Opposing Copy Number Variation Dynamics Shape Adaptation to Glucose and Galactose in Diploid Yeast

Prachitha Nagendra[1], Saket Choudhary[2], Supreet Saini[*,1]

[1]Department of Chemical Engineering, Indian Institute of Technology Bombay, Mumbai, India 400 076

[2]Koita Centre for Digital Health, Indian Institute of Technology Bombay, Mumbai, India 400 076

* Corresponding Author. Email: saini@che.iitb.ac.in, Phone: +91 22 2576 7216

Running Title: *CNVs Drive Yeast Adaptation*

Abstract

Adaptation to constant environments is often thought to proceed through point mutations that fine-tune gene function. However, structural variation, such as gene duplications and deletions, can also reshape genomes and drive rapid phenotypic change. Yet, how these classes of mutations jointly influence long-term adaptation remains unclear. Here we evolve six replicate diploid *Saccharomyces cerevisiae* populations for 1200 generations in either glucose or galactose and show that adaptation is dominated by large-scale copy number variations (CNVs) rather than single-nucleotide polymorphisms. In glucose, adaptation proceeds through early, extensive telomeric deletions that target carbon-use modules, followed by later compensatory duplications that restore metabolic breadth. In galactose, adaptation is marked instead by persistent telomeric and subtelomeric duplications that reinforce specialization. Opposing selection acted on overlapping sets of genes, with loci deleted during glucose adaptation becoming duplicated during galactose adaptation, linking structural remodeling to divergent physiological strategies. Contrary to models emphasizing SNP accumulation, these findings demonstrate that predictable CNV trajectories dominate genome evolution in stable environments, and that the direction and persistence of such structural changes are constrained by both the ancestral genotype and the regulatory architecture of the cell.

Keywords: Experimental evolution; Saccharomyces cerevisiae; Copy Number Variation (CNV); Structural Variants.

Introduction

Natural environments are rarely constant[1]. Organisms frequently encounter transient changes in nutrient availability, temperature, or biotic interactions, which require populations not only to adapt to their current environment but also to maintain resilience against future fluctuations[2–4]. A central question in evolutionary biology is whether the capacity to tolerate such variability is encoded in the genetic architecture of populations, even when they have evolved in a single, stable environment[5–7]. Understanding these constraints is key to predicting evolutionary trajectories[8–11].

Even under stable conditions, evolution can take multiple molecular routes to the same phenotypic outcome, raising questions about predictability, constraint, and the genetic mechanisms that dominate over long timescales[5,6,11,12]. Classical models emphasize point mutations as the main substrate of adaptation[13–16], yet structural variants such as gene duplications and deletions can produce much larger phenotypic effects and respond rapidly to selection[7,17–19]. Diploid genomes, in particular, add complexity by allowing recessive alleles to hide and by permitting large-scale copy number variation that can rewire gene dosage and network architecture[20–23]. Disentangling how these mutation classes contribute to adaptation, and whether their trajectories are predictable across environments that engage distinct regulatory and metabolic modules, remains a key open problem[10,24,25].

Experimental evolution in microbes provides a tractable system to address these questions, as populations can be propagated under controlled conditions, and phenotypic and genetic changes can be monitored over hundreds of generations[16,24,26,27]. In this study, we evolved twelve populations of diploid *Saccharomyces cerevisiae* (*S. cerevisiae*) in either glucose or galactose for 1200 generations, environments that differ in metabolic requirements and gene regulation. In recent years, many studies have examined the physiological responses and adaptive strategies of yeast in glucose and galactose environments[28–35]. By assaying fitness in the home environment, both when sub-cultured from home environment and when brought after growth in the alternative environment, we quantify specialization, pleiotropic costs, and history-dependent effects.

Here, we systematically track CNVs and SNPs in glucose- and galactose-evolved populations across four time points, combining genome sequencing with phenotypic assays. This design allows us to ask: (1) How does adaptation in an environment affect fitness in an alternative environment across time? (2) Which genomic changes underlie specialist and generalist phenotypes? (3) Are there recurrent or predictable patterns of CNVs that correspond to fitness trajectories?

Our results show that adaptation is primarily driven by CNVs rather than SNPs. Glucose-evolved populations initially specialize, displaying early fitness costs that are later mitigated, while galactose-evolved populations accumulate persistent duplications that reinforce specialization. Many of the same genes are targeted in opposing ways in the two environments, revealing contrasting strategies of genomic remodeling in similar environments. Overall, these findings highlight the central role of

CNVs in shaping predictable adaptive trajectories and environment-dependent trade-offs, providing insight into the genetic architecture underlying specialization and generalist evolution.

