAP1 amplification alleles are selected with remarkably reproducible dynamics in the nine glutamine-limited populations (**Figure 2.2C**). CNVs are predominantly duplications (two copies), but quantification of fluorescence suggests that many cells contain three or more copies of the GAP1 locus (**Figure 2.S4**).

We quantified the dynamics of CNVs in each population evolved in glutamine-limited chemostats using metrics defined by Lang et al. (Lang, Botstein, and Desai 2011). CNVs are detected by generation 70-75 (average = 72.8) in all 9 populations ($T_{up}$) (**Table 2.1**). To estimate the fitness of all CNV lineages relative to the mean population fitness, we calculated $S_{up}$, the rate of increase in the abundance of the CNV subpopulation (see **methods**). The average relative fitness of the CNV subpopulation is 1.077 ($S_{up}$) and CNV alleles are at frequencies greater than 75% in all populations by 250 generations (**Table 2.1**). Thus, in all replicated glutamine-limited selection experiments, GAP1 amplifications emerge early, increase in frequency rapidly, and are maintained in each population throughout the selection.

GAP1 CNVs undergo two distinct phases of population dynamics. The initial dynamics with which CNV subpopulations emerge and increase in frequency are highly reproducible in independent evolving populations. However, after 125 generations, the trajectories of the CNV subpopulation in the different replicate populations diverge. Many populations maintain a high

frequency of GAP1 amplification alleles, but in some populations they decrease in frequency. In one population, GAP1 CNV alleles are nearly lost from the population before subsequently increasing to an appreciable frequency (gln_07).

**Table 2.1. Summary statistics of** GAP1 **CNV dynamics in glutamine-limited chemostats.** $T_{up}$ is the number of elapsed generations before CNVs are reliably detected (>7% frequency, see **methods**). $S_{up}$ is the rate of increase in CNV abundance during the initial expansion of the CNV subpopulation. The frequency of CNVs in the population at generation 150 and generation 250, when genome sequencing was performed, is also reported.

| Population | $T_{up}$ | $1 + S_{up} \pm SE$ | g150% | g250% |
|---|---|---|---|---|
| gln_01 | 70 | 1.066 ± 0.0038 | 62 | 77 |
| gln_02 | 75 | 1.071 ± 0.0034 | 57 | 87 |
| gln_03 | 70 | 1.071 ± 0.0037 | 88 | 94 |
| gln_04 | 70 | 1.079 ± 0.0036 | 80 | 95 |
| gln_05 | 75 | 1.077 ± 0.0041 | 74 | 89 |
| gln_06 | 70 | 1.082 ± 0.0043 | 91 | 75 |
| gln_07 | 75 | 1.094 ± 0.0048 | 18 | 78 |
| gln_08 | 75 | 1.090 ± 0.0052 | 90 | 82 |
| gln_09 | 75 | 1.066 ± 0.0050 | 48 | 93 |
| **AVG ± STD** | 72.8 ± 2.6 | 1.077 ± 0.01 | 68 ± 24 | 86 ± 8 |

## 2.3.3: GAP1 CNV alleles are diverse within and between replicate populations

Based on prior studies (Payen et al. 2014; Hong and Gresham 2014), we hypothesized that multiple CNV alleles exist within each population. To characterize the diversity of GAP1 CNVs, we isolated a total of 29 clones containing increased fluorescence from glutamine-limited chemostats at 150 and 250 generations for whole genome sequencing. We used read depth to calculate GAP1 copy number and to estimate CNV boundaries (**Figure 2.3A** and **methods**). We

find that GAP1 copy number estimated by sequencing read depth correlates with the fluorescent signal for individual clones (**Figure 2.3B**), indicating that fluorescent signal is predictive of copy number. In 3 clones, we find increased read depth across the entirety of chromosome XI consistent with aneuploidy. Thus, the CNV reporter is able to detect aneuploid chromosomes as well as subchromosomal CNVs.

We identified diverse GAP1 CNVs between and within populations (**Figure 2.3C**). In the majority of populations (6/9) different clones had different CNVs. For example, in population gln_01 at generation 150, we identified a large GAP1 CNV that includes the entire right arm of chromosome XI and another clone that was aneuploid for chromosome XI. At generation 250, clones isolated from population gln_01 have CNV alleles that are distinct from each other and from those observed at generation 150. Clones from the 8 additional glutamine-limited populations show evidence for CNV diversity within and between the two time points analyzed (**Figure 2.3C**) suggesting the presence of multiple CNV lineages within evolving populations. Furthermore, the diversity of GAP1 CNVs indicates that they are not predominantly formed through a recurrent mechanism as might be anticipated by the presence of proximate repetitive elements.

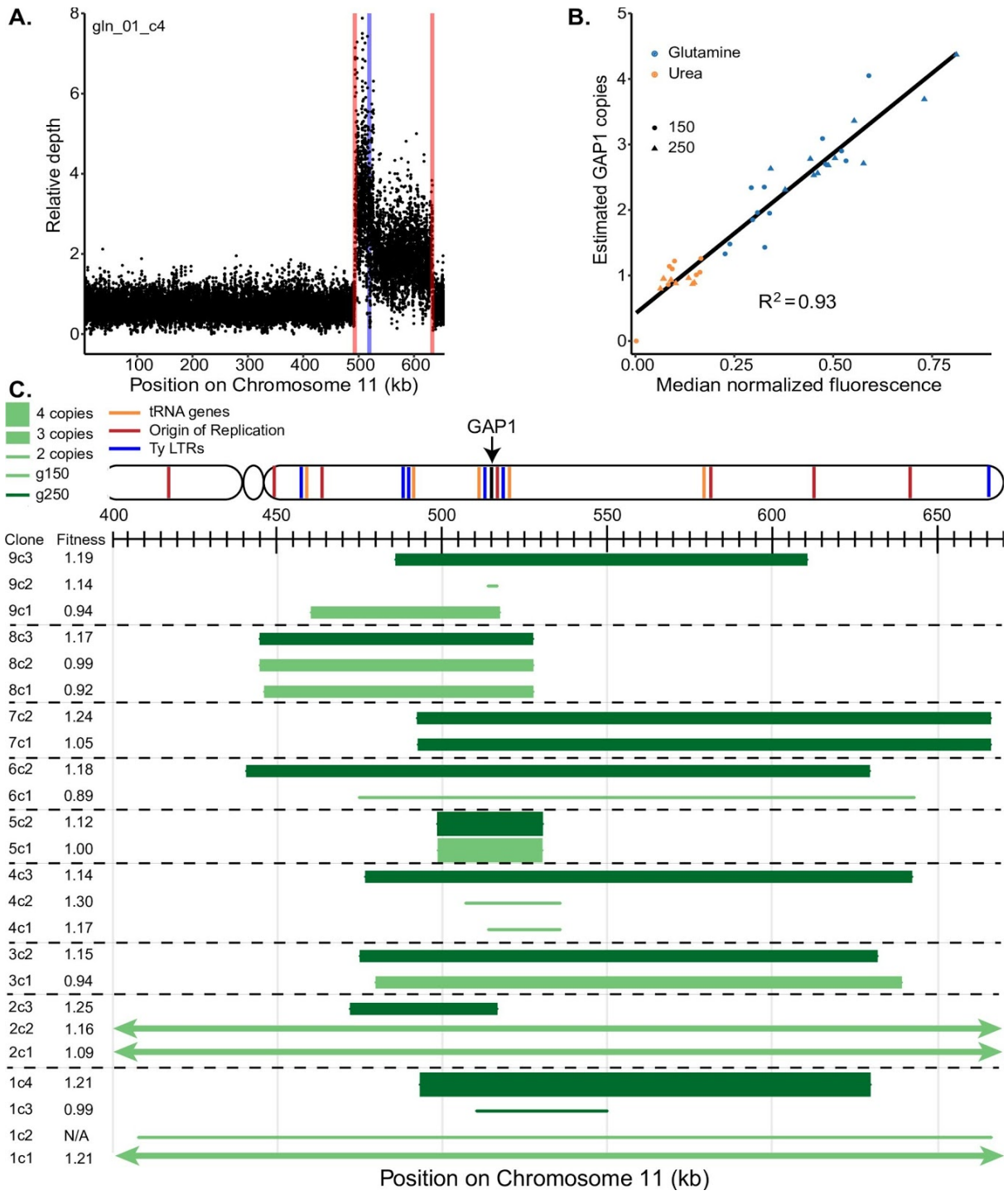


**Figure 2.3. Diversity and fitness effects of** GAP1 **CNVs.** (A) Representative sequence read depth plot from a glutamine-limited clone (gln_01_c4). The nucleotide coordinates of GAP1 in our CNV reporter strain are chromosome XI: 518438-520246 (blue line). Estimated breakpoint boundaries are shown in red. Read depth was normalized to the average read depth on chromosome XI. Reads at each nucleotide position were randomly downsampled for presentation purposes. (B) Read depth based estimates of

GAP1 copy number are positively correlated with median fluorescence of glutamine-limited clones, indicating that fluorescence is informative about the copy number of de novo CNVs. (C) Schematic representation of CNVs identified in clones isolated from glutamine-limited populations. The relative fitness of each clone is also indicated. Copy number and CNV boundaries were estimated using read depth. This schematic is simplified for presentation purposes: the reported copy number refers specifically to the GAP1 coding sequence and does not necessarily reflect copy number throughout the entire CNV, which may vary.

We used pulsed-field gel electrophoresis and Southern blotting to confirm CNV structures (**Figure 2.S5**). Using GAP1 and CEN11 probes for Southern blotting, we identified size shifts in some samples consistent with the large CNVs ( >140 kilobases) we identified in several clones. Interestingly, in some cases, we identified two discrete bands in our GAP1 Southern blot, indicating that the additional copies of GAP1 were not contained on chromosome XI. The GAP1 Southern also provided further evidence for the GAP1 deletion in a clone isolated from urea-limitation. Importantly, while control populations evolving in glutamine-limited chemostats did not show evidence for GAP1 CNVs on the basis of fluorescence, sequence and Southern blotting analysis identified GAP1 amplifications in lineages isolated from these populations (**Figure 2.S5**). As one and two copy control strains do not have the GAP1 CNV reporter, this suggests that GAP1 CNV formation and selection is not affected by the reporter. Moreover, we find no evidence that the molecular features of GAP1 CNVs are affected by the presence of the CNV reporter.

We determined the fitness of GAP1 CNV-containing clones using pair-wise competitive fitness assays in glutamine-limited chemostats (**Figure 2.S6, Figure 2.S7** and **Figure 2.3C)**. Four independent competition assays with the ancestral strain containing the GAP1 CNV reporter showed no significant differences in fitness compared to the isogenic non-fluorescent reference strain. The majority of evolved clones (18/28) have higher relative fitness than the ancestor, indicating that GAP1 CNVs typically confer large fitness benefits. Several clones have

neutral (8/28) or lower (2/28) relative fitness, which indicates that either 1) the fitness effect of GAP1 CNVs may be context-specific or 2) not all GAP1 CNVs confer a fitness benefit.

### 2.3.4: DUR3 CNVs are repeatedly selected during urea limitation

We analyzed the genome sequences of 21 clones that were randomly isolated from urea-limited populations at generation 150 and generation 250 and identified multiple CNVs at the DUR3 locus (**Figure 2.S8A**). DUR3 encodes a high affinity urea transporter and we have previously reported DUR3 amplifications during experimental evolution in a urea-limited chemostat (Hong and Gresham 2014). We compared properties of GAP1 and DUR3 amplifications and found that the average copy number for clones with GAP1 CNVs is 3 (**Figure 2.S8B**) whereas clones with DUR3 CNVs contain significantly more copies with an average copy number of 5 (**Figure 2.S8C,** t-test, p-value < 0.01). Copy number within clones does not significantly increase between 150 and 250 generations at either locus. DUR3 CNV alleles (average of 26 kilobases) are also significantly smaller than GAP1 CNVs (average of 105 kilobases) (**Figure 2.S8D-E,** t-test, p-value < 0.01). Thus, comparison of GAP1 and DUR3 CNVs suggests differences in the properties of selected CNVs as a function of locus and the selective condition.

### 2.3.5: CNV breakpoints are characterized by short, interrupted inverted repeats

To resolve CNV breakpoint sequences, we generated a pipeline integrating CNV calls from multiple existing CNV detection methods (CNVnator, Pindel, Lumpy, and SvABA (K. Ye et al. 2009; Layer et al. 2014; Wala et al. 2017; Abyzov et al. 2011)) and optimized their

performance on synthetic yeast genome data. Although these algorithms perform well using simulated data, we found that they had a high false positive and false negative rate when applied to real data and in general were not informative about the novel sequence formed at CNV boundaries. Therefore, we developed a breakpoint detection pipeline that integrates information from read depth, discordant reads and split reads. To define the breakpoint sequence, we performed de novo assembly using split reads and aligned the resulting contig against the reference genome (**methods**). In addition to GAP1 and DUR3 CNVs, we identified 3 structural variants in our clonal sequencing data using this method. A read-depth based approach was also used to characterize CNVs genome-wide and calculate rDNA and CUP1 copy number, which exhibit variation among lineages.

We analyzed 29 lineages containing GAP1 CNVs and inferred the underlying mechanisms for 19 (66%) of them on the basis of copy number and breakpoint sequences (**methods**). Of the 19 GAP1 CNVs that can be reliably resolved, 3 are the result of aneuploidies and 2 are the result of non-reciprocal interchromosomal translocations. Translocations were confirmed using pulsed-field gel electrophoresis and Southern blot analysis (**Figure 2.S5**), which clearly shows that the second copy of GAP1 is located on a different chromosome. Southern blotting also indicates that an additional 3 GAP1 CNVs are the result of partial (i.e. segmental) aneuploidies, which include the chromosome XI centromere (CEN11) but are smaller than the ancestral chromosome XI (**Figure 2.S5**). At least 4 GAP1 CNVs appear to be the result of a tandem duplication mediated by non-allelic homologous recombination (NAHR). For two of these CNVs, novel junction sequences were obtained that included a hybrid sequence composed of half of each flanking long terminal repeat (YKRCdelta11/YKRCdelta12),

similar to our previous report (Gresham et al. 2010). This mechanism is also likely to underlie the GAP1 deletion that we identified in one urea-limited population.

For 12 out of 29 (41%) GAP1 CNVs and 8 out of 9 (89%) DUR3 CNVs, we identified a pair of short, interrupted, inverted repeats proximate to at least one breakpoint (**Figure 2.4**). We were able to resolve breakpoints at both ends of the CNV for 12 of the 20 CNVs. Analysis of these breakpoints indicates that inverted repeat sequences range in length from 4-24 base pairs (**Figure 2.4D**) and are typically separated by 40 base pairs (**Figure 2.4E**). Microhomology at breakpoint junctions is characteristic of replication-based CNV formation, including microhomology-mediated break-induced replication (MMBIR) and origin-dependent inverted-repeat amplification (ODIRA). ODIRA has several other requirements including the presence of at least one replication origin within the CNV, an internal inversion, and an odd copy number. The identification of inverted sequence relative to the reference at all identified breakpoint junctions is consistent with an inverted structure. We find that 6/29 GAP1 CNVs and 8/9 DUR3 CNVs meet these criteria and thus are likely the result of ODIRA. In cases where the CNV lacks an odd copy number (see **methods**) we cannot reliably infer the mechanism. Interestingly, in one case (ure_07_c1) the CNV meets all the requirements of ODIRA, but does not contain a DNA replication origin (see **discussion**).

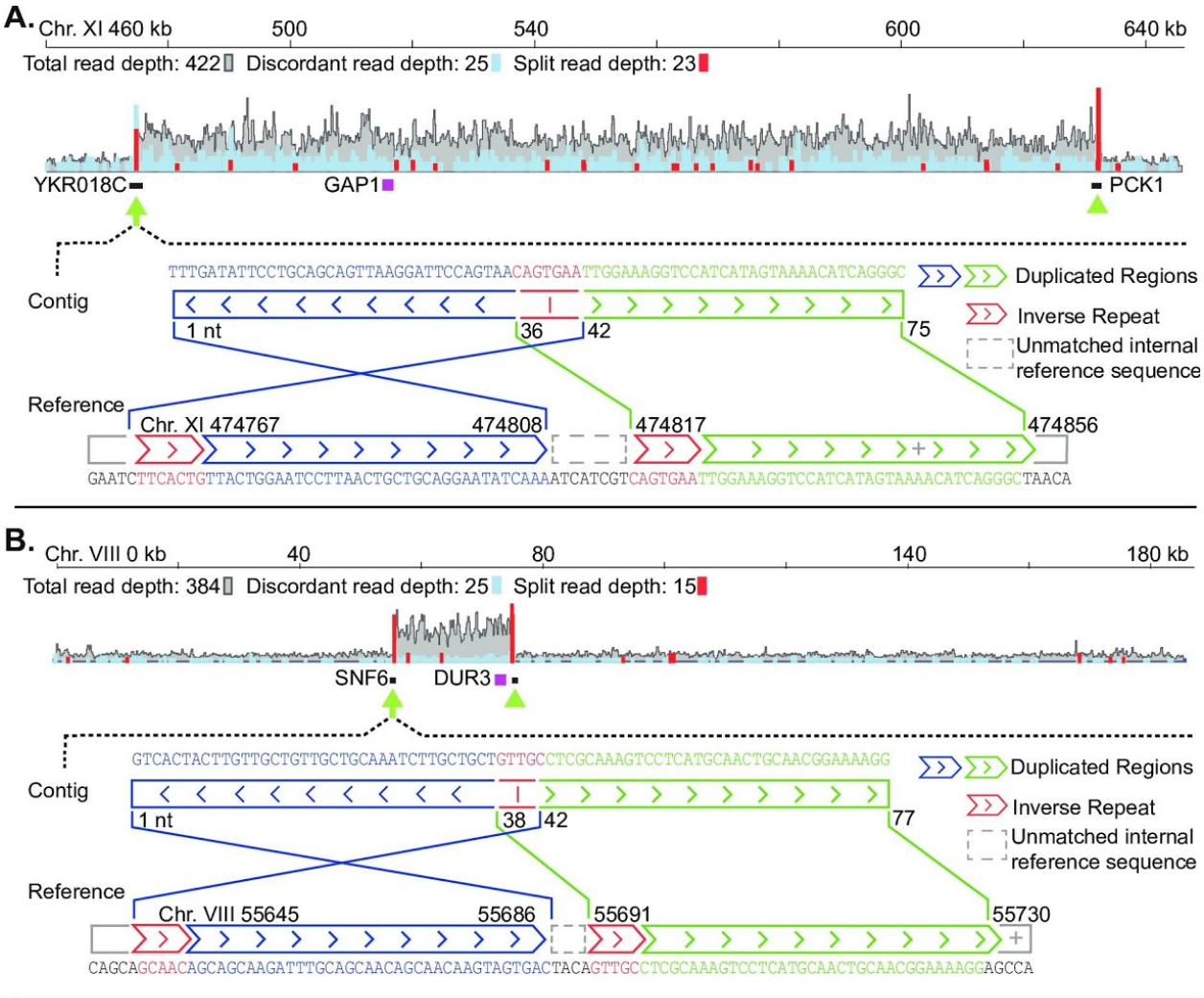
A.
Chr. XI 460 kb
500
540
600
640 kb
Total read depth: 422
Discordant read depth: 25
Split read depth: 23
YKR018C
GAP1
PCK1
TTTGATATTCCTGCAGCAGTTAAGGATTCCAGTAACAGTGAATTGGAAAGGTCCATCATAGTAAAACATCAGGGC
Contig
1 nt
36
42
75
Duplicated Regions
Inverse Repeat
Unmatched internal reference sequence
Reference
Chr. XI 474767
474808
474817
474856
GAATCTTCACTGTTACTGGAATCCTTAACTGCTGCAGGAATATCAAAATCATCGTCAGTGAATTGGAAAGGTCCATCATAGTAAAACATCAGGGCTAACA
B.
Chr. VIII 0 kb
40
80
140
180 kb
Total read depth: 384
Discordant read depth: 25
Split read depth: 15
SNF6
DUR3
GTCACTACTTGTTGCTGTTGCTGCAAATCTTGCTGCTGTTGCCTCGCAAAGTCCTCATGCAACTGCAACGGAAAAGG
Contig
1 nt
38
42
77
Duplicated Regions
Inverse Repeat
Unmatched internal reference sequence
Reference
Chr. VIII 55645
55686
55691
55730
CAGCAGCAACAGCAGCAAGATTTGCAGCAACAGCAACAAGTAGTGACTACAGTTGCCTCGCAAAGTCCTCATGCAACTGCAACGGAAAAGGAGCCA
C.
Replication Fork
Leading strand
Lagging strand
Fork progression
Annealing at Inverted Repeat
Fork regression
Original sequence
CNV sequence
Breakpoint sequence
D.
Inverted repeat length
Median 8
Number of Breakpoints
0
1
2
3
4
5
4 5 6 7 8 9 10 11 12 13 14 15 24
Length (nt)
E.
Length between inverted repeats
Median 40
Number of Breakpoints
0
1
2
3
4
5
6
7
8
<10 10 20 30 40 50 60 70 80 90 100 110 180
Length (nt)

**Figure 2.4. Inverted repeats mediate CNV formation.** Nucleotide resolution of CNV breakpoints for (**A**) GAP1 and (**B**) DUR3 CNVs were identified using a combination of discordant and split reads. To characterize novel sequence, we identified all supporting split reads, performed de novo assembly and aligned the resulting sequence against the reference genome. Sequences in the reference genome (blue) are inversely oriented in the assembled contig, suggesting an inverted structure within CNVs. (**C**) Schematic representation of replication-based CNV formation. After fork stalling, fork regression results in the newly replicated inverted repeat sequence annealing to the complementary sequence and ligating to the lagging strand. (D-E) Distribution of sequence features across 28 breakpoints at the GAP1 and DUR3 loci that contain inverted repeats.

### 2.3.6: Whole genome population sequencing provides insight into population heterogeneity

To comprehensively characterize genomic variation in populations, we performed whole population, whole genome sequencing of glutamine-, urea-, and glucose-limited populations at generations 150 and 250. Analysis of relative sequence read depth is consistent with high frequency GAP1 CNVs in glutamine-limited populations. Population sequencing also confirmed the fixation of a GAP1 deletion (ure_05) in a urea-limited population. Relative sequence read depth at the GAP1 locus correlates well with the normalized fluorescence of the GAP1 CNV reporter in populations (**Figure 2.S9**) providing additional evidence for the utility of the CNV reporter. In glutamine-limited chemostats, GAP1 copy number estimated within populations (which is a function of copy number within clones and allele frequencies) ranges from 2-4 copies, with a trend towards increased copy number over time (**Figure 2.S9**).

We performed single nucleotide variant (SNV) analysis using genome sequencing data from populations and clones at generations 150 and 250. More non-synonymous SNVs were identified in glucose-limited populations than the glutamine- and urea-limited populations (**Table 2.2**), which contained GAP1 and DUR3 amplifications at high frequencies at 150 and 250 generations. In contrast to previous studies (Brown, Todd, and Rosenzweig 1998; Kao and

Sherlock 2008), we did not identify CNVs at the HXT6/7 locus in glucose-limited populations. Increased nucleotide variation within these populations may reflect alternative adaptive strategies in glucose-limited populations.

**Table 2.2. Summary of single nucleotide variation in three different selection conditions.** Populations were sequenced at 150 and 250 generations. For variants that were identified at both time points, we determined whether they increased (↑) or decreased (↓) in frequency between generation 150 and 250.

| | Glucose (n=10) | | | | Urea (n=11) | | | | Glutamine (n=11) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Total | | Trend | | Total | | Trend | | Total | | Trend | |
| **Predicted effect** | **g150** | **g250** | ↑ | ↓ | **g150** | **g250** | ↑ | ↓ | **g150** | **g250** | ↑ | ↓ |
| Non-coding | 4 | 9 | 2 | 1 | 8 | 12 | 1 | 0 | 6 | 5 | 0 | 0 |
| Missense | 47 | 61 | 17 | 4 | 22 | 34 | 6 | 2 | 12 | 22 | 4 | 1 |
| Frameshift | 4 | 5 | 1 | 1 | 5 | 6 | 2 | 0 | 2 | 2 | 0 | 0 |
| Synonymous | 1 | 7 | 0 | 0 | 4 | 10 | 2 | 0 | 4 | 3 | 0 | 0 |
| Stop gained | 4 | 10 | 2 | 0 | 7 | 4 | 0 | 2 | 0 | 5 | 0 | 0 |
| Start lost | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Splice variant | 2 | 3 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Mito. genome | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| Inframe insertion | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Total variants | 62 | 96 | 22 | 7 | 46 | 68 | 11 | 4 | 24 | 39 | 4 | 1 |

We find several genes with multiple independent, non-synonymous variation in glutamine-limited populations (**Table 2.3**) including MCK1, a protein kinase with potential roles in NHEJ; SOG2, a member of the RAM signaling pathway and regulator of bud separation after mitosis; and TAO3, another member of the RAM network. We previously reported mutations in MCK1 from selection in glutamine- and arginine-limited chemostats (Hong and Gresham 2014), suggesting that it is a recurrent target of selection in these conditions. Changes in cell morphology are potentially adaptive in nutrient-poor conditions, which may result from defects in cell cycle progression and bud separation associated with mutations in the RAM pathway

(Cullen, Sprague, and Jr. 2012). However, the effect of these mutations on bud separation is likely to be minor as we did not observe increases in forward scatter (which varies with cell size) in flow cytometry data, except in one glucose-limited population (**Figure 2.S3**).

In the nine urea-limited populations, we identified 14 independent non-synonymous variants in DUR1,2 (**Table 2.3**). DUR1,2, encodes urea amidolyase, which metabolizes urea to ammonium. At two different nucleotide positions, we find that the same nucleotide was mutated multiple times independently. In a third location, we identified a SNV at the exact nucleotide position as we previously reported (Hong and Gresham 2014). Thus, a subset of variants in DUR1,2 appear to be uniquely beneficial and recurrently selected in urea-limited environments.

**Table 2.3. Genes with multiple, independent, non-synonymous acquired mutations.** Variants found at greater than 5% frequency within each population.

| **Glucose-limitation** | | **Urea-limitation** | | **Glutamine-limitation** | |
|---|---|---|---|---|---|
| Gene Name | Total Variants | Gene Name | Total Variants | Gene Name | Total Variants |
| TRK1 | 11 | DUR1,2 | 14 | MCK1 | 3 |
| SVF1 | 2 | | | SOG2 | 3 |
| CDC48 | 3 | | | TAO3 | 2 |
| WHI2 | 3 | | | GPB2 | 2 |

In glucose-limited populations, we identified multiple, independent mutations in four genes (**Table 2.3**): TRK1, a component of the potassium transport system; SVF1, which is important for the diauxic growth shift and is implicated in cell survival during aneuploidy (Torres et al. 2010); CDC48, an AAA ATPase; and WHI2, which is a mediator of the cellular stress response. Previous studies have identified loss-of-function mutations in WHI2 suggesting it is a general target of selection across different conditions (Hong and Gresham 2014; Payen et al. 2016; Gresham et al. 2008).

Analysis of clonal samples was largely consistent with population sequencing. Interestingly, we identified two cases in which SNVs occurred within GAP1 CNVs. These SNVs are present at frequencies of 53% in a lineage containing a GAP1 duplication and 30% in a lineage containing a GAP1 triplication indicating that they are present on only one of the copies within the CNV. We also identified polymorphisms within DUR3 amplifications. This suggests that individual copies of a gene within a CNV can accumulate additional nucleotide variation even in relatively short-term evolutionary scenarios. Interestingly, 8 of the 9 clones with DUR3 amplifications also acquired a variant in DUR1,2, which may be indicative of a synergistic relationship between CNVs and single nucleotide variation.

### 2.3.7: Lineage tracking reveals extensive clonal interference among CNV lineages

The reproducible dynamics of CNV lineages observed during glutamine-limited experimental evolution may be due to two non-exclusive reasons: either 1) a high supply rate of de novo CNVs or 2) pre-existing CNVs in the ancestral population (**Figure 2.S10**). In both scenarios, a single CNV or multiple, competing CNVs may underlie the reproducible dynamics. Sequence analysis of clonal lineages suggests at least two, and as many as four, CNV lineages may co-exist in populations (**Figure 2.3**); however, genome sequencing is uninformative about the total number of lineages for two key reasons. First, the recurrent formation of CNVs confounds distinguishing CNVs that are identical by state from those that are identical by descent. Second, CNVs that arise de novo may subsequently diversify over time resulting in distinct alleles that are derived from a common event.

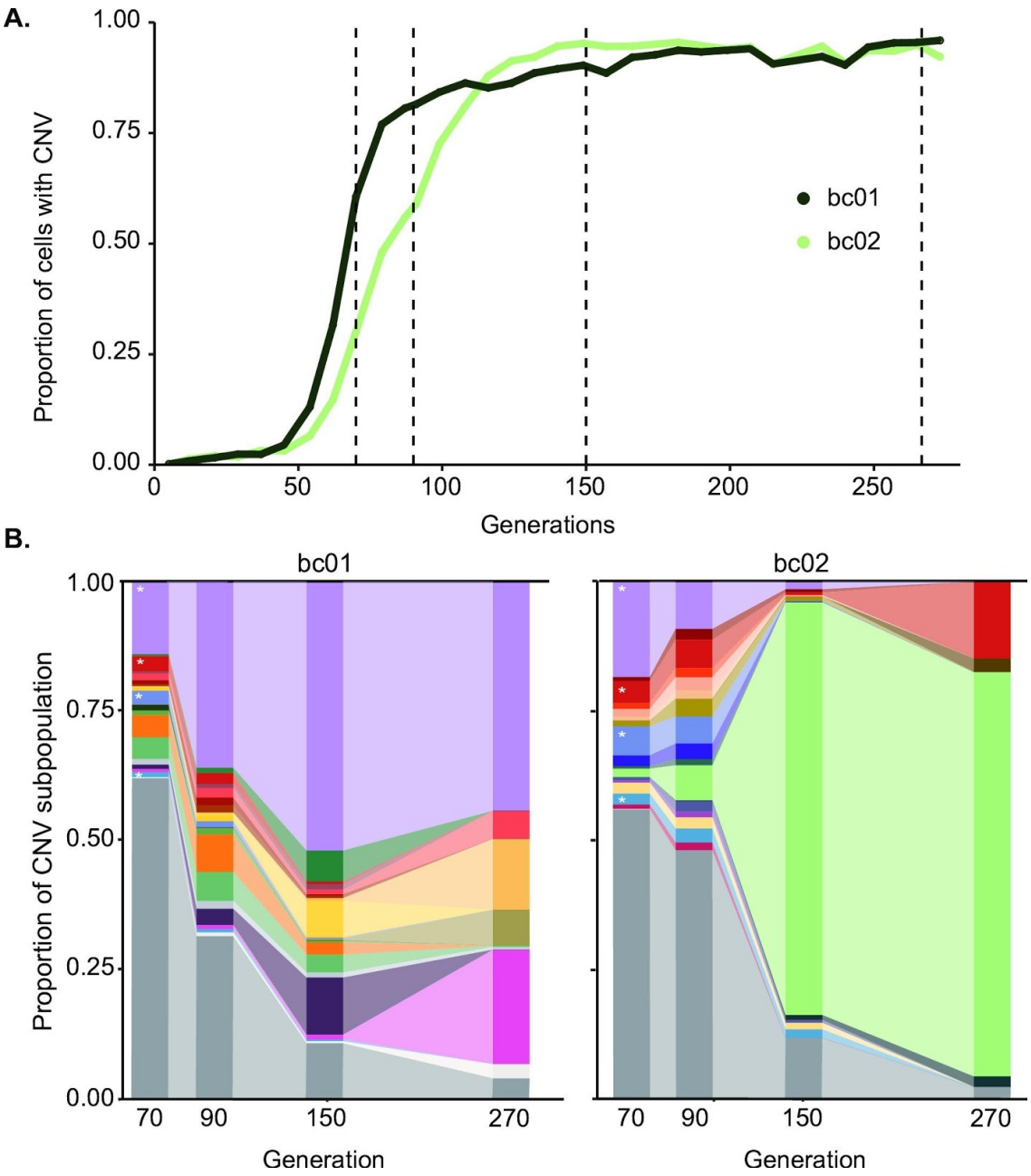


**Figure 2.5. Lineage tracking reveals extensive clonal interference among CNV-containing lineages**. (**A**) We used fluorescence-activated cell sorting (FACS) to fractionate cells containing GAP1 CNVs from two populations at four time points (dashed black lines) and performed barcode sequencing. (**B**) Using a sample- and time point-specific false positive correction, we identified 7067, 973, 131, and 76 barcodes in one population, bc01 (left), and 5305, 5351, 583, and 28 barcodes in another population, bc02 (right), at generations 70, 90, 150 and 270 respectively. Each barcode found at >1% frequency in at least one time point is represented by a unique color in the plot, for a total of 21 barcodes in bc01 and 18 barcodes in bc02. All other lineages that are never detected at >1% frequency are shown in grey. Lineages denoted by a * are found at >1% frequency in both populations.

To quantify the number, relationship and dynamics of individual CNV lineages, we constructed a lineage tracking library using random DNA barcodes (Levy et al. 2015). We constructed a library of ~80,000 unique barcodes (**Figure 2.S11**) in the background of the GAP1 CNV reporter and performed six independent replicate experiments in glutamine-limited chemostats. Real time monitoring of CNV dynamics using the GAP1 CNV reporter recapitulated the dynamics of our original experiment (**Figure 2.5A**, **Figure 2.S12A)** although CNV lineages appeared significantly earlier in these populations ($T_{up}$; t-test p-value < 0.01). As the lineage tracking strain was independently derived from the strain used in our original experiment, these results indicate that selection of GAP1 CNVs in glutamine-limited chemostats is reproducible and independent of genetic background.

**Table 2.4. Estimation of CNV lineages in evolving populations across time.** We determined the number of GAP1 CNV containing lineages by correcting the number of identified barcodes by the estimated false positive rate associated with CNV isolation using FACS. High confidence GAP1 CNV lineages are defined as those that are found at two or more consecutive timepoints.

| Population | Generation | Number of detected barcodes | False positive rate (FP) | FP corrected barcode count | Barcodes identified at >1 time point |
|---|---|---|---|---|---|
| bc01 | 70 | 9650 | 0.27 | 7067 | 891 |
| bc01 | 90 | 1064 | 0.09 | 973 | 891 |
| bc01 | 150 | 136 | 0.04 | 131 | 131 |
| bc01 | 270 | 79 | 0.04 | 76 | 38 |
| bc02 | 70 | 7243 | 0.27 | 5305 | 2676 |
| bc02 | 90 | 5851 | 0.09 | 5351 | 2710 |
| bc02 | 150 | 606 | 0.04 | 583 | 162 |
| bc02 | 270 | 29 | 0.04 | 28 | 22 |

To quantify individual lineages, we isolated the subpopulation containing CNVs from two populations (bc01 and bc02) at multiple timepoints (generations 70, 90, 150, and 270). Isolation

of the CNV subpopulation was performed by fluorescence activated cell sorting (FACS) using gates based on one and two copy control populations (**Figure 2.5A**). We sequenced barcodes from the CNV subpopulation at each time point and determined the number of unique lineages ((L. Zhao et al. 2017) and **methods**). To account for variation in the purity of the FACS-isolated CNV subpopulation, we analyzed individual clones using a flow cytometer. Using these data, we estimated a false positive rate, which we find varies between time points (**Figure 2.S12B** and **methods**), and applied this correction to barcode counts (**Table 2.4**).

Strikingly, we detect thousands of independent GAP1 CNV lineages at generation 70 indicating that a large number of independent GAP1 CNVs are generated and selected in the early stages of the evolution experiments **(Figure 2.5B)**. Applying a conservative false positive correction, we identified 7,067 GAP1 CNV lineages in bc01 and 5,305 GAP1 CNV lineages in bc02 at generation 70 **(Table 2.4)**. If we only consider lineages detected in the CNV subpopulation at multiple time points, we identify 891 CNV lineages in bc01 and 2,676 CNV lineages at generation 70 **(Table 2.4)**. Thus, $10^2$-$10^4$ independent CNV lineages initially compete within each population that are on the order of ~$10^8$ cells. The overall diversity of CNV lineages decreases with time, consistent with decreases in lineage diversity observed in other evolution experiments (Levy et al. 2015; Blundell et al. 2017). By generation 270, we detect only 76 CNV lineages in bc01 and 28 CNV lineages in bc02. To determine the dominant lineages in each population, we identified barcodes that reached greater than 1% frequency in the CNV subpopulation in at least one time point: 21 independent lineages are found at greater than 1% frequency in bc01 and 18 independent lineages are found at greater than 1% frequency in bc02 (**Figure 2.5B**). These results indicate the presence and persistence of multiple GAP1 CNVs

across hundreds of generation of selection during which there is a continuous reduction in the overall diversity of CNV lineages.

Although CNVs rise to high frequencies in both populations (**Figure 2.5A**), the composition of competing CNV lineages is dramatically different: in bc02, a single lineage dominates the population by generation 150 (**Figure 2.5B**), whereas in bc01, there is much greater diversity at later time points. In both populations, several CNV lineages that comprise a large fraction of the CNV subpopulation at early generations (generations 70, 90, or 150) are extinct by generation 270. Thus, within populations, individual CNV lineages do not increase in frequency with uniform dynamics despite the consistent and reproducible dynamics of the entire CNV subpopulations (**Figure 2.5A** and **Figure 2.2**). Differences in fitness between individual CNV lineages, possibly as a result of variation in copy number, CNV size and secondary adaptive mutations, are likely to contribute to these dynamics.

### 2.3.8: CNV subpopulations comprise de novo and pre-existing CNV alleles

To distinguish the contribution of pre-existing genetic variation (i.e. CNVs introduced to the population before chemostat inoculation; **Figure 2.S10**) and de novo variation (i.e. CNVs introduced to the population following chemostat inoculation) to CNV lineage dynamics, we assessed whether barcodes were shared between CNV lineages in independent populations. We identified four barcodes at greater than 1% frequency that are common to both populations (**Figure 2.5B**). At generation 70, one of these barcodes (indicated in light purple) was present at 14% and 19% in bc01 and bc02, respectively. We find that the barcode for this lineage was over-represented in the ancestral unselected population (an initial frequency of 0.014%, which is one order of magnitude greater than the average starting frequency of 0.0011%; **Figure 2.S11**).

Although there is a possibility that de novo CNVs formed independently in this barcode lineage, it is more likely that this lineage contained a pre-existing CNV in the ancestral population. While this lineage represented a sizable fraction of the CNV subpopulation in both replicate populations, it was only maintained at high frequency in one of them (bc01). Only one of the four pre-existing CNV lineages persists throughout the experiment in both populations. By contrast, in each population, we identified 17 and 14 unique high frequency CNV lineages that are most likely new CNVs. These results indicate that both pre-existing CNVs and de novo CNVs that arise during glutamine limitation contribute to adaptive evolution.

## 2.4: Discussion

Copy number variants are an important class of genetic variation and adaptive potential. In this study, we sought to understand the short-term fate of CNVs as they are generated and selected in evolving populations. Previous work from our laboratory and others has shown that the defined, strong selective conditions of a chemostat provides an ideal system for studying CNVs. We used nitrogen limitation to establish conditions that select for amplification and deletion of the gene GAP1, which encodes the general amino acid permease, in S. cerevisiae.

### 2.4.1: A GAP1 CNV reporter reveals the dynamics of selection

To determine the dynamics with which CNVs are selected at the GAP1 locus, we inserted a constitutively expressed fluorescent gene adjacent to GAP1 and tracked changes in single cell fluorescence over time. While one and two copy control strains with mCitrine at neutral loci maintain a steady fluorescent signal over 250 generations of selection, all glutamine-limited populations with the GAP1 CNV reporter show increased fluorescence by

generation 75. Importantly, the structure and breakpoints of CNVs within and between populations are different, indicating independent formation of CNVs. Control strains were inoculated independently, and have different genetic backgrounds, but also form CNVs at the GAP1 locus as determined by whole genome sequencing and Southern blot analysis. These data indicate that GAP1 CNVs are positively selected early and repeatedly in glutamine-limited environments.

While the majority of evolved clones with GAP1 CNVs (18/28) have higher relative fitness in glutamine-limited chemostats compared to the ancestor, several clones have neutral (8/28) or lower (2/28) relative fitness. CNV-containing clones were selected on the basis of increased fluorescence, which does not necessarily mean the clone had higher fitness than the ancestor. The fitness effect of a CNV within the chemostat environment is context-specific, and may depend on factors such as frequency-dependent selection. In addition, if GAP1 CNVs are generated at a high rate as we have hypothesized, neutral or deleterious CNVs could be present for several generations before these lineages are purged from the population or acquire additional adaptive mutations. The data presented here explicitly demonstrate that there is a selective advantage provided by GAP1 CNVs: 1) GAP1 CNVs provide a fitness benefit for 64% of the evolved clones tested, 2) GAP1 CNVs are repeatedly selected during independently replicated evolution experiments, and 3) GAP1 CNVs are selected regardless of genetic background.

### 2.4.2: Inferences of CNV formation mechanisms

Whole genome sequencing of GAP1 CNV lineages isolated on the basis of increased fluorescence uncovered a wide range of CNV structures within and between populations. We

found cases in which distinct alleles were identified within populations at different time points and cases in which we identified the same CNV allele 100 generations later. GAP1 CNV alleles are 105 kilobases on average, but can include the entire right arm of chromosome XI (260 kilobases). A previous study in bacteria showed that there is a cost to gene duplication, with a fitness reduction of 0.15% per kilobase (Adler et al. 2014). Therefore, we hypothesized that CNVs would decrease in size over evolutionary time through a refinement process in order to reduce the fitness burden. However, we failed to detect a significant reduction in CNV allele size over time. This may be because increased CNV size does not confer a fitness cost in yeast, the fitness benefit of the GAP1 CNV outweighs this cost, or because there are other genes within the CNV whose amplification confers a fitness benefit.

Our reporter detects increases in gene copy number that result from a variety of processes including aneuploidy, non-reciprocal translocation, tandem duplication, and complex copy number variants including inverted triplications. The ability to track and isolate these diverse gene amplifications allows us to enumerate the frequency of each type and characterize the mechanisms underlying their formation. Combining our approach with molecular techniques allowed us to further understand the nature of these GAP1 CNVs. Three particularly interesting GAP1 CNV-containing clones appear to have partial (i.e. segmental) aneuploidies that encompass centromere XI (**Figure 2.S5**). As the presence of two centromeres in one chromosome is extremely unlikely, it is plausible that these exist as independent, supernumerary chromosomes (Natesuntorn et al. 2015). Similar adaptive rearrangements occur in other yeast species: isochromosome formation, potentially mediated by the presence of inverted repeats, has been observed during treatment of Candida albicans with antifungal drugs (A. Selmecki, Forche, and Berman 2006). The use of a CNV reporter should facilitate

determination of the frequency with which these and other complex mechanisms give rise to CNVs at a given locus.

Breakpoint analysis provided further insight into the mechanisms underlying CNV formation. We identified breakpoints within LTRs and other repetitive elements for 4 unique glutamine-limited clones that have 2 copies of GAP1. These findings suggest that these CNVs were formed by a tandem duplication mediated through NAHR. Of these, 3 GAP1 gene amplifications (3/28) are formed after NAHR between flanking LTRs YKRCdelta11 and YKRCdelta12. The GAP1 deletion, which occurred in one population undergoing urea limitation, also had breakpoints in these flanking elements consistent with NAHR-mediated gene deletion. NAHR may drive the non-reciprocal translocations we identified, and additional unresolved events with breakpoints adjacent to LTRs. Surprisingly, we did not find evidence for the selection of $GAP1^{circle}$ CNVs in any population. Thus, it may be that circular elements containing beneficial genes only exist transiently in cells and may rapidly resolve to chromosomal amplifications via homologous recombination-mediated reintegration.

We identified 9 GAP1 CNVs and 8 DUR3 CNVs containing breakpoints that comprise closely-spaced inverted repeat sequences. Of these, the majority (14/17) also had an odd copy number, and contained an origin of replication consistent with the ODIRA mechanism (Brewer et al. 2011, 2015). However, we also identified one DUR3 CNV that does not include a replication origin (ure_07_c1), although the origin is nearby (<1 kilobase). This could result from a distinct replication-based mechanism of CNV generation. For example, microhomology-mediated break-induced replication (MMBIR) is a RAD51-independent process that relies on short stretches of homology (“microhomology”) to restart a stalled replication fork (Hastings, Ira, et al. 2009). While we cannot explicitly distinguish between these models, the short stretches of

homology in the inverted repeats is inconsistent with formation of this CNV by NAHR. Thus, while NAHR plays an important role in CNV formation, our results suggest that replication-based mechanisms may be a major source of gene amplification in yeast. This is consistent with increasing evidence for replication-based CNV formation in diverse organisms including yeast, mice, and humans (Feng Zhang et al. 2009; Ottaviani, LeCain, and Sheer 2014; Arlt et al. 2012; Sakofsky et al. 2015).

Comparison between DUR3 and GAP1 CNVs identified quantitative differences in CNV formation at the two loci. We primarily identified CNVs with 2 or 3 copies of GAP1 in glutamine-limited clones, but urea-limited clones always contained 5 copies of DUR3. The size (average of 26 kilobases) of DUR3 CNVs was also significantly smaller than GAP1 CNVs. Molecular characterization revealed a diverse range of processes underlying GAP1 CNV formation, whereas DUR3 CNVs are all characterized by inversions mediated by short, interrupted, inverted repeats. These data suggest that generation and selection of CNVs varies as a function of locus and selective condition. The CNV reporter can readily be integrated throughout the genome to further test whether there are fundamental differences in CNV formation mechanisms at different loci and how these differences change the temporal dynamics of CNV selection.

### 2.4.3: Clonal interference underlies CNV dynamics

By combining a CNV reporter with lineage tracking, we identified a surprisingly large number of independent CNV lineages. Whereas clonal isolation and sequencing suggested at least four independent lineages within populations, lineage tracking indicates that hundreds to thousands of individual CNV lineages emerge within less than 100 generations. Most of these

lineages do not achieve high frequency, as we identified only 18-21 lineages present at >1% frequency in the CNV subpopulation. The number of independent CNV lineages we identified is remarkable. Although we have attempted to account for technical factors that may inflate this number, unanticipated aspects of barcode transformation and library construction, cell sorting, and barcode sequencing and identification may impact this estimation. Conversely, the exact number of CNV lineages may be underestimated, as the unselected barcode library was not maximally diverse and each unique barcode was shared by multiple founding cells.

While we found lineages that were common to both populations (at least one of which is likely to contain a pre-existing CNV), ancestral CNV lineages do not drive the evolutionary dynamics. Pre-existing CNV lineages have different dynamics in each population, and do not prevent the emergence of unique de novo CNV lineages. This demonstrates that the ultimate fate of a CNV lineage depends on multiple factors, and a high frequency at an early generation does not guarantee that a lineage will persist in the population. Thus, CNV dynamics result from pre-existing and de novo variation and are characterized by extensive clonal interference and replacement among competing CNV lineages.

The large number of CNV lineages identified in our study indicates that they occur at a high rate. Recent studies have suggested that adaptive mutations may be stimulated by the environment. Stress can lead to increases in genome-wide mutation rates in both bacteria and yeast (Foster 2007; Galhardo, Hastings, and Rosenberg 2007; Shor, Fox, and Broach 2013) and replicative stress can lead directly to increased formation of CNVs (Chen et al. 2015; Wilson et al. 2015). Other groups have proposed an interplay between transcription and CNV generation, and that active transcription units might even be “hotspots” of CNV formation (Thomas and Rothstein 1989; Skourti-Stathaki and Proudfoot 2014; Aguilera and Gaillard

2014). These hotspots, often designated as common fragile sites, may occur in long, late replicating genes, with large inter-origin distances (Wilson et al. 2015). Local transcription at the rDNA locus leads to rDNA amplification, and is thought to be regulated in response to the environment (Jack et al. 2015; Mansisidor et al. 2018). Transcription of the CUP1 locus in response to environmental copper leads to promoter activity that further destabilizes stalled replication forks and generates CNVs (Hull et al. 2017). Given the high level of GAP1 transcription in nitrogen limited chemostats (Airoldi et al. 2016) it is tempting to speculate that this condition may promote the formation of GAP1 CNVs. Further studies are required to understand the full extent of processes that underlie CNV formation at the GAP1 locus and how these different mechanisms may contribute to the fitness and overall success of CNV lineages.

The frequency of GAP1 CNVs can be attributed to a combination of factors including: a high mutation supply rate due in part to the large chemostat population size (~$10^8$), the strength of selection, and the fitness benefit typically conferred by GAP1 amplification. Together, these factors contribute to an early, deterministic phase, during which CNVs are formed at a high rate and thousands of lineages with CNVs rapidly increase in frequency. During a second phase, the dynamics are more variable as competition from different types of adaptive lineages, and additional acquired variation, influence evolutionary trajectories of individual CNV lineages. This phenomenon has recently been observed in other evolution experiments, where early events are driven by multiple competing single-mutant lineages (Blundell et al. 2017), but later dynamics are influenced by stochastic factors and secondary mutations (Levy et al. 2015).

The high degree of clonal interference observed among a single class of adaptive mutations may have important implications for adaptive evolution. CNVs are alleles of large effect that can simultaneously change the dosage of multiple protein-coding genes and

subsequently lead to changes in cell physiology. Epistatic relationships between CNVs and other adaptive mutations could therefore dramatically alter the fitness landscape (Kvitek and Sherlock 2011). Additionally, CNVs can confer a fitness benefit per se but also serve to increase the amount of DNA in the genome that can accumulate mutations. Therefore, CNVs can potentially increase the rate of adaptive evolution by increasing the target size for adaptive mutations. In this study, we found evidence for polymorphisms within individual CNVs and potential epistasis between SNVs and CNV alleles, two phenomena that require further exploration as we continue to define the role of CNVs in driving rapid adaptive evolution.

## 2.5: Conclusion

The combined use of a fluorescent CNV reporter and barcode lineage tracking provides unprecedented insight into this important class of mutation. Previous studies have tracked specific mutations and their fitness effects (Lang, Botstein, and Desai 2011), but ours is the first single-cell based approach to identify an entire class of mutations and follow evolutionary trajectories with high resolution. While barcode tracking alone provides information about the number of adaptive lineages and their fitness effects, the CNV reporter enables us to specifically determine the number of unique CNV events. In addition, the reporter provides an estimate of the total proportion of CNVs in the population, which we can use to inform our understanding of lineage dynamics. Using these tools, we have shown that CNVs are generated at a high rate through diverse mechanisms including homologous recombination and replication-based errors. These processes lead to the formation of many distinct CNV alleles segregating within populations. One limitation of our approach is that a complex copy number variant could be the product of multiple, independent events (for example, a duplication followed by a subsequent

triplication). Evolution experiments that start with a pre-existing CNV would be informative for studying how CNVs diversify when maintained under selection.

Our results demonstrate an important role for CNVs in driving rapid adaptive evolution in microbial populations, but could be broadly applicable to plants, animals, and humans. Our system provides a facile means for studying the molecular processes underlying CNV generation as well as evolutionary aspects of CNVs including: whether there are fundamental differences in CNV formation and selection at different loci, the impact of a high rate of CNV formation on the evolutionary dynamics of other adaptive lineages, how CNVs are maintained or refined over longer evolutionary timescales, how CNVs interact with other adaptive mutations to influence fitness landscapes, whether there are consequences and tradeoffs in alternative environments, and how the formation of CNVs impacts gene expression and genome architecture. Extension of this method is likely to be useful for addressing additional fundamental questions regarding the evolutionary and pathogenic role of CNVs in diverse systems.

## 2.6: Methods

### 2.6.1: Strains and media

We used FY4 and FY4/5, haploid and diploid derivatives of the reference strain S288c, for all experiments. To generate fluorescent strains, we performed high efficiency yeast transformation (Gietz and Schiestl 2007) with an mCitrine gene under control of the constitutively expressed ACT1 promoter (ACT1pr::mCitrine::ADH1term) and marked by the KanMX G418-resistance cassette (TEFpr::KanMX::TEFterm). The entire construct, which we refer to as the mCitrine

CNV reporter, is 3,375 base pairs. For control strains, the mCitrine reporter was integrated at two neutral loci: HO (YDL227C) on chromosome IV and the dubious ORF, YLR123C on chromosome XII. Diploid control strains containing 3 and 4 copies of the mCitrine CNV reporter were generated using a combination of backcrossing and mating. We constructed the GAP1 CNV reporter by integrating the mCitrine construct at an intergenic region 1,118 base pairs upstream of GAP1 (integration coordinates, chromosome XI: 513945-517320). PCR and Sanger sequencing were used to confirm integration of the GAP1 CNV reporter at each location. Transformants were subsequently backcrossed and sporulated, and the resulting segregants were genotyped.

For the purpose of lineage tracking, we constructed a strain containing a landing pad and the GAP1 CNV reporter by segregation analysis after mating the original GAP1 CNV reporter strain to a landing pad strain (derived from BY4709) (Levy et al. 2015). As the kanMX cassette is present at two loci in this cross, we performed tetrad dissection and identified four spore tetrads that exhibited 2:2 G418 resistance. A segregant with the correct genotype (G418 resistant, ura-) was identified and confirmed using a combination of PCR and fluorescence analysis. We introduced a library of random barcodes by transformation and selection on SC-ura plates (Levy et al. 2015). We plated an average of 500 transformants on 200 petri plates and estimated 78,000 independent transformants.

Nitrogen limiting media (glutamine and urea limitations) contained 800 uM nitrogen regardless of molecular form and 1 g/L $CaCl_2$-$2H_2O$, 1 g/L of NaCl, 5 g/L of $MgSO_4$-$7H_2O$, 10 g/L $KH_2PO_4$, 2% glucose and trace metals and vitamins as previously described (Hong and Gresham 2014).

Glucose limiting media contained 0.08% glucose, 1 g/L $CaCl_2$-$2H_2O$, 1 g/L of NaCl, 5 g/L of $MgSO_4$-$7H_2O$, 10 g/L $KH_2PO_4$, 50g/L $(NH_4)_2SO_4$ and trace metals and vitamins (Brauer et al. 2008).

### 2.6.2: Long-term experimental evolution

We inoculated the GAP1 CNV reporter strain into 20mL ministat vessels (A. W. Miller et al. 2013) containing either glutamine-, urea-, or glucose-limited media. Control populations containing either one or two copies of the CNV reporter at neutral loci (HO and YLR123C) were also inoculated in ministat vessels for each media condition. Ministats were maintained at 30°C in aerobic conditions and diluted at a rate of 0.12 $hr^{-1}$ (corresponding to a population doubling time of 5.8 hours). Steady state populations of 3 x $10^8$ cells were maintained in continuous mode for 270 generations (65 days). Every 30 generations, we archived 2 mL population samples at -80°C in 15% glycerol.

### 2.6.3: Flow cytometry sampling and analysis

To monitor the dynamics of CNVs, we sampled 1mL from each population every ~8 generations. We performed sonication to disrupt any cellular aggregates and immediately analyzed the samples on an Accuri flow cytometer, measuring 100,000 cells per population for mCitrine fluorescence signal (excitation = 516nm, emission = 529nm, filter = 514/20nm), cell size (forward scatter) and cell complexity (side scatter). We generated a modified version of our laboratory flow cytometry pipeline for this analysis (https://github.com/GreshamLab/flow), which uses the R package flowCore (Ellis et al. 2016). We used forward scatter height (FSC-H) and forward scatter area (FSC-A) to filter out doublets, and FSC-A and side scatter area (SSC-A) to

filter debris. We quantified fluorescence for each cell and divided this value by the forward scatter measurement for the cell to account for differences in cell size. To determine population frequencies of cells with zero, one, two, and three plus copies of GAP1, we used one and two copy control strains grown in glutamine-limited chemostats to define gates and perform manual gating. We used a conservative gating approach to reduce the number of false positive CNV calls by first manually drawing a liberal gate for the one copy control strain, followed by a non-overlapping gate for the two copy control strain.

### 2.6.4: Quantification of CNV dynamics

To quantify the dynamics of CNVs in evolving populations, we defined summary statistics as in (Lang, Botstein, and Desai 2011). $T_{up}$ is the generation at which CNVs are initially detected and $S_{up}$ is the slope of the linear fit during initial population expansion of CNVs. We first determined the proportion of cells with a CNV and the proportion of cells without CNVs at each time point, using the manually defined gates. To calculate $T_{up}$, we defined a false positive rate for CNV detection in evolving one copy control strains from generations 1-153 (defined as the average plus one standard deviation = 7.1%). We designate $T_{up}$ once an experimental population surpasses this threshold. To calculate $S_{up}$, we plotted the natural log of the ratio of the proportion of cells with and without a CNV against time and calculated the linear fit during initial population expansion of CNVs. We defined the linear phase on the basis of $R^2$ values. $S_{up}$ can also be defined as the percent increase in CNVs per generation, which is an approximation for the relative average fitness of all CNV alleles in the population.

### 2.6.5: Isolation and analysis of evolved clones

Clonal isolates were obtained from each glutamine- and urea-limited population at generation 150 and generation 250. We isolated clones by plating cells onto rich media (YPD) and randomly selecting individual colonies. We inoculated each clone into 96 well plates containing the limited media used for evolution experiments and analyzed them on an Accuri flow cytometer following 24 hours of growth. We compared fluorescence to unevolved ancestral strains and evolved 1 and 2 copy controls grown under the same conditions, and chose a subset of clones for whole genome sequencing.

To measure the fitness coefficient of evolved clones, we performed pairwise competitive fitness assays in glutamine-limited chemostats using the same, glutamine-limited conditions as our evolution experiments (Hong and Gresham 2014). We co-cultured our fluorescent evolved strains with a non-fluorescent, unevolved reference strain (FY4). We determined the relative abundance of each strain every 2-3 generations for approximately 15 generations using flow cytometry. We performed linear analysis of the natural log of the ratio of the two genotypes against time and estimated the fitness, and associated error, relative to the ancestral strain.

### 2.6.5: Plug preparation, pulsed-field gel electrophoresis, and Southern blotting

Evolved clones were grown overnight in glutamine-limited media and embedded in agarose using Bio-rad plug molds. Plugs were incubated in zymolyase T100 (200µg/mL) overnight at

37°C, proteinase K (4 mg/mL) overnight at 50°C, and PMSF (1 mM) for one hour at 4°C. PMSF was removed by washing plugs with 1 mL of CHEF TE three times for thirty minutes. Plugs were subsequently run in a 1X TAE, 1% agarose gel using a Bio-rad CHEF-DR II. Southern blotting was performed by alkaline transfer using Hybond-XL membranes. Blots were subsequently probed with $^{32}$P-labeled DNA complementary to GAP1 or CEN11. Probes were created using nested PCR. Signal from blots was detected using FujiFilm imaging plates and imaged using Typhoon FLA9000.

### 2.6.6: Genome sequencing

For both population and clonal samples, we performed genomic DNA extraction using a modified Hoffman-Winston protocol (Hoffman and Winston 1987). We used SYBR Green I to measure gDNA concentration, standardized each sample to 2.5 ng/μL, and constructed libraries using tagmentation following a modified Illumina Nextera library preparation protocol (Baym et al. 2015). To perform PCR clean-up and size selection, we used an Agilent Bravo liquid handling robot. We measured the concentration of purified libraries using SYBR Green I and pooled libraries by balancing their concentrations. We measured fragment size with an Agilent TapeStation 2200 and performed qPCR to determine the final library concentration.

DNA libraries were sequenced using a paired-end (2x75) protocol on an Illumina NextSeq 500. Standard metrics were used to assess data quality (Q30 and %PF). To remove reads from a potentially contaminating organism that was introduced after recovery from the chemostats, we filtered any reads that aligned to Pichia kudriavzevii. Given the evolutionary divergence between these species, the majority of filtered reads belonged to rDNA and similar, deeply conserved

sequences. The median percent contamination was 1.165%. We modified the S. cerevisiae reference genome from NCBI (assembly R64) to include the entire GAP1 CNV reporter and aligned all reads to this reference. We aligned reads using bwa mem ((Li and Durbin 2010), version 0.7.15) and generated BAM files using samtools ((Li et al. 2009), version 1.3.1). FASTQ files for all sequencing are available from the SRA (accession SRP142330).

### 2.6.7: Sequence read depth and breakpoint analysis

To manually estimate CNVs boundaries we used a read-depth based approach. For each sample sequenced, we used samtools (Li et al. 2009) to determine the read depth for each nucleotide in the genome. We liberally defined CNVs by identifying ≥300 base pairs of contiguous sequence when read depth was ≥3 times the standard deviation across chromosome XI for GAP1 or chromosome VIII for DUR3. These boundaries were further refined by visual inspection of contiguous sequence ≥100 base pairs with read depth ≥3 times the standard deviation. These analyses were only performed on sequenced clones because population samples are likely to have multiple CNVs and breakpoints thereby confounding read-depth based approaches.

To determine CNV breakpoints at nucleotide resolution, we extracted split and discordant reads from bam files using samblaster (Faust and Hall 2014). Both split reads and discordant reads were used to identify breakpoints using a weighted scoring method; wherein a split read was worth 1 and discordant reads were worth 3. Positively identified breakpoints required at least 4 split reads and a combined score of at least 9. Breakpoint sequences were generated by making local assemblies of breakpoint associated split reads using MAFFT, EMBOSS, and

velvet (Katoh and Standley 2013; P. Rice, Longden, and Bleasby 2000; Baxevanis et al. 2002). The relationship between breakpoint sequences and the reference genome was determined using BLAST+ (Camacho et al. 2009), with blastn and blastn-short using default settings.

To infer the underlying mechanism by which CNVs were formed, we applied the following criteria. If at least one of the two CNV boundaries contained inverted repeat sequences, and we estimated an odd number of copies in the CNV, we classified the mechanism as ODIRA (Brewer et al. 2011, 2015; Payen et al. 2014). If both of the CNV boundaries occurred within repetitive sequence elements (LTRs or telomeres) and had two copies, we inferred tandem duplication by NAHR (Hastings, Lupski, et al. 2009). Aneuploids were defined on the basis of increased read depth throughout the entire chromosome, but no detected novel sequence junctions. Translocations were identified by LUMPY and Southern blot analysis. All breakpoints that failed to meet these criteria were defined as unresolved.

In addition to CNVs at GAP1 and DUR3 we also identified additional structural variants and CNVs. Structural variants were identified using the split and discordant read approach described above. Additional CNVs were identified using a two pass genome wide read depth approach. In the initial pass, each sample was scanned for regions (400 nucleotide minimum size) with read depth higher than 3 standard deviations relative to the genome. During the second pass, the read depth of each candidate is normalized by the median read depth of that region, as calculated using a subset of clones that lack a candidate in that region. This normalization allows for the correction of sequencing artifacts, batch effects, and the removal of CNV regions

that are not substantially different between the evolved and ancestral clones (ie. rDNA, Ty elements, etc.)

### 2.6.8: SNV and variant identification

SNVs and indel variants were first identified using GATK4's Mutect2 (McKenna et al. 2010), which allows for the identification of variants in evolved samples ('Tumor') after filtering using matched unevolved samples ('Normal') and pool of normals (PON). The PON was constructed using six sequenced ancestral clones, while the paired normal was a single, deeply sequenced, ancestor. Variants were further filtered using GATK's FilterMutectCalls to remove low quality predictions; only variants flagged as 'passed' or 'germline risk' were retained. Given the haploid nature of the evolved population, and further downstream filtering of 'too-recurrent' mutations, we allowed germline risk variants to be retained. Variants were further filtered if they occurred in low complexity sequence; i.e. variants were filtered if the SNV or indel occurred in, or generated, a homogenous nucleotide stretch of five or more of the same nucleotide. Variants from within populations that were detected at less than 5% frequency were considered low confidence and excluded. Finally, variants were filtered if they were found to be 'too-recurrent'; that is, if the exact nucleotide variant was identified in more than three independently evolved lineages, we deemed it more parsimonious to assume that the variant was present in the ancestor at low frequency.

### 2.6.9: Quantifying the number of CNV lineages

We inoculated the lineage tracking library into 20mL ministat vessels (A. W. Miller et al. 2013) containing glutamine-limited media. Control populations containing either zero, one or two

copies of the GAP1 CNV reporter at neutral loci (HO and YLR123C) were also inoculated in ministat vessels for each media condition. Control populations did not contain lineage tracking barcodes. Ministat vessels were maintained and archived as above. Samples were taken for flow cytometry every ~8 generations and analyzed as previously described.

We used fluorescence activated cell sorting (FACS) to isolate the subpopulation of cells containing two or more copies of the the mCitrine CNV reporter using a FACSAria. We defined our gates using zero, one, and two copy mCitrine control strains sampled from ministat vessels at the corresponding timepoints: 70, 90, 150, and 265 generations. Depending on the sample, we isolated 500,000-1,000,000 cells with increased fluorescence, corresponding to two or more copies of the reporter. We grew the isolated subpopulation containing CNVs for 48 hours in glutamine-limited media and performed genomic DNA extraction using a modified Hoffman-Winston protocol (Hoffman and Winston 1987). We verified FACS isolation of true CNVs by isolating clones from subpopulations sorted at generation 70, 90, and 150 (sorted from all lineage tracking populations, bc01-06) and performing independent flow cytometry analysis using an Accuri. We estimated the average false positive rate of CNV isolation at each time point as the percent of clones from a population with FL1 less than one standard deviation above the median FL1 in the one copy control strain. Only subpopulations with fluorescence measurements for at least 25 clones were included in calculations of false positive rate.

We performed a sequential PCR protocol to amplify DNA barcodes and purified the products using a Nucleospin PCR clean-up kit (Levy et al. 2015). We quantified DNA concentrations by qPCR before balancing and pooling libraries. DNA libraries were sequenced using a paired-end

(2x150) protocol on an Illumina MiSeq 300 Cycle v2. Standard metrics were used to assess data quality (Q30 and %PF). However, the reverse read failed due to over-clustering, so all analyses were performed only using the forward read. We used the Bartender algorithm with UMI handling to account for PCR duplicates and to cluster sequences with merging decisions based solely on distance except in cases of low coverage (<500 reads/barcode), for which the default cluster merging threshold was used (L. Zhao et al. 2017). Clusters with a size less than four or with high entropy (>0.75 quality score) were discarded. We estimated relative abundance of barcodes using the number of unique reads supporting a cluster compared to total library size.

## 2.7: Supplemental figures

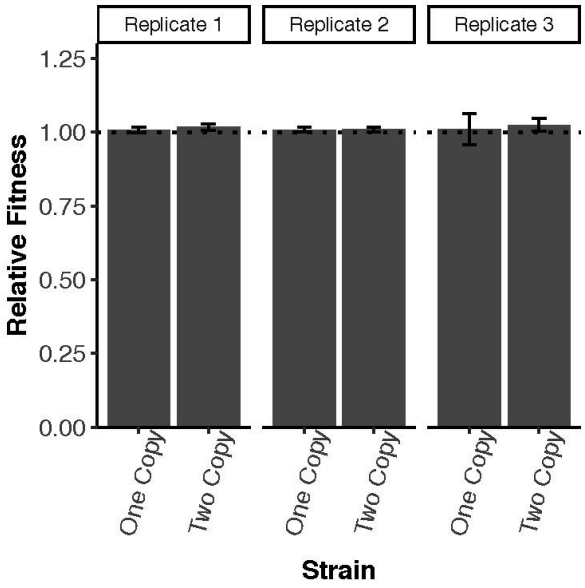


**Figure 2.S1**. **Assessment of CNV reporter fitness effects.** The fitness of strains carrying one (DGY500) or two copies (DGY1315) of a constitutively expressed mCitrine gene was assayed. Fluorescent strains were co-cultured with the non-fluorescent, unevolved reference strain (FY4). We performed three independent competitive fitness assays in glutamine-limited chemostats using the same conditions as evolution experiments. No significant fitness defect was observed for either strain indicating that constitutive expression of one or two copies of the fluorescent gene does not confer a fitness cost in these conditions. Error bars are 95% confidence intervals.

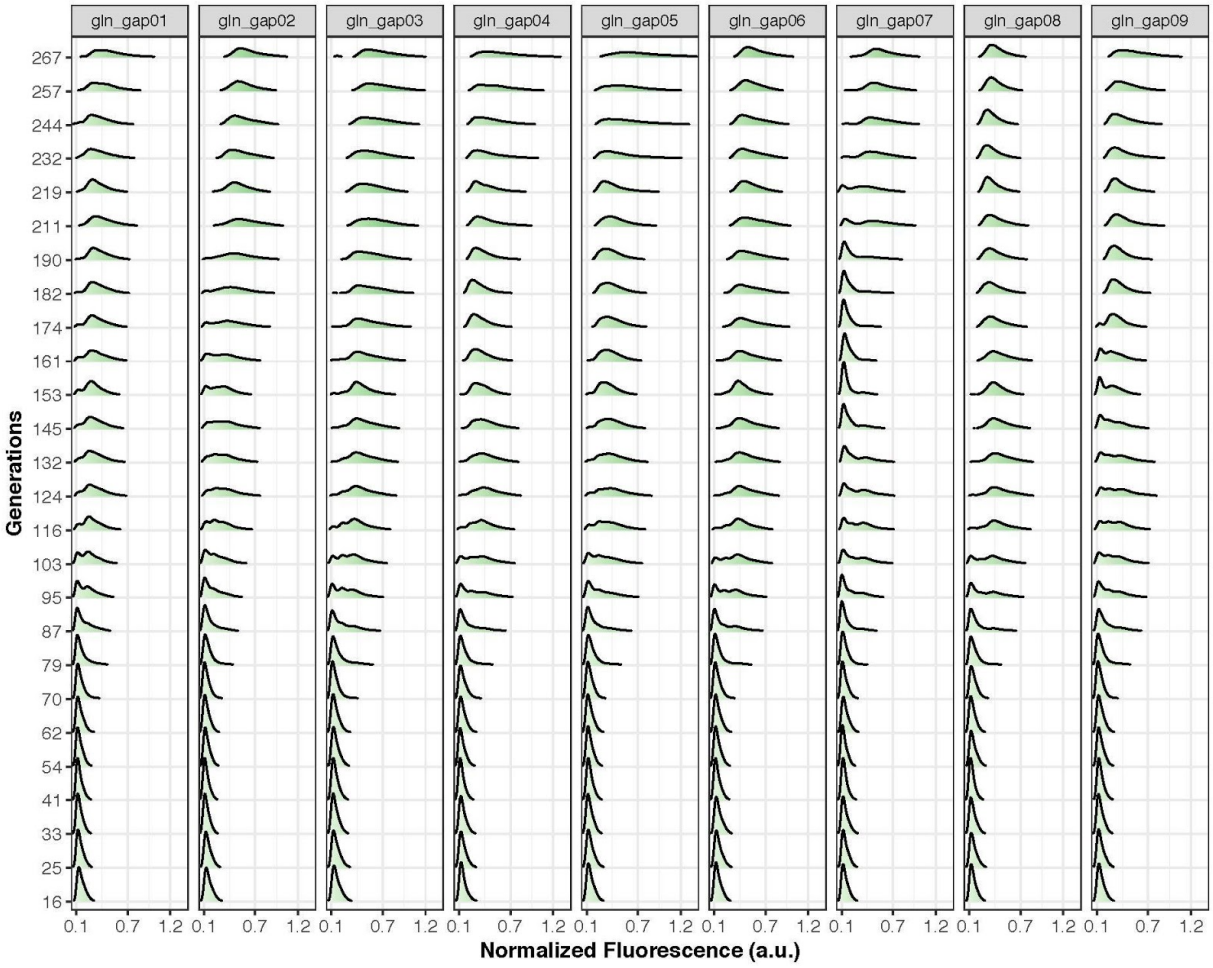


**Figure 2.S2**. **The** GAP1 **CNV reporter indicates the emergence of** GAP1 **CNVs in all glutamine-limited populations.** Distributions of single-cell fluorescence over time for all glutamine-limited experimental populations. Fluorescent signal is normalized by forward scatter, which varies as a function of cell size. Each distribution is based on 100,000 single cell measurements.

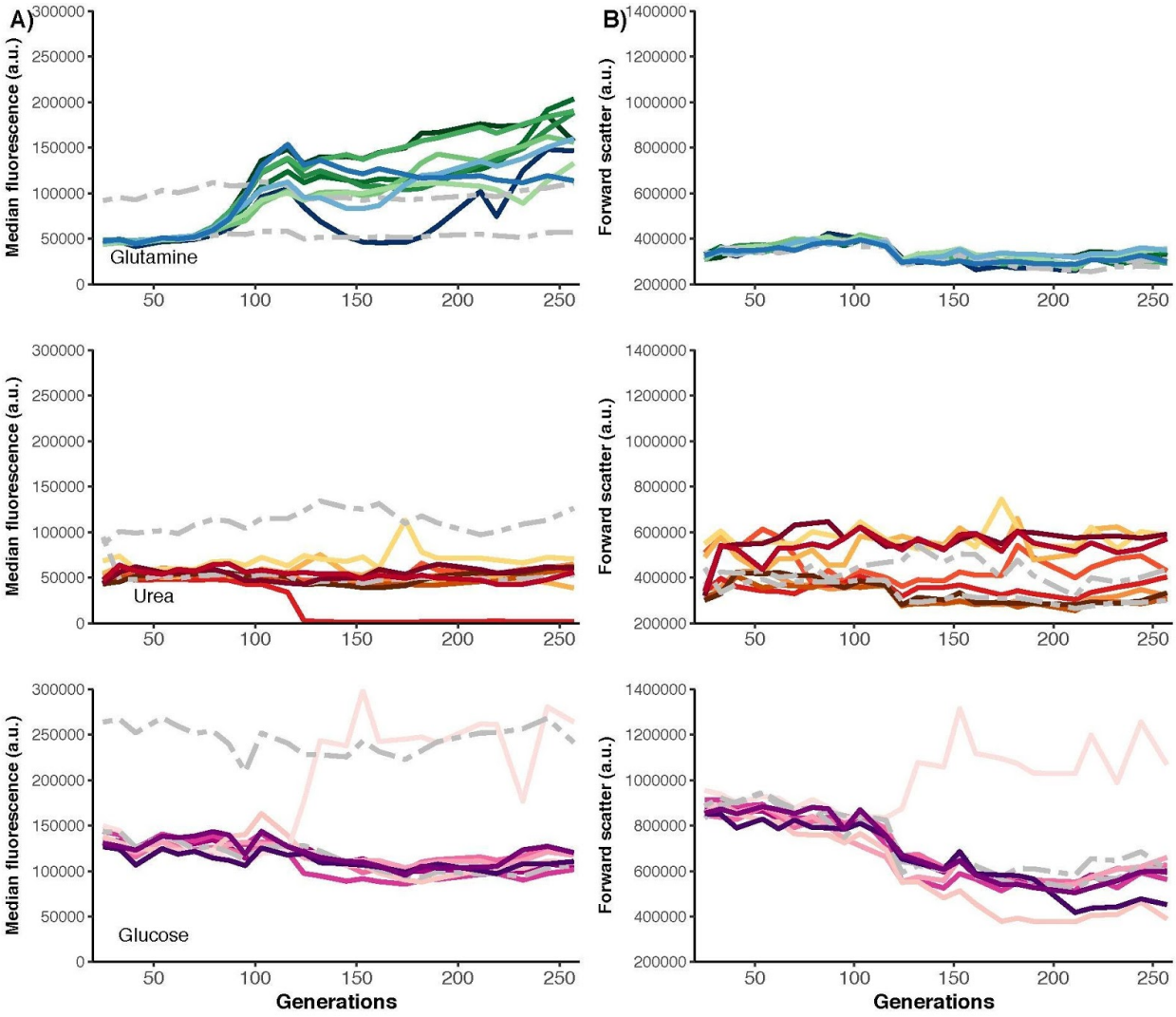


**Figure 2.S3. Normalization by forward scatter mitigates effects of cell physiology and morphology variation on CNV reporter signal.** Dashed grey lines represent one and two copy control populations. (**A**) Median unnormalized fluorescence across time for all evolving populations. (**B**) Median forward scatter over time for all populations. One glucose-limited population (pink) developed a bud separation defect, resulting in a cell aggregation phenotype and large forward scatter and fluorescence measurements. Normalizing by forward scatter accounts for this issue and other changes in overall cell physiology during the evolution experiments (see **Fig 2.2B**).

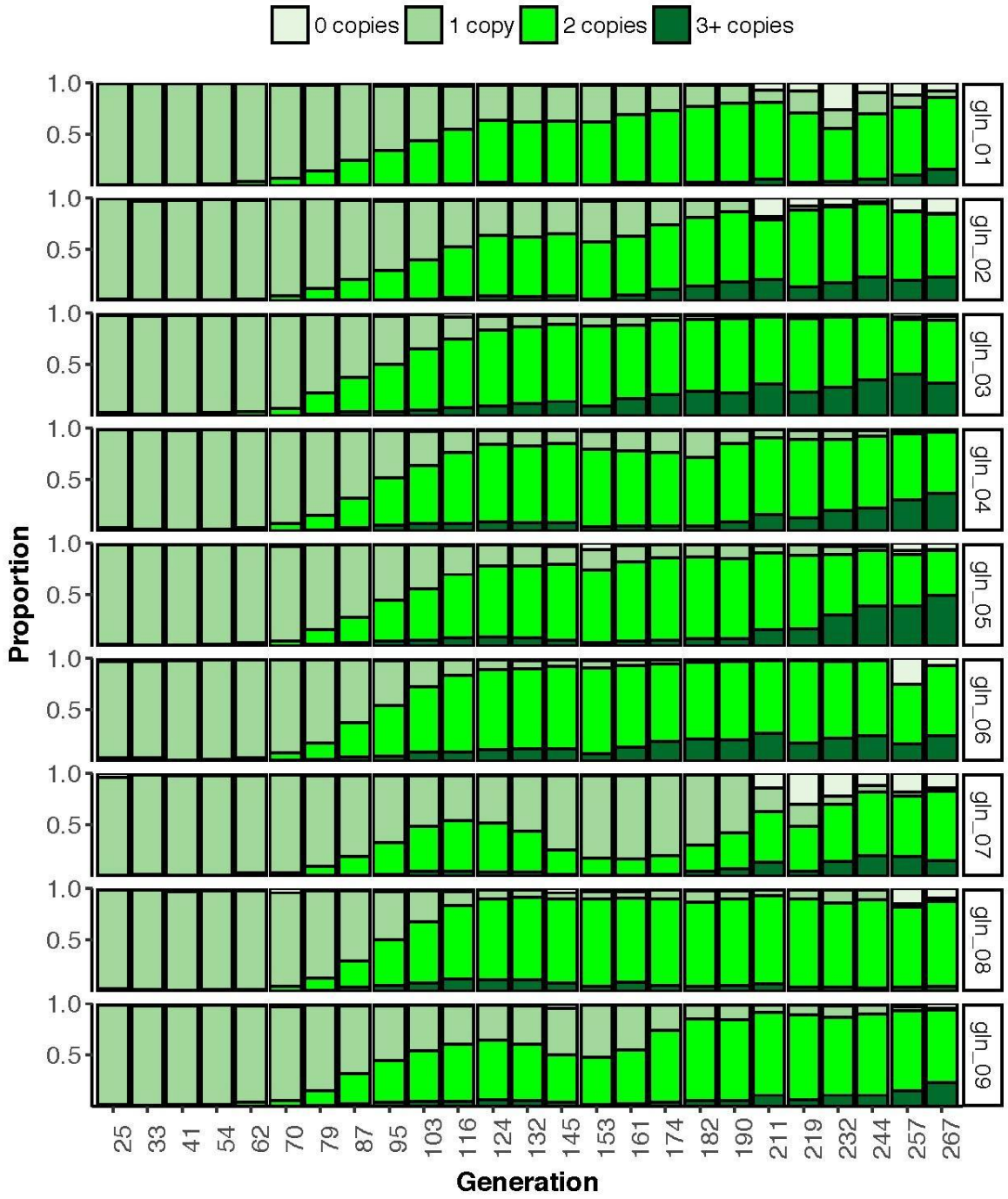


**Figure 2.S4. Gating flow cytometry data enables estimation of CNV alleles that contain more than two copies.** The proportion of cells with zero, one, two, and three or more copies of GAP1 in each glutamine-limited experimental population. Proportions were calculated after generating gating criteria based on one and two copy control populations.

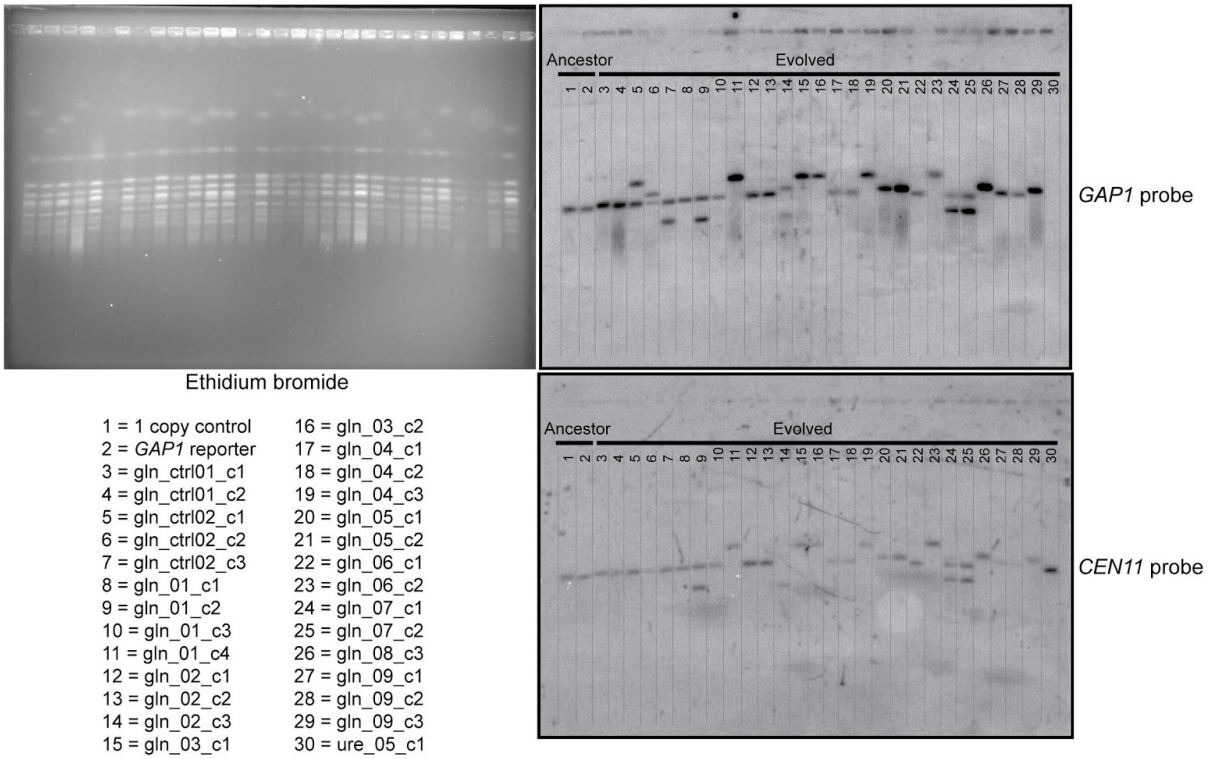


**Figure 2.S5. Pulsed-field gel electrophoresis enables molecular characterization of** GAP1 **CNVs.** Analysis of ancestral and evolved clones. Whole chromosomes were visualized by ethidium bromide staining (left) then probed for GAP1 and CEN11 (right). In the majority of cases, the CEN11 probe correlates with GAP1 probe signal, indicating that these GAP1 amplifications are located on chromosome XI. Instances where the CEN11 and GAP1 probes do not correlate are indicative of non-reciprocal translocations. Duplication of CEN11 may indicate segmental aneuploidy.