Results

Adaptive dynamics and genomic architecture of glucose- and galactose-evolved populations. We evolved six replicate diploid *Saccharomyces cerevisiae* populations in either 0.5% glucose or 0.5% galactose for 1200 generations. Fitness was assayed every 300 generations, while whole-genome sequencing of all six glucose- and six galactose-adapted populations, each at four time points was performed to identify the underlying genetic basis of adaptation **(Figure 1A)**.

In both environments, populations initially experienced a fitness increase in their home environment, which later fluctuated mildly. Upon assaying fitness by preculturing in the same medium, we see that by 1200 generations, glucose-adapted populations grew ~20–30% faster than their ancestor in glucose, while galactose-adapted populations improved ~15–25% in galactose **(Figure 1B and Supplement Figure S1)**.

Fitness changes correspond to environment-specific copy number variation dynamics. Despite the different selective environments, we observed single-nucleotide polymorphisms (SNPs) to be rare across all lines and time points (**Figure 1C and Supplement Sheets 1 and 2**). Contrastingly, we found the dominant mode of genomic change was copy-number variation (CNV), which occurred repeatedly and independently in multiple replicate lines **(Figure 1D-E and Supplement Sheets 3 and 4)**. Thus, adaptation in both glucose and galactose populations was primarily driven by copy number variations rather than point mutations.

Fitness gains were accompanied by widespread CNVs whose signatures differed sharply between the two environments. Overall, glucose lines carried roughly equal numbers of deletions and duplications, while galactose lines accumulated several-fold more duplications than deletions, a bias that remained consistent across all four time points **(Supplement Sheets 5 and 6)**.

However, duplications and deletions exhibited strong contrasts, depending on the environment in which the populations evolved. In glucose-evolved populations, deletions were frequent, particularly in telomeric regions of all chromosomes, except chromosomes I and II **(Figure 2 and Supplement Figure S2)**. The extent of these deletions was most pronounced in the early intervals (first 300 generations) but declined with further evolution, suggesting that early adaptive gains involved a large number of deletions that were later refined or compensated by subsequent changes.

In contrast, galactose-evolved populations exhibited widespread telomeric and sub-telomeric duplications **(Figure 3 and Supplement Figure S3)** except in chromosome II. These duplications were evident from as early as 300 generations and were maintained or expanded throughout 1200 generations, indicating persistent selective benefit in the galactose environment. Strikingly, 217 genes that were recurrently deleted in glucose-evolved lines were found duplicated in galactose-evolved lines during some stage of the evolution experiment, highlighting the opposing selective pressures

acting on the same genetic loci as populations adapted in the two environments **(Figure 4, Supplement Figure S4a and S4b)**.

Specialization trajectories are mirrored in genomic remodeling.

Adaptation to glucose increases fitness. However, we asked if there were particular costs associated with this adaptation. Specifically, we investigated the cost of adaptation in the ability of the cells to switch environments and exhibit growth. To quantify this, we compared the glucose-evolved lines' growth to the glucose ancestor, when the cells were first pre-cultured in galactose. Glucose-evolved populations showed a strong penalty or growth defect at 300 generations because of prior exposure to galactose, consistent with the cost of specialization on glucose **(Figure 5)**. However, this penalty declined sharply by generation 900, reaching nearly zero, before increasing slightly again later in the experiment. This dynamic trajectory suggested a transient phase of specialization followed by the acquisition of mutations conferring generalist performance.

The genomic data shows that early adaptation coincided with extensive telomeric deletions, including in regions encompassing genes involved in galactose metabolism and alternative carbon source utilization. Pathway-level analyses revealed frequent targeting of galactose metabolism, pyruvate metabolism, fructose and mannose metabolism, and the glyoxylate pathway **(Supplement Figure S5)**. These deletions likely conferred rapid specialization to glucose but incurred fitness costs when preculture history involved galactose. Later in evolution, however, the extent of deletions decreased, suggesting the appearance of compensatory CNVs (including duplications) that may have restored broader metabolic flexibility and enabled the observed decline in fitness penalties.