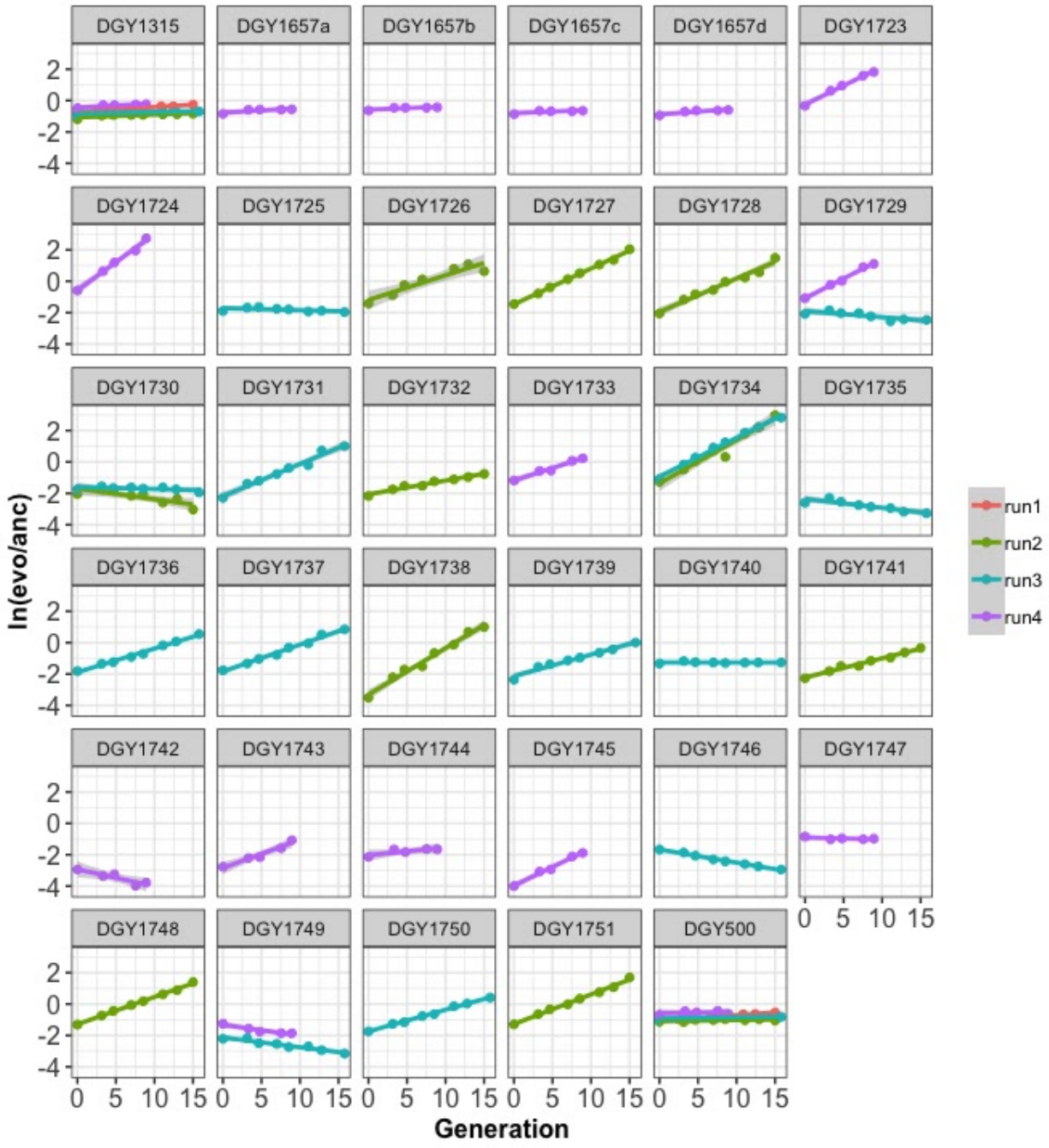


**Figure 2.S6. Fitness is calculated using linear regression after performing pairwise competition assays.** Evolved lineages with GAP1 CNVs were co-cultured with a non-fluorescent, unevolved reference strain (FY4) in glutamine-limited chemostats. We determined the proportion of fluorescent cells over time (for either 10 or 15 generations) using flow cytometry. The slope of the linear fit is equivalent to the fitness, relative to the ancestor, for each evolved lineage. In cases where two independent fitness measurements for a single lineage were in rough agreement, we included only the replicate with the lowest standard error. Measurements for DGY1729 were conflicting and are not reported in **Figure 2.S7**.

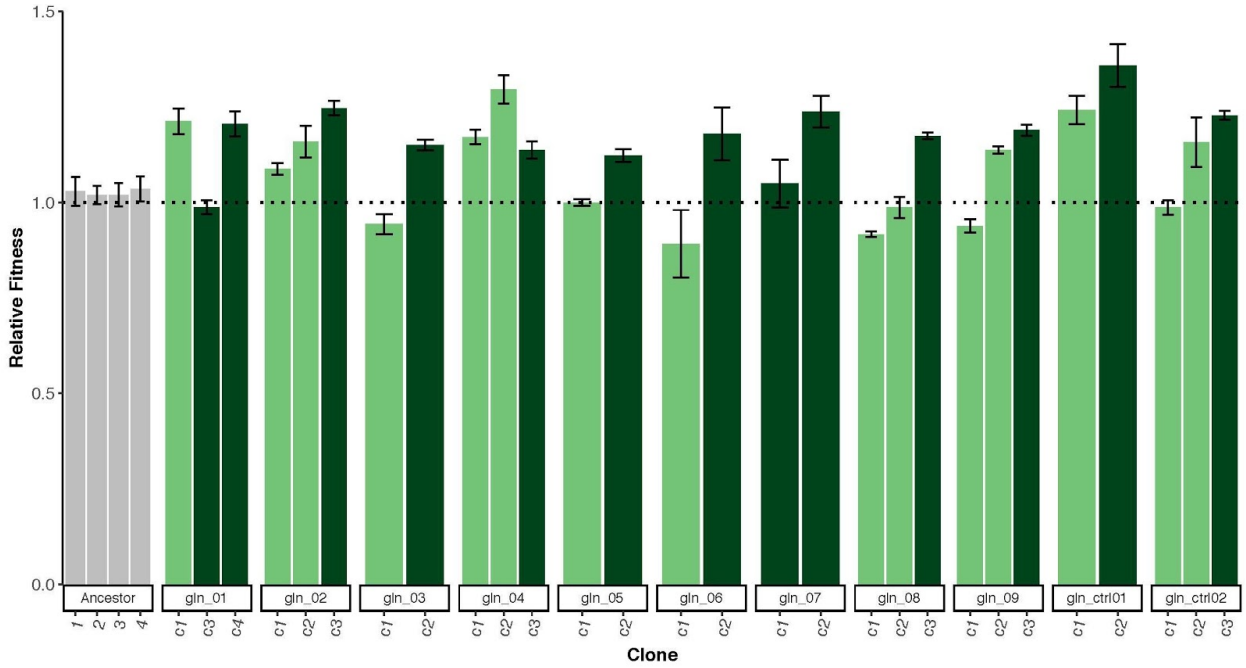


**Figure 2.S7.** GAP1 **CNV-containing lineages have a higher relative fitness than the ancestral strain.** The fitness of evolved lineages containing GAP1 CNVs was determined by pairwise competition experiments with a non-fluorescent, unevolved reference strain (FY4) in glutamine-limited chemostats. The majority (18/28) of evolved CNV-containing lineages have significantly higher fitness (t-test, bonferroni-corrected p.val < 0.00156) than the ancestor. Decreased (2/28), or insignificant fitness differences (8/28), may reflect context-specific fitness effects of GAP1 CNV-containing lineages. Error bars are 95% confidence intervals.

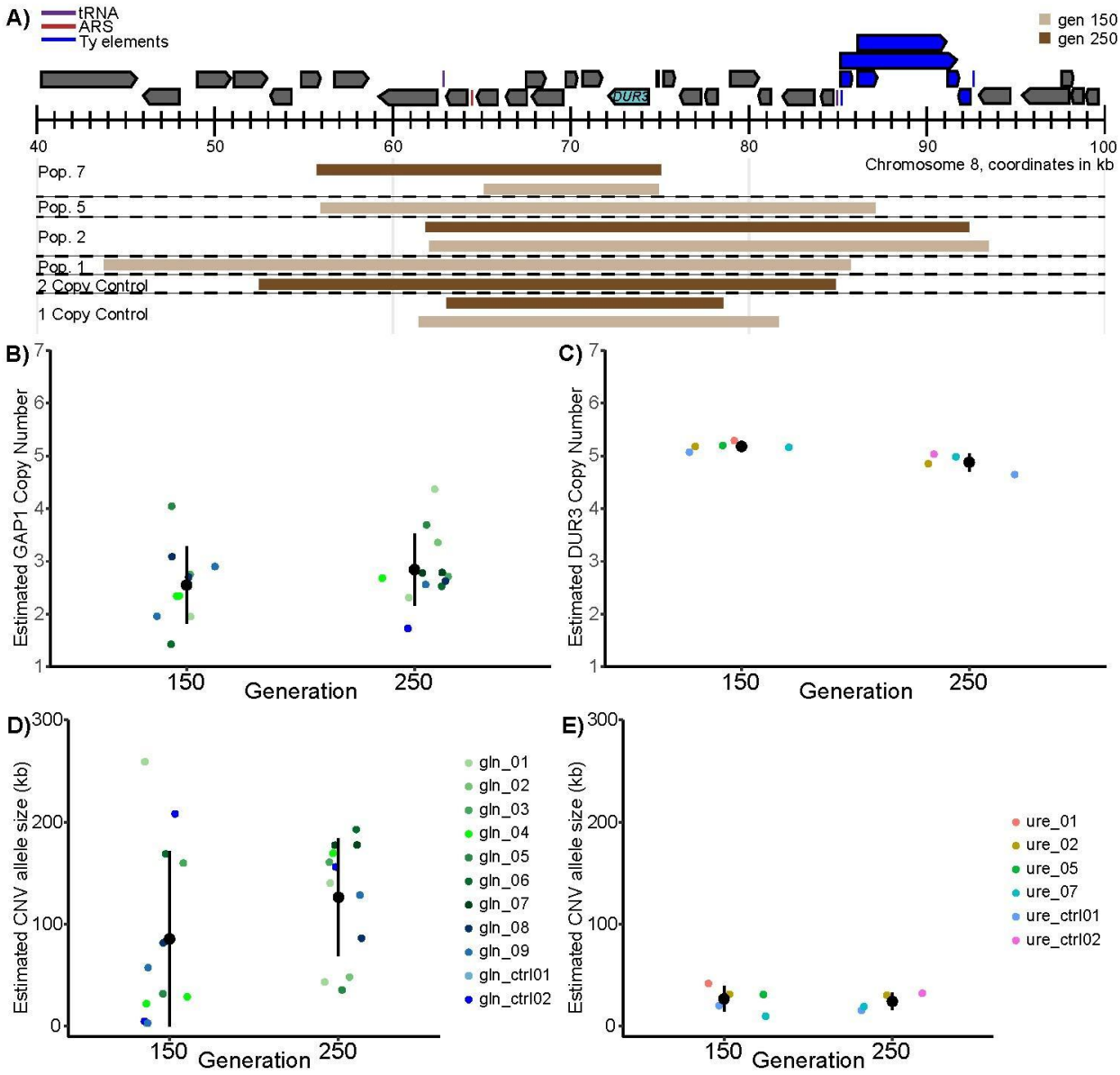


**Figure 2.S8**. **Identification of CNV alleles at the** DUR3 **locus.** (**A**) A schematic illustrating the genomic context and estimated breakpoints for clones containing DUR3 CNVs isolated from urea-limited chemostats at generation 150 and generation 250. Breakpoint boundaries were estimated using a read depth based approach. Compared to (**B**) clones isolated from glutamine-limited chemostats containing GAP1 CNVs, (**C**) clones isolated from urea-limited chemostats have significantly higher copy number (t-test p.val < 0.01). (**D**) GAP1 CNV alleles are significantly larger than (**E**) DUR3 CNV alleles (t-test p.val < 0.01). ARS = autonomously replicating sequence.

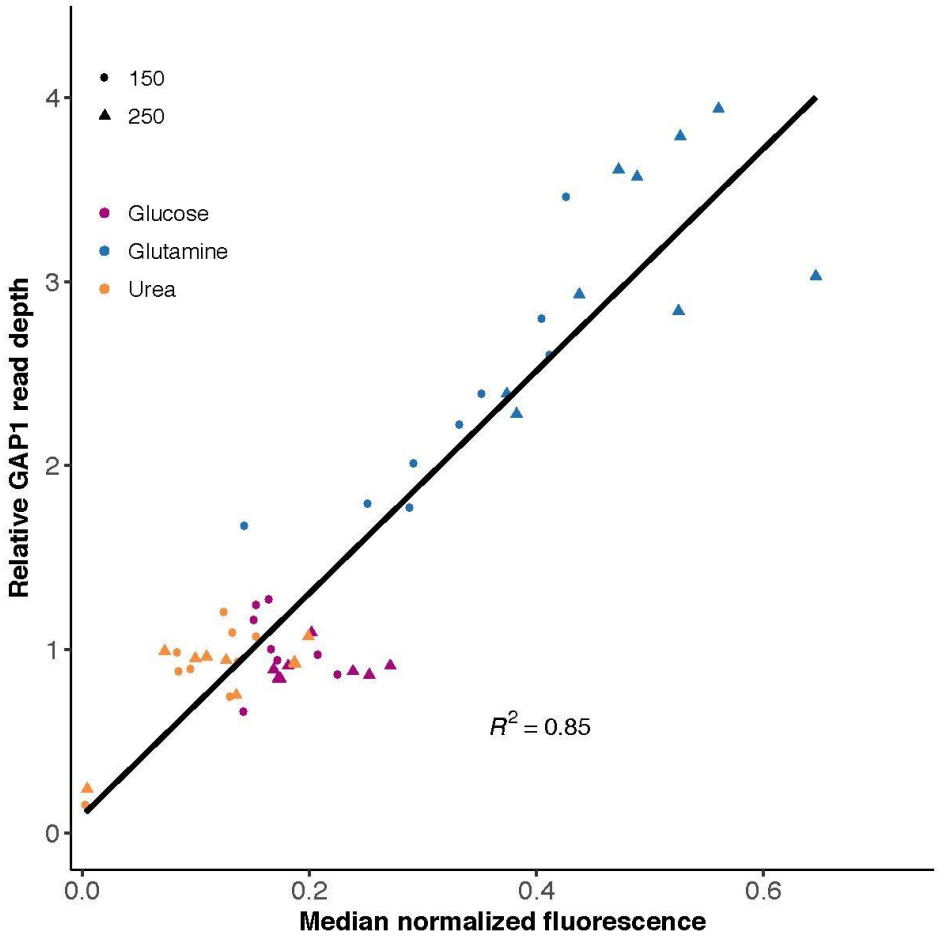


**Figure 2.S9**. **Population estimates of** GAP1 **copy number by CNV reporter and quantitative sequencing are linearly correlated and increase with time of adaptive evolution.** Relative depth at the GAP1 locus, calculated from whole genome sequencing data, is strongly correlated with the median normalized fluorescence of the GAP1 CNV reporter in populations. Glutamine-limited populations measured at generation 250 tend to have higher fluorescence and higher relative read depth at the GAP1 locus than at generation 150.

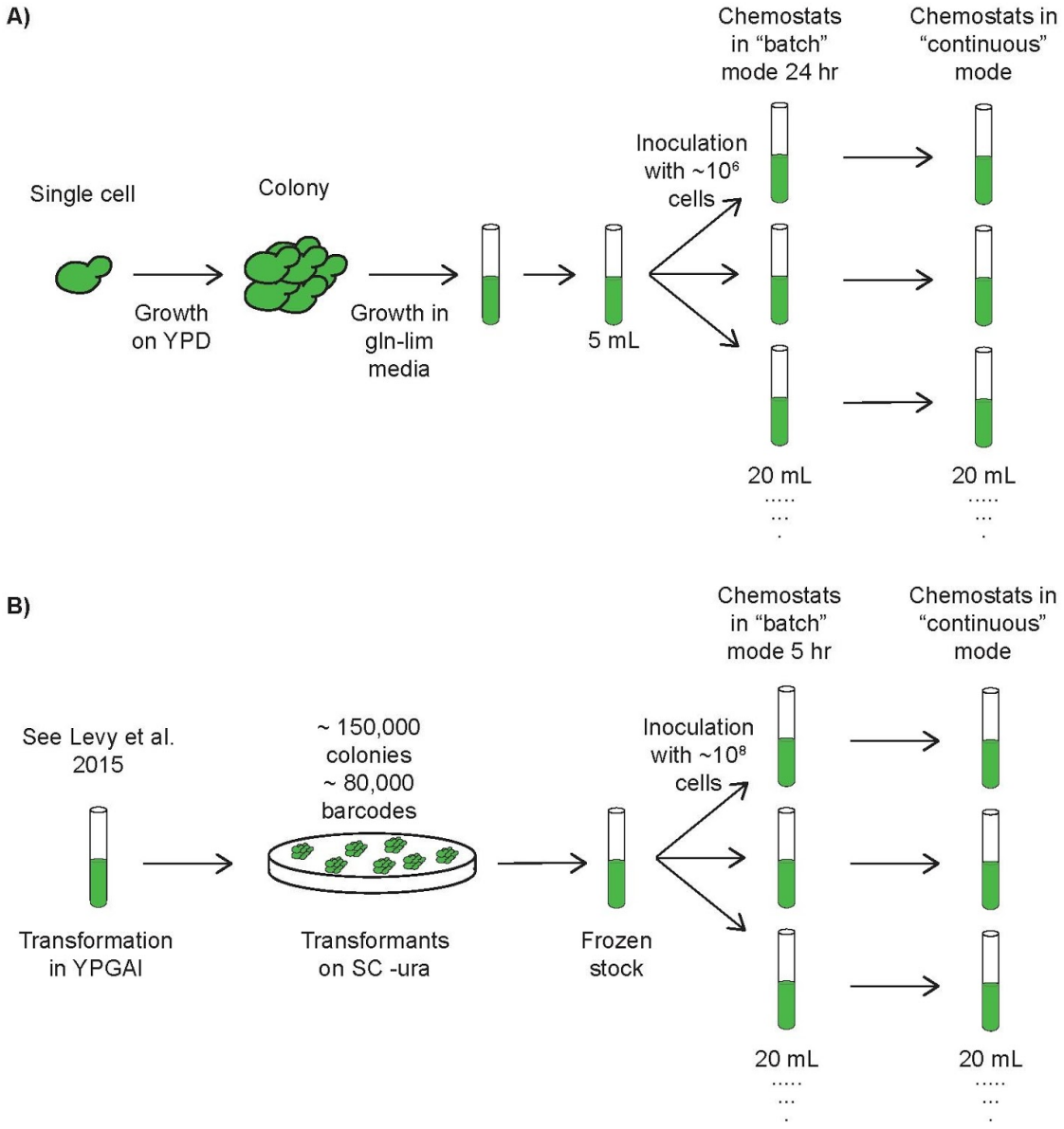


**Figure 2.S10. Population prehistory of independent evolution experiments.** All independent populations share a common history prior to founding of individual populations. The prehistory of experiments using the GAP1 CNV reporter (**A**) differ with respect to the size of the founding population in experiments using a lineage tracking library (**B**). Any variation that is introduced prior to founding of individual populations may contribute to the evolution of all populations. Variation that is introduced after separation into individual populations contributes to evolutionary outcomes in that population only. YPD = yeast extract-peptone-dextrose (rich media). Gln-lim = glutamine limited. YPGAl = yeast extract-peptone-galactose.

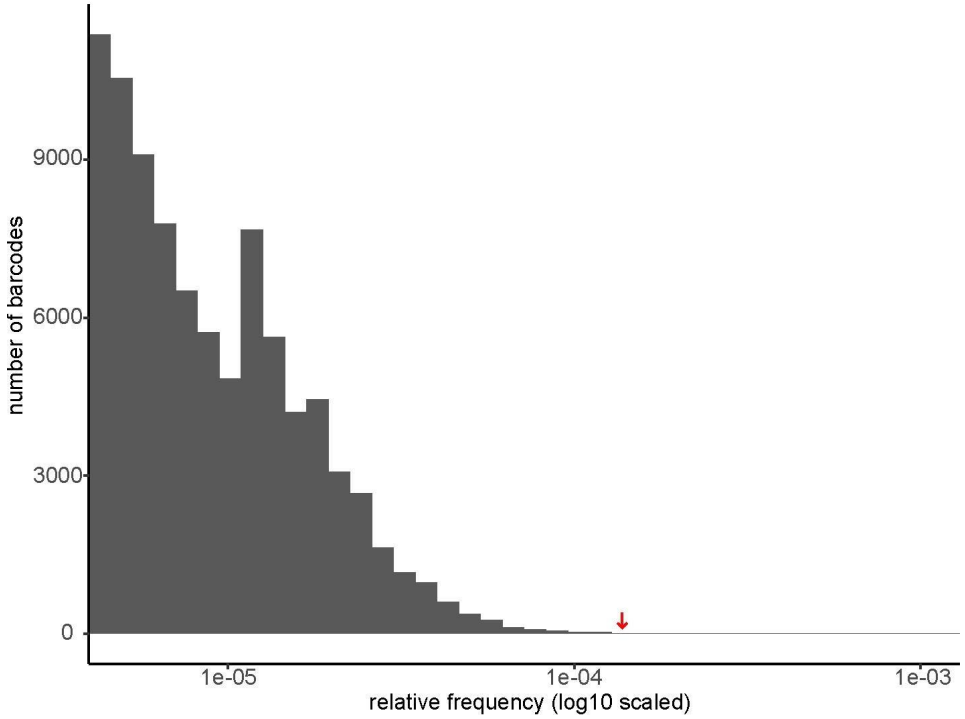


**Figure 2.S11**. **Distribution of barcode counts in ancestral populations**. We determined the distribution of read counts supporting each unique barcode in the ancestral population, after filtering out low confidence clusters. The relative frequencies of barcodes vary by over an order of magnitude and we observe a long tail with a few barcodes significantly overrepresented in the ancestral population. The red arrow indicates an overrepresented barcode in the ancestral population that was identified in the CNV subpopulation in both independent barcoded evolution experiments (indicated in purple in **Fig 5B**). This distribution is consistent with that found in other barcode lineage tracking experiments (Levy et al. 2015).

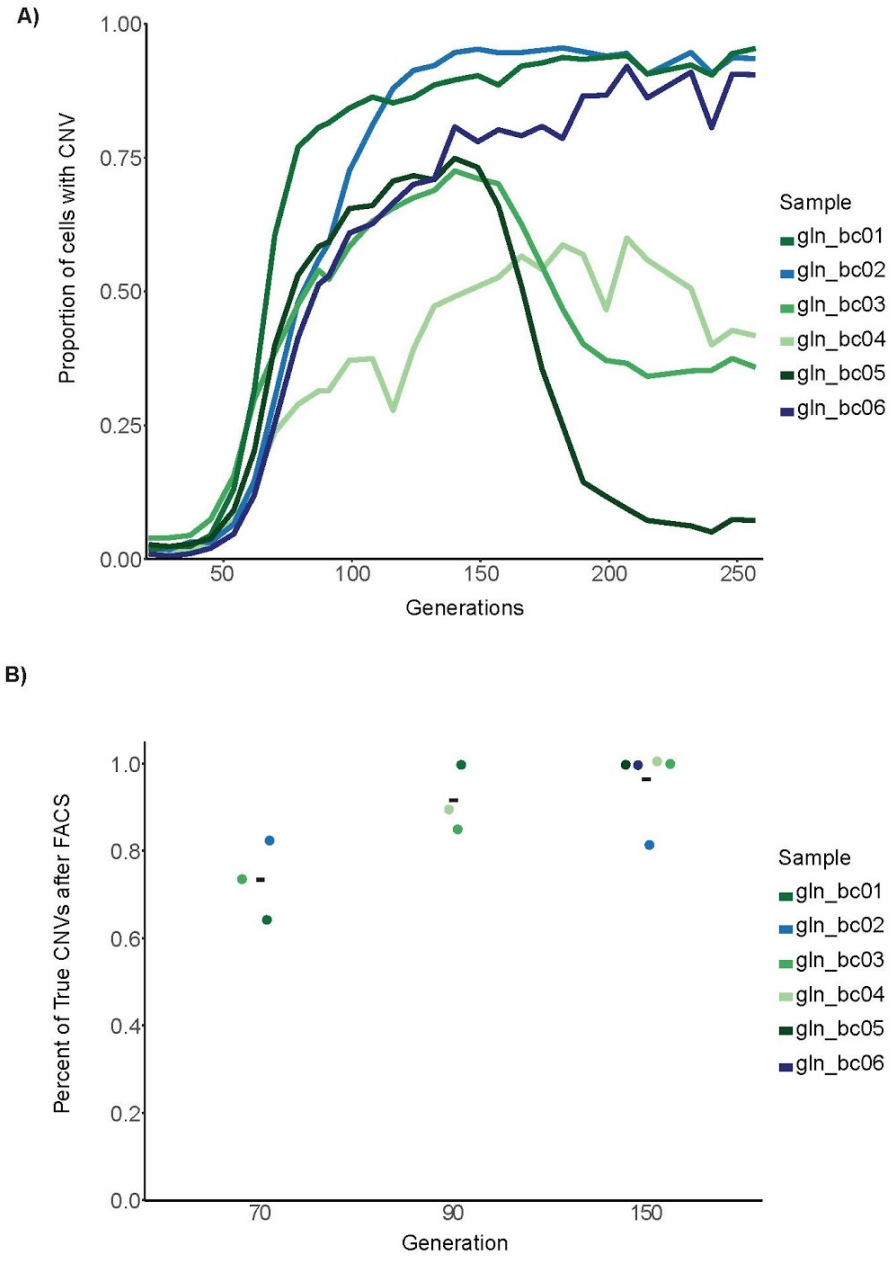


**Figure 2.S12. Identification of barcoded** GAP1 **CNV-lineages in evolving populations.** (**A**) GAP1 CNV dynamics in barcoded populations assayed using a CNV reporter. **(B)** Estimation of true positive rate of CNV isolation by FACS at generations 70, 90, and 150. CNV subpopulations were isolated by FACS at each timepoint and clones isolated by plating for single colonies. The percent of cells containing a CNV in the fractionated subpopulation was estimated using at least 25 clones. A one copy control strain was used to define gates.

# Chapter 3: Comparing CNV dynamics across loci using a fluorescent CNV reporter

## 3.1: Abstract

Copy number variants (CNVs) are a common outcome during adaptive evolution in nutrient-limited environments, but the dynamics with which they are generated and selected across different environmental conditions are largely unknown. To overcome the inherent challenge in detecting CNVs at low frequency, we previously developed a reporter construct that uses fluorescence to distinguish single cells that acquire CNVs at a specific locus. An advantage of this approach is that the CNV reporter can be integrated throughout the genome, allowing comparative analysis across multiple loci and selective conditions. To determine the dynamics with which CNVs emerge and are selected across loci, we inserted the CNV reporter at three different locations: between the high-affinity glucose transporters HXT6 and HXT7, adjacent to the ammonium permease MEP2, and adjacent to the general amino acid permease GAP1. In agreement with our previous findings, we observe CNVs arising early and predictably at the GAP1 locus. We also identified CNVs at the HXT6/7 locus, but CNV formation was more variable and CNV dynamics were complex. We did not observe CNVs at the MEP2 locus. Our results show that the CNV reporter enables quantitative, high-resolution temporal measurements of CNV dynamics across three different nutrient limitations and loci under selection, providing the first comparative analysis of the emergence, selection, and maintenance of adaptive gene amplifications in evolving populations.

## 3.2: Introduction

Copy number variants (CNVs) are duplications or deletions of DNA sequence that result in polymorphisms among individuals of a population. In evolution, CNVs can drive rapid phenotypic diversification and adaptation. During short-term evolutionary scenarios, such as niche specialization or domestication, changes in DNA copy number can result in strong selective advantages (Clop, Vidal, and Amills 2012; Żmieńko et al. 2014; Greenblum, Carr, and Borenstein 2015; Meyer and Purugganan 2013; Ramirez et al. 2014; Dhami, Hartwig, and Fukami 2016). Across longer evolutionary timescales, duplication of genes allows diversification and evolution of gene families through neofunctionalization and subfunctionalization (Conant and Wolfe 2008). For example, CNVs played a significant role in human evolution, especially in dietary adaptations and in advancing brain function and development (Barreiro et al. 2008; Iskow et al. 2012). Despite the prevalence and importance of CNVs for phenotypic variation and adaptive evolution, fundamental questions regarding their rate of formation, the molecular mechanisms underlying their formation, and their ultimate fate in evolving populations remain unresolved.

CNVs are commonly observed in experimentally evolving lineages of microbial populations. In studies that use a limited nutrient as the selective agent, evolved strains often have amplification alleles for the transporter of that specific nutrient. This phenomenon was first observed in experiments with Escherichia coli limited for lactose. These evolved lineages produce large amounts of B-galactosidase as a result of amplifications in the lac operon (Horiuchi, Horiuchi, and Novick 1963). Early studies in Saccharomyces cerevisiae identified

amplification alleles of the gene that hydrolyzes organic phosphates in phosphate-limited environments (Hansche 1975). Subsequently, multiple, independent studies have identified amplifications of the high-affinity sulfur transporter SUL1 (Gresham et al. 2008; Payen et al. 2014) and the high affinity glucose transporter genes HXT6 and HXT7 in glucose-limited media (Brown, Todd, and Rosenzweig 1998; Kao and Sherlock 2008; Gresham et al. 2008). HXT6 and HXT7 have 99.7% sequence similarity and unequal crossover by non-allelic homologous recombination (NAHR) can lead to their amplification (Brown, Todd, and Rosenzweig 1998).

Amplification alleles for the corresponding limiting nutrient have been identified in the high-affinity proline transporter PUT4, the urea transporter DUR3, the allantoin permease DAL4, and the ammonia permease MEP2 (Hong and Gresham 2014; Hong et al. 2018). We have also previously shown that CNV formation occurs early and repeatedly at the GAP1 locus during selection in nitrogen-limited environments (Hong and Gresham 2014; Gresham et al. 2010; Lauer et al. 2018). GAP1 is a general amino-acid permease that can transport all natural L-amino acids, several D-amino acids, related compounds, and toxic analogs (Stanbrough and Magasanik 1995).

One mechanism of CNV formation contributing to GAP1 amplification in nitrogen-limited environments is generation of a self-propagating extra-chromosomal circle that contains the GAP1 locus and an origin of replication (Gresham et al. 2010). This amplification event is mediated by intrachromosomal recombination between long terminal repeats (LTRs) flanking GAP1. CNVs often occur in response to DNA damage, such as double strand breaks (DSBs). Typically, DSBs are repaired by homologous recombination (HR) and do not result in structural

or genomic alterations. However, LTRs and other repetitive elements increase the probability that NAHR will occur.

LTRs are common in the yeast genome and include retrotransposon elements from the Ty1, Ty2, Ty3, Ty4, and Ty5 families. A total of 483 Ty element insertions have been identified, comprising ~3% of the yeast genome (Carr, Bensasson, and Bergman 2012). GAP1 is relatively unique in that it has flanking LTRs and flanking tRNA genes, which have also been implicated in genome instability (Tran et al. 2017). This means GAP1 could be a potential “hotspot” of repeated CNV formation. This is contrast to other loci including MEP2, which does not have repetitive sequence elements in close proximity. As a result, CNV formation at the MEP2 locus may rely on other mechanisms.

Several mechanisms of CNV formation require little or no homology. These include microhomology-mediated repair (MMR) and origin-dependent inverted-repeat amplification (ODIRA). First observed for amplifications in the SUL1 locus, ODIRA requires a nearby origin of replication and inverted-repeats that are closely spaced, no more than a distance the length of Okazaki fragments (Brewer et al. 2011, 2015). Replication fork slippage occurs and results in broken forks, which regress and anneal to their complementary repeat. The invading strand becomes ligated to the adjacent Okazaki fragment, resulting in a “closed” fork (Brewer et al. 2011, 2015). This forms an extra-chromosomal circle that can replicate autonomously because it contains an origin of replication. Alternatively, it can stably reintegrate into the genome through homologous recombination (Brewer et al. 2011, 2015). Recent evidence suggests that

extrachromosomal circles are quite common, even in the absence of selection, and may be a major source of CNVs (Møller et al. 2015; K. M. Turner et al. 2017).

A central challenge to quantifying CNV dynamics is the development of accurate and sensitive methods to detect and analyze CNVs in heterogeneous evolving populations. To overcome this challenge, we previously developed a novel fluorescent reporter to detect and isolate subpopulations of cells that acquire de novo CNVs. Our new assay allows detection of CNVs with unprecedented temporal resolution and provides an opportunity to perform comparative analysis across different selective conditions and loci. To determine the effect of local genomic features and varying selective pressures on the rate and molecular mechanisms of CNV formation, we analyzed CNV dynamics across three distinct loci under three different environmental conditions. We find that temporal dynamics differ, with CNV formation occurring at the GAP1 and HXT6/7 loci, while MEP2 amplifications were not observed.

## 3.3: Results

### 3.3.1: Stability of a fluorescent reporter during laboratory evolution

For a fluorescent marker to accurately report on CNV formation, fluorescent signal must remain stable during long term laboratory evolution. To determine the stability of the fluorescent reporter, I performed three independent evolution experiments with a strain that has mCitrine integrated at an endogenous neutral locus (HO) in FY4, a haploid derivative of the reference strain S288c. The mCitrine reporter is under control of the constitutively expressed ACT1 promoter and marked by the KanMX G418-resistance cassette. Populations were maintained in ammonium-limited media on continuous mode (dilution rate = 0.12 culture volumes/hr;

population doubling time = 5.8 hours) for 170, 174, and 120 generations respectively. Flow cytometry was used to measure fluorescence of 100,000 cells per sample. Fluorescence remains constant for the duration of each replicated evolution experiment (**Figure 3.1**). Maintenance of protein fluorescence is consistent with the absence of a detectable fitness cost associated with one copy of the CNV reporter in ammonium-limited chemostats. A small population of cells have high fluorescence (generation 11 in population 1), which we believe is a direct result of aberrant changes to cell size during sampling (**see Figure 3.10**).

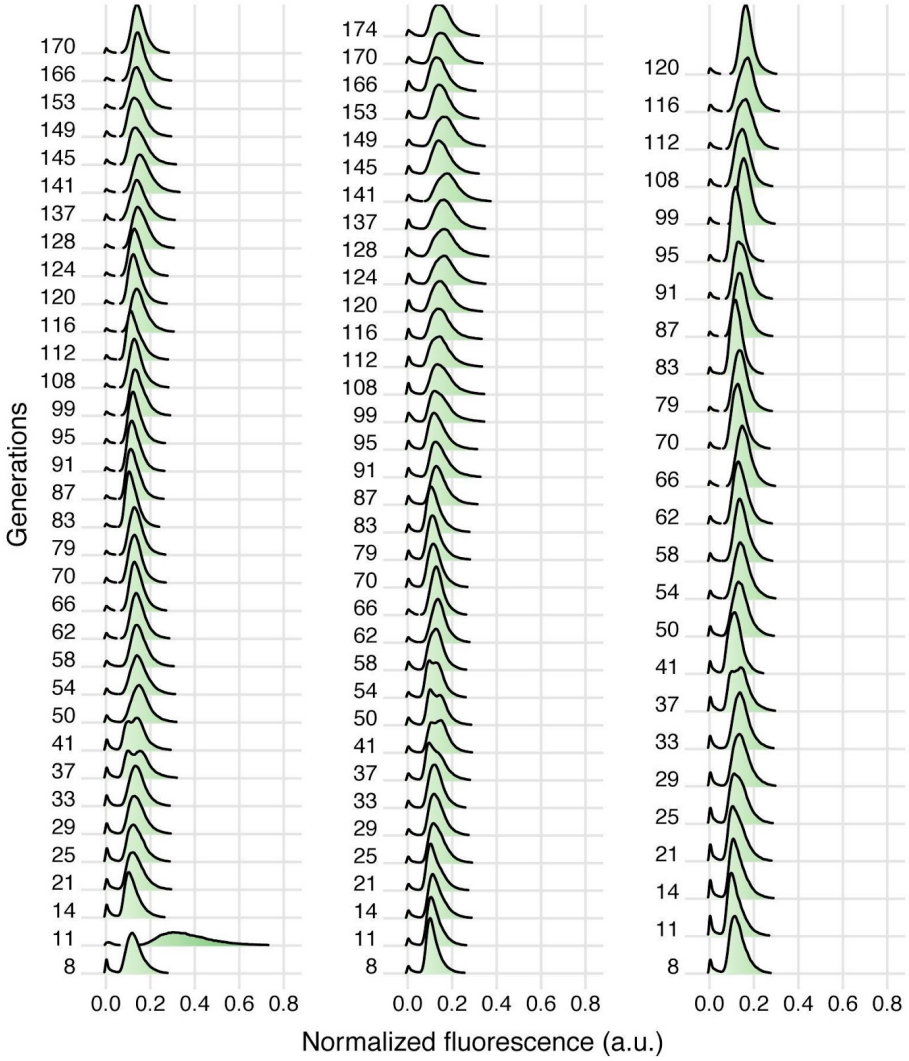

**Figure 3.1. Fluorescent signal of an mCitrine reporter integrated at a neutral locus remains constant during adaptive evolution.** Normalized distributions of single-cell fluorescence over time for a one-copy reporter strain evolving in three independent ammonium-limited chemostats. The mCitrine cassette is integrated at an endogenous neutral locus, HO, in S. cerevisiae. Single cell fluorescence is normalized by the forward scatter measurement of the cell, and 100,000 single cells were measured to generate each distribution. Cells from population 1 had abnormal cell sizes at the time of sampling (generation 11) which affected the normalized fluorescence value.

## 3.3.2: Creation of a CNV reporter

To study the effect of genomic context on the rate and diversity of CNVs, I inserted the CNV reporter at three distinct loci: 1) the high-affinity glucose transporters HXT6 and HXT7, 2) the ammonium permease MEP2, and 3) the general amino-acid permease GAP1. Each locus has distinct genomic features (**Figure 3.2)**: HXT6 and HXT7 are 99.7% identical gene sequences that lie proximal to an autonomously replicating sequence (ARS) whereas GAP1 is adjacent to an ARS and flanked by Ty1 long terminal repeats (LTRs) and tRNA genes (not shown in the diagram). By contrast, MEP2 does not have large flanking repetitive sequences or a nearby ARS. The mCitrine reporter construct was stably integrated into FY4 at intergenic regions to avoid promoters and 5' and 3'-UTRs. For GAP1 and MEP2, this location is approximately 1-1.2 kilobases upstream of the coding sequence (integration coordinates, chromosome XI: 513945 and chromosome XIV: 356282, respectively). For HXT6/7, the construct is inserted between the two genes, avoiding the ARS sequence (integration coordinates, chromosome VI: 1157590).

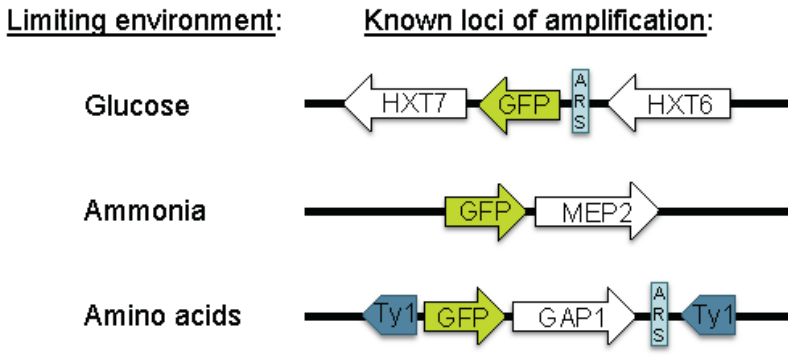

**Figure 3.2. Schematic representation of mCitrine integration.** Reporter strains were generated by integration of mCitrine into nearby intergenic regions for loci of interest: HXT6/7, MEP2, and GAP1.

### 3.3.3: CNV dynamics for MEP2 in ammonium-limited chemostats

To test detection of de novo CNVs in evolving populations, I performed a long-term evolution experiment using the MEP2-mCitrine strain in three ammonium-limited chemostats for 199 or 215 generations using the same continuous culturing mode as described above. Using flow cytometric analysis in the flowCore R package (Ellis et al. 2016), I plotted the distribution of fluorescence for 100,000 single cells, each normalized by their forward scatter measurement (a proxy for cell size). During ammonium-limitation, changes in fluorescent signal were not observed (**Figure 3.3**). In fact, fluorescence remains low and constant, similar to what was observed with the 1 copy control strain in ammonium-limited chemostats (**Figure 3.1**). This is consistent with previous data from our laboratory, which demonstrates that while MEP2 CNVs have been observed, single nucleotide variants are also a common mode of adaptation during ammonium limitation in chemostats (Hong and Gresham 2014; Hong et al. 2018). There are three non-exclusive explanations for the observed data: 1) CNVs form at the MEP2 locus, but CNV formation is rare and three (relatively) short-term replicates simply don't provide enough data points to make any conclusions, 2) CNVs form at the MEP2 locus, but the MEP2-mCitrine reporter does not accurately report on their formation, and 3) MEP2 CNVs form but are quickly outcompeted and eliminated from the population by adaptive mutations with higher fitness benefits (see **discussion**).

### 3.3.4: CNV dynamics of HXT6/7 in glucose-limited chemostats

In addition to the three ammonium-limited chemostat experiments, three additional replicates in glucose-limited media were performed simultaneously. These evolution experiments were performed as described above, but using the HXT6/7-mCitrine strain, and were carried out for 240 generations. I detected a subpopulation of cells with increased fluorescence at generation 87 in population Gluc1 and at generation 95 in Gluc3 (**Figure 3.4**). These data are consistent with generation of one or more adaptive CNV(s) at the HXT6/7 locus that are under strong positive selection. In Gluc3, a small subset of the population maintained increased fluorescence for the duration of the experiment. In Gluc1, the increase in fluorescence was more dramatic and more variable, which could be indicative of multiple subpopulations with 3, 4 or more copies of mCitrine.

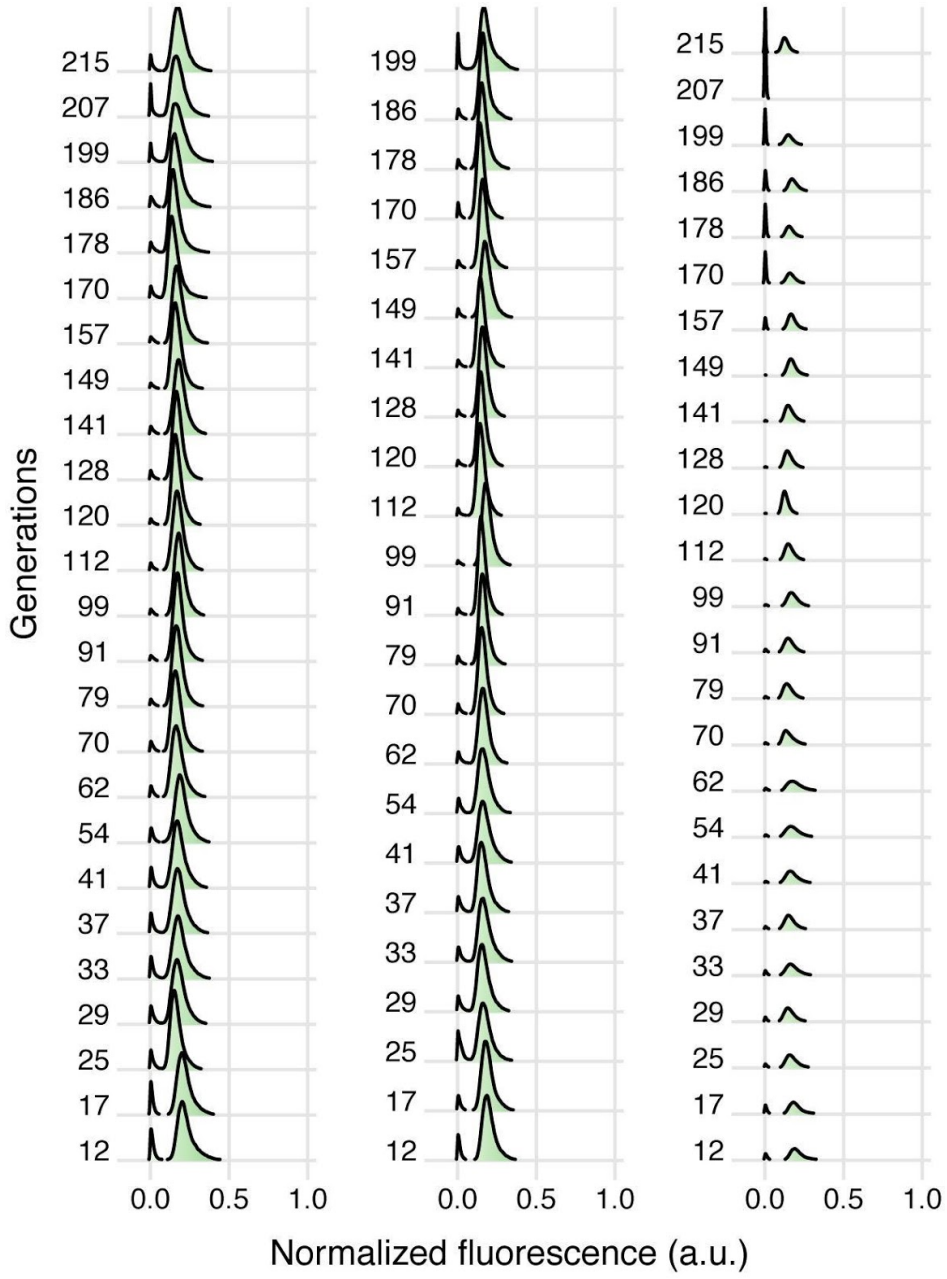


**Figure 3.3. Fluorescent signal of a** MEP2-mCitrine **reporter does not increase during hundreds of generations of ammonium limitation.** Normalized distributions of single-cell fluorescence over time for the MEP2-mCitrine reporter strain evolving in three independent ammonium-limited chemostats. Single cell fluorescence is normalized by the forward scatter measurement of the cell, and 100,000 single cells were measured to generate each distribution.

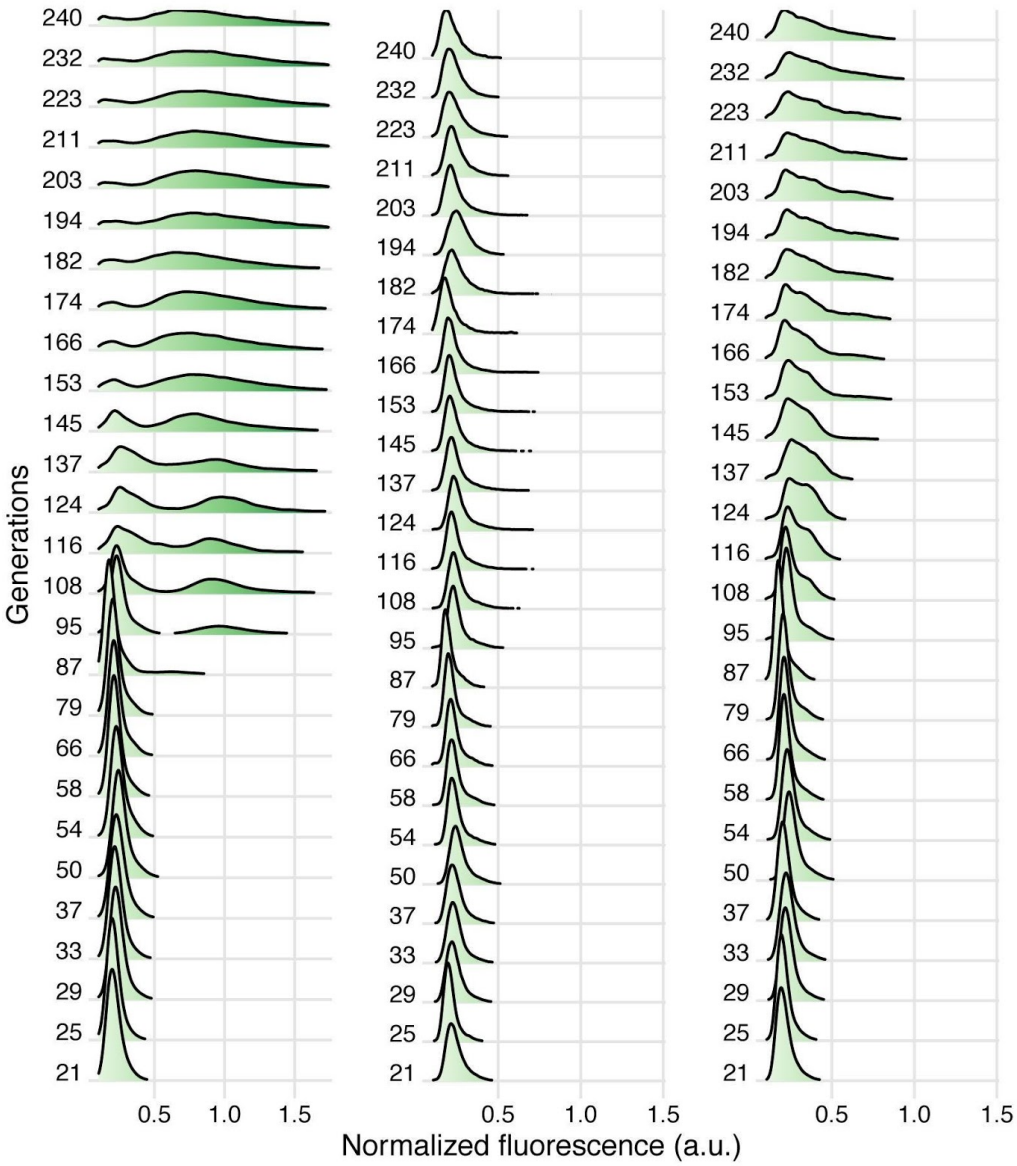


**Figure 3.4**. **Increases in fluorescence for the** HXT6/7-mCitrine **CNV reporter suggest the presence of** HXT6/7 **amplifications in glucose-limited populations.** Distributions of single-cell fluorescence over time for three independent glucose-limited experimental populations. Fluorescent signal is normalized by forward scatter, which varies as a function of cell size. Each distribution is based on 100,000 single cell measurements. The x-axis has been modified to remove a large population of non-fluorescing cells (see **Figure 3.6** for comparison).

Using fluorescence activated cell sorting (FACS), I fractionated the evolved population from Gluc1 into two distinct groups: those with a single copy of the CNV reporter, and cells with increased fluorescent signal that have two, or more, copies of the HXT6/7 locus. To confirm the purity of the sorted population, fractionated samples were analyzed using an Accuri flow cytometer (**Figure 3.5**). Importantly, the CNV subpopulation was isolated with high accuracy, and shows our ability to enrich for single cells with 2 or more copies of the reporter.

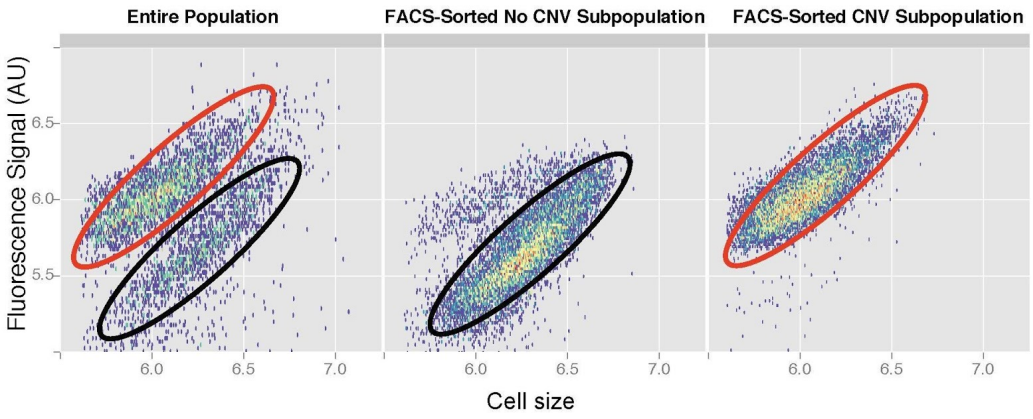


**Figure 3.5**. **Fluorescence activated cell sorting (FACS) can be used to fractionate the subpopulation containing CNVs from heterogeneous evolving populations**. An Accuri flow cytometer was used to determine the fluorescence of 100,000 fractionated cells from each sample sorted using FACS. Samples were fractionated from evolved population Gluc1.

While these experiments demonstrate the utility of the method for monitoring CNV dynamics and isolating subpopulations of CNV-containing cells, we identified additional concerns. Surprisingly, there were large populations of non-fluorescing cells in all 3 glucose-limited populations, which could be indicative of a high rate of loss of the fluorescent reporter gene (**Figure 3.6; see discussion**).

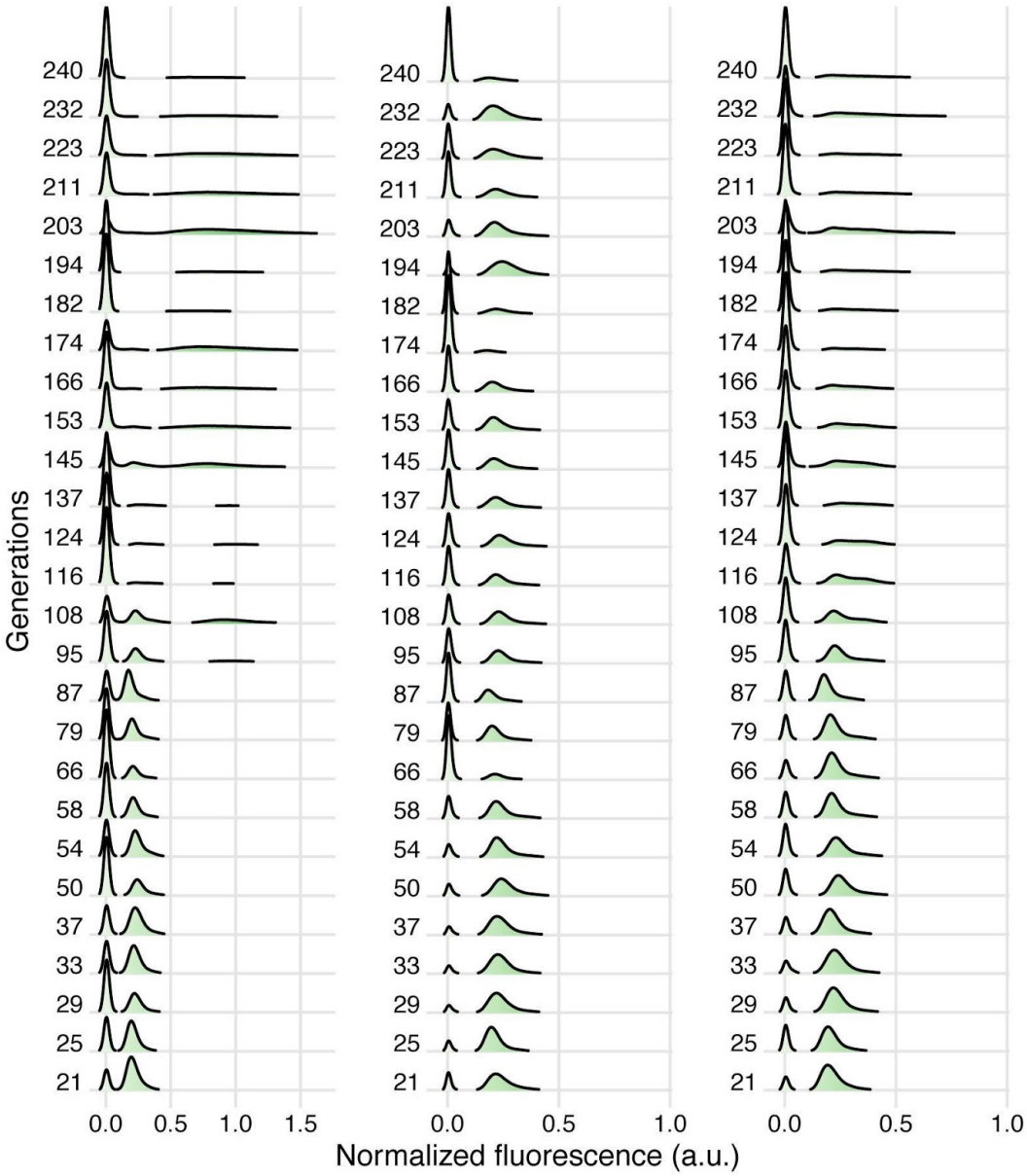


**Figure 3.6**. **Large populations of non-fluorescing cells are maintained throughout the duration of glucose-limited experimental evolution.** Despite increases in fluorescence indicating the presence of HXT6/7 CNVs**,** a subpopulation of non-fluorescing cells is present at high frequency. This figure presents the same data as **Figure 3.4**, but with an unmodified x-axis. Distributions of single-cell fluorescence over time for three independent glucose-limited experimental populations. Fluorescent signal is normalized by forward scatter, which varies as a function of cell size. Each distribution is based on 100,000 single cell measurements.

To summarize CNV dynamics, the median fluorescence of each sample is plotted by time point. However, plotting this value for the glucose-limited population yielded trajectories with large fluctuations in fluorescence at alternating time points (**Figure 3.7A**). Since this phenomenon was somewhat mitigated by plotting the mean fluorescence at each time point, it indicates that there were cells with extreme outlying fluorescence values. In addition, after generation 175, the median forward scatter more than doubled, indicating significant changes to cell size (**Figure 3.7B**). Cell aggregation, also called flocculation, is common in chemostat experiments and may partially explain increases in forward scatter (Hope et al. 2017). We discuss these technical challenges and other issues further below. Since increases in fluorescence were observed many generations earlier than the increases in forward scatter, we believe that these anomalies do not change our overall interpretation of CNV formation at this locus.

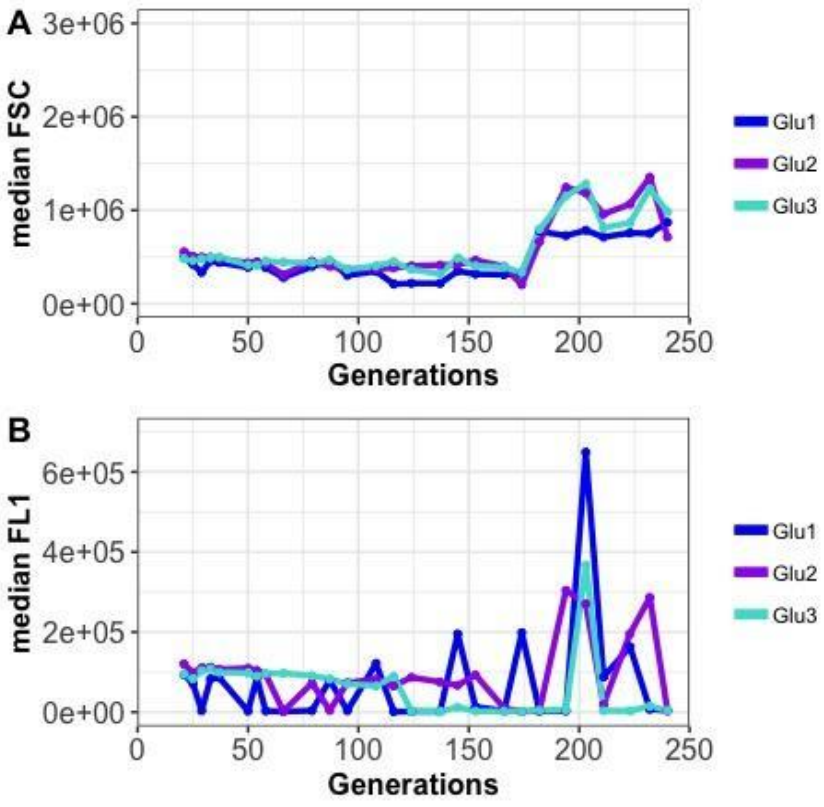

**Figure 3.7**. **Glucose-limited populations experience extreme fluctuations in forward scatter and fluorescent signal.** For each glucose-limited population, the median forward scatter (A) and median fluorescence (B) are plotted at each time point sampled during glucose limitation. Forward scatter is typically used as a proxy for cell size.

### 3.3.5: CNV dynamics in glutamine-limited chemostats

In addition to the experiments performed above, I also inoculated the GAP1-mCitrine strain into three glutamine-limited chemostats during a separate round of adaptive laboratory evolution. We previously used the GAP1-mCitrine reporter for experiments in smaller format miniature chemostats or “ministats” (Lauer et al. 2018). These evolution experiments were conducted in large-format, standard chemostats: populations were maintained in glutamine-limited media on continuous mode (dilution rate = 0.12 culture volumes/hr; population doubling time = 5.8 hours) for either 219 or 182 generations. Flow cytometry was used to measure fluorescence of 100,000 cells per sample. For all three replicates, increases in fluorescence were identified between generation 83 and 95, suggesting the emergence of de novo CNVs at the GAP1 locus (**Figure 3.8**). These dynamics are indicative of striking parallelism among populations. However, the subsequent dynamics of CNVs were markedly different for each population: GAP1 CNVs in replicate Gln2 decrease in frequency after generation 112 and are subsequently outcompeted, while GAP1 CNVs are maintained for the duration of the experiment in Gln1 and Gln3. These data are consistent with our previous report (Lauer et al. 2018).

### 3.3.6: Comparative dynamics across loci

While only the HXT6/7-mCitrine glucose-limited and MEP2-mCitrine ammonium-limited experiments were performed concurrently, additional comparisons can be performed across loci.

Increases in fluorescence indicative of the emergence of de novo CNVs were identified for both the GAP1 and HXT6/7 reporter strains. GAP1 amplifications may occur earlier than HXT6/7 amplifications, suggesting there is a higher rate of CNV formation at the GAP1 locus. HXT6/7 populations have larger increases in fluorescent signal, indicative of higher copy number (**Figure 3.9**). In both cases, 2/3 replicates maintained increased fluorescent signal for the duration of the experiment. Additional replicate experiments should be performed simultaneously in both conditions to determine if there are any statistically significant differences in GAP1 and HXT6/7 CNV dynamics.

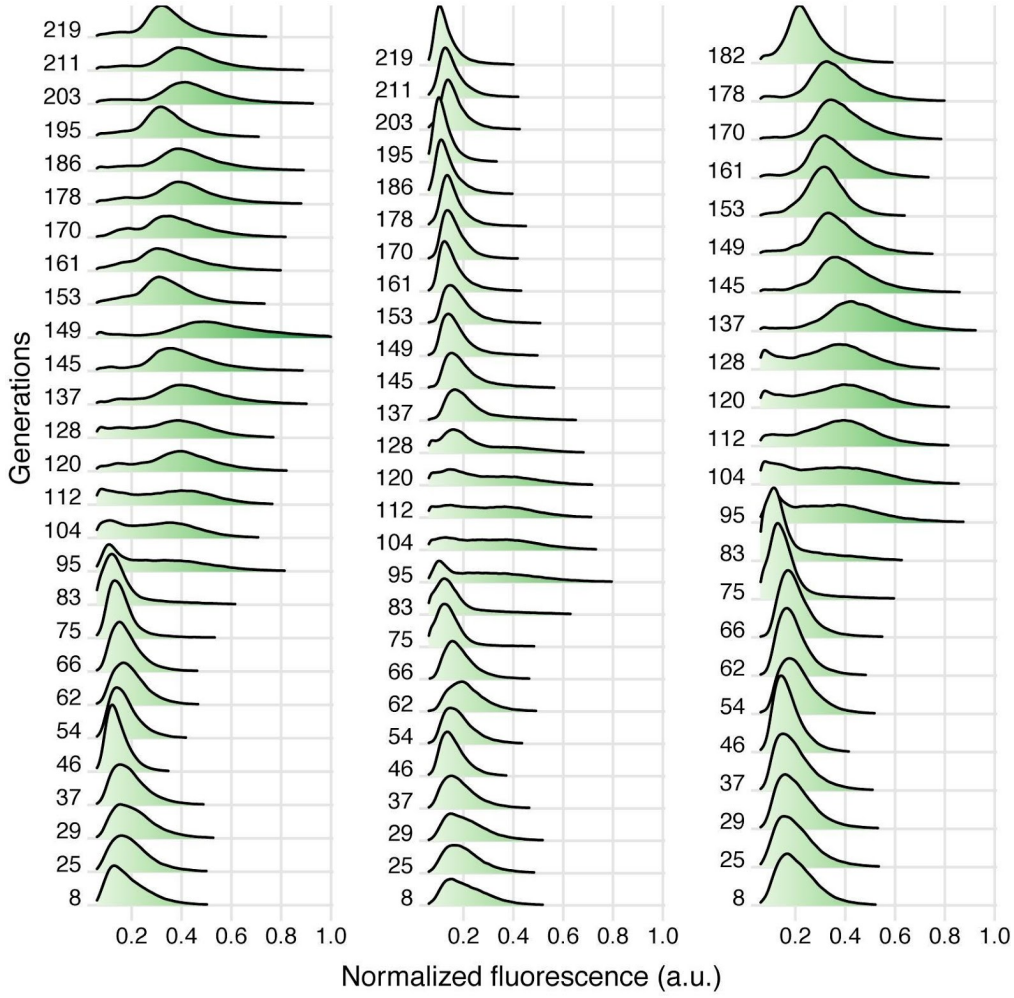

**Figure 3.8. The** GAP1 **CNV reporter indicates the emergence of** GAP1 **CNVs in all glutamine-limited populations.** Distributions of single-cell fluorescence over time for all glutamine-limited experimental populations. Fluorescent signal is normalized by forward scatter, which varies as a function of cell size. Each distribution is based on 100,000 single cell measurements.

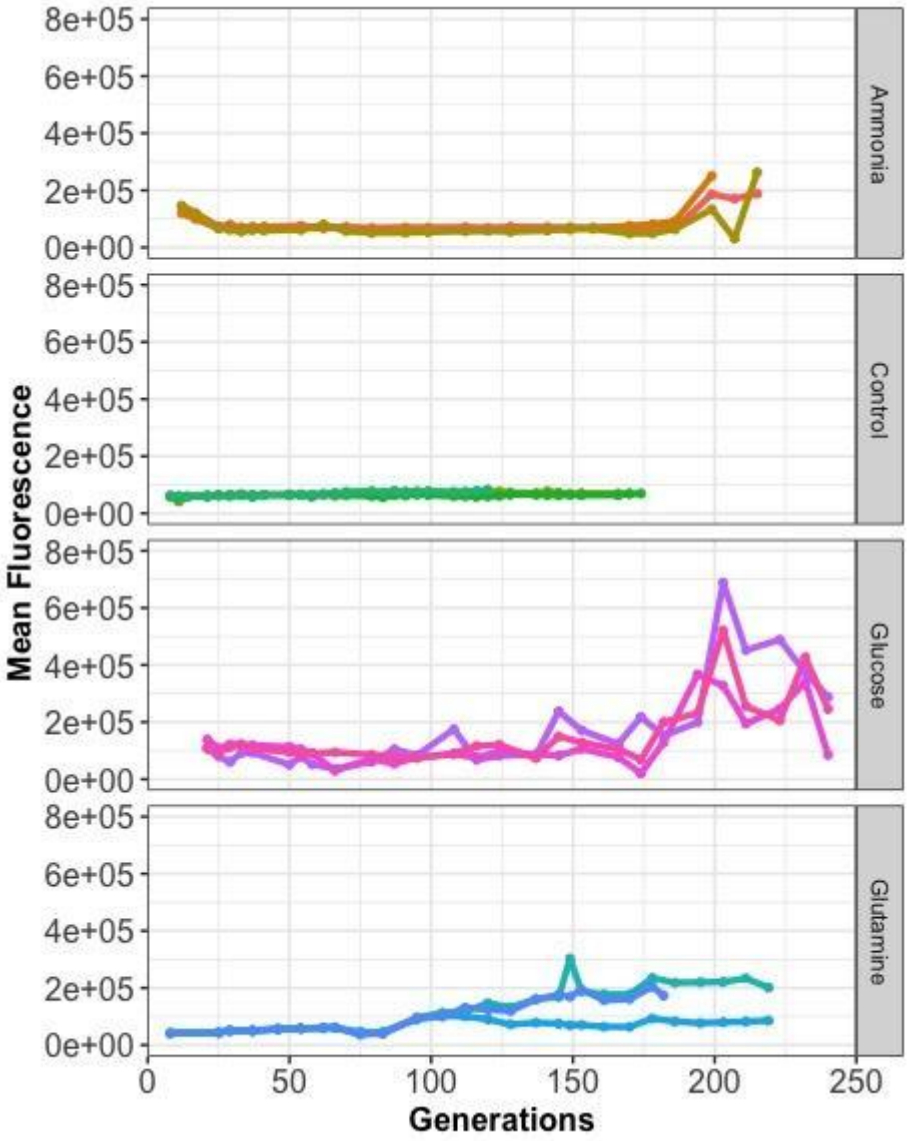


**Fig 3.9. Comparative CNV dynamics across all evolving populations.** Mean fluorescence over time for each population evolving in ammonium (MEP2 reporter, n=3), ammonium (HO reporter/control, n=3), glucose (HXT6/7 reporter, n=3) and glutamine (GAP1 reporter, n=3) limitation. Fluorescence signal is not normalized by forward scatter. Any observed increases in fluorescence for ammonium-limited populations are due to changes in forward scatter (see **Figure 3.10**).

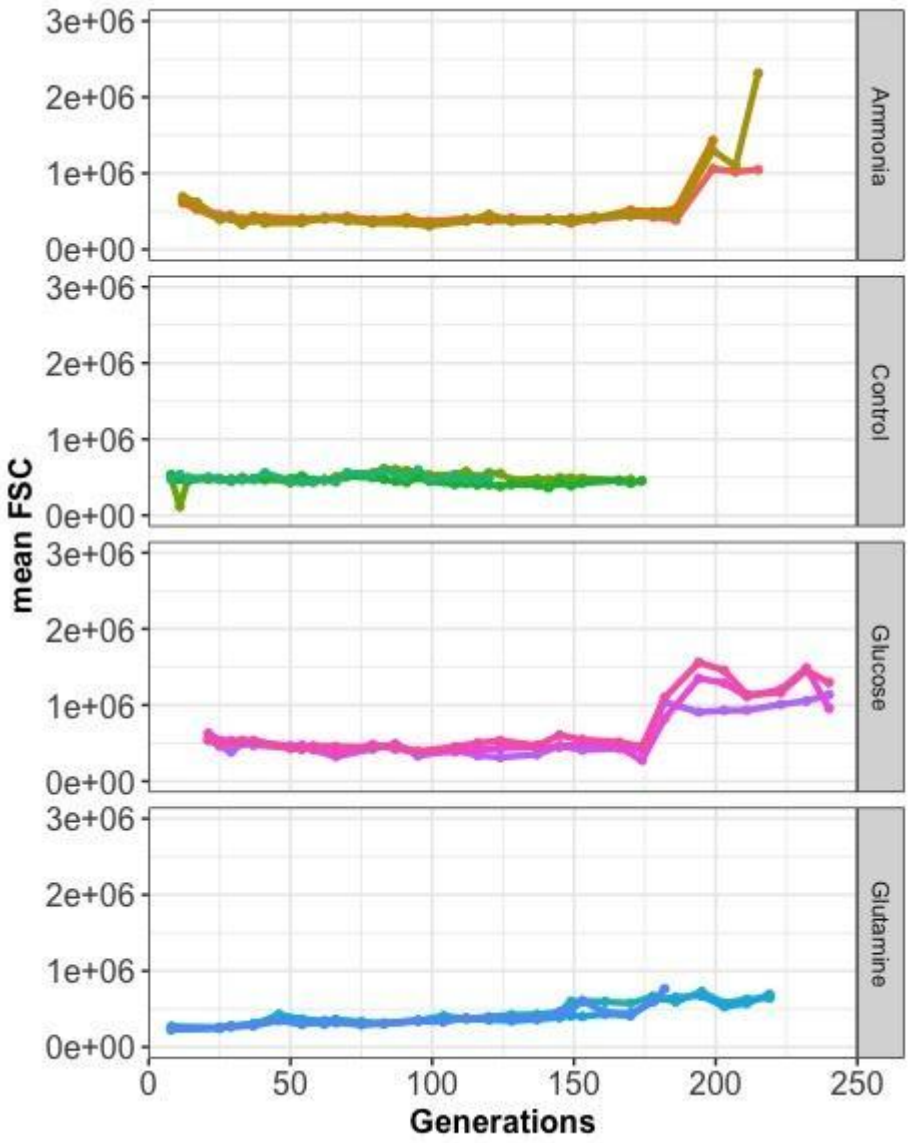


**Fig 3.10. Changes to forward scatter during adaptive evolution in glucose- and ammonium- limited chemostats.** Mean forward scatter over time for each population evolving in ammonium (MEP2 reporter, n=3), ammonium (HO reporter/control, n=3), glucose (HXT6/7 reporter, n=3) and glutamine (GAP1 reporter, n=3) limitation. Normalization by forward scatter is typically used to mitigate effects of cell physiology and morphology variation on CNV reporter signal. However, the dramatic FSC increases observed here are potentially indicative of diploidization or systemic changes during an evolution experiment.

As described above, glucose-limited populations experience a dramatic increase in forward scatter by generation 182. A similar increase in forward scatter was observed for the

ammonium-limited populations by generation 199 **(Figure 3.10)**. These experiments were performed at the same time using the same chemostat devices, and these changes to forward scatter could therefore be the result of systematic changes to the chemostat devices themselves (see **discussion**).

## 3.4: Discussion

The results presented here are the first characterization of CNV dynamics across different loci and nutrient-limited environments. Determining the dynamics with which CNVs are generated and selected in these diverse conditions can provide further insight into the role of genomic context and selective condition on CNV formation.

CNVs are formed at the GAP1 and HXT6/7 loci during glutamine- and glucose-limitation, respectively. Amplifications were identified at the GAP1 locus in all three replicate populations, and as a result, may form at a higher rate than HXT6/7 amplifications. While we lacked the proper controls to directly quantify the proportion of cells with 2, 3, and more copies, it is possible that HXT6/7 amplifications were present in higher copy number than GAP1 amplifications since we identified larger and more variable increases in fluorescence. These data could reflect fundamental differences in CNV generation and selection at these loci.

We originally hypothesized that GAP1 CNVs occur primarily through non-allelic homologous recombination (NAHR) between flanking Ty1 LTRs, but subsequent studies revealed that GAP1 CNVs are generated through diverse mechanisms (Lauer et al. 2018). HXT6 and HXT7 share 99.7% sequence identity and non-allelic homologous recombination (NAHR) between the two

loci has been reported (Brown, Todd, and Rosenzweig 1998). Repeated recombination events that further increase copy number could therefore be driving the large and variable increases in fluorescence. Additional replicate experiments for both reporter constructs and conditions (performed simultaneously) are required to determine if there are any statistically significant differences in CNV formation between the two loci. While whole genome sequencing was performed for Glu1, it was relatively uninformative since we used a whole-population sample (where any increases in read depth would be obscured by the high frequency of non-fluorescing/non-CNV containing cells). Genome sequencing of isolated clones would provide additional information about differences in CNV allele size, CNV copy number, and any differences in the molecular mechanisms underlying CNV formation.

As discussed in Chapter 2, we previously used the GAP1-mCitrine reporter to perform evolution experiments in glutamine-limited chemostats (Lauer et al. 2018). While the earlier experiments were performed in miniature chemostats (20mL media instead of 300mL media), all other parameters were consistent. Direct comparisons between these experiments are not possible (we did not use 1 and 2 copy controls in the large-format chemostats), but the overall dynamics we observed are consistent. CNV dynamics include an early, reproducible stage where CNVs form in all replicate populations, and a stochastic phase where dynamics are more variable and CNVs sometimes decrease in frequency. While CNVs may have emerged slightly earlier in the miniature chemostats, this could also be due to more accurate CNV detection enabled by the use of matched control populations. Importantly, these results are an independent verification of the findings from Chapter 2, further supporting the early increase in GAP1 CNVs during adaptive evolution in glutamine-limited chemostats.

We did not identify increases in fluorescence consistent with adaptive MEP2 CNVs in any of the three ammonium-limited chemostats. While it is possible that MEP2 CNVs formed but were quickly outcompeted from the population, we find this scenario unlikely. Previous results from our laboratory show that MEP2 amplifications are common and have large fitness benefits (Hong et al. 2018). In this study, MEP2 amplifications were only detected after genome sequencing at generation 250 (Hong et al. 2018), which could indicate that they do not form at a high rate, and that our experiments were too short-term to observe them. Unlike GAP1 and HXT6/7, MEP2 is not adjacent to any large repetitive sequences that facilitate NAHR, and as a result, CNV formation may proceed by an alternative mechanism (such as a microhomology mediated mechanism) that does not occur as readily. Additionally, since the MEP2 reporter is ~1 kilobase upstream of the gene, it would not capture partial MEP2 CNVs or small CNVs, which potentially form at this locus (personal communication with Farah Abdul-Rahman). Redesign of the CNV reporter could facilitate more accurate identification of small MEP2 CNVs and is ongoing in the laboratory. Further studies are needed to determine the frequency of CNV formation at this locus and the dynamics with which CNVs are selected and maintained in ammonium-limited chemostats.

Glucose-limited populations had a high proportion of non-fluorescing cells and cells with extreme or outlying fluorescent values. Frequent and rapid NAHR between HXT6 and HXT7 could result in high rates of loss for the reporter gene (personal communication with Grace Avecilla). In addition, we observed increases in forward scatter for all three glucose-limited populations consistent with significant changes to cell size. Flocculation can lead to cell-cell adhesion or cell-surface adhesion and biofilm formation; this is a common adaptive strategy in

the chemostat as it can prevent dilution of cells from the vessel (Hope et al. 2017). Flocculation also presents technical challenges as it can lead to clogs in outflow tubing. These unforeseen issues may have led to the extreme fluctuations in fluorescence and cell size we observed. To avoid these potential issues in the future, we can perform evolution experiments in a FLO1 deletion background (Hope et al. 2017).

Cell size changes may have been systematic, as cells in the ammonium-limited chemostats running concurrently also exhibited increases in forward scatter. Uncontrolled changes to temperature and oxygen availability could cause cells in both media conditions to exhibit similar physiological changes. Another possibility is that cells in all the replicates simultaneously underwent diploidization, leading to concomitant increases in cell size. However, this seems like an unlikely explanation. The strains used during this evolution experiment (DGY1588 and DGY1592 in ammonium- and glucose-limited chemostats respectively) were later identified as strains that form petite colonies when plated. Petite colonies typically denote respiratory deficiency (i.e. cells cannot metabolize non-fermentable carbon sources). Therefore, any issues observed during this laboratory evolution experiment could be the result of using strains with defective mitochondria. These strains have since been back-crossed to the wild-type strain (FY5).

While the experiments described here provide the first evidence for fundamental differences in CNV dynamics across loci and selective condition, further studies are required. In order to directly test the role of genomic context on CNV formation, we plan to re-engineer the surrounding features of GAP1, MEP2 and HXT6/7. For example, we could remove the Ty1 LTRs

and the ARS from the GAP1-mCitrine strain and the ARS from the HXT6/7-mCitrine strain while inserting repetitive elements near MEP2 (**Figure 3.11**). Evolution experiments would be repeated with these modified strains and compared directly to the dynamics observed for unmodified strains. By re-engineering these loci, we will test how repetitive sequence elements and replication origins influence the rate of CNV formation. Genome sequence analysis in the presence or absence of these elements will provide insights into the role of various mechanisms in generating CNVs. Understanding the rate at which CNVs are formed, the dynamics with which they are selected, and how they interact with other adaptive variants will provide critical insight into this complex class of large-effect alleles. These findings will have broad implications for addressing evolutionary dynamics and mechanisms of adaptive evolution in natural populations, and may lead to a better understanding of how CNVs contribute to phenotypic diversity and human disease.

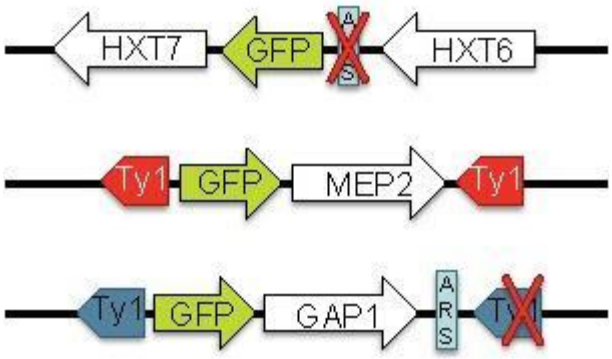


**Fig 3.11. Schematic depiction of potential modifications to** HXT6/7**,** MEP2 **and** GAP1 **loci.** Re-engineering these loci permits direct testing of the hypothesis that genomic context contributes to the rate and mechanisms underlying CNV formation.