In contrast, galactose-evolved populations showed steadily increasing penalties across the 1200 generations, consistent with a trajectory of deepening specialization **(Figure 5)**. The underlying CNV patterns reinforced this phenotype: galactose lines accumulated persistent duplications in telomeric and subtelomeric regions, often in loci involved in glycolysis/gluconeogenesis, starch and sucrose metabolism, and galactose metabolism **(Supplement Figure S6)**. Rather than being compensated, these duplications were maintained and amplified over time, reinforcing a genetic architecture that promoted strong specialization and increasing costs of cross-environment exposure.

Physiological history dependence and its genomic correlates.

We further tested the influence of short-term growth history on fitness. We grew galactose-evolved populations briefly in glucose (~15 generations) and then returned them to galactose media. The transferred population displayed reduced growth rates in galactose, but this penalty was mitigated when, following growth in glucose, cells were re-exposed to galactose for 7 or 15 generations **(Figure 6A)**. This plasticity suggests that the specialized physiology of galactose-adapted lines is sensitive to immediate history but can be restored with sufficient exposure to galactose. Genomic signatures provide a plausible explanation: the duplications enriched in galactose-evolved lines were heavily

biased toward genes encoding metabolic enzymes and transporters, creating a cellular state optimized for galactose but disrupted by prior glucose exposure.

Glucose-evolved populations behaved differently. During the first 600 generations, the duration of glucose preculture (7 vs. 15 generations) influenced measured growth rates, consistent with continued sensitivity to short-term history **(Figure 6B)**. However, beyond generation 600, this effect disappeared: fitness became independent of preculture duration. This history-independence parallels the genomic trend in which extensive deletions present early in evolution diminished later, with glucose lines acquiring duplications that likely stabilized metabolism and reduced sensitivity to environmental history.

Predictability of evolutionary outcomes.

To investigate the contributors to fitness of the evolved populations in different conditions, we analyzed our dataset using a linear model incorporating ancestor fitness, preconditioning environment, recovery duration, evolutionary time, and their interactions. We ran an ANOVA analysis which showed fitness of the ancestral strain as the strongest predictor of evolved fitness ($\eta^2 \approx 0.71$, $p < 10^{-62}$), demonstrating that the initial genotype largely sets the baseline for subsequent adaptation. Evolutionary time also contributed significantly to fitness changes ($\eta^2 \approx 0.22$, $p < 10^{-28}$), consistent with the gradual accumulation of adaptive changes observed in both glucose- and galactose-evolved lines.

Other main effects, including preconditioning environment and recovery duration, were individually non-significant. However, modestly significant interactions with evolutionary time were detected (preconditioning × evolution: $p \approx 0.0014$; recovery × evolution: $p \approx 0.027$), indicating that environmental context can subtly modulate the pace or trajectory of adaptation. Higher-order interactions contributed minimally to overall variance in fitness.

Linear regression analysis further supported these conclusions. The coefficient for ancestral fitness was large and highly significant ($\beta = 1.324$, $p < 0.001$), whereas the effect of evolutionary time per generation was small but detectable ($\beta = 8.7 \times 10^{-5}$, $p < 0.001$). Interaction terms were largely negligible, emphasizing that while environmental history and short-term physiological conditions can influence fitness to a limited extent, the initial genetic background dominates subsequent evolutionary outcomes.

These findings suggest that the evolutionary response in our experiment is highly predictable at the level of gross fitness changes: the starting genotype establishes a strong constraint on the magnitude and direction of adaptation, while environmental factors such as preconditioning and recovery modulate the response only subtly. When interpreted alongside our genomic analyses, these results imply that predictable structural changes, particularly environment-specific CNVs in key metabolic loci,

act within the bounds set by the ancestral genotype to produce incremental but repeatable fitness gains.

Discussion.

Our genomic and phenotypic analyses reveal two fundamentally different routes to adaptation in glucose and galactose. In glucose, populations initially specialize through extensive telomeric deletions affecting galactose- and carbon-metabolism loci, but these early losses are later countered by compensatory duplications that restore metabolic breadth and generalist growth. In contrast, galactose-adapted populations exhibit persistent subtelomeric duplications across evolution, reinforcing specialization and maintaining high costs upon environmental switching. Over two hundred loci were altered in opposite directions, deleted in glucose yet duplicated in galactose, demonstrating that the two environments impose antagonistic selective pressures on shared genomic targets.

Specialization thus emerges as a common first step in adaptation, but only glucose-evolved populations escape it through subsequent structural remodeling. In galactose, sustained duplication-driven amplification traps populations in a progressively specialized state. Together, these patterns illustrate