## 3.5: Materials and Methods

We used FY4, a haploid derivative of the reference strain S288c, for all experiments. To generate fluorescent strains, we performed high efficiency yeast transformation (Gietz and Schiestl 2007) with an mCitrine gene under control of the constitutively expressed ACT1 promoter (ACT1pr::mCitrine::ADH1term) and marked by the KanMX G418-resistance cassette (TEFpr::KanMX::TEFterm). The entire construct, which we refer to as the mCitrine CNV reporter, is 3,375 base pairs. To generate a one copy control strain, the mCitrine reporter was integrated at a neutral locus: HO (YDL227C) on chromosome IV. We constructed the MEP2 CNV reporter by integrating the mCitrine construct at an intergenic region 1,171 base pairs upstream of MEP2 (integration coordinates, chromosome XIV: 356282). We constructed the HXT6/7 reporter by integrating the mCitrine construct at an intergenic region 1,662 base pairs upstream of HXT7 (integration coordinates, chromosome VI: 1157590). We constructed the GAP1 CNV reporter by integrating the mCitrine construct at an intergenic region 1,118 base pairs upstream of GAP1 (integration coordinates, chromosome XI: 513945). PCR and Sanger sequencing were used to confirm integration of the CNV reporter at each location. For the GAP1-mCitrine strain used in this study, transformants were subsequently backcrossed and sporulated, and the resulting segregants were genotyped.

Nitrogen limiting media (glutamine and urea limitations) contained 800 μM nitrogen regardless of molecular form and 1 g/L $CaCl_2$-$2H_2O$, 1 g/L of NaCl, 5 g/L of $MgSO_4$-$7H_2O$, 10 g/L $KH_2PO_4$, 2% glucose and trace metals and vitamins as previously described (Hong and Gresham 2014). Glucose limiting media contained 0.08% glucose, 1 g/L $CaCl_2$-$2H_2O$, 1 g/L of NaCl, 5 g/L of

$MgSO_4$-$7H_2O$, 10 g/L $KH_2PO_4$, 50g/L $(NH_4)_2SO_4$ and trace metals and vitamins (Brauer et al. 2008).

We inoculated the CNV reporter strains, DGY1588, DGY1592, and DGY1657 into 300mL chemostat vessels containing either ammonium-, glucose, or glutamine-limited media, respectively. The control strain with the CNV reporter at the HO locus was inoculated into ammonium-limited chemostats. Chemostats were maintained at 30°C in aerobic conditions and diluted at a rate of 0.12 $hr^{-1}$ (corresponding to a population doubling time of 5.8 hours). Steady state populations were maintained in continuous mode for a range of generations, between 120-240 generations, depending on condition. Typical population sizes were 1.5 x $10^{10}$ cells in ammonium-, 3 x $10^9$ cells in glucose-, and 6 x $10^9$ cells in glutamine-limited chemostats. Every 30 generations, we archived 1 mL population samples at -80°C in 15% glycerol.

To monitor the dynamics of CNVs, we sampled 1mL from each population every ~4-12 generations. We stored samples in phosphate-buffered saline at 4°C for up to 5 days before performing the analysis. We used sonication to disrupt any cellular aggregates and subsequently analyzed the samples on an Accuri flow cytometer, measuring 100,000 cells per population for mCitrine fluorescence signal (excitation = 516nm, emission = 529nm, filter = 514/20nm), cell size (forward scatter) and cell complexity (side scatter). We quantified fluorescence for each cell and divided this value by the forward scatter measurement for the cell to account for differences in cell size.

# Chapter 4: Determining the rate of intrachromosomal recombination and GAP1 deletion

## 4.1: Abstract

Copy number variants (CNVs) are duplications and deletions of DNA sequence that can lead to rapid adaptation and maladaptive phenotypes such as cancer and other human diseases. Understanding the full range of processes that contribute to CNV formation in diverse systems is an understudied but important problem in biology. Previous work from our laboratory demonstrates that CNVs are generated at a high rate, and that there are multiple mechanisms underlying their formation. Here, we aim to dissect the contribution of one particular mechanism: non-allelic homologous recombination (NAHR). NAHR can occur when the wrong template is used for homologous repair, a process that occurs more frequently when DNA sequences share high levels of sequence identity. The general amino acid permease (GAP1) is a model locus for studying the role of NAHR in CNV formation because it has two flanking Ty1 long terminal repeats that can undergo recombination. Recombination between these sequences leads to excision of the GAP1 locus and the formation of a hybrid LTR. To determine the rate at which this particular event occurs, we constructed a reporter for intrachromosomal recombination and subsequent GAP1 deletion. We tested the efficiency of the reporter and find that it does not accurately reflect GAP1 deletions. We discuss the potential issues with our reporter system and suggest alternative strategies as well as additional follow-up experiments.

## 4.2: Introduction

Copy number variants are the substrate for myriad evolutionary processes. Seminal contributions by Susumo Ohno and others first described the importance of gene duplication for generating evolutionary novelty and diversification through modification of existing heritable material (Ohno 1970; Michael Lynch and Conery 2000; Austin L. Hughes 1994; R. P. Anderson and Roth 1977). Now that detection of gene duplications and deletions from genome sequence data has vastly improved, we know copy number variants (CNVs) are extremely common in natural populations. In humans, de novo CNVs introduced each generation are more numerous than point mutations (Itsara et al. 2010), and among individuals, between 4.8-9.5% of the genome contains CNVs (Zarrei et al. 2015). CNVs are important sources of phenotypic diversification and adaptation in animals, plants, and microbes (Barreiro et al. 2008; Iskow et al. 2012; Clop, Vidal, and Amills 2012; Żmieńko et al. 2014; Greenblum, Carr, and Borenstein 2015).

A prevailing view in evolutionary biology is that mutations are randomly generated and then selection acts on them to dictate their ultimate fate in a population. However, recent studies suggest that CNV mutation is directly linked to DNA repair, replication, and transcription (Chen et al. 2015; Wilson et al. 2015; Thomas and Rothstein 1989; Skourti-Stathaki and Proudfoot 2014). Induction of replicative stress leads to increased formation of CNVs (Foster 2007; Galhardo, Hastings, and Rosenberg 2007; Shor, Fox, and Broach 2013). Active transcription units may be “hotspots” for CNV formation, as collisions between the replisome and RNA polymerase (as well as other transcription-mediated events) can lead to DNA damage and a

higher probability of improper DNA repair (Wilson et al. 2015). Gene duplication may generally occur in response to environmental stimuli, a process that has been observed at the rDNA and the CUP1 locus in yeast (Hull et al. 2017; Jack et al. 2015). In addition, nearby genomic features may contribute to genome instability at certain loci. For example, long terminal repeats and tRNA genes can lead to fork stalling (Labib et al. 2007; Bermudez-Santana et al. 2010; Di Rienzi et al. 2009; Tran et al. 2017). Further empirical tests are needed to disentangle the respective contributions of transcription and selective environment on stimulating CNV formation.

CNVs can be generated by a variety of mechanisms, but are often preceded by formation of DNA double strand breaks (DSBs). Damaging exogenous agents, including ionizing radiation and toxic chemicals, can lead to DSBs (Mehta and Haber 2014). Spontaneous DSBs can also occur during replication after the formation of atypical DNA structures, collisions with the transcription machinery, or collisions with transcription factors and other DNA-binding proteins (Mehta and Haber 2014). Spontaneous DSBs have been estimated at one DSB per $10^8$ base pairs (Vilenchik and Knudson 2003; Coïc et al. 2008).

DSBs are routinely repaired by homologous recombination and do not result in CNV formation. However, non-allelic homologous recombination (NAHR) can generate CNVs when the incorrect template is used for repair, which is often an outcome for repetitive sequences such as long terminal repeats (LTRs). NAHR frequency depends on the extent of homology, the distance between sequences, and sequence length (Dittwald et al. 2013; Peng et al. 2015). NAHR can occur between sequences on sister chromatids, homologous chromosomes, and even

sequences on the same chromatid (known as intrachromosomal recombination).

Recurrent CNVs, which repeatedly occur in specific regions of the genome, typically underlie re-occurring germline mutations and human disease (Itsara et al. 2009; Girirajan, Campbell, and Eichler 2011). Formation of recurrent CNVs is driven by NAHR between repetitive sequence elements such as segmental duplications, also known as low copy repeats (LCRs), which are >1 kilobase in size and have >90% sequence identity (Sharp et al. 2005). Segmental duplications and other repetitive regions can act as "hotspots," generating spontaneous CNVs at the same site in different individuals. Several microduplication and deletion syndromes resulting from NAHR have been characterized in humans including Williams-Beuren, Prader-Willi, Smith-Magenis and Potocki-Lupski syndromes and Charcot-Marie-Tooth disease, among others (Lupski 2009).

The gene encoding the general amino acid permease, GAP1, in Saccharomyces cerevisiae is a model locus for understanding the role of NAHR in CNV formation. We have previously shown that GAP1 is amplified under amino acid limitation and deleted under urea limitation (Gresham et al. 2010; Hong and Gresham 2014). We have determined that GAP1 CNVs can be generated by a diverse range of processes (Lauer et al. 2018), one of which is NAHR between flanking Ty1 LTRs, YKRCδ11 and YKRCδ12. There is 85% homology between YKRCδ11 and YKRCδ12, which are each 300 base pairs in length. Intrachromosomal recombination between these elements is hypothesized to result in excision of a $GAP1^{circle}$ and the formation of a hybrid YKRCδ11/YKRCδ12 LTR. Since NAHR occurs recurrently between these flanking LTRs (Lauer

et al. 2018), GAP1 is also an ideal locus for further dissecting the potential effects of transcription and environment on CNV stimulation.

To determine the spontaneous rate of intrachromosomal recombination and GAP1 deletion, we constructed a GAP1 deletion reporter. We inserted a URA3 gene ~1 kilobase upstream from the GAP1 coding sequence and hypothesized that any deletion event occurring between YKRCδ11 and YKRCδ12 would subsequently result in loss of the URA3 gene. Loss-of-function (LOF) in the URA3 gene can then be identified after plating cells on 5-fluoroorotic acid (5-FOA), which is converted to toxic 5-fluorouracil by the URA3 gene product, orotidine-5'-phosphate decarboxylase. To distinguish between GAP1 deletion and URA3 inactivation by other methods (e.g. point mutations or indels), we sought to confirm simultaneous GAP1 and URA3 LOF. The results of this study were largely inconclusive, as we were unable to confirm GAP1 LOF for the majority of isolated ura3- mutants. We highlight several potential issues with our approach and discuss improvements to the method as well as future experiments.

## 4.3: Results

### 4.3.1: Creation of a reporter for the GAP1 deletion

GAP1 is flanked by two Ty1 LTRs: YKRCδ11 and YKRCδ12. These LTRs share 85% sequence identity, making them an effective substrate for NAHR. We have previously shown that NAHR between YKRCδ11 and YKRCδ12 can lead to both tandem duplications and deletions of the GAP1 gene (Lauer et al. 2018). GAP1 deletion may be accompanied by excision of a self-propagating $GAP1^{circle}$. After GAP1 deletion or excision, a hybrid LTR remains in the genome: YKRCδ11/YKRCδ12. To determine the role of NAHR in CNV formation at the GAP1

locus, I sought to generate a reporter strain for the specific deletion event that removes all intervening sequence between YKRCδ11 and YKRCδ12, including GAP1 (indicated as gap1Δ$^{LTR-LTR}$). The principle and design of the reporter is depicted in **Figure 4.1.**

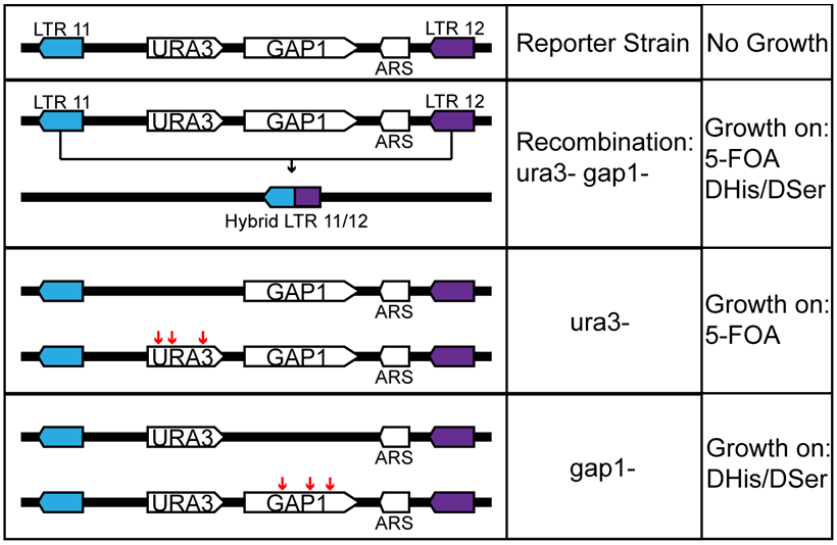


**Fig 4.1. Schematic representation of experimental design for a** GAP1 **deletion reporter.** A URA3 gene is inserted ~1 kilobase upstream of the GAP1 open reading frame. Intrachromosomal recombination between the LTRs YKRCδ11 and YKRCδ12 results in formation of a hybrid LTR and simultaneous deletion of both URA3 and GAP1. Spontaneous NAHR between YKRCδ11 and YKRCδ12 should therefore lead to the formation of mutant colonies that grow on 5-FOA and D-Histidine/D-Serine media. However, additional mutational events can occur (red arrows), rendering either URA3 or GAP1 nonfunctional. Bonafide gap1Δ$^{LTR-LTR}$ mutants occurring as a result of intrachromosomal recombination should therefore be ura3- and gap1-.

To generate the GAP1 deletion reporter, I inserted a URA3 gene 1,118 base pairs upstream of the GAP1 start codon (integration coordinates, ChrXI: 513945; **Figure 4.1**). The transformation was performed in a strain background where the endogenous open reading frame for the URA3 gene is deleted and replaced by the KanMX G418 resistance cassette (DGY1428). Five independent deletion reporter strains were constructed: two by transformation (DGY2010 and DGY2011), and three by transformation with a subsequent backcross to FY4, sporulation, and tetrad dissection analysis (DGY2012, DGY2013, and DGY2014). Insertion of URA3 was

confirmed by PCR. The strains generated here and all other strains used as controls in this study are listed in **Table 4.1.** This panel of strains was used to perform all the assays described below.

**Table 4.1. Yeast strains used in Luria Delbrück fluctuation assays.**

| Strain | Use | Growth on 5-FOA | Growth on D-His/D-Ser |
|---|---|---|---|
| FY4 (DGY1, wild-type) | Negative control | No | No |
| DGY41 (ura3-52) | Positive control | Yes | No |
| DGY1428 (ura3Δ0) | Positive control | Yes | No |
| DGY138 (gap1Δ) | Positive control | No | Yes |
| DGY1764 (gap1Δ$^{LTR\text{-}LTR}$) | Positive control | No | Yes |
| DGY2010 | Reporter | No | No |
| DGY2011 | Reporter | No | No |
| DGY2012 | Reporter | No | No |
| DGY2013 | Reporter | No | No |
| DGY2014 | Reporter | No | No |
| DGY2041 (gap1Δ0) | Positive control | No | Yes |
| DGY2047 (gap1Δ$^{LTR\text{-}LTR}$) | Positive control | No | Yes |

## 4.3.2: Determining spontaneous URA3 mutation rate in the reporter strain

To determine the spontaneous frequency of GAP1 deletion, we used Luria Delbrück fluctuation assays where cells are grown in rich, non-selective media, than plated to 5-fluoroorotic acid (5-FOA). 5-FOA is converted to toxic 5-fluorouracil by the URA3 gene product,

orotidine-5'-phosphate decarboxylase. Plating on 5-FOA selects for cells with URA3 loss of function (LOF), which should include all URA3 gene deletions generated concurrently with intrachromosomal recombination and GAP1 deletion (**Figure 4.1**). GAP1 deletions can be confirmed by plating on D-Histidine/D-Serine media (D-His/D-Ser). GAP1 LOF confers resistance to toxic D-amino acids by preventing their uptake. A $ura3^{-}$ $gap1^{-}$ phenotype (D-His/$DSer^{R}$ and 5-$FOA^{R}$) should therefore specifically report on GAP1 deletions that encompass the entire genomic region flanked by Ty1 LTRs ($gap1\Delta^{LTR\text{-}LTR}$). This phenotype is important for distinguishing between deletion events resulting from intrachromosomal recombination and inactivation of either URA3 or GAP1 by point mutations or indels.

Several important controls were used during this study (**Table 4.1**): two positive controls where GAP1 is inactivated and growth occurs on D-His/D-Ser (DGY138 and DGY1764) and two positive controls where URA3 is deleted and growth occurs on 5-FOA (DGY41 and DGY1428). The GAP1 deletion strains were confirmed to grow successfully on D-His/D-Ser media. While DGY138 has an inactivating frameshift mutation in GAP1, the deletion in DGY1764 includes GAP1 and all intervening sequence from YKRCδ11 to YKRCδ12 ($gap1\Delta^{LTR\text{-}LTR}$). The wild-type strain FY4 (DGY1) is an important negative control, as wild-type GAP1 and URA3 genes prevent growth on both D-His/D-Ser and 5-FOA. FY4 mutants growing on 5-FOA after the Luria Delbrück fluctuation assay report on the basal mutation rate for the endogenous URA3 gene.

All strains were grown overnight in 5mL of non-selective media (yeast extract-peptone-dextrose or YPD) at 30°C and then 1 million, 10 million, or 100 million cells were plated on 5-FOA. Colonies were counted after 3 days of growth at 30°C. Five independent replicate experiments

were performed, but data is only shown for four of these replicates (**Figures 4.2-4.5**; only 1 million cells were plated for each strain in the first replicate and very few colonies were counted). Controls with the URA3 gene deletion (DGY41 and DGY1428) consistently formed lawns on 5-FOA after overnight growth in YPD and colonies were not counted.

Unsurprisingly, the basal mutation rate for the endogenous URA3 gene is low, with the largest number of mutant colonies for FY4 occurring during replicate five: 47 mutant colonies/$10^8$ cells plated. Only a coarse estimate of the phenotypic mutation rate for 5-FOA resistance can be obtained from this data: ~3 x $10^{-8}$ mutations/generation. For the deletion reporter strains (DGY2010, DGY2011, DGY2012, DGY2013, and DGY2014), the amount of colonies growing on 5-FOA plates varied depending on the replicate. This is consistent with URA3 mutations occurring randomly during overnight growth, where an early mutation leads to a “jackpot” and higher total mutant colonies. In the majority of cases, each test strain formed more colonies than FY4 by an order of magnitude. This result demonstrates that the mutation rate for URA3 in the deletion reporter is higher than the mutation rate at the endogenous URA3 locus. Within replicates, colony counts increased proportionally with the amount of total cells plated (see **Figure 4.4**), indicating that the total cells plated on 5-FOA and the subsequent colony counts were performed accurately.

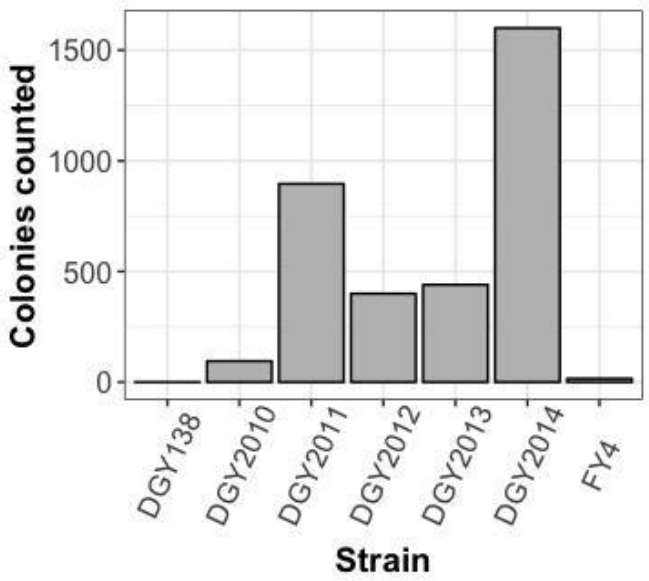


**Fig 4.2. Replicate two colony counts after performing a Luria Delbrück fluctuation assay.** Colonies were counted after plating $10^8$ cells on 5-FOA media.

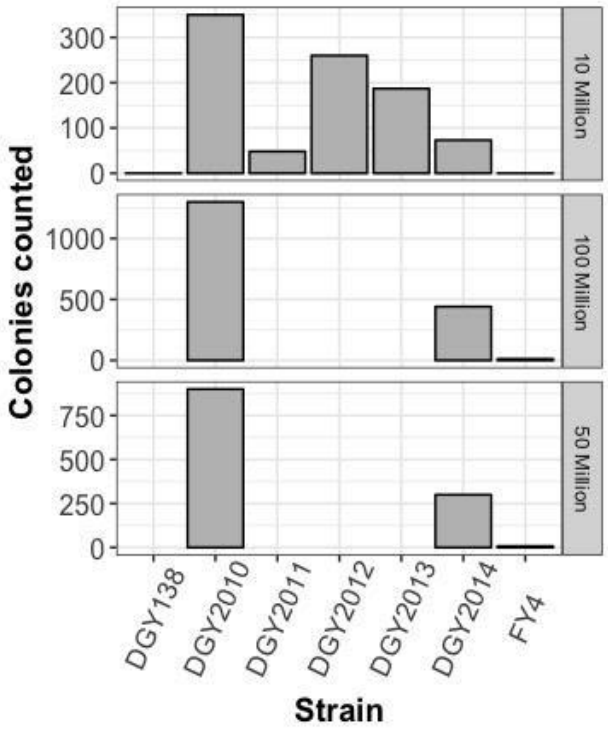


**Fig 4.3. Replicate three colony counts after performing a Luria Delbrück fluctuation assay.** Colonies were counted after plating $10^7$-$10^8$ cells on 5-FOA media. All strains were plated with 10 million cells, but only strains DGY2010, DGY2014, and FY4 were plated with 100 or 50 million cells.

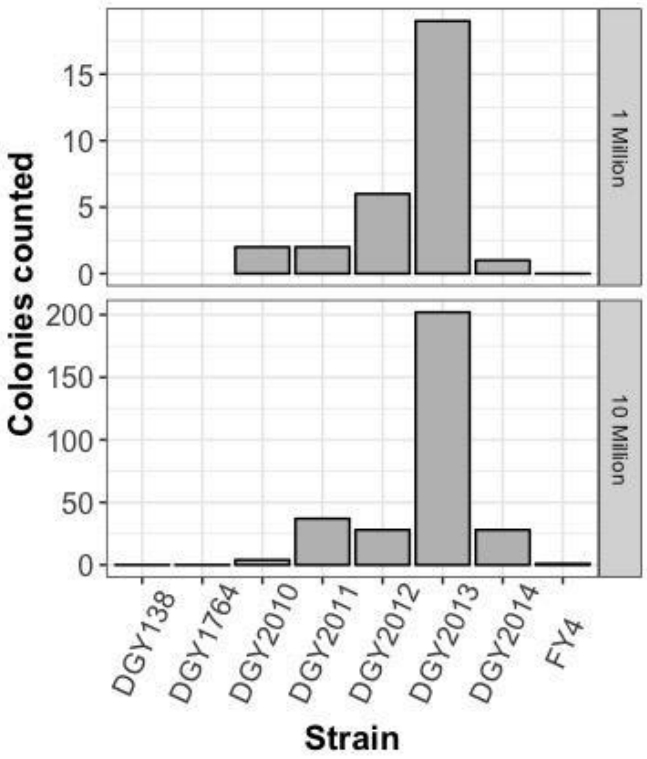


**Fig 4.4. Replicate four colony counts after performing a Luria Delbrück fluctuation assay.** Colonies were counted after plating between $10^6$-$10^7$ cells on 5-FOA media. DGY138 and DGY1764 were only plated with 10 million cells.

Across the first four replicates of the experiment, gap1Δ controls DGY138 and DGY1764 had a total of 0 mutant colonies on 5-FOA plates. This result was surprising, and suggests synthetic lethality between gap1 LOF and ura3 LOF. To rule out this possibility, we used two independent gap1Δ strains, DGY2041 and DGY2047 during the fifth and final replicate of the Luria Delbrück fluctuation assay. While the colonies that grew on 5-FOA were smaller overall, colony formation occurred at a similar rate to FY4 (**Figure 4.5**). These results demonstrate that there may be an interaction between GAP1 and URA3, but that this could also be a direct result of the specific strains that were used as positive controls (see **discussion**).

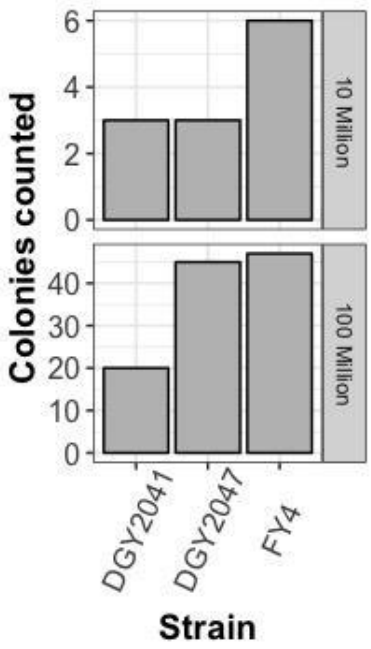


**Fig 4.5. Replicate five colony counts after performing a Luria Delbrück fluctuation assay.** Colonies were counted after plating between $10^7$-$10^8$ cells on 5-FOA media. Two independently generated gap1Δ strains (DGY2041 and DGY2047) were used as controls.

### 4.3.3: Assessment of reporter accuracy

Spontaneous LOF mutations in URA3 can also include point mutations, indels, and other mutations (**Figure 4.1**). To determine what fraction of ura3- mutants are bona fide GAP1 deletions (ura3- gap1-), we simultaneously assayed for GAP1 LOF by replica plating to D-histidine/D-serine media. Since D-amino acids are toxic to the cell, GAP1 LOF prevents their uptake, and cells can grow in their presence. Importantly, D-Histidine/D-Serine plates are composed of minimal media where proline is the only available nitrogen source. Proline is a poor nitrogen source and cells grow slowly, typically requiring longer periods of growth at 30°C (up to 5 days). Since D-His/D-Ser selection is not particularly strong, positive and negative controls must always be plated simultaneously. Secondly, in this Luria Delbrück fluctuation

assay, we specifically selected mutants that lose the ability to grow on uracil. During standard preparation of D-His/D-Ser media, there is no uracil present. To remedy this situation, D-His/D-Ser plates used for this assay must be supplemented with uracil.

During Luria Delbrück replicates two and three, I isolated a subset of 5-FOA$^{R}$ mutants. To confirm that these 5-FOA$^{R}$ mutants were also gap1Δ$^{LTR\text{-}LTR}$, I used a 3-primer PCR technique and Sanger sequencing. The 3-primer PCR results in a band regardless of genotype but band size depends on the presence or absence of the gap1Δ$^{LTR\text{-}LTR}$ allele (**Figure 4.6**). The 3-primer PCR assay confirmed a gap1Δ$^{LTR\text{-}LTR}$ genotype for a total of 3 independently generated mutants (for an example, see **Figure 4.7**). Sanger sequencing identified hybrid YKRCδ11/YKRCδ12 LTRs in the genome, with each mutant having distinct nucleotide breakpoints. While these results demonstrated that 5-FOA$^{R}$ could accurately report on intrachromosomal recombination and GAP1 deletion, we noticed a higher incidence of false positives than expected.

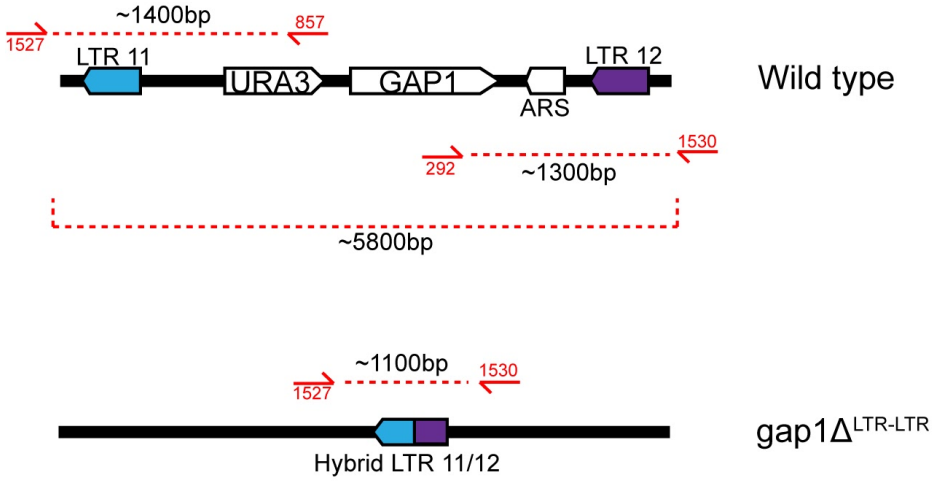


**Fig 4.6. Schematic representation of 3-primer PCR assay.** Flanking primers 1527 and 1530 are combined with primer 857 to assay for the presence of URA3 and primer 292 to determine the presence of GAP1. These are 2 separate PCR reactions. Primers 1527 and 1530 do not form a product unless GAP1 is deleted (when a 2 minute extension time is used) and the 2 PCR reactions both result in a ~1100 base pair band. However, if URA3 and GAP1 are present, the 2 PCR reactions will instead result in a ~1400 and ~1300 base pair band respectively.

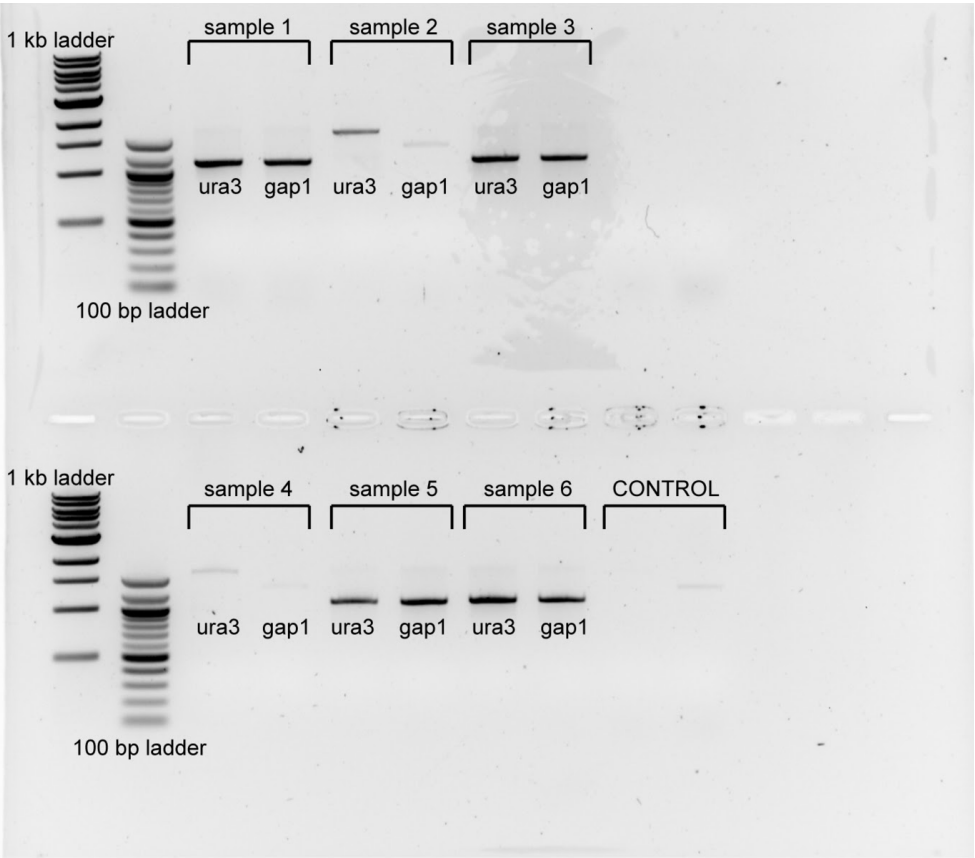


**Fig 4.7. Representative PCR result from Luria Delbrück replicate two.** Samples 1, 3, 5, and 6 have two bands of the same size, indicating the presence of the gap1$\Delta^{LTR\text{-}LTR}$ allele in these samples (**see Figure 4.6**). Sanger sequencing was used to confirm the genotype. Nucleotide resolution shows that the breakpoints for all four samples are identical. These clones are likely to be descended from a single event since they were isolated from the same plate. The control sample used here was a negative control: strain DGY2010.

To determine the proportion of 5-FOA$^R$ cells that are also gap1-, 30 total 5-FOA$^R$ mutants were isolated and grown overnight in YPD. These mutants were isolated from Luria Delbrück replicates four and five, and included 5-FOA mutants in the following strain backgrounds: reporter strains (DGY2010, DGY2011, DGY2012, DGY2013, and DGY2014), gap1Δ strains (DGY2041 and DGY2047), and FY4. I performed a 10-fold dilution series and plated $10^7$-$10^2$ cells on a variety of media conditions: rich media with a glucose carbon source (YPD), rich

media with a glycerol carbon source (YPG), synthetic complete media lacking uracil (sc-ura), 5-FOA media and D-His/D-Ser media. The 30 5-FOA mutants were plated across a series of 5 plates (5x5 or 25 total plates; **Fig 4.8-Fig 4.12**). Each plate included ura3Δ and gap1Δ strains that never underwent 5-FOA selection as controls. All control strains grew as expected on the various media conditions.

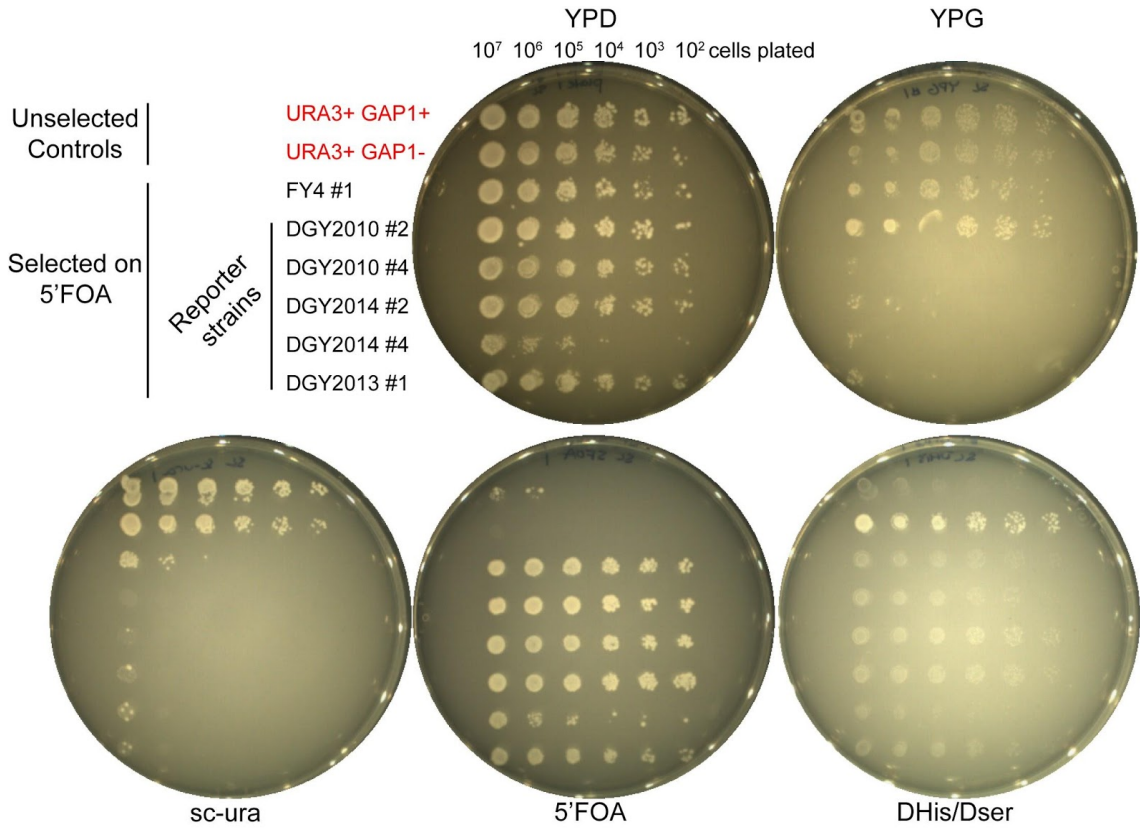


**Fig 4.8. Results of spot dilution assay for plate one.** Unselected controls (shown in red) grow as expected: growth occurs on sc-ura but not 5-FOA media. The gap1- control is the only strain that grows successfully on D-His/D-Ser media.

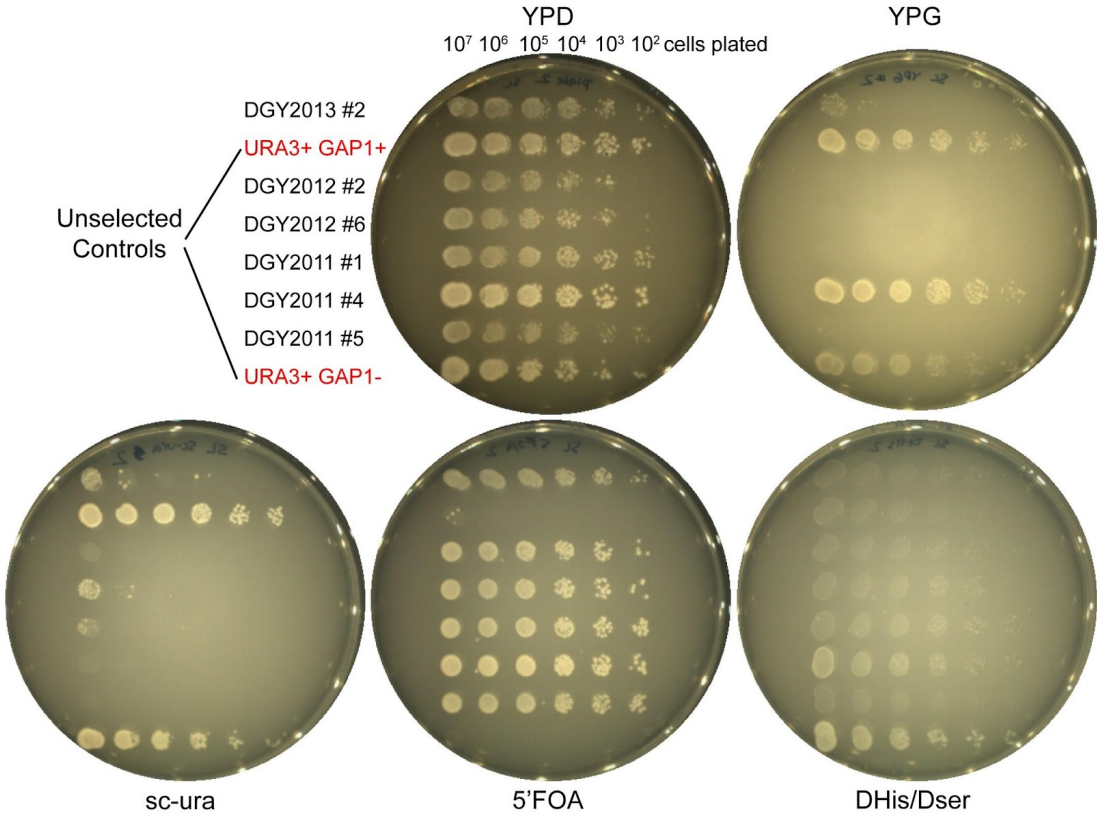


**Fig 4.9. Results of spot dilution assay for plate two.** Unselected controls (shown in red) grow as expected: growth occurs on sc-ura but not 5-FOA media. All other mutants were isolated after selection on 5-FOA plates. Mutant DGY2011 #4 grows similarly to the gap1- control on D-His/D-Ser media, indicating that it may be a bonafide ura3- gap1- mutant. Growth on D-His/D-Ser for mutant 2011 #4 was independently confirmed by replica plating (data not shown).

Of the 30 mutants originally selected on 5-FOA, 29/30 (97%) were confirmed to grow on 5-FOA. Since we specifically selected for URA3 LOF, these 29 mutants should all be ura3-, incapable of growing on sc-ura media. However, there were 4/29 (14%) cases in which 5-FOA$^R$ mutants grew on sc-ura media (see **Figure 4.10** for an example). All 4 cases occurred in gap1Δ strain backgrounds (DGY2041 and DGY2047). This result suggests that there is an additional mechanism enabling growth on 5-FOA media (see **discussion**).

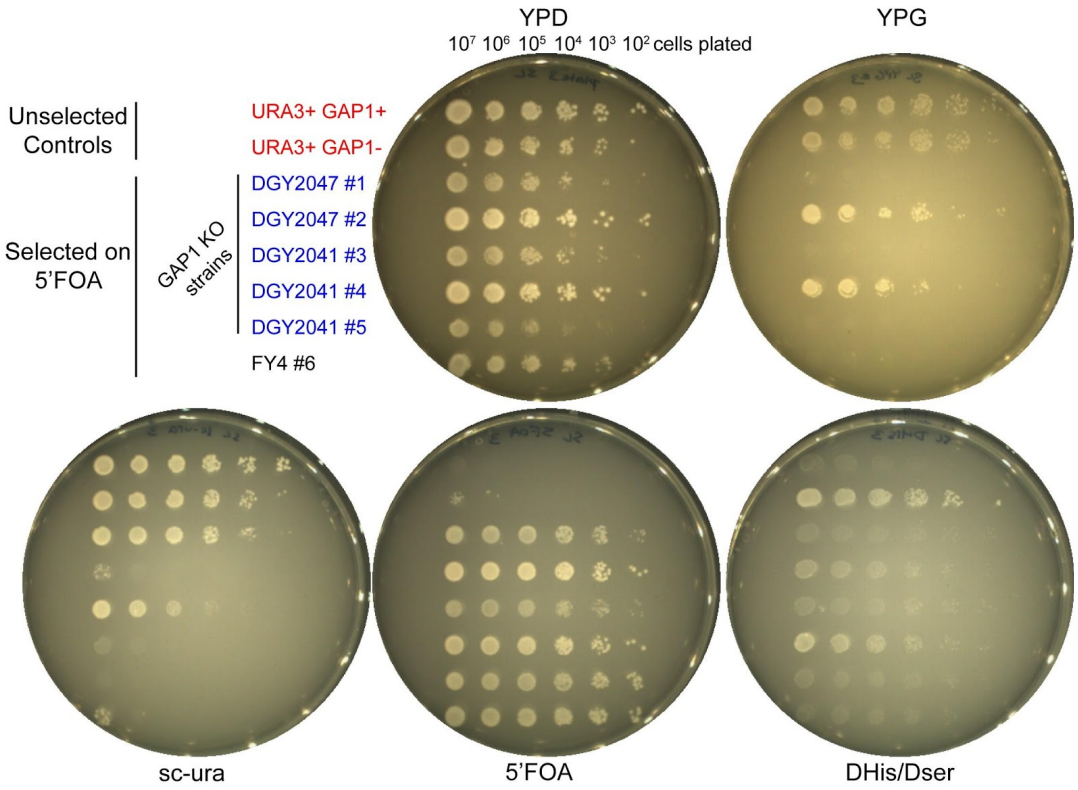


**Fig 4.10. Results of spot dilution assay for plate three.** Unselected controls (shown in red) grow as expected: growth occurs on sc-ura but not 5-FOA media. In this assay, a subset of mutants that are in a gap1Δ background were isolated from 5-FOA plates (shown in blue). All of these mutants should grow on D-His/D-Ser media, but only mutant DGY2041 #4 grows similarly to the gap1- control.

Surprisingly, of the 29 confirmed 5-FOA mutants, only 3 of these were able to grow on D-His/D-Ser media. These results raise important concerns, especially because a third of the mutants isolated (9/30) after the Luria Delbrück fluctuation assay were from gap1Δ strains. While these mutants were selected on 5-FOA, they should also still be able to grow on D-His/D-Ser since GAP1 is deleted from the genome. However, only 2/9 (22%) were able to grow on D-His/D-Ser (see **Figure 4.10** for an example). One possibility is that these mutants lost the ability to grow on proline, and it would be useful to include a proline plate without D-His/D-Ser as a control in the future. The third mutant able to grow on D-His/D-ser is in the

reporter background, and could represent a bonafide ura3- gap1- strain (**Figure 4.9**). These results demonstrate that reporter accuracy is extremely low, as only 1/12 (.08%) ura3- mutants were also gap1- mutants. Future follow-up experiments should also include confirmation by 3-primer PCR and Sanger sequencing, as D-His/D-ser growth can be difficult to interpret, potentially as a result from inconsistencies in making and storing D-His/D-Ser plates. Unfortunately, 3-primer PCR requires further optimization and was uninformative during a large-scale attempt to confirm gap1- mutants.

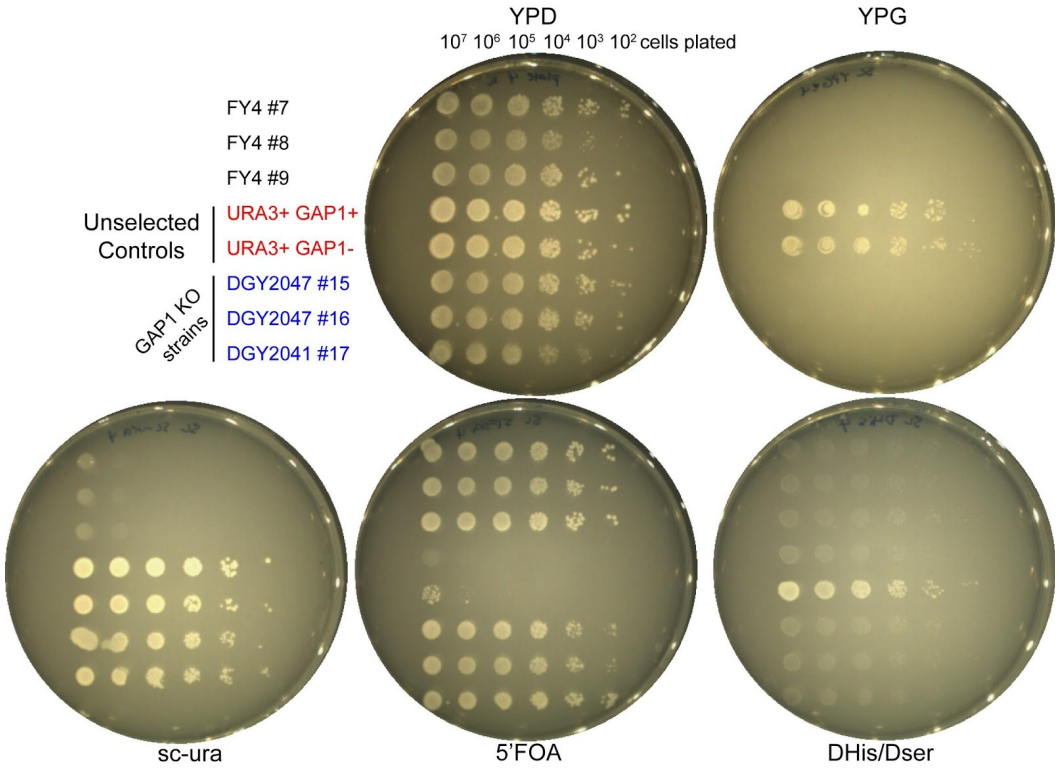


**Fig 4.11. Results of spot dilution assay for plate four.** Unselected controls (shown in red) grow as expected: growth occurs on sc-ura but not 5-FOA media. All other mutants were isolated after selection on 5-FOA plates. In this assay, a subset of mutants that are in a gap1Δ background were isolated from 5-FOA plates (shown in blue). All of these mutants should grow on D-His/D-Ser media, but none of them grow comparably to the gap1- control. None of the other mutants isolated from 5-FOA plates grow on YPG.

While the results presented here indicate that the reporter strain is inaccurate for detecting GAP1 deletions by NAHR, I observed a phenomenon that could prove informative for further studies. Since 23/30 (77%) 5-FOA mutants did not grow on YPG, these results suggest that selection on 5-FOA media is associated with absence of growth on YPG. One possible explanation is that selection on 5-FOA simultaneously selects for the formation of petite cells that have dysfunctional mitochondria and cannot metabolize glycerol (see **discussion**). Interestingly, 3 of the 7 mutants (43%) that grew on YPG were also able to grow on D-His/D-Ser, suggesting a correlation between growth on these two media types (see **Figure 4.10** for an example). Therefore, an additional selection for growth on YPG after plating cells to 5-FOA might prove informative for determining the number of bona fide GAP1 deletions.

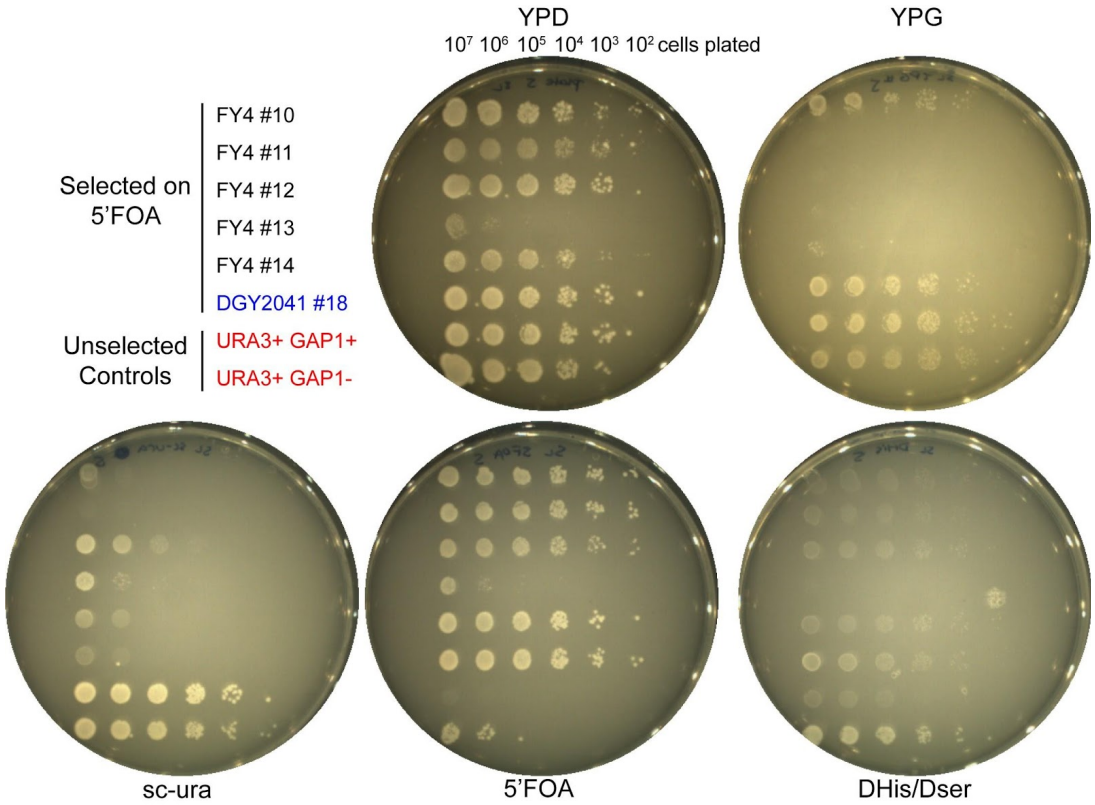


**Fig 4.12. Results of spot dilution assay for plate five.** Unselected controls (shown in red) grow as expected: growth occurs on sc-ura but not 5-FOA media. In this assay, one mutant from a gap1Δ background was isolated from 5-FOA plates (shown in blue), and only this mutant grows comparably to the gap1- control.

## 4.4: Discussion

Non-allelic homologous recombination (NAHR) is an important mechanism underlying CNV formation. Repetitive elements, including segmental duplications, mediate CNV formation by NAHR in diverse scenarios including microduplication and microdeletion syndromes in humans. To dissect the relative contribution of NAHR to spontaneous CNV formation, we constructed a CNV reporter for a specific deletion event. This event is driven by NAHR between flanking long terminal repeats at the GAP1 locus in yeast. Design and implementation of a deletion reporter, which includes integration of a URA3 gene and Luria Delbrück fluctuation assays to measure the rate of URA3 loss of function (LOF), are tractable in principle. However, reporter accuracy is low in practice. While the results of this assay were somewhat inconclusive, we have identified several unexpected avenues for further exploration. We discuss alternative strategies and potential follow-up experiments below.

To determine the spontaneous mutation rate for the GAP1 deletion reporter compared to the endogenous URA3 locus, we performed Luria Delbrück fluctuation assays. Mutations at the endogenous URA3 locus are relatively low, as indicated by the small number of mutant colonies for the wild-type strain (FY4) growing on 5-FOA. We were able to provide a coarse estimate of phenotypic resistance to 5-FOA: ~3 x $10^{-8}$ mutations/generation. A previous estimate of the phenotypic mutation rate for 5-FOA resistance is $5.43 \times 10^{-8}$ (Lang and Murray 2008). In order to provide a more accurate estimate of the mutation rate, future experiments should be performed with higher throughput, using at least 96 parallel cultures and either the $p_0$ or MSS-maximum likelihood method to estimate the mutation rate and 95% confidence intervals (Lang 2018).

The number of mutant colonies growing on 5-FOA was consistently higher (usually by an order of magnitude) for the deletion reporter. This indicates that URA3 LOF mutations are occurring more frequently than at the endogenous URA3 locus. This result is surprising considering the failure of our reporter to accurately predict simultaneous deletions at the GAP1 locus (discussed further below). There are two possible explanations that are not mutually exclusive: 1) the GAP1 locus is a “hotspot” for mutational events and the URA3 mutation rate is similarly increased by proximity, resulting in a position effect or 2) the URA3 promoter and terminator sequences in the reporter strain recombine with endogenous URA3 promoter and terminator sequences, leading to partial translocations or truncations that inactive the gene. In the first scenario, mutational events could include recombination with other Ty1 elements in addition to recombination between YKRCδ11 and YKRCδ12 (Winston et al. 1984). A useful control that would help in distinguishing between position effects due to proximity of the GAP1 gene would be a GAP1 deletion in the background of our reporter. In order to rule out the second possibility, the GAP1 deletion reporter should be reconstructed in a background where the promoter and terminator sequences are removed in addition to the URA3 open reading frame.

Surprisingly, GAP1 deletion control strains (DGY138 and DGY1764) had a total of 0 mutant colonies across four replicate Luria Delbrück fluctuation assays, even when 100 million cells were plated. In addition, when we plated a subset of 5-FOA mutants isolated from GAP1 deletion backgrounds onto D-His/D-Ser media, only 2/9 were D-His/D-Ser$^{R}$. Future experiments should include proline plates without D-His/D-Ser as a control, since it is possible that these 5-FOA mutants lost the ability to grow on proline. These data raised concerns about potential

synthetic lethality between GAP1 and URA3. One group previously observed a synthetic interaction between GAP1 and URA3, but growth was only affected on certain nitrogen sources including tryptophan and citrulline (Iraqui et al. 1999). Plating on minimal media supplemented with ammonium resulted in normal growth during their experiment, and our 5-FOA media uses ammonium sulfate as the nitrogen source.

When the experiment was repeated for a fifth time, two different GAP1 deletion control strains were used. While the colonies that formed were smaller than expected, mutant colony counts on 5-FOA plates were relatively similar to those for FY4. These data could be indicative of a strain effect, where GAP1 deletion alone is not responsible for the observed phenotype. The strains used in the first four Luria Delbrück replicates, DGY138 and DGY1764, were isolated from evolution experiments in urea-limited chemostats. Therefore, it is possible that they acquired additional adaptive mutations that somehow either 1) prevent growth on 5-FOA plates and/or 2) require URA3 function for survival. To confirm that there is no synthetic interaction between gap1- and ura3- strains, additional experiments in various strain backgrounds should be performed.

5-FOA resistance primarily arises through URA3 loss of function mutations. However, in this study, we identified 4/29 (14%) cases in which 5-FOA mutants grew on sc-ura media. All 4 cases occurred in GAP1 deletion strains (DGY2041 and DGY2047). This result further implies an interaction between GAP1 and URA3, but also suggests that either: 1) Mechanisms other than URA3 LOF enable growth on 5-FOA media and/or 2) there is a high level of background noise, where many “mutants” are not actually 5-FOA$^{R}$. In the second scenario, the URA3+

mutants we observe should have a no-growth phenotype on 5-FOA, which is not the case in our experiment. These results suggest a potential mechanism of 5-FOA resistance other than URA3 LOF. For example, over-expression of LOG1 has been shown to mitigate the effects of 5-FOA on cellular toxicity (Ko, Nishihama, and Pringle 2008). 5-FOA is metabolized to 5-fluorouracil (5-FU) by the cell, and there are several genes implicated in 5-FU resistance, including LOG1 (Carlsson, Hu, and Ronne 2018). Another study suggests that growth in YPD supplemented with 5-FOA does not prevent growth of all URA3+ strains (Boeke et al. 1987). Using a minimal media for the overnight growth step could therefore reduce identification of false positives by our reporter.

After plating confirmed 5-FOA$^R$ mutants on D-His/D-Ser media, we determined that reporter accuracy is extremely low (.08%). While a multitude of factors may contribute to this process as described above, I believe that off-target recombination events between 1) other Ty1 elements and/or 2) the endogenous URA3 promoter and terminator sequences are primarily responsible. While it may be difficult to rule out the first scenario, the second can be amended by reporter reconstruction. In addition, we identified a correlation between growth on YPG and D-His/D-Ser media. One possible explanation is that selection on 5-FOA simultaneously selects for the formation of petite cells that have defects in respiration and cannot grow on non-fermentable carbon sources such as glycerol. Similarly, 5-FOA could be mutagenic and lead to damaged mitochondria, or there could simply be a high spontaneous rate of petite formation. But another possibility is that slow-growth phenotypes conferred by petite formation enable 5-FOA$^R$ without requiring URA3 LOF. Depending on which scenario is relevant during our experimental conditions, it could be feasible to increase the efficiency of our reporter by simultaneously

replica plating to YPG or by performing the overnight culture in YPG. This process could reduce the high number of false positives we observed.

Once the accuracy of our deletion reporter is improved, there are many potential applications. Determining the spontaneous rate of GAP1 deletion by intrachromosomal NAHR is still a primary goal. But after calculating the spontaneous rate, we could directly test how the rate is affected by transcription at the target locus. GAP1 is highly expressed in the presence of non-preferred nitrogen sources such as proline, but strongly repressed in rich media. By comparing the mutation rates obtained from cells grown overnight in rich media to cells grown overnight in repressive conditions, we could determine the effect of environmental condition on CNV formation. We could also keep the environment constant and directly assess how the rate changes when 1) transcription is induced under conditions where GAP1 is normally repressed, and 2) transcription is repressed at the GAP1 locus when it is normally expressed. A variety of methods could be used to perform these experiments, including an estradiol-inducible system, a Tet-on/Tet-off system (using doxycycline to either repress transcription or activate gene expression), and the use of nuclease-deficient dCas9 to block GAP1 transcription. These experiments would provide further insight into the mechanistic role of NAHR by directly testing the effect of transcription and environmental condition on CNV formation. To confirm that NAHR is the main driver of intrachromosomal recombination, we could also compare the rates observed in our reporter strain to those observed in a rad52Δ background. Dissecting the relative contribution of NAHR in generating CNVs is an important next step in the field that will provide further insight into the molecular basis of CNV formation.

## 4.5: Materials and Methods

We used FY4, a haploid derivative of the reference strain S288c, for all experiments. To generate the GAP1 deletion reporter, we performed high efficiency yeast transformation (Gietz and Schiestl 2007) in a strain where the open reading frame of URA3 is deleted and replaced with the KanMX G418 resistance cassette (DGY1428). We integrated the URA3 gene at an intergenic region 1,118 base pairs upstream of GAP1 (integration coordinates, chromosome XI: 513945). Two independent transformations were performed and synthetic complete media lacking uracil (sc-ura) was used to select successful transformants. Transformants were confirmed by PCR. The strain isolated from the first transformation (DGY1829) was identified as petite (respiratory deficient), and was subsequently backcrossed with FY4 to produce a diploid (DGY1826). After sporulation and tetrad dissection, the resulting segregants were genotyped for $G418^R$, the ability to grow on sc-ura media, and the ability to grow on YPGlycerol (which selects for non-petite cells). Segregants with the correct genotype were confirmed by PCR. Three independent GAP1 deletion reporter strains were produced: DGY2012, DGY2013, and DGY2014. Two additional strains (DGY2010 and DGY2011) were generated by a second round of transformation and confirmed by PCR.

Plates containing 5-fluoroorotic acid (5-FOA) consisted of synthetic complete base media with 2% glucose, 50µg/mL uracil, and 5-FOA dissolved in DMSO and filter-sterilized for a final concentration of 0.1%. Plates containing D-Histidine and D-Serine (D-His/D-Ser) consisted of synthetic complete base media with 10mM proline, 10mM D-Histine and 5mM D-Serine, 2% glucose, and 50µg/mL uracil.

Several important positive and negative controls were used in this study (**see Table 4.1**) including FY4 as a negative control for growth on both 5-FOA and D-His/D-Ser and DGY1428 as a positive control for growth on 5-FOA. DGY41 which contains the ura-52 allele was also used as a positive control for growth on 5-FOA. GAP1 deletion strains used as positive controls for growth on D-His/D-Ser include: DGY138 and DGY1764 (isolated from urea-limited chemostats during evolution experiments) and DGY2041 and DGY2047 (generated by cloning).

To perform Luria Delbrück fluctuation assays, we plated for single colonies on rich, non-selective media (yeast-peptone dextrose or YPD) and inoculated one colony/strain into 5mL liquid YPD. Liquid cultures were grown overnight at 30°C. For each overnight culture, we counted cells per milliliter using a Z2 Beckman Coulter Counter. Depending on the replicate experiment, between $10^6$ and $10^8$ total cells were plated on 5-FOA media. After 3 days of growth at 30°C, colonies were counted manually when possible (<200 colonies) or counted using a Synbiosis aCOLyte and accompanying software.

A 3-primer PCR was used to confirm ura3Δ (DGP1527, DGP1530, DGP857) and gap1Δ (DGP1527, DGP1530, DGP292) genotypes. A bona fide gap1$\Delta^{LTR\text{-}LTR}$ mutant resulting from intrachromosomal recombination between YKRCδ11 and YKRCδ12 leads to a 1.1-1.2 kilobase product for both the URA3 and GAP1 PCRs. The wild-type genotype for URA3 and GAP1 results in a 1.4 and 1.3 kilobase product for the URA3 and GAP1 PCRs, respectively (**see Figure 4.6** and **Figure 4.7**).

To determine the fraction of 5-FOA$^R$ mutants that were also gap1-, we randomly chose a subset of 5-FOA$^R$ colonies for further analysis. 12 mutants were chosen from Luria Delbrück replicate four, individually streaked onto YPD, then a single colony was chosen for overnight growth in liquid YPD at 30°C. 18 additional mutants were isolated from Luria Delbrück replicate five, but only grown overnight in liquid YPD at 30°C (not re-streaked). All 30 mutants and unselected control strains (here, DGY2010 or DGY2012 and DGY138) were grown together in 200µL volume in 96 well plates. From these overnight cultures, 10-fold serial dilutions were performed and ranged from $10^7$ to $10^2$ cells. A frogger was used to spot each dilution onto a series of 5 plates: YPD, YPG (yeast-peptone-glycerol), sc-ura (synthetic complete media lacking uracil), 5-FOA, and D-His/D-Ser.

# Chapter 5: Conclusion

A central challenge in biology is understanding the genetic basis of both adaptive evolution and disease. Copy number variants (CNVs) are an important class of genetic variation underlying both adaptive and maladaptive phenotypes, but are understudied compared to single nucleotide variants (SNVs). While SNVs affect 1 base pair of DNA, a single CNV can affect $10^2$-$10^6$ base pairs. By duplicating or deleting such large genomic segments, which are likely to contain multiple protein-coding genes and regulatory regions, CNVs can have immediate consequences on organismal fitness. CNVs have been shown to directly alter gene dosage and protein abundance, but they can also affect global levels of transcription and even chromatin organization. CNVs can also provide new functionality through position effects that result in the formation of chimeric genes and modification of regulatory elements such as promoters and

enhancers. The diversity of possible functional effects of CNVs can result in a range of phenotypic effects. In Chapter 1, I highlighted the diverse roles of CNVs in driving rapid adaptive evolution in animals, plants, microbes, and pathogens. I also discussed the challenges in studying CNVs, including their potential for widespread pleiotropic and epistatic interactions, which can result in fitness costs and trade-offs in alternative environments. Importantly, CNVs segregating within populations can be polymorphic among individuals. Detecting CNVs in heterogeneous evolving populations remains difficult, and as a result, the dynamics with which they are generated, selected, and maintained were previously unknown. The goal of this dissertation research was to determine both the temporal dynamics of CNVs under selection and the diversity of CNV alleles that arise during adaptive evolution.

To study the temporal dynamics with which CNVs are generated and selected, I designed a novel phenotypic reporter that detects de novo gene duplications and deletions on the basis of changes in fluorescent signal. I used the reporter to study CNV dynamics and allelic diversity at a model locus in S. cerevisiae: the gene encoding the general amino acid permease (GAP1). The results of this study are presented in Chapter 2. Under the strong selective conditions of nutrient-limited chemostats, GAP1 undergoes both gene duplication and deletion depending on the type of limiting nitrogen source present. During glutamine limitation, GAP1 amplification alleles are repeatedly generated and selected with remarkably reproducible dynamics: they emerge early, they rapidly increase in frequency, and they are maintained in each population throughout the selection. This consistently early increase in GAP1 CNV frequency, indicative of a high rate of GAP1 CNV formation, was independently confirmed by results from Chapter 3. Using barcode lineage tracking, we found that hundreds to thousands of CNV lineages initially

compete, resulting in extreme clonal interference. Competition among CNV-containing lineages results in subpopulation dynamics that are distinct between populations despite the overall reproducibility of whole-population dynamics. Since CNVs are alleles of large effect, this high degree of clonal interference has important implications for the field of evolution. Epistatic relationships between CNVs and other adaptive mutations can dramatically alter the fitness landscape, leading to antagonistic pleiotropy and fitness consequences in alternate environments. CNVs also serve to increase the amount of DNA substrate that can accumulate mutations, and additional polymorphisms could subsequently increase the rate of adaptive evolution. Both of these phenomena are important to consider as we continue to define the role of CNVs in evolutionary processes.

During urea-limited chemostat evolution, we identified a GAP1 deletion using our CNV reporter. While we did not specifically track or isolate CNVs at other loci, we used genome sequencing to detect amplifications at another locus: DUR3, the high-affinity urea permease. We detected quantitative differences in copy number and CNV allele size between the two loci, which suggested that there are fundamental differences in generation and selection of CNVs depending on locus and selective condition. Because the CNV reporter can be integrated throughout the genome, we sought to test how different features of a genomic locus change the temporal dynamics of CNV selection. This work is presented in Chapter 3. We used the CNV reporter to track gene amplification at two loci in addition to GAP1: the high-affinity glucose transporters HXT6 and HXT7 and the ammonium permease MEP2. This comparative analysis of adaptive gene amplification across environments demonstrates that the temporal dynamics of CNV selection are distinct for each locus. While additional replicates should be performed to

determine if there are statistically significant differences, GAP1 amplification appears to occur earlier and at a faster rate than HXT6/7 amplification. However, HXT6/7 amplification may reach a higher copy number. Increases in fluorescence were not observed at the MEP2 locus, consistent with an absence of de novo amplification events. Unlike GAP1 and HXT6/7, MEP2 does not have an origin of replication or repetitive elements in close proximity to the open reading frame. Because these features appear to be important for facilitating fast and reversible CNV formation, MEP2 amplification may represent a secondary mode of adaptation in ammonium-limited chemostats that we were unable to detect during our relatively short-term experiment.

The use of a CNV reporter can easily be extended for future applications. Follow-up experiments using fluctuating or complex environments in chemostats are particularly tractable. For example, analyzing GAP1 CNV dynamics in a nitrogen-limited chemostat where glutamine and urea are both present in limiting concentrations would allow us to simultaneously track divergent ecological niches and adaptive strategies (i.e. the presence of GAP1 amplifications vs. GAP1 deletions in independent replicates). Implementing a two-color system (for example, mCitrine at the GAP1 locus and mCherry at the DUR3 locus) would enable direct comparisons of CNV dynamics at these two loci in this type of complex environment. A two-color system could also be useful for differentiating between diploidization or aneuploidy and amplification of a specific gene of interest. Diversification of a previously identified GAP1 CNV allele could also be determined after generating a barcode library, inoculating the library into chemostats, and performing lineage tracking. In addition to the experiments described here which could be

performed in yeast or other microbes, the CNV reporter could be feasibly integrated into cell lines to study tumorigenesis.

Our ability to track and isolate cells with gene amplifications enabled us to characterize the mechanisms underlying CNV formation in Chapter 2. I performed genome sequencing of whole populations and isolated clones and uncovered a diverse range of processes contributing to CNV formation including: aneuploidy, non-reciprocal translocation, tandem duplication, inverted triplication, and segmental aneuploidies potentially resulting from the formation of neo-chromosomes. CNV formation is mediated by both recombination and replication-mediated mechanisms that require extensive sequence homology and microhomology (4-24 nucleotides), respectively. These findings suggest that errors in DNA replication may be a previously underappreciated source of non-recurrent structural variation at the GAP1 locus. Dissecting the relative contributions of each mechanism to CNV formation is therefore an important next step that we sought to address in Chapter 4. We generated a reporter for intrachromosomal recombination between the flanking LTRs at the GAP1 locus, a process which results in excision of the GAP1 gene and formation of a hybrid LTR. Our reporter construct relies on the simultaneous deletion of a URA3 gene integrated ~1 kilobase upstream of the GAP1 open reading frame. While we determined that the reporter was inaccurate for determining the number of bona fide GAP1 deletions, we identified alternative solutions and new strategies for reporter design.

While determining the spontaneous rate of gene duplication and deletion is still an outstanding question in the field, important follow-up experiments are needed to further define the

mechanisms underlying CNV formation. While we believe that replication errors are important contributors to de novo CNV generation, this should be tested directly by removal of key repair enzymes such as RAD51 and RAD52. If CNV formation can proceed in the absence of homologous recombination, we will have identified replication-mediated errors as the main driver of CNV formation in S. cerevisiae. Secondly, recent studies strongly suggest a role for environmental stimulation in CNV generation, potentially through elevated rates of transcription at a locus under selection. Ideally, future studies would directly test the role of transcription in mediating CNV stimulation. While this is a potentially difficult avenue of research, the findings would have broad implications for challenging our historical belief that natural selection is driven by genotypic variation arising as a result of random mutation.

# REFERENCES


Abyzov, A., A. E. Urban, M. Snyder, and M. Gerstein. 2011. “CNVnator: An Approach to Discover, Genotype, and Characterize Typical and Atypical CNVs from Family and Population Genome Sequencing.” Genome Research 21 (6): 974–84.

Adam, D., N. Dimitrijevic, and M. Schartl. 1993. “Tumor Suppression in Xiphophorus by an Accidentally Acquired Promoter.” Science 259 (5096): 816–19.

Adamo, Giusy M., Marina Lotti, Markus J. Tamás, and Stefania Brocca. 2012. “Amplification of the CUP1 Gene Is Associated with Evolution of Copper Tolerance in Saccharomyces Cerevisiae.” Microbiology 158 (Pt 9): 2325–35.

Adler, Marlen, Mehreen Anjum, Otto G. Berg, Dan I. Andersson, and Linus Sandegren. 2014. “High Fitness Costs and Instability of Gene Duplications Reduce Rates of Evolution of New Genes by Duplication-Divergence Mechanisms.” Molecular Biology and Evolution 31 (6): 1526–35.

Aguilera, Andrés, and Hélène Gaillard. 2014. “Transcription and Recombination: When RNA Meets DNA.” Cold Spring Harbor Perspectives in Biology 6 (8): a016543–a016543.

Aigner, Johanna, Sergi Villatoro, Raquel Rabionet, Jaume Roquer, Jordi Jiménez-Conde, Eulàlia Martí, and Xavier Estivill. 2013. “A Common 56-Kilobase Deletion in a Primate-Specific Segmental Duplication Creates a Novel Butyrophilin-like Protein.” BMC Genetics 14 (July): 61.

Airoldi, Edoardo M., Darach Miller, Rodoniki Athanasiadou, Nathan Brandt, Farah Abdul-Rahman, Benjamin Neymotin, Tatsu Hashimoto, Tayebeh Bahmani, and David Gresham. 2016. “Steady-State and Dynamic Gene Expression Programs in Saccharomyces Cerevisiae in Response to Variation in Environmental Nitrogen.” Molecular

Biology of the Cell 27 (8): 1383–96.
Anderson, P., and J. Roth. 1981. “Spontaneous Tandem Genetic Duplications in Salmonella Typhimurium Arise by Unequal Recombination between rRNA (rrn) Cistrons.” Proceedings of the National Academy of Sciences of the United States of America 78 (5): 3113–17.
Anderson, R. P., and J. R. Roth. 1977. “Tandem Genetic Duplications in Phage and Bacteria.” Annual Review of Microbiology 31 (1): 473–505.
Arguello, J. Roman, Ying Chen, Shuang Yang, Wen Wang, and Manyuan Long. 2006. “Origination of an X-Linked Testes Chimeric Gene by Illegitimate Recombination in Drosophila.” PLoS Genetics 2 (5): e77.
Arlt, Martin F., Jennifer G. Mulle, Valerie M. Schaibley, Ryan L. Ragland, Sandra G. Durkin, Stephen T. Warren, and Thomas W. Glover. 2009. “Replication Stress Induces Genome-Wide Copy Number Changes in Human Cells That Resemble Polymorphic and Pathogenic Variants.” American Journal of Human Genetics 84 (3): 339–50.
Arlt, Martin F., Alev Cagla Ozdemir, Shanda R. Birkeland, Thomas E. Wilson, and Thomas W. Glover. 2011. “Hydroxyurea Induces de Novo Copy Number Variants in Human Cells.” Proceedings of the National Academy of Sciences of the United States of America 108 (42): 17360–65.
Arlt, Martin F., Sountharia Rajendran, Shanda R. Birkeland, Thomas E. Wilson, and Thomas W. Glover. 2012. “De Novo CNV Formation in Mouse Embryonic Stem Cells Occurs in the Absence of Xrcc4-Dependent Nonhomologous End Joining.” PLoS Genetics 8 (9): e1002981.
———. 2014. “Copy Number Variants Are Produced in Response to Low-Dose Ionizing Radiation in Cultured Cells.” Environmental and Molecular Mutagenesis 55 (2): 103–13.
Avise, J. C., and G. B. Kitto. 1973. “Phosphoglucose Isomerase Gene Duplication in the Bony Fishes: An Evolutionary History.” Biochemical Genetics 8 (2): 113–32.
Axelsson, Erik, Abhirami Ratnakumar, Maja-Louise Arendt, Khurram Maqbool, Matthew T. Webster, Michele Perloski, Olof Liberg, Jon M. Arnemo, Ake Hedhammar, and Kerstin Lindblad-Toh. 2013. “The Genomic Signature of Dog Domestication Reveals Adaptation to a Starch-Rich Diet.” Nature 495 (7441): 360–64.
Barnes, D. A., S. J. Foote, D. Galatis, D. J. Kemp, and A. F. Cowman. 1992. “Selection for High-Level Chloroquine Resistance Results in Deamplification of the pfmdr1 Gene and Increased Sensitivity to Mefloquine in Plasmodium Falciparum.” The EMBO Journal 11 (8): 3067–75.
Barreiro, Luis B., Guillaume Laval, Hélène Quach, Etienne Patin, and Lluís Quintana-Murci. 2008. “Natural Selection Has Driven Population Differentiation in Modern Humans.” Nature Genetics 40: 340–45.
Baxevanis, Andreas D., Daniel B. Davison, Roderic D. M. Page, Gregory A. Petsko, Lincoln D. Stein, and Gary D. Stormo, eds. 2002. “Using the Velvet de Novo Assembler for Short-Read Sequencing Technologies.” In Current Protocols in Bioinformatics, 18:810. Hoboken, NJ, USA: John Wiley & Sons, Inc.
Bayer, Avraham, Greg Brennan, and Adam P. Geballe. 2018. “Adaptation by Copy Number Variation in Monopartite Viruses.” Current Opinion in Virology 33 (July): 7–12.
Baym, Michael, Sergey Kryazhimskiy, Tami D. Lieberman, Hattie Chung, Michael M. Desai, Roy Kishony, G. W. Tyson, et al. 2015. “Inexpensive Multiplexed Library Preparation for Megabase-Sized Genomes.” PloS One 10 (5): e0128036–e0128036.

Bell, Michael A. 1987. “Interacting Evolutionary Constraints in Pelvic Reduction of Threespine Sticklebacks, Gasterosteus Aculeatus (Pisces, Gasterosteidae).” Biological Journal of the Linnean Society. Linnean Society of London 31 (4): 347–82.
Bell, Michael A., and Susan Adlai Foster. 1994. The Evolutionary Biology of the Threespine Stickleback. Oxford University Press.
Benelli, Giovanni. 2015. “Research in Mosquito Control: Current Challenges for a Brighter Future.” Parasitology Research 114 (8): 2801–5.
Bermudez-Santana, Clara, Camille Attolini, Toralf Kirsten, Jan Engelhardt, Sonja J. Prohaska, Stephan Steigele, Peter F. Stadler, et al. 2010. “Genomic Organization of Eukaryotic tRNAs.” BMC Genomics 11 (1): 270–270.
Blount, Zachary D., Jeffrey E. Barrick, Carla J. Davidson, and Richard E. Lenski. 2012. “Genomic Analysis of a Key Innovation in an Experimental Escherichia Coli Population.” Nature 489 (7417): 513–18.
Blount, Zachary D., Christina Z. Borland, and Richard E. Lenski. 2008. “Historical Contingency and the Evolution of a Key Innovation in an Experimental Population of Escherichia Coli.” Proceedings of the National Academy of Sciences of the United States of America 105 (23): 7899–7906.
Blundell, Jamie R., Katja Schwartz, Danielle Francois, Daniel S. Fisher, Gavin J. Sherlock, and Sasha F. Levy. 2017. “The Dynamics of Adaptive Genetic Diversity during the Early Stages of Clonal Evolution.” bioRxiv. https://doi.org/10.1101/170589.
Boeke, J. D., J. Trueheart, G. Natsoulis, and G. R. Fink. 1987. “5-Fluoroorotic Acid as a Selective Agent in Yeast Molecular Genetics.” Methods in Enzymology 154: 164–75.
Brauer, Matthew J., Curtis Huttenhower, Edoardo M. Airoldi, Rachel Rosenstein, John C. Matese, David Gresham, Viktor M. Boer, Olga G. Troyanskaya, and David Botstein. 2008. “Coordination of Growth Rate, Cell Cycle, Stress Response, and Metabolic Activity in Yeast.” Molecular Biology of the Cell 19 (1): 352–67.
Brewer, Bonita J., Celia Payen, Sara C. Di Rienzi, Megan M. Higgins, Giang Ong, Maitreya J. Dunham, and M. K. Raghuraman. 2015. “Origin-Dependent Inverted-Repeat Amplification: Tests of a Model for Inverted DNA Amplification.” PLoS Genetics 11 (12): e1005699–e1005699.
Brewer, Bonita J., Celia Payen, M. K. Raghuraman, and Maitreya J. Dunham. 2011. “Origin-Dependent Inverted-Repeat Amplification: A Replication-Based Model for Generating Palindromic Amplicons.” PLoS Genetics 7 (3): e1002016–e1002016.
Brochet, Mathieu, Elisabeth Couvé, Mohamed Zouine, Claire Poyart, and Philippe Glaser. 2008. “A Naturally Occurring Gene Amplification Leading to Sulfonamide and Trimethoprim Resistance in Streptococcus Agalactiae.” Journal of Bacteriology 190 (2): 672–80.
Brown, C. J., K. M. Todd, and R. F. Rosenzweig. 1998. “Multiple Duplications of Yeast Hexose Transport Genes in Response to Selection in a Glucose-Limited Environment.” Molecular Biology and Evolution 15 (8): 931–42.
Bruder, Carl E. G., Arkadiusz Piotrowski, Antoinet A. C. J. Gijsbers, Robin Andersson, Stephen Erickson, Teresita Diaz de Ståhl, Uwe Menzel, et al. 2008. “Phenotypically Concordant and Discordant Monozygotic Twins Display Different DNA Copy-Number-Variation Profiles.” American Journal of Human Genetics 82 (3): 763–71.
Brügger, Kim, Peter Redder, Qunxin She, Fabrice Confalonieri, Yvan Zivanovic, and Roger A. Garrett. 2002. “Mobile Elements in Archaeal Genomes.” FEMS Microbiology Letters 206

(2): 131–41.
Brynildsrud, Ola, Snorre Gulla, Edward J. Feil, Simen Foyn Nørstebø, and Linda D. Rhodes. 2016. “Identifying Copy Number Variation of the Dominant Virulence factorsmsaandp22within Genomes of the Fish pathogenRenibacterium Salmoninarum.” Microbial Genomics 2 (4): e000055.
Burke, Molly K., Joseph P. Dunham, Parvin Shahrestani, Kevin R. Thornton, Michael R. Rose, and Anthony D. Long. 2010. “Genome-Wide Analysis of a Long-Term Evolution Experiment with Drosophila.” Nature 467 (7315): 587–90.
Bussotti, Giovanni, Evi Gouzelou, Mariana Cortes Boite, Ihcen Kherachi, and Gerald F. Spath. 2018. “Leishmania Genome Dynamics during Environmental Adaptation Reveals Strain-Specific Differences in Gene Copy Number Variation, Karyotype Instability, and Telomeric Amplification.” mBio, September. http://dx.doi.org/.
Cairns, J., and P. L. Foster. 1991. “Adaptive Reversion of a Frameshift Mutation in Escherichia Coli.” Genetics 128 (4): 695–701.
Cai, Xuyu, Gilad D. Evrony, Hillel S. Lehmann, Princess C. Elhosary, Bhaven K. Mehta, Annapurna Poduri, and Christopher A. Walsh. 2014. “Single-Cell, Genome-Wide Sequencing Identifies Clonal Somatic Copy-Number Variation in the Human Brain.” Cell Reports 8 (5): 1280–89.
Camacho, Christiam, George Coulouris, Vahram Avagyan, Ning Ma, Jason Papadopoulos, Kevin Bealer, and Thomas L. Madden. 2009. “BLAST+: Architecture and Applications.” BMC Bioinformatics 10 (December): 421.
Cardoso-Moreira, Margarida, J. Roman Arguello, and Andrew G. Clark. 2012. “Mutation Spectrum of Drosophila CNVs Revealed by Breakpoint Sequencing.” Genome Biology 13 (12): R119–R119.
Cardoso-Moreira, Margarida, J. Roman Arguello, Srikanth Gottipati, L. G. Harshman, Jennifer K. Grenier, and Andrew G. Clark. 2016. “Evidence for the Fixation of Gene Duplications by Positive Selection in Drosophila.” Genome Research 26 (6): 787–98.
Carlsson, Mattias, Guo-Zhen Hu, and Hans Ronne. 2018. “Gene Dosage Effects in Yeast Support Broader Roles for the LOG1, HAM1 and DUT1 Genes in Detoxification of Nucleotide Analogues.” PloS One 13 (5): e0196840.
Carr, Martin, Douda Bensasson, and Casey M. Bergman. 2012. “Evolutionary Genomics of Transposable Elements in Saccharomyces Cerevisiae.” PloS One 7 (11): e50978.
Carvalho, Claudia M. B., and James R. Lupski. 2016. “Mechanisms Underlying Structural Variant Formation in Genomic Disorders.” Nature Reviews. Genetics 17 (4): 224–38.
Chain, Frédéric J. J., Philine G. D. Feulner, Mahesh Panchal, Christophe Eizaguirre, Irene E. Samonte, Martin Kalbe, Tobias L. Lenz, et al. 2014a. “Extensive Copy-Number Variation of Young Genes across Stickleback Populations.” PLoS Genetics 10 (12): e1004830.
Chan, Yingguang Frank, Melissa E. Marks, Felicity C. Jones, Guadalupe Villarreal Jr, Michael D. Shapiro, Shannon D. Brady, Audrey M. Southwick, et al. 2010. “Adaptive Evolution of Pelvic Reduction in Sticklebacks by Recurrent Deletion of a Pitx1 Enhancer.” Science 327 (5963): 302–5.
Chen, Lu, Weichen Zhou, Cheng Zhang, James R. Lupski, Li Jin, and Feng Zhang. 2015. “CNV Instability Associated with DNA Replication Dynamics: Evidence for Replicative Mechanisms in CNV Mutagenesis.” Human Molecular Genetics 24 (6): 1574–83.
Chou, Hsin-Hung, Hsuan-Chao Chiu, Nigel F. Delaney, Daniel Segrè, and Christopher J. Marx.

2011. “Diminishing Returns Epistasis among Beneficial Mutations Decelerates Adaptation.” Science 332: 1190–92.
Chow, Eve W. L., Carl A. Morrow, Julianne T. Djordjevic, Ian A. Wood, and James A. Fraser. 2012. “Microevolution of Cryptococcus Neoformans Driven by Massive Tandem Gene Amplification.” Molecular Biology and Evolution 29 (8): 1987–2000.
Christ, Daniel, and Jason W. Chin. 2008. “Engineering Escherichia Coli Heat-Resistance by Synthetic Gene Amplification.” Protein Engineering, Design & Selection: PEDS 21 (2): 121–25.
Clewell, D. B., Y. Yagi, and B. Bauer. 1975. “Plasmid-Determined Tetracycline Resistance in Streptococcus Faecalis: Evidence for Gene Amplification during Growth in Presence of Tetracycline.” Proceedings of the National Academy of Sciences of the United States of America 72 (5): 1720–24.
Clop, A., O. Vidal, and M. Amills. 2012. “Copy Number Variation in the Genomes of Domestic Animals.” Animal Genetics 43 (5): 503–17.
Cohen, S., and D. Segal. 2009. “Extrachromosomal Circular DNA in Eukaryotes: Possible Involvement in the Plasticity of Tandem Repeats.” Cytogenetic and Genome Research 124 (3-4): 327–38.
Coïc, Eric, Taya Feldman, Allison S. Landman, and James E. Haber. 2008. “Mechanisms of Rad52-Independent Spontaneous and UV-Induced Mitotic Recombination in Saccharomyces Cerevisiae.” Genetics 179 (1): 199–211.
Colosimo, Pamela F., Kim E. Hosemann, Sarita Balabhadra, Guadalupe Villarreal Jr, Mark Dickson, Jane Grimwood, Jeremy Schmutz, Richard M. Myers, Dolph Schluter, and David M. Kingsley. 2005. “Widespread Parallel Evolution in Sticklebacks by Repeated Fixation of Ectodysplasin Alleles.” Science 307 (5717): 1928–33.
Conant, Gavin C., and Kenneth H. Wolfe. 2008. “Turning a Hobby into a Job: How Duplicated Genes Find New Functions.” Nature Reviews. Genetics 9 (12): 938–50.
Cone, Kelsey R., Zev N. Kronenberg, Mark Yandell, and Nels C. Elde. 2017. “Emergence of a Viral RNA Polymerase Variant during Gene Copy Number Amplification Promotes Rapid Evolution of Vaccinia Virus.” Journal of Virology 91 (4). https://doi.org/10.1128/JVI.01428-16.
Connallon, Tim, and Andrew G. Clark. 2011. “The Resolution of Sexual Antagonism by Gene Duplication.” Genetics 187 (3): 919–37.
Conrad, Donald F., Dalila Pinto, Richard Redon, Lars Feuk, Omer Gokcumen, Yujun Zhang, Jan Aerts, et al. 2010. “Origins and Functional Impact of Copy Number Variation in the Human Genome.” Nature 464 (7289): 704–12.
Cooke, J., M. A. Nowak, M. Boerlijst, and J. Maynard-Smith. 1997. “Evolutionary Origins and Maintenance of Redundant Gene Expression during Metazoan Development.” Trends in Genetics: TIG 13 (9): 360–64.
Copley, Shelley D. 2012. “Toward a Systems Biology Perspective on Enzyme Evolution.” The Journal of Biological Chemistry 287 (1): 3–10.
Cowell, Annie N., Eva S. Istvan, Amanda K. Lukens, Maria G. Gomez-Lorenzo, Manu Vanaerschot, Tomoyo Sakata-Kato, Erika L. Flannery, et al. 2018. “Mapping the Malaria Parasite Druggable Genome by Using in Vitro Evolution and Chemogenomics.” Science 359 (6372): 191–99.
Cullen, Paul J., George F. Sprague, and Jr. 2012. “The Regulation of Filamentous Growth in

Yeast.” Genetics 190 (1): 23–49.
DeBolt, Seth. 2010. “Copy Number Variation Shapes Genome Diversity in Arabidopsis over Immediate Family Generational Scales.” Genome Biology and Evolution 2 (July): 441–53.
Dhami, Manpreet K., Thomas Hartwig, and Tadashi Fukami. 2016. “Genetic Basis of Priority Effects: Insights from Nectar Yeast.” Proceedings. Biological Sciences / The Royal Society 283 (1840). https://doi.org/10.1098/rspb.2016.1455.
Di Rienzi, Sara C., David Collingwood, M. K. Raghuraman, and Bonita J. Brewer. 2009. “Fragile Genomic Sites Are Associated with Origins of Replication.” Genome Biology and Evolution 1: 350–63.
Dittwald, Piotr, Tomasz Gambin, Przemyslaw Szafranski, Jian Li, Stephen Amato, Michael Y. Divon, Lisa Ximena Rodríguez Rojas, et al. 2013. “NAHR-Mediated Copy-Number Variants in a Clinical Population: Mechanistic Insights into Both Genomic Disorders and Mendelizing Traits.” Genome Research 23 (9): 1395–1409.
Dolatabadian, Aria, Dhwani Apurva Patel, David Edwards, and Jacqueline Batley. 2017. “Copy Number Variation and Disease Resistance in Plants.” TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik 130 (12): 2479–90.
Dorsey, M., C. Peterson, K. Bray, and C. E. Paquin. 1992. “Spontaneous Amplification of the ADH4 Gene in Saccharomyces Cerevisiae.” Genetics 132 (4): 943–50.
Dulmage, Keely A., Cynthia L. Darnell, Angie Vreugdenhil, and Amy K. Schmid. 2018. “Copy Number Variation Is Associated with Gene Expression Change in Archaea.” Microbial Genomics, August. https://doi.org/10.1099/mgen.0.000210.
Dumetz, F., B. Cuypers, H. Imamura, D. Zander, E. D’Haenens, I. Maes, M. A. Domagalska, J. Clos, J-C Dujardin, and G. De Muylder. 2018. “Molecular Preadaptation to Antimony Resistance in Leishmania Donovani on the Indian Subcontinent.” mSphere 3 (2). https://doi.org/10.1128/mSphere.00548-17.
Dumont, Beth L. 2015. “Interlocus Gene Conversion Explains at Least 2.7% of Single Nucleotide Variants in Human Segmental Duplications.” BMC Genomics 16 (June): 456.
Durkin, Keith, Wouter Coppieters, Cord Drögemüller, Naima Ahariz, Nadine Cambisano, Tom Druet, Corinne Fasquelle, et al. 2012. “Serial Translocation by Means of Circular Intermediates Underlies Colour Sidedness in Cattle.” Nature 482 (7383): 81–84.
Durkin, Sandra G., Ryan L. Ragland, Martin F. Arlt, Jennifer G. Mulle, Stephen T. Warren, and Thomas W. Glover. 2008. “Replication Stress Induces Tumor-like Microdeletions in FHIT/FRA3B.” Proceedings of the National Academy of Sciences of the United States of America 105 (1): 246–51.
Eichler, E. E. 2001. “Recent Duplication, Domain Accretion and the Dynamic Mutation of the Human Genome.” Trends in Genetics: TIG 17 (11): 661–69.
Elde, Nels C., Stephanie J. Child, Michael T. Eickbush, Jacob O. Kitzman, Kelsey S. Rogers, Jay Shendure, Adam P. Geballe, and Harmit S. Malik. 2012. “Poxviruses Deploy Genomic Accordions to Adapt Rapidly against Host Antiviral Defenses.” Cell 150 (4): 831–41.
Ellis, B., P. Haaland, F. Hahne, N. L. Meur, N. Gopalakrishnan, and Spidlen J And Jiang. 2016. flowCore: flowCore: Basic Structures for Flow Cytometry Data.
Emerson, J. J., Margarida Cardoso-Moreira, Justin O. Borevitz, and Manyuan Long. 2008. “Natural Selection Shapes Genome-Wide Patterns of Copy-Number Polymorphism in Drosophila Melanogaster.” Science 320 (5883): 1629–31.
Faik, Imad, Elisandra Grangeiro de Carvalho, and Jürgen Fj Kun. 2009. “Parasite-Host

Interaction in Malaria: Genetic Clues and Copy Number Variation." Genome Medicine 1 (9): 82.
Farslow, James C., Kendra J. Lipinski, Lucille B. Packard, Mark L. Edgley, Jon Taylor, Stephane Flibotte, Donald G. Moerman, Vaishali Katju, and Ulfar Bergthorsson. 2015. "Rapid Increase in Frequency of Gene Copy-Number Variants during Experimental Evolution in Caenorhabditis Elegans." BMC Genomics 16 (December): 1044.
Faust, Gregory G., and Ira M. Hall. 2014. "SAMBLASTER: Fast Duplicate Marking and Structural Variant Read Extraction." Bioinformatics 30 (17): 2503–5.
Ferris, S. D., and G. S. Whitt. 1979. "Evolution of the Differential Regulation of Duplicate Genes after Polyploidization." Journal of Molecular Evolution 12 (4): 267–317.
Feuk, Lars, Andrew R. Carson, and Stephen W. Scherer. 2006. "Structural Variation in the Human Genome." Nature Reviews. Genetics 7 (2): 85–97.
Feulner, Philine G. D., Frédéric J. J. Chain, Mahesh Panchal, Christophe Eizaguirre, Martin Kalbe, Tobias L. Lenz, Marvin Mundry, et al. 2012. "Genome-Wide Patterns of Standing Genetic Variation in a Marine Population of Three-Spined Sticklebacks." Molecular Ecology 22 (3): 635–49.
Fogel, S., J. W. Welch, G. Cathala, and M. Karin. 1983. "Gene Amplification in Yeast: CUP1 Copy Number Regulates Copper Resistance." Current Genetics 7 (5): 347–55.
Fontanesi, L., F. Beretti, V. Riggio, E. Gómez González, S. Dall'Olio, R. Davoli, V. Russo, and B. Portolano. 2009. "Copy Number Variation and Missense Mutations of the Agouti Signaling Protein (ASIP) Gene in Goat Breeds with Different Coat Colors." Cytogenetic and Genome Research 126 (4): 333–47.
Force, A., M. Lynch, F. B. Pickett, A. Amores, Y. L. Yan, and J. Postlethwait. 1999. "Preservation of Duplicate Genes by Complementary, Degenerative Mutations." Genetics 151 (4): 1531–45.
Foster, Patricia L. 2007. "Stress-Induced Mutagenesis in Bacteria." Critical Reviews in Biochemistry and Molecular Biology 42 (5): 373–97.
Franke, Martin, Daniel M. Ibrahim, Guillaume Andrey, Wibke Schwarzer, Verena Heinrich, Robert Schöpflin, Katerina Kraft, et al. 2016. "Formation of New Chromatin Domains Determines Pathogenicity of Genomic Duplications." Nature 538 (7624): 265–69.
Frickel, Jens, Philine G. D. Feulner, Emre Karakoc, and Lutz Becks. 2018. "Population Size Changes and Selection Drive Patterns of Parallel Evolution in a Host-Virus System." Nature Communications 9 (1): 1706.
Galhardo, Rodrigo S., P. J. Hastings, and Susan M. Rosenberg. 2007. "Mutation as a Stress Response and the Regulation of Evolvability." Critical Reviews in Biochemistry and Molecular Biology 42 (5): 399–435.
Gamazon, Eric R., Dan L. Nicolae, and Nancy J. Cox. 2011. "A Study of CNVs as Trait-Associated Polymorphisms and as Expression Quantitative Trait Loci." PLoS Genetics 7 (2): e1001292.
Gamazon, Eric R., and Barbara E. Stranger. 2015. "The Impact of Human Copy Number Variation on Gene Expression." Briefings in Functional Genomics 14 (5): 352–57.
Gao, Yuxia, Huayao Zhao, Yin Jin, Xiaoyu Xu, and Guan-Zhu Han. 2017. "Extent and Evolution of Gene Duplication in DNA Viruses." Virus Research 240 (August): 161–65.
Garcia-Diaz, Miguel, and Thomas A. Kunkel. 2006. "Mechanism of a Genetic Glissando: Structural Biology of Indel Mutations." Trends in Biochemical Sciences 31 (4): 206–14.

Geiger, Tamar, Juergen Cox, and Matthias Mann. 2010. “Proteomic Changes Resulting from Gene Copy Number Variations in Cancer Cells.” PLoS Genetics 6 (9): e1001090–e1001090.
Gelbart, W. M., and A. Chovnick. 1979. “Spontaneous Unequal Exchange in the Rosy Region of Drosophila Melanogaster.” Genetics 92 (3): 849–59.
Gietz, R. Daniel, and Robert H. Schiestl. 2007. “High-Efficiency Yeast Transformation Using the LiAc/SS Carrier DNA/PEG Method.” Nature Protocols 2: 31–34.
Girirajan, Santhosh, Zoran Brkanac, Bradley P. Coe, Carl Baker, Laura Vives, Tiffany H. Vu, Neil Shafer, et al. 2011. “Relative Burden of Large CNVs on a Range of Neurodevelopmental Phenotypes.” PLoS Genetics 7 (11): e1002334–e1002334.
Girirajan, Santhosh, Catarina D. Campbell, and Evan E. Eichler. 2011. “Human Copy Number Variation and Complex Genetic Disease.” Annual Review of Genetics 45 (August): 203–26.
Giuffra, Elisabetta, Anna Törnsten, Stefan Marklund, Erik Bongcam-Rudloff, Patrick Chardon, James M. H. Kijas, Susan I. Anderson, Alan L. Archibald, and Leif Andersson. 2002. “A Large Duplication Associated with Dominant White Color in Pigs Originated by Homologous Recombination between LINE Elements Flanking KIT.” Mammalian Genome: Official Journal of the International Mammalian Genome Society 13 (10): 569–77.
Good, Benjamin H., Michael J. McDonald, Jeffrey E. Barrick, Richard E. Lenski, and Michael M. Desai. 2017. “The Dynamics of Molecular Evolution over 60,000 Generations.” Nature 551 (7678): 45–50.
Gordon, David J., Benjamin Resio, and David Pellman. 2012. “Causes and Consequences of Aneuploidy in Cancer.” Nature Reviews. Genetics 13 (3): 189–203.
Greenblum, Sharon, Rogan Carr, and Elhanan Borenstein. 2015. “Extensive Strain-Level Copy-Number Variation across Human Gut Microbiome Species.” Cell 160 (4): 583–94.
Grenson, M., C. Hou, and M. Crabeel. 1970. “Multiplicity of the Amino Acid Permeases in Saccharomyces Cerevisiae. IV. Evidence for a General Amino Acid Permease.” Journal of Bacteriology 103 (3): 770–77.
Gresham, David, Michael M. Desai, Cheryl M. Tucker, Harry T. Jenq, Dave A. Pai, Alexandra Ward, Christopher G. DeSevo, David Botstein, and Maitreya J. Dunham. 2008. “The Repertoire and Dynamics of Evolutionary Adaptations to Controlled Nutrient-Limited Environments in Yeast.” PLoS Genetics, 2008.
Gresham, David, and Maitreya J. Dunham. 2014. “The Enduring Utility of Continuous Culturing in Experimental Evolution.” Genomics 104 (6): 399–405.
Gresham, David, Renata Usaite, Susanne Manuela Germann, Michael Lisby, David Botstein, and Birgitte Regenberg. 2010. “Adaptation to Diverse Nitrogen-Limited Environments by Deletion or Extrachromosomal Element Formation of the GAP1 Locus.” Proceedings of the National Academy of Sciences of the United States of America 107 (43): 18551–56.
Griesbeck, O., G. S. Baird, R. E. Campbell, D. A. Zacharias, and R. Y. Tsien. 2001. “Reducing the Environmental Sensitivity of Yellow Fluorescent Protein. Mechanism and Applications.” The Journal of Biological Chemistry 276 (31): 29188–94.
Gruber, Jonathan D., Kara Vogel, Gizem Kalay, and Patricia J. Wittkopp. 2012. “Contrasting Properties of Gene-Specific Regulatory, Coding, and Copy Number Mutations in Saccharomyces Cerevisiae: Frequency, Effects, and Dominance.” PLoS Genetics 8 (2): e1002497.
Gu, Zhenglong, Lars M. Steinmetz, Xun Gu, Curt Scharfe, Ronald W. Davis, and Wen-Hsiung

Li. 2003. “Role of Duplicate Genes in Genetic Robustness against Null Mutations.” Nature 421 (6918): 63–66.
Hansche, P. E. 1975. “Gene Duplication as a Mechanism of Genetic Adaptation in Saccharomyces Cerevisiae.” Genetics 79 (4): 661–74.
Harris, Rebecca, Andrew Sackman, and Jeffrey D. Jensen. 2018. “On the Unfounded Enthusiasm for Soft Selective Sweeps II: Examining Recent Evidence from Humans, Flies, and Viruses.” bioRxiv. https://doi.org/10.1101/443051.
Hartasánchez, Diego A., Oriol Vallès-Codina, Marina Brasó-Vives, and Arcadi Navarro. 2014. “Interplay of Interlocus Gene Conversion and Crossover in Segmental Duplications under a Neutral Scenario.” G3 4 (8): 1479–89.
Hastings, P. J., H. J. Bull, J. R. Klump, and S. M. Rosenberg. 2000. “Adaptive Amplification: An Inducible Chromosomal Instability Mechanism.” Cell 103 (5): 723–31.
Hastings, P. J., Grzegorz Ira, James R. Lupski, A. J. Iafrate, L. Feuk, M. N. Rivera, M. L. Listewnik, et al. 2009. “A Microhomology-Mediated Break-Induced Replication Model for the Origin of Human Copy Number Variation.” PLoS Genetics 5 (1): e1000327–e1000327.
Hastings, P. J., James R. Lupski, Susan M. Rosenberg, and Grzegorz Ira. 2009. “Mechanisms of Change in Gene Copy Number.” Nature Reviews. Genetics 10 (8): 551–64.
Helyar, S. J., J. Hemmer-Hansen, D. Bekkevold, M. I. Taylor, R. Ogden, M. T. Limborg, A. Cariani, et al. 2011. “Application of SNPs for Population Genetics of Nonmodel Organisms: New Opportunities and Challenges.” Molecular Ecology Resources 11 Suppl 1 (March): 123–36.
Henrichsen, Charlotte N., Nicolas Vinckenbosch, Sebastian Zöllner, Evelyne Chaignat, Sylvain Pradervand, Frédéric Schütz, Manuel Ruedi, Henrik Kaessmann, and Alexandre Reymond. 2009. “Segmental Copy Number Variation Shapes Tissue Transcriptomes.” Nature Genetics 41 (4): 424–29.
Hermisson, Joachim, and Pleuni S. Pennings. 2005. “Soft Sweeps: Molecular Population Genetics of Adaptation from Standing Genetic Variation.” Genetics 169 (4): 2335–52.
Hirase, Shotaro, Haruka Ozaki, and Wataru Iwasaki. 2014. “Parallel Selection on Gene Copy Number Variations through Evolution of Three-Spined Stickleback Genomes.” BMC Genomics 15 (August): 735.
Hoffman, Charles S., and Fred Winston. 1987. “A Ten-Minute DNA Preparation from Yeast Efficiently Releases Autonomous Plasmids for Transformaion of Escherichia Coli.” Gene 57 (2): 267–72.
Hollox, Edward J., and Boon-Peng Hoh. 2014. “Human Gene Copy Number Variation and Infectious Disease.” Human Genetics 133 (10): 1217–33.
Hong, Jungeui, Nathan Brandt, Farah Abdul-Rahman, Ally Yang, Tim Hughes, and David Gresham. 2018. “An Incoherent Feedforward Loop Facilitates Adaptive Tuning of Gene Expression.” eLife 7 (April). https://doi.org/10.7554/eLife.32323.
Hong, Jungeui, and David Gresham. 2014. “Molecular Specificity, Convergence and Constraint Shape Adaptive Evolution in Nutrient-Poor Environments.” PLoS Genetics 10 (1): e1004041–e1004041.
Hope, Elyse A., Clara J. Amorosi, Aaron W. Miller, Kolena Dang, Caiti Smukowski Heil, and Maitreya J. Dunham. 2017. “Experimental Evolution Reveals Favored Adaptive Routes to Cell Aggregation in Yeast.” Genetics 206 (2): 1153–67.
Hopkinson, D. A., Y. H. Edwards, and H. Harris. 1976. “The Distributions of Subunit Numbers

and Subunit Sizes of Enzymes: A Study of the Products of 100 Human Gene Loci." Annals of Human Genetics 39 (4): 383–411.
Horiuchi, T., S. Horiuchi, and A. Novick. 1963. "The Genetic Basis of Hyper-Synthesis of Beta-Galactosidase." Genetics 48: 157–69.
Hughes, A. L. 1994. "The Evolution of Functionally Novel Proteins after Gene Duplication." Proceedings. Biological Sciences / The Royal Society 256 (1346): 119–24.
Hughes, Austin L. 1994. "The Evolution of Functionally Novel Proteins after Gene Duplication." Proceedings of the Royal Society of London B: Biological Sciences 256.
Hughes, Julie M., Brian K. Lohman, Gail E. Deckert, Eric P. Nichols, Matt Settles, Zaid Abdo, and Eva M. Top. 2012. "The Role of Clonal Interference in the Evolutionary Dynamics of Plasmid-Host Adaptation." mBio 3: e00077–12.
Hull, Ryan M., Cristina Cruz, Carmen V. Jack, and Jonathan Houseley. 2017. "Environmental Change Drives Accelerated Adaptation through Stimulated Copy Number Variation." PLoS Biology 15 (6): e2001333.
Hu, Tina T., Pedro Pattyn, Erica G. Bakker, Jun Cao, Jan-Fang Cheng, Richard M. Clark, Noah Fahlgren, et al. 2011. "The Arabidopsis Lyrata Genome Sequence and the Basis of Rapid Genome Size Change." Nature Genetics 43 (5): 476–81.
Iafrate, A. John, Lars Feuk, Miguel N. Rivera, Marc L. Listewnik, Patricia K. Donahoe, Ying Qi, Stephen W. Scherer, and Charles Lee. 2004. "Detection of Large-Scale Variation in the Human Genome." Nature Genetics 36 (9): 949–51.
Iakoubov, Leonid, Malgorzata Mossakowska, Malgorzata Szwed, Zhibing Duan, Federico Sesti, and Monika Puzianowska-Kuznicka. 2013. "A Common Copy Number Variation (CNV) Polymorphism in the CNTNAP4 Gene: Association with Aging in Females." PloS One 8 (11): e79790.
Iantorno, Stefano A., Caroline Durrant, Asis Khan, Mandy J. Sanders, Stephen M. Beverley, Wesley C. Warren, Matthew Berriman, David L. Sacks, James A. Cotton, and Michael E. Grigg. 2017. "Gene Expression in Leishmania Is Regulated Predominantly by Gene Dosage." mBio 8 (5). https://doi.org/10.1128/mBio.01393-17.
Iraqui, I., S. Vissers, B. André, and A. Urrestarazu. 1999. "Transcriptional Induction by Aromatic Amino Acids in Saccharomyces Cerevisiae." Molecular and Cellular Biology 19 (5): 3360–71.
Iskow, Rebecca C., Omer Gokcumen, Alexej Abyzov, Joanna Malukiewicz, Qihui Zhu, Ann T. Sukumar, Athma A. Pai, et al. 2012. "Regulatory Element Copy Number Differences Shape Primate Expression Profiles." Proceedings of the National Academy of Sciences of the United States of America 109: 12656–61.
Itokawa, Kentaro, Osamu Komagata, Shinji Kasai, Yoshika Okamura, Masahiro Masada, and Takashi Tomita. 2010. "Genomic Structures of Cyp9m10 in Pyrethroid Resistant and Susceptible Strains of Culex Quinquefasciatus." Insect Biochemistry and Molecular Biology 40 (9): 631–40.
Itsara, Andy, Gregory M. Cooper, Carl Baker, Santhosh Girirajan, Jun Li, Devin Absher, Ronald M. Krauss, et al. 2009. "Population Analysis of Large Copy Number Variants and Hotspots of Human Genetic Disease." American Journal of Human Genetics 84 (2): 148–61.
Itsara, Andy, Hao Wu, Joshua D. Smith, Deborah A. Nickerson, Isabelle Romieu, Stephanie J. London, and Evan E. Eichler. 2010. "De Novo Rates and Selection of Large Copy Number Variation." Genome Research 20 (11): 1469–81.

Izutsu, Minako, Jun Zhou, Yuzo Sugiyama, Osamu Nishimura, Tomoyuki Aizu, Atsushi Toyoda, Asao Fujiyama, Kiyokazu Agata, and Naoyuki Fuse. 2012. “Genome Features of ‘Dark-Fly’, a Drosophila Line Reared Long-Term in a Dark Environment.” PloS One 7 (3): e33288.
Jack, Carmen V., Cristina Cruz, Ryan M. Hull, Markus A. Keller, Markus Ralser, and Jonathan Houseley. 2015. “Regulation of Ribosomal DNA Amplification by the TOR Pathway.” Proceedings of the National Academy of Sciences of the United States of America 112 (31): 9674–79.
Jalvingh, Kirsten M., Peter L. Chang, Sergey V. Nuzhdin, and Bregje Wertheim. 2014. “Genomic Changes under Rapid Evolution: Selection for Parasitoid Resistance.” Proceedings. Biological Sciences / The Royal Society 281 (1779): 20132303.
Jensen, Jeffrey D. 2014. “On the Unfounded Enthusiasm for Soft Selective Sweeps.” Nature Communications 5 (October): 5281.
Jensen, R. A. 1976. “Enzyme Recruitment in Evolution of New Function.” Annual Review of Microbiology 30 (1): 409–25.
Jones, Felicity C., Manfred G. Grabherr, Yingguang Frank Chan, Pamela Russell, Evan Mauceli, Jeremy Johnson, Ross Swofford, et al. 2012. “The Genomic Basis of Adaptive Evolution in Threespine Sticklebacks.” Nature 484 (7392): 55–61.
Kafri, Moshe, Eyal Metzl-Raz, Ghil Jona, and Naama Barkai. 2016. “The Cost of Protein Production.” Cell Reports 14 (1): 22–31.
Kang, Lin, Dau Dayal Aggarwal, Eugenia Rashkovetsky, Abraham B. Korol, and Pawel Michalak. 2016. “Rapid Genomic Changes in Drosophila Melanogaster Adapting to Desiccation Stress in an Experimental Evolution System.” BMC Genomics 17 (1). https://doi.org/10.1186/s12864-016-2556-y.
Kao, Katy C., and Gavin Sherlock. 2008. “Molecular Characterization of Clonal Interference during Adaptive Evolution in Asexual Populations of Saccharomyces Cerevisiae.” Nature Genetics 40 (12): 1499–1504.
Katju, Vaishali, and Ulfar Bergthorsson. 2013. “Copy-Number Changes in Evolution: Rates, Fitness Effects and Adaptive Significance.” Frontiers in Genetics 4 (December): 273.
Katoh, Kazutaka, and Daron M. Standley. 2013. “MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability.” Molecular Biology and Evolution 30 (4): 772–80.
Kern, Andrew D., and Matthew W. Hahn. 2018. “The Neutral Theory in Light of Natural Selection.” Molecular Biology and Evolution, May. https://doi.org/10.1093/molbev/msy092.
Kettler, Gregory C., Adam C. Martiny, Katherine Huang, Jeremy Zucker, Maureen L. Coleman, Sebastien Rodrigue, Feng Chen, et al. 2007. “Patterns and Implications of Gene Gain and Loss in the Evolution of Prochlorococcus.” PLoS Genetics 3 (12): e231.
Khan, Aisha I., Duy M. Dinh, Dominique Schneider, Richard E. Lenski, and Tim F. Cooper. 2011. “Negative Epistasis between Beneficial Mutations in an Evolving Bacterial Population.” Science 332: 1193–96.
Knouse, Kristin A., Jie Wu, and Angelika Amon. 2016. “Assessment of Megabase-Scale Somatic Copy Number Variation Using Single-Cell Sequencing.” Genome Research 26 (3): 376–84.
Kondrashov, Fyodor A. 2012. “Gene Duplication as a Mechanism of Genomic Adaptation to a Changing Environment.” Proceedings. Biological Sciences / The Royal Society 279 (1749): 5048–57.

Kondrashov, Fyodor A., Igor B. Rogozin, Yuri I. Wolf, and Eugene V. Koonin. 2002. “Selection in the Evolution of Gene Duplications.” Genome Biology 3: RESEARCH0008.
Koning, A. P. Jason de, and Bianca D. De Sanctis. 2018. “The Rate of Observable Molecular Evolution When Mutation May Not Be Weak.” https://doi.org/10.1101/259507.
Ko, Nolan, Ryuichi Nishihama, and John R. Pringle. 2008. “Control of 5-FOA and 5-FU Resistance by Saccharomyces Cerevisiae YJL055W.” Yeast 25 (2): 155–60.
Korbel, Jan O., Alexander Eckehart Urban, Jason P. Affourtit, Brian Godwin, Fabian Grubert, Jan Fredrik Simons, Philip M. Kim, et al. 2007. “Paired-End Mapping Reveals Extensive Structural Variation in the Human Genome.” Science 318 (5849): 420–26.
Koszul, Romain, Sandrine Caburet, Bernard Dujon, and Gilles Fischer. 2004. “Eucaryotic Genome Evolution through the Spontaneous Duplication of Large Chromosomal Segments.” The EMBO Journal 23 (1): 234–43.
Kugelberg, Elisabeth, Eric Kofoid, Andrew B. Reams, Dan I. Andersson, and John R. Roth. 2006. “Multiple Pathways of Selected Gene Amplification during Adaptive Mutation.” Proceedings of the National Academy of Sciences of the United States of America 103 (46): 17319–24.
Kvitek, Daniel J., and Gavin Sherlock. 2011. “Reciprocal Sign Epistasis between Frequently Experimentally Evolved Adaptive Mutations Causes a Rugged Fitness Landscape.” PLoS Genetics 7 (4): e1002056–e1002056.
Labib, Karim, Ben Hodgson, A. Admire, L. Shanks, N. Danzl, M. Wang, U. Weier, et al. 2007. “Replication Fork Barriers: Pausing for a Break or Stalling for Time?” EMBO Reports 8 (4): 346–53.
Lam, Kwan-Wood G., and Alec J. Jeffreys. 2006. “Processes of Copy-Number Change in Human DNA: The Dynamics of {alpha}-Globin Gene Deletion.” Proceedings of the National Academy of Sciences of the United States of America 103 (24): 8921–27.
Lang, Gregory I. 2018. “Measuring Mutation Rates Using the Luria-Delbrück Fluctuation Assay.” In Genome Instability: Methods and Protocols, edited by Marco Muzi-Falconi and Grant W. Brown, 21–31. New York, NY: Springer New York.
Lang, Gregory I., David Botstein, and Michael M. Desai. 2011. “Genetic Variation and the Fate of Beneficial Mutations in Asexual Populations.” Genetics 188: 647–61.
Lang, Gregory I., and Andrew W. Murray. 2008. “Estimating the per-Base-Pair Mutation Rate in the Yeast Saccharomyces Cerevisiae.” Genetics 178 (1): 67–82.
Lang, Gregory I., Daniel P. Rice, Mark J. Hickman, Erica Sodergren, George M. Weinstock, David Botstein, and Michael M. Desai. 2013. “Pervasive Genetic Hitchhiking and Clonal Interference in Forty Evolving Yeast Populations.” Nature 500: 571–74.
Langridge, J. 1969. “Mutations Conferring Quantitative and Qualitative Increases in β-Galactosidase Activity inEscherichia Coli.” Molecular & General Genetics: MGG 105 (1): 74–83.
Lauer, Stephanie, Grace Avecilla, Pieter Spealman, Gunjan Sethia, Nathan Brandt, Sasha Levy, and David Gresham. 2018. “Single-Cell Copy Number Variant Detection Reveals the Dynamics and Diversity of Adaptation.” bioRxiv. https://doi.org/10.1101/381590.
Layer, Ryan M., Colby Chiang, Aaron R. Quinlan, Ira M. Hall, C. Alkan, B. P. Coe, E. E. Eichler, et al. 2014. “LUMPY: A Probabilistic Framework for Structural Variant Discovery.” Genome Biology 15 (6): R84–R84.
Lee, Dae-Hee, and Bernhard Ø. Palsson. 2010. “Adaptive Evolution of Escherichia Coli K-12

MG1655 during Growth on a Nonnative Carbon Source, L-1,2-Propanediol." Applied and Environmental Microbiology 76: 4158–68.
Lee, Heun-Sik, Sanghoon Moon, Jun Ho Yun, Meehee Lee, Mi Yeong Hwang, Young-Jin Kim, Bok-Ghee Han, Jeong-Min Kim, and Bong-Jo Kim. 2014. "Genome-Wide Copy Number Variation Study Reveals KCNIP1 as a Modulator of Insulin Secretion." Genomics 104 (2): 113–20.
Lee, Jennifer A., Claudia M. B. Carvalho, and James R. Lupski. 2007. "A DNA Replication Mechanism for Generating Nonrecurrent Rearrangements Associated with Genomic Disorders." Cell 131: 1235–47.
Lee, Ming-Chun, and Christopher J. Marx. 2012. "Repeated, Selection-Driven Genome Reduction of Accessory Genes in Experimental Populations." PLoS Genetics 8 (5): e1002651.
Leinonen, Tuomas, R. J. Scott McCairns, Robert B. O'Hara, and Juha Merilä. 2013. "QST–FST Comparisons: Evolutionary and Ecological Insights from Genomic Heterogeneity." Nature Reviews. Genetics 14 (3): 179–90.
Lenski, Richard E., Michael R. Rose, Suzanne C. Simpson, and Scott C. Tadler. 1991. "Long-Term Experimental Evolution in Escherichia Coli. I. Adaptation and Divergence During 2,000 Generations." The American Naturalist 138: 1315–41.
Leprohon, Philippe, Danielle Légaré, Frédéric Raymond, Eric Madore, Gary Hardiman, Jacques Corbeil, and Marc Ouellette. 2009. "Gene Expression Modulation Is Associated with Gene Amplification, Supernumerary Chromosomes and Chromosome Loss in Antimony-Resistant Leishmania Infantum." Nucleic Acids Research 37 (5): 1387–99.
Levy, Sasha F., Jamie R. Blundell, Sandeep Venkataram, Dmitri A. Petrov, Daniel S. Fisher, and Gavin Sherlock. 2015. "Quantitative Evolutionary Dynamics Using High-Resolution Lineage Tracking." Nature 519 (7542): 181–86.
Li, Heng, and Richard Durbin. 2010. "Fast and Accurate Long-Read Alignment with Burrows-Wheeler Transform." Bioinformatics 26 (5): 589–95.
Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. "The Sequence Alignment/Map Format and SAMtools." Bioinformatics 25 (16): 2078–79.
Lipinski, Kendra J., James C. Farslow, Kelly A. Fitzpatrick, Michael Lynch, Vaishali Katju, and Ulfar Bergthorsson. 2011. "High Spontaneous Rate of Gene Duplication in Caenorhabditis Elegans." Current Biology: CB 21 (4): 306–10.
Lowe, Craig B., Nicelio Sanchez-Luege, Timothy R. Howes, Shannon D. Brady, Rhea R. Daugherty, Felicity C. Jones, Michael A. Bell, and David M. Kingsley. 2018. "Detecting Differential Copy Number Variation between Groups of Samples." Genome Research 28 (2): 256–65.
Lupiáñez, Darío G., Katerina Kraft, Verena Heinrich, Peter Krawitz, Francesco Brancati, Eva Klopocki, Denise Horn, et al. 2015. "Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions." Cell 161 (5): 1012–25.
Lupiáñez, Darío G., Malte Spielmann, and Stefan Mundlos. 2016. "Breaking TADs: How Alterations of Chromatin Domains Result in Disease." Trends in Genetics: TIG 32 (4): 225–37.
Lupski, James R. 2009. "Genomic Disorders Ten Years on." Genome Medicine 1 (4): 42.

Lynch, M., and A. Force. 2000. “The Probability of Duplicate Gene Preservation by Subfunctionalization.” Genetics 154 (1): 459–73.
Lynch, Michael, and John S. Conery. 2000. “The Evolutionary Fate and Consequences of Duplicate Genes.” Science 290 (5494).
Lynch, Michael, Way Sung, Krystalynne Morris, Nicole Coffey, Christian R. Landry, Erik B. Dopman, W. Joseph Dickinson, et al. 2008. “A Genome-Wide View of the Spectrum of Spontaneous Mutations in Yeast.” Proceedings of the National Academy of Sciences of the United States of America 105 (27): 9272–77.
MacDonald, Jeffrey R., Robert Ziman, Ryan K. C. Yuen, Lars Feuk, and Stephen W. Scherer. 2014. “The Database of Genomic Variants: A Curated Collection of Structural Variation in the Human Genome.” Nucleic Acids Research 42 (Database issue): D986–92.
Mackerras, Ian Murray. 1936. The Sheep Blowfly Problem in Australia: Results of Some Recent Investigations.
Maddamsetti, R., R. E. Lenski, and J. E. Barrick. 2015. “Adaptation, Clonal Interference, and Frequency-Dependent Interactions in a Long-Term Evolution Experiment with Escherichia Coli.” Genetics. https://doi.org/10.1534/genetics.115.176677.
Maharjan, Ram P., Joël Gaffé, Jessica Plucain, Martin Schliep, Lei Wang, Lu Feng, Olivier Tenaillon, Thomas Ferenci, and Dominique Schneider. 2013. “A Case of Adaptation through a Mutation in a Tandem Duplication during Experimental Evolution in Escherichia Coli.” BMC Genomics 14 (July): 441.
Mansisidor, Andres R., Temistocles Molinar, Priyanka Srivastava, Hannah Blitzblau, Hannah Klein, and Andreas Hochwagen. 2018. “Genomic Copy-Number Loss Is Rescued by Self-Limiting Production of DNA Circles.” bioRxiv. https://doi.org/10.1101/255471.
Marichal, P., H. Vanden Bossche, F. C. Odds, G. Nobels, D. W. Warnock, V. Timmerman, C. Van Broeckhoven, S. Fay, and P. Mose-Larsen. 1997. “Molecular Biological Characterization of an Azole-Resistant Candida Glabrata Isolate.” Antimicrobial Agents and Chemotherapy 41 (10): 2229–37.
Marshall, Christian R., and Stephen W. Scherer. 2012. “Detection and Characterization of Copy Number Variation in Autism Spectrum Disorder.” Methods in Molecular Biology 838: 115–35.
Mary, C., F. Faraut, M. Deniau, J. Dereure, K. Aoun, S. Ranque, and R. Piarroux. 2010. “Frequency of Drug Resistance Gene Amplification in Clinical Leishmania Strains.” International Journal of Microbiology 2010 (July). https://doi.org/10.1155/2010/819060.
May, Jürgen, Jennifer A. Evans, Christian Timmann, Christa Ehmen, Wibke Busch, Thorsten Thye, Tsiri Agbenyega, and Rolf D. Horstmann. 2007. “Hemoglobin Variants and Disease Manifestations in Severe Falciparum Malaria.” JAMA: The Journal of the American Medical Association 297 (20): 2220–26.
Mayo, Sonia, Sandra Monfort, Mónica Roselló, Carmen Orellana, Silvestre Oltra, Alfonso Caro-Llopis, and Francisco Martínez. 2017. “Chimeric Genes in Deletions and Duplications Associated with Intellectual Disability.” International Journal of Genomics and Proteomics 2017 (May): 4798474.
McCarroll, Steven A., Alan Huett, Petric Kuballa, Shannon D. Chilewski, Aimee Landry, Philippe Goyette, Michael C. Zody, et al. 2008. “Deletion Polymorphism Upstream of IRGM Associated with Altered IRGM Expression and Crohn’s Disease.” Nature Genetics 40 (9): 1107–12.

McConnell, Michael J., Michael R. Lindberg, Kristen J. Brennand, Julia C. Piper, Thierry Voet, Chris Cowing-Zitron, Svetlana Shumilina, et al. 2013. “Mosaic Copy Number Variation in Human Neurons.” Science 342 (6158): 632–37.
McKenna, Aaron, Matthew Hanna, Eric Banks, Andrey Sivachenko, Kristian Cibulskis, Andrew Kernytsky, Kiran Garimella, et al. 2010. “The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data.” Genome Research 20 (9): 1297–1303.
Mehta, Anuja, and James E. Haber. 2014. “Sources of DNA Double-Strand Breaks and Models of Recombinational DNA Repair.” Cold Spring Harbor Perspectives in Biology 6 (9): a016428.
Merla, Giuseppe, Cédric Howald, Charlotte N. Henrichsen, Robert Lyle, Carine Wyss, Marie-Thérèse Zabot, Stylianos E. Antonarakis, and Alexandre Reymond. 2006. “Submicroscopic Deletion in Patients with Williams-Beuren Syndrome Influences Expression Levels of the Nonhemizygous Flanking Genes.” American Journal of Human Genetics 79 (2): 332–41.
Messer, Philipp W., and Dmitri A. Petrov. 2013. “Population Genomics of Rapid Adaptation by Soft Selective Sweeps.” Trends in Ecology & Evolution 28 (11): 659–69.
Meyer, Rachel S., and Michael D. Purugganan. 2013. “Evolution of Crop Species: Genetics of Domestication and Diversification.” Nature Reviews. Genetics 14 (12): 840–52.
Miller, Aaron W., Corrie Befort, Emily O. Kerr, and Maitreya J. Dunham. 2013. “Design and Use of Multiplexed Chemostat Arrays.” Journal of Visualized Experiments: JoVE, no. 72 (February): e50262.
Miller, Craig T., Sandra Beleza, Alex A. Pollen, Dolph Schluter, Rick A. Kittles, Mark D. Shriver, and David M. Kingsley. 2007. “Cis-Regulatory Changes in Kit Ligand Expression and Parallel Evolution of Pigmentation in Sticklebacks and Humans.” Cell 131 (6): 1179–89.
Mockenhaupt, Frank P., Stephan Ehrhardt, Sabine Gellert, Rowland N. Otchwemah, Ekkehart Dietz, Sylvester D. Anemana, and Ulrich Bienzle. 2004. “Alpha(+)-Thalassemia Protects African Children from Severe Malaria.” Blood 104 (7): 2003–6.
Molina, Jessica, Paulina Carmona-Mora, Jacqueline Chrast, Paola M. Krall, César P. Canales, James R. Lupski, Alexandre Reymond, and Katherina Walz. 2008. “Abnormal Social Behaviors and Altered Gene Expression Rates in a Mouse Model for Potocki-Lupski Syndrome.” Human Molecular Genetics 17 (16): 2486–95.
Møller, Henrik D., Kaj S. Andersen, and Birgitte Regenberg. 2013. “A Model for Generating Several Adaptive Phenotypes from a Single Genetic Event: Saccharomyces Cerevisiae GAP1 as a Potential Bet-Hedging Switch.” Communicative & Integrative Biology 6 (3): e23933.
Møller, Henrik D., Lance Parsons, Tue S. Jørgensen, David Botstein, and Birgitte Regenberg. 2015. “Extrachromosomal Circular DNA Is Common in Yeast.” Proceedings of the National Academy of Sciences of the United States of America 112 (24): E3114–22.
Monte-Neto, Rubens, Marie-Claude N. Laffitte, Philippe Leprohon, Priscila Reis, Frédéric Frézard, and Marc Ouellette. 2015. “Intrachromosomal Amplification, Locus Deletion and Point Mutation in the Aquaglyceroporin AQP1 Gene in Antimony Resistant Leishmania (Viannia) Guyanensis.” PLoS Neglected Tropical Diseases 9 (2): e0003476.
Mouches, C., N. Pasteur, J. Berge, O. Hyrien, M. Raymond, B. de Saint Vincent, M. de Silvestri, and G. Georghiou. 1986. “Amplification of an Esterase Gene Is Responsible for Insecticide

Resistance in a California Culex Mosquito.” Science 233 (4765): 778–80.
Mulla, Wahid, Jin Zhu, and Rong Li. 2014. “Yeast: A Simple Model System to Study Complex Phenomena of Aneuploidy.” FEMS Microbiology Reviews 38 (2): 201–12.
Nair, Shalini, Becky Miller, Marion Barends, Anchalee Jaidee, Jigar Patel, Mayfong Mayxay, Paul Newton, François Nosten, Michael T. Ferdig, and Tim J. C. Anderson. 2008. “Adaptive Copy Number Evolution in Malaria Parasites.” PLoS Genetics 4 (10): e1000243.
Nair, Shalini, Denae Nash, Daniel Sudimack, Anchalee Jaidee, Marion Barends, Anne-Catrin Uhlemann, Sanjeev Krishna, François Nosten, and Tim J. C. Anderson. 2007. “Recurrent Gene Amplification and Soft Selective Sweeps during Evolution of Multidrug Resistance in Malaria Parasites.” Molecular Biology and Evolution 24 (2): 562–73.
Näsvall, Joakim, Lei Sun, John R. Roth, and Dan I. Andersson. 2012. “Real-Time Evolution of New Genes by Innovation, Amplification, and Divergence.” Science 338 (6105): 384–87.
Natesuntorn, Waranya, Kotaro Iwami, Yuki Matsubara, Yu Sasano, Minetaka Sugiyama, Yoshinobu Kaneko, and Satoshi Harashima. 2015. “Genome-Wide Construction of a Series of Designed Segmental Aneuploids in Saccharomyces Cerevisiae.” Scientific Reports 5 (July): 12510.
Newcomb, Richard D., Dianne M. Gleeson, Catherine G. Yong, Robyn J. Russell, and John G. Oakeshott. 2005. “Multiple Mutations and Gene Duplications Conferring Organophosphorus Insecticide Resistance Have Been Selected at the Rop-1 Locus of the Sheep Blowfly, Lucilia Cuprina.” Journal of Molecular Evolution 60 (2): 207–20.
Nilsson, Annika I., Anna Zorzet, Anna Kanth, Sabina Dahlström, Otto G. Berg, and Dan I. Andersson. 2006. “Reducing the Fitness Cost of Antibiotic Resistance by Amplification of Initiator tRNA Genes.” Proceedings of the National Academy of Sciences of the United States of America 103 (18): 6976–81.
Normark, S., T. Edlund, T. Grundström, S. Bergström, and H. Wolf-Watz. 1977. “Escherichia Coli K-12 Mutants Hyperproducing Chromosomal Beta-Lactamase by Gene Repetitions.” Journal of Bacteriology 132 (3): 912–22.
Norris, Belinda J., and Vicki A. Whan. 2008. “A Gene Duplication Affecting Expression of the Ovine ASIP Gene Is Responsible for White and Black Sheep.” Genome Research 18 (8): 1282–93.
O’Farrell, C. L., and M. S. Strom. 1999. “Differential Expression of the Virulence-Associated Protein p57 and Characterization of Its Duplicated Gene Msa in Virulent and Attenuated Strains of Renibacterium Salmoninarum.” Diseases of Aquatic Organisms 38 (2): 115–23.
Ohno, Susumu. 1970. Evolution by Gene Duplication. Berlin, Heidelberg: Springer Berlin Heidelberg.
Oromendia, Ana B., and Angelika Amon. 2014. “Aneuploidy: Implications for Protein Homeostasis and Disease.” Disease Models & Mechanisms 7 (1): 15–20.
Ottaviani, Diego, Magdalena LeCain, and Denise Sheer. 2014. “The Role of Microhomology in Genomic Structural Variation.” Trends in Genetics: TIG 30 (3): 85–94.
Pan, Deng, and Liqing Zhang. 2007. “Quantifying the Major Mechanisms of Recent Gene Duplications in the Human and Mouse Genomes: A Novel Strategy to Estimate Gene Duplication Rates.” Genome Biology 8 (8): R158.
Parkes, Miles, Jeffrey C. Barrett, Natalie J. Prescott, Mark Tremelling, Carl A. Anderson, Sheila A. Fisher, Roland G. Roberts, et al. 2007. “Sequence Variants in the Autophagy Gene IRGM and Multiple Other Replicating Loci Contribute to Crohn’s Disease Susceptibility.”

Nature Genetics 39 (7): 830–32.
Paulander, Wilhelm, Dan I. Andersson, and Sophie Maisnier-Patin. 2010. “Amplification of the Gene for Isoleucyl-tRNA Synthetase Facilitates Adaptation to the Fitness Cost of Mupirocin Resistance in Salmonella Enterica.” Genetics 185 (1): 305–12.
Pavelka, Norman, Giulia Rancati, Jin Zhu, William D. Bradford, Anita Saraf, Laurence Florens, Brian W. Sanderson, Gaye L. Hattem, and Rong Li. 2010. “Aneuploidy Confers Quantitative Proteome Changes and Phenotypic Variation in Budding Yeast.” Nature 468 (7321): 321–25.
Payen, Celia, Sara C. Di Rienzi, Giang T. Ong, Jamie L. Pogachar, Joseph C. Sanchez, Anna B. Sunshine, M. K. Raghuraman, Bonita J. Brewer, and Maitreya J. Dunham. 2014. “The Dynamics of Diverse Segmental Amplifications in Populations of Saccharomyces Cerevisiae Adapting to Strong Selection.” G3  4 (3): 399–409.
Payen, Celia, Romain Koszul, Bernard Dujon, Gilles Fischer, J. A. Bailey, E. E. Eichler, A. J. Sharp, et al. 2008. “Segmental Duplications Arise from Pol32-Dependent Repair of Broken Forks through Two Alternative Replication-Based Mechanisms.” PLoS Genetics 4 (9): e1000175–e1000175.
Payen, Celia, Anna B. Sunshine, Giang T. Ong, Jamie L. Pogachar, Wei Zhao, and Maitreya J. Dunham. 2016. “High-Throughput Identification of Adaptive Mutations in Experimentally Evolved Yeast Populations.” PLoS Genetics 12 (10): e1006339.
Pecoraro, Vito, Karolin Zerulla, Christian Lange, and Jörg Soppa. 2011. “Quantification of Ploidy in Proteobacteria Revealed the Existence of Monoploid, (mero-)oligoploid and Polyploid Species.” PloS One 6 (1): e16392.
Peichel, Catherine, and David Kingsley. 2006. “The Molecular Geneticsof Evolutionary Changein Sticklebacks.” In Marine Biology, 41–81.
Peng, Zhen, Weichen Zhou, Wenqing Fu, Renqian Du, Li Jin, and Feng Zhang. 2015. “Correlation between Frequency of Non-Allelic Homologous Recombination and Homology Properties: Evidence from Homology-Mediated CNV Mutations in the Human Genome.” Human Molecular Genetics 24 (5): 1225–33.
Perlman, D., and R. H. Rownd. 1975. “Transition of R Factor NR1 in Proteus Mirabilis: Molecular Structure and Replication of NR1 Deoxyribonucleic Acid.” Journal of Bacteriology 123 (3): 1013–34.
Perry, George H., Nathaniel J. Dominy, Katrina G. Claw, Arthur S. Lee, Heike Fiegler, Richard Redon, John Werner, et al. 2007. “Diet and the Evolution of Human Amylase Gene Copy Number Variation.” Nature Genetics 39 (10): 1256–60.
Pettersson, Mats E., Song Sun, Dan I. Andersson, and Otto G. Berg. 2009. “Evolution of New Gene Functions: Simulation and Analysis of the Amplification Model.” Genetica 135 (3): 309–24.
Polley, Shamik, Natalie Prescott, Elaine Nimmo, Colin Veal, Ida Vind, Pia Munkholm, Peder Fode, et al. 2016. “Copy Number Variation of Scavenger-Receptor Cysteine-Rich Domains within DMBT1 and Crohn’s Disease.” European Journal of Human Genetics: EJHG. https://doi.org/10.1038/ejhg.2015.280.
Pränting, Maria, and Dan I. Andersson. 2011. “Escape from Growth Restriction in Small Colony Variants of Salmonella Typhimurium by Gene Amplification and Mutation.” Molecular Microbiology 79 (2): 305–15.
Ramirez, Oscar, Iñigo Olalde, Jonas Berglund, Belen Lorente-Galdos, Jessica

Hernandez-Rodriguez, Javier Quilez, Matthew T. Webster, et al. 2014. “Analysis of Structural Diversity in Wolf-like Canids Reveals Post-Domestication Variants.” BMC Genomics 15 (1): 465–465.
Reams, Andrew B., Eric Kofoid, Michael Savageau, and John R. Roth. 2010. “Duplication Frequency in a Population of Salmonella Enterica Rapidly Approaches Steady State with or without Recombination.” Genetics 184 (4): 1077–94.
Reams, Andrew B., and Ellen L. Neidle. 2004. “Gene Amplification Involves Site-Specific Short Homology-Independent Illegitimate Recombination in Acinetobacter Sp. Strain ADP1.” Journal of Molecular Biology 338 (4): 643–56.
Reams, Andrew B., and John R. Roth. 2015. “Mechanisms of Gene Duplication and Amplification.” Cold Spring Harbor Perspectives in Biology 7 (2): a016592.
Reiter, Taylor, Evelyn Jagoda, and Terence D. Capellini. 2016. “Dietary Variation and Evolution of Gene Copy Number among Dog Breeds.” PloS One 11 (2): e0148899.
Remolina, Silvia C., Peter L. Chang, Jeff Leips, Sergey V. Nuzhdin, and Kimberly A. Hughes. 2012. “Genomic Basis of Aging and Life-History Evolution in Drosophila Melanogaster.” Evolution; International Journal of Organic Evolution 66 (11): 3390–3403.
Rezelj, Veronica V., Laura I. Levi, and Marco Vignuzzi. 2018. “The Defective Component of Viral Populations.” Current Opinion in Virology 33 (August): 74–80.
Rhodes, Linda D., Alison M. Coady, and Rebecca K. Deinhard. 2004. “Identification of a Third Msa Gene in Renibacterium Salmoninarum and the Associated Virulence Phenotype.” Applied and Environmental Microbiology 70 (11): 6488–94.
Rice, Alan M., and Aoife McLysaght. 2017. “Dosage Sensitivity Is a Major Determinant of Human Copy Number Variant Pathogenicity.” Nature Communications 8 (February): 14366.
Rice, P., I. Longden, and A. Bleasby. 2000. “EMBOSS: The European Molecular Biology Open Software Suite.” Trends in Genetics: TIG 16 (6): 276–77.
Riehle, M. M., A. F. Bennett, and A. D. Long. 2001. “Genetic Architecture of Thermal Adaptation in Escherichia Coli.” Proceedings of the National Academy of Sciences of the United States of America 98 (2): 525–30.
Rigau, Maria, David Juan, Alfonso Valencia, and Daniel Rico. 2018. “Intronic CNVs Cause Gene Expression Variation in Human Population.” bioRxiv. https://doi.org/10.1101/171165.
Rippey, Caitlin, Tom Walsh, Suleyman Gulsuner, Matt Brodsky, Alex S. Nord, Molly Gasperini, Sarah Pierce, et al. 2013. “Formation of Chimeric Genes by Copy-Number Variation as a Mutational Mechanism in Schizophrenia.” American Journal of Human Genetics 93 (4): 697–710.
Riveron, Jacob M., Sulaiman S. Ibrahim, Emmanuel Chanda, Themba Mzilahowa, Nelson Cuamba, Helen Irving, Kayla G. Barnes, Miranda Ndula, and Charles S. Wondji. 2014. “The Highly Polymorphic CYP6M7 Cytochrome P450 Gene Partners with the Directionally Selected CYP6P9a and CYP6P9b Genes to Expand the Pyrethroid Resistance Front in the Malaria Vector Anopheles Funestus in Africa.” BMC Genomics 15 (1): 817.
Riveron, Jacob M., Helen Irving, Miranda Ndula, Kayla G. Barnes, Sulaiman S. Ibrahim, Mark J. I. Paine, and Charles S. Wondji. 2013. “Directionally Selected Cytochrome P450 Alleles Are Driving the Spread of Pyrethroid Resistance in the Major Malaria Vector Anopheles Funestus.” Proceedings of the National Academy of Sciences of the United States of America 110 (1): 252–57.
Rodrigo, Guillermo, and Mario A. Fares. 2018. “Intrinsic Adaptive Value and Early Fate of Gene

Duplication Revealed by a Bottom-up Approach." eLife 7 (January). https://doi.org/10.7554/eLife.29739.
Roth, John R., and D. I. Andersson. 2012. "Poxvirus Use a 'Gene Accordion' to Tune out Host Defenses." Cell 150 (4): 671–72.
Roudnitzky, Natacha, Davide Risso, Dennis Drayna, Maik Behrens, Wolfgang Meyerhof, and Stephen P. Wooding. 2016. "Copy Number Variation inTAS2RBitter Taste Receptor Genes: Structure, Origin, and Population Genetics." Chemical Senses 41 (8): 649–59.
Rownd, R., and S. Mickel. 1971. "Dissociation and Reassociation of RTF and R-Determinants of the R-Factor NR1 in Proteus Mirabilis." Nature: New Biology 234 (45): 40–43.
Rozycki, Torsten von, and Dietrich H. Nies. 2009. "Cupriavidus Metallidurans: Evolution of a Metal-Resistant Bacterium." Antonie van Leeuwenhoek 96 (2): 115–39.
Sakofsky, Cynthia J., Sandeep Ayyar, Angela K. Deem, Woo-Hyun Chung, Grzegorz Ira, and Anna Malkova. 2015. "Translesion Polymerases Drive Microhomology-Mediated Break-Induced Replication Leading to Complex Chromosomal Rearrangements." Molecular Cell 60 (6): 860–72.
Sandegren, Linus, and Dan I. Andersson. 2009. "Bacterial Gene Amplification: Implications for the Evolution of Antibiotic Resistance." Nature Reviews. Microbiology 7 (8): 578–88.
Schrider, Daniel R., Fabio C. P. Navarro, Pedro A. F. Galante, Raphael B. Parmigiani, Anamaria A. Camargo, Matthew W. Hahn, and Sandro J. de Souza. 2013. "Gene Copy-Number Polymorphism Caused by Retrotransposition in Humans." PLoS Genetics 9 (1): e1003242.
Schughart, K., C. Kappen, and F. H. Ruddle. 1989. "Duplication of Large Genomic Regions during the Evolution of Vertebrate Homeobox Genes." Proceedings of the National Academy of Sciences of the United States of America 86 (18): 7067–71.
Scott, Amber L., Phillip A. Richmond, Robin D. Dowell, and Anna M. Selmecki. 2017. "The Influence of Polyploidy on the Evolution of Yeast Grown in a Sub-Optimal Carbon Source." Molecular Biology and Evolution 34 (10): 2690–2703.
Sebat, Jonathan, B. Lakshmi, Jennifer Troge, Joan Alexander, Janet Young, Pär Lundin, Susanne Månér, et al. 2004. "Large-Scale Copy Number Polymorphism in the Human Genome." Science 305 (5683): 525–28.
Sellis, Diamantis, Daniel J. Kvitek, Barbara Dunn, Gavin Sherlock, and Dmitri A. Petrov. 2016. "Heterozygote Advantage Is a Common Outcome of Adaptation in Saccharomyces Cerevisiae." Genetics 203 (3): 1401–13.
Selmecki, Anna, Anja Forche, and Judith Berman. 2006. "Aneuploidy and Isochromosome Formation in Drug-Resistant Candida Albicans." Science 313 (5785): 367–70.
Selmecki, Anna, Maryam Gerami-Nejad, Carsten Paulson, Anja Forche, and Judith Berman. 2008. "An Isochromosome Confers Drug Resistance in Vivo by Amplification of Two Genes, ERG11 and TAC1." Molecular Microbiology 68 (3): 624–41.
Selmecki, Anna M., Yosef E. Maruvka, Phillip A. Richmond, Marie Guillet, Noam Shoresh, Amber L. Sorenson, Subhajyoti De, et al. 2015. "Polyploidy Can Drive Rapid Adaptation in Yeast." Nature 519 (7543): 349–52.
Shapira, Stuart K., and Victoria G. Finnerty. 1986. "The Use of Genetic Complementation in the Study of Eukaryotic Macromolecular Evolution: Rate of Spontaneous Gene Duplication at Two Loci ofDrosophila Melanogaster." Journal of Molecular Evolution 23 (2): 159–67.
Sharp, Andrew J., Devin P. Locke, Sean D. McGrath, Ze Cheng, Jeffrey A. Bailey, Rhea U. Vallente, Lisa M. Pertz, et al. 2005. "Segmental Duplications and Copy-Number Variation in

the Human Genome." American Journal of Human Genetics 77 (1): 78–88.
Sheltzer, Jason M., Eduardo M. Torres, Maitreya J. Dunham, and Angelika Amon. 2012. "Transcriptional Consequences of Aneuploidy." Proceedings of the National Academy of Sciences of the United States of America 109 (31): 12644–49.
Shlien, Adam, and David Malkin. 2009. "Copy Number Variations and Cancer." Genome Medicine 1 (6): 62–62.
Shor, Erika, Catherine A. Fox, and James R. Broach. 2013. "The Yeast Environmental Stress Response Regulates Mutagenesis Induced by Proteotoxic Stress." PLoS Genetics 9 (8): e1003680.
Sidhu, Amar Bir Singh, Anne-Catrin Uhlemann, Stephanie G. Valderramos, Juan-Carlos Valderramos, Sanjeev Krishna, and David A. Fidock. 2006. "Decreasing pfmdr1 Copy Number in Plasmodium Falciparum Malaria Heightens Susceptibility to Mefloquine, Lumefantrine, Halofantrine, Quinine, and Artemisinin." The Journal of Infectious Diseases 194 (4): 528–35.
Siguier, Patricia, Edith Gourbeyre, and Mick Chandler. 2014. "Bacterial Insertion Sequences: Their Genomic Impact and Diversity." FEMS Microbiology Reviews 38 (5): 865–91.
Skourti-Stathaki, Konstantina, and Nicholas J. Proudfoot. 2014. "A Double-Edged Sword: R Loops as Threats to Genome Integrity and Powerful Regulators of Gene Expression." Genes & Development 28 (13): 1384–96.
Slack, Andrew, P. C. Thornton, Daniel B. Magner, Susan M. Rosenberg, and P. J. Hastings. 2006. "On the Mechanism of Gene Amplification Induced under Stress in Escherichia Coli." PLoS Genetics 2 (4): e48.
Slager, Jelle, Morten Kjos, Laetitia Attaiech, and Jan-Willem Veening. 2014. "Antibiotic-Induced Replication Stress Triggers Bacterial Competence by Increasing Gene Dosage near the Origin." Cell 157 (2): 395–406.
Slechta, E. Susan, Kim L. Bunny, Elisabeth Kugelberg, Eric Kofoid, Dan I. Andersson, and John R. Roth. 2003. "Adaptive Mutation: General Mutagenesis Is Not a Programmed Response to Stress but Results from Rare Coamplification of dinB with Lac." Proceedings of the National Academy of Sciences of the United States of America 100 (22): 12847–52.
Smith, J. M., and J. Haigh. 1974. "The Hitch-Hiking Effect of a Favourable Gene." Genetical Research 23 (1): 23–35.
Sonti, R. V., and J. R. Roth. 1989. "Role of Gene Duplications in the Adaptation of Salmonella Typhimurium to Growth on Limiting Carbon Sources." Genetics 123 (1): 19–28.
Stanbrough, M., and B. Magasanik. 1995. "Transcriptional and Posttranslational Regulation of the General Amino Acid Permease of Saccharomyces Cerevisiae." Journal of Bacteriology 177 (1): 94–102.
———. 1996. "Two Transcription Factors, Gln3p and Nil1p, Use the Same GATAAG Sites to Activate the Expression of GAP1 of Saccharomyces Cerevisiae." Journal of Bacteriology 178 (8): 2465–68.
Stankiewicz, Paweł, and James R. Lupski. 2002. "Genome Architecture, Rearrangements and Genomic Disorders." Trends in Genetics: TIG 18 (2): 74–82.
Stapley, Jessica, Julia Reger, Philine G. D. Feulner, Carole Smadja, Juan Galindo, Robert Ekblom, Clair Bennison, Alexander D. Ball, Andrew P. Beckerman, and Jon Slate. 2010. "Adaptation Genomics: The next Generation." Trends in Ecology & Evolution 25 (12): 705–12.

Starlinger, P. 1977. “DNA Rearrangements in Procaryotes.” Annual Review of Genetics 11: 103–26.
Steinrueck, Magdalena, and Călin C. Guet. 2017. “Complex Chromosomal Neighborhood Effects Determine the Adaptive Potential of a Gene under Selection.” eLife 6 (July). https://doi.org/10.7554/eLife.25100.
Stenberg, Per, and Jan Larsson. 2011. “Buffering and the Evolution of Chromosome-Wide Gene Regulation.” Chromosoma 120 (3): 213–25.
Stingele, Silvia, Gabriele Stoehr, Karolina Peplowska, Jürgen Cox, Matthias Mann, and Zuzana Storchova. 2012. “Global Analysis of Genome, Transcriptome and Proteome Reveals the Response to Aneuploidy in Human Cells.” Molecular Systems Biology 8: 608.
Stranger, Barbara E., Matthew S. Forrest, Mark Dunning, Catherine E. Ingle, Claude Beazley, Natalie Thorne, Richard Redon, et al. 2007. “Relative Impact of Nucleotide and Copy Number Variation on Gene Expression Phenotypes.” Science 315 (5813): 848–53.
Stratton, Michael R., Peter J. Campbell, and P. Andrew Futreal. 2009. “The Cancer Genome.” Nature 458 (7239): 719–24.
Stuber, C. W., and M. M. Goodman. 1983. “Inheritance, Intracellular Localization, and Genetic Variation of Phosphoglucomutase Isozymes in Maize (Zea Mays L.).” Biochemical Genetics 21 (7-8): 667–89.
Sturtevant, A. H. 1925. “The Effects of Unequal Crossing over at the Bar Locus in Drosophila.” Genetics 10 (2): 117–47.
Sun, Song, Otto G. Berg, John R. Roth, and Dan I. Andersson. 2009. “Contribution of Gene Amplification to Evolution of Increased Antibiotic Resistance in Salmonella Typhimurium.” Genetics 182 (4): 1183–95.
Sun, Song, Rongqin Ke, Diarmaid Hughes, Mats Nilsson, and Dan I. Andersson. 2012. “Genome-Wide Detection of Spontaneous Chromosomal Rearrangements in Bacteria.” PloS One 7 (8): e42639.
Sutton, Tim, Ute Baumann, Julie Hayes, Nicholas C. Collins, Bu-Jun Shi, Thorsten Schnurbusch, Alison Hay, et al. 2007. “Boron-Toxicity Tolerance in Barley Arising from Efflux Transporter Amplification.” Science 318 (5855): 1446–49.
Suwanarusk, R., M. Chavchich, B. Russell, A. Jaidee, F. Chalfein, M. Barends, B. Prasetyorini, et al. 2008. “Amplification of pvmdr1 Associated with Multidrug-Resistant Plasmodium Vivax.” The Journal of Infectious Diseases 198 (10): 1558–64.
Suzuki, Yo, Robert P. St Onge, Ramamurthy Mani, Oliver D. King, Adrian Heilbut, Vyacheslav M. Labunskyy, Weidong Chen, et al. 2011. “Knocking out Multigene Redundancies via Cycles of Sexual Assortment and Fluorescence Selection.” Nature Methods 8 (2): 159–64.
Tang, Yun-Chi, and Angelika Amon. 2013. “Gene Copy-Number Alterations: A Cost-Benefit Analysis.” Cell 152 (3): 394–405.
Taylor, John S., and Jeroen Raes. 2004. “Duplication and Divergence: The Evolution of New Genes and Old Ideas.” Annual Review of Genetics 38: 615–43.
Thomas, Barbara J., and Rodney Rothstein. 1989. “Elevated Recombination Rates in Transcriptionally Active DNA.” Cell 56 (4): 619–30.
Torres, Eduardo M., Noah Dephoure, Amudha Panneerselvam, Cheryl M. Tucker, Charles A. Whittaker, Steven P. Gygi, Maitreya J. Dunham, and Angelika Amon. 2010. “Identification of Aneuploidy-Tolerating Mutations.” Cell 143 (1): 71–83.
Torres, Eduardo M., Tanya Sokolsky, Cheryl M. Tucker, Leon Y. Chan, Monica Boselli, Maitreya

J. Dunham, and Angelika Amon. 2007. “Effects of Aneuploidy on Cellular Physiology and Cell Division in Haploid Yeast.” Science 317 (5840): 916–24.
Tran, Phong Lan Thao, Thomas J. Pohl, Chi-Fu Chen, Angela Chan, Sebastian Pott, and Virginia A. Zakian. 2017. “PIF1 Family DNA Helicases Suppress R-Loop Mediated Genome Instability at tRNA Genes.” Nature Communications 8 (April): 15025.
Tsai, Chia-Ti, Chia-Shan Hsieh, Sheng-Nan Chang, Eric Y. Chuang, Kwo-Chang Ueng, Chin-Feng Tsai, Tsung-Hsien Lin, et al. 2016. “Genome-Wide Screening Identifies a KCNIP1 Copy Number Variant as a Genetic Predictor for Atrial Fibrillation.” Nature Communications 7 (February): 10190.
Turner, Daniel J., Marcos Miretti, Diana Rajan, Heike Fiegler, Nigel P. Carter, Martyn L. Blayney, Stephan Beck, and Matthew E. Hurles. 2008. “Germline Rates of de Novo Meiotic Deletions and Duplications Causing Several Genomic Disorders.” Nature Genetics 40 (1): 90–95.
Turner, Kristen M., Viraj Deshpande, Doruk Beyter, Tomoyuki Koga, Jessica Rusert, Catherine Lee, Bin Li, et al. 2017. “Extrachromosomal Oncogene Amplification Drives Tumour Evolution and Genetic Heterogeneity.” Nature 543 (7643): 122–25.
Turner, Thomas L., Elizabeth C. Bourne, Eric J. Von Wettberg, Tina T. Hu, and Sergey V. Nuzhdin. 2010. “Population Resequencing Reveals Local Adaptation of Arabidopsis Lyrata to Serpentine Soils.” Nature Genetics 42 (3): 260–63.
Turner, Thomas L., Andrew D. Stewart, Andrew T. Fields, William R. Rice, and Aaron M. Tarone. 2011. “Population-Based Resequencing of Experimentally Evolved Populations Reveals the Genetic Basis of Body Size Variation in Drosophila Melanogaster.” PLoS Genetics 7 (3): e1001336.
Usakin, Lev A., Galina L. Kogan, Alla I. Kalmykova, and Vladimir A. Gvozdev. 2005. “An Alien Promoter Capture as a Primary Step of the Evolution of Testes-Expressed Repeats in the Drosophila Melanogaster Genome.” Molecular Biology and Evolution 22 (7): 1555–60.
Valdivia-Anistro, Jorge A., Luis E. Eguiarte-Fruns, Gabriela Delgado-Sapién, Pedro Márquez-Zacarías, Jaime Gasca-Pineda, Jennifer Learned, James J. Elser, Gabriela Olmedo-Alvarez, and Valeria Souza. 2015. “Variability of rRNA Operon Copy Number and Growth Rate Dynamics of Bacillus Isolated from an Extremely Oligotrophic Aquatic Ecosystem.” Frontiers in Microbiology 6: 1486.
Vilenchik, Michael M., and Alfred G. Knudson. 2003. “Endogenous DNA Double-Strand Breaks: Production, Fidelity of Repair, and Induction of Cancer.” Proceedings of the National Academy of Sciences of the United States of America 100 (22): 12871–76.
Vincent, Isabel M., Darren Creek, David G. Watson, Mohammed A. Kamleh, Debra J. Woods, Pui Ee Wong, Richard J. S. Burchmore, and Michael P. Barrett. 2010. “A Molecular Mechanism for Eflornithine Resistance in African Trypanosomes.” PLoS Pathogens 6 (11): e1001204.
Wala, Jeremiah, Pratiti Bandopadhayay, Noah Greenwald, Ryan O’Rourke, Ted Sharpe, Chip Stewart, Steven E. Schumacher, et al. 2017. “Genome-Wide Detection of Structural Variants and Indels by Local Assembly.” bioRxiv. https://doi.org/10.1101/105080.
Walsh, Bruce. 2003. “Population-Genetic Models of the Fates of Duplicate Genes.” Genetica 118 (2-3): 279–94.
Walsh, J. B. 1987. “Sequence-Dependent Gene Conversion: Can Duplicated Genes Diverge Fast Enough to Escape Conversion?” Genetics 117 (3): 543–57.
———. 1995. “How Often Do Duplicated Genes Evolve New Functions?” Genetics 139 (1).

Whoriskey, S. K., V. H. Nghiem, P. M. Leong, J. M. Masson, and J. H. Miller. 1987. “Genetic Rearrangements and Gene Amplification in Escherichia Coli: DNA Sequences at the Junctures of Amplified Gene Fusions.” Genes & Development 1 (3): 227–37.
Williams, Thomas N., Sammy Wambua, Sophie Uyoga, Alex Macharia, Jedidah K. Mwacharo, Charles R. J. C. Newton, and Kathryn Maitland. 2005. “Both Heterozygous and Homozygous Alpha+ Thalassemias Protect against Severe and Fatal Plasmodium Falciparum Malaria on the Coast of Kenya.” Blood 106 (1): 368–71.
Wilson, Thomas E., Martin F. Arlt, So Hae Park, Sountharia Rajendran, Michelle Paulsen, Mats Ljungman, and Thomas W. Glover. 2015. “Large Transcription Units Unify Copy Number Variants and Common Fragile Sites Arising under Replication Stress.” Genome Research 25 (2): 189–200.
Winfield, S. L., and J. O. Falkinham 3rd. 1981. “Effect of recA and polA Mutations on Gene Duplication in Salmonella Typhimurium.” Mutation Research 91 (1): 15–20.
Winston, F., D. T. Chaleff, B. Valent, and G. R. Fink. 1984. “Mutations Affecting Ty-Mediated Expression of the HIS4 Gene of Saccharomyces Cerevisiae.” Genetics 107 (2): 179–97.
Wright, Dominic, Henrik Boije, Jennifer R. S. Meadows, Bertrand Bed’hom, David Gourichon, Agathe Vieaud, Michèle Tixier-Boichard, et al. 2009. “Copy Number Variation in Intron 1 of SOX5 Causes the Pea-Comb Phenotype in Chickens.” PLoS Genetics 5 (6): e1000512–e1000512.
Würschum, Tobias, Philipp H. G. Boeven, Simon M. Langer, C. Friedrich H. Longin, and Willmar L. Leiser. 2015. “Multiply to Conquer: Copy Number Variations at Ppd-B1 and Vrn-A1 Facilitate Global Adaptation in Wheat.” BMC Genetics 16 (July): 96.
Xi, Ruibin, Angela G. Hadjipanayis, Lovelace J. Luquette, Tae-Min Kim, Eunjung Lee, Jianhua Zhang, Mark D. Johnson, et al. 2011. “Copy Number Variation Detection in Whole-Genome Sequencing Data Using the Bayesian Information Criterion.” Proceedings of the National Academy of Sciences of the United States of America 108 (46): E1128–36.
Yagi, Y., and D. B. Clewell. 1976. “Plasmid-Determined Tetracycline Resistance in Streptococcus Faecalis: Tandemly Repeated Resistance Determinants in Amplified Forms of pAMalpha1 DNA.” Journal of Molecular Biology 102 (3): 583–600.
Yang, Fan, David A. Pecina, Shelly D. Kelly, Sang-Hoon Kim, Kenneth M. Kemner, David T. Long, and Terence L. Marsh. 2010. “Biosequestration via Cooperative Binding of Copper by Ralstonia Pickettii.” Environmental Technology 31 (8-9): 1045–60.
Yant, Levi, and Kirsten Bomblies. 2017. “Genomic Studies of Adaptive Evolution in Outcrossing Arabidopsis Species.” Current Opinion in Plant Biology 36 (April): 9–14.
Ye, Christine J., Sarah Regan, Guo Liu, Sarah Alemara, and Henry H. Heng. 2018. “Understanding Aneuploidy in Cancer through the Lens of System Inheritance, Fuzzy Inheritance and Emergence of New Genome Systems.” Molecular Cytogenetics 11 (May): 31.
Ye, Kai, Marcel H. Schulz, Quan Long, Rolf Apweiler, and Zemin Ning. 2009. “Pindel: A Pattern Growth Approach to Detect Break Points of Large Deletions and Medium Sized Insertions from Paired-End Short Reads.” Bioinformatics 25 (21): 2865–71.
Yoder, Anne D., Jelmer Poelstra, George P. Tiley, and Rachel Williams. 2018. “Neutral Theory Is the Foundation of Conservation Genetics.” Molecular Biology and Evolution, April. https://doi.org/10.1093/molbev/msy076.
Zarrei, Mehdi, Jeffrey R. MacDonald, Daniele Merico, and Stephen W. Scherer. 2015. “A Copy

Number Variation Map of the Human Genome.” Nature Reviews. Genetics 16 (3): 172–83.
Zerulla, Karolin, and Jörg Soppa. 2014. “Polyploidy in Haloarchaea: Advantages for Growth and Survival.” Frontiers in Microbiology 5 (June): 274.
Zhang, Feng, Mehrdad Khajavi, Anne M. Connolly, Charles F. Towne, Sat Dev Batish, and James R. Lupski. 2009. “The DNA Replication FoSTeS/MMBIR Mechanism Can Generate Genomic, Genic and Exonic Complex Rearrangements in Humans.” Nature Genetics 41 (7): 849–53.
Zhang, F., W. Gu, M. E. Hurles, and J. R. Lupski. 2009. “Copy Number Variation in Human Health, Disease, and Evolution.” Annual Review of Genomics and Human Genetics 10: 451–81.
Zhao, Lu, Zhimin Liu, Sasha F. Levy, and Song Wu. 2017. “Bartender: A Fast and Accurate Clustering Algorithm to Count Barcode Reads.” Bioinformatics , October. https://doi.org/10.1093/bioinformatics/btx655.
Zhao, Ying, Pooja K. Strope, Stanislav G. Kozmin, John H. McCusker, Fred S. Dietrich, Robert J. Kokoska, and Thomas D. Petes. 2014. “Structures of Naturally Evolved CUP1 Tandem Arrays in Yeast Indicate That These Arrays Are Generated by Unequal Nonhomologous Recombination.” G3  4 (11): 2259–69.
Zhou, Dan, Nitin Udpa, Merril Gersten, Deeann W. Visk, Ali Bashir, Jin Xue, Kelly A. Frazer, et al. 2011. “Experimental Selection of Hypoxia-Tolerant Drosophila Melanogaster.” Proceedings of the National Academy of Sciences of the United States of America 108 (6): 2349–54.
Zhu, Jie, Stephen Pearce, Adrienne Burke, Deven Robert See, Daniel Z. Skinner, Jorge Dubcovsky, and Kimberly Garland-Campbell. 2014. “Copy Number and Haplotype Variation at the VRN-A1 and Central FR-A2 Loci Are Associated with Frost Tolerance in Hexaploid Wheat.” TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik 127 (5): 1183–97.
Zichner, Thomas, David A. Garfield, Tobias Rausch, Adrian M. Stütz, Enrico Cannavó, Martina Braun, Eileen E. M. Furlong, and Jan O. Korbel. 2013. “Impact of Genomic Structural Variation in Drosophila Melanogaster Based on Population-Scale Sequencing.” Genome Research 23 (3): 568–79.
Żmieńko, Agnieszka, Anna Samelak, Piotr Kozłowski, and Marek Figlerowicz. 2014. “Copy Number Polymorphism in Plant Genomes.” TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik 127 (1): 1–18.
Zuellig, Matthew P., and Andrea L. Sweigart. 2018. “Gene Duplicates Cause Hybrid Lethality between Sympatric Species of Mimulus.” PLoS Genetics 14 (4): e1007130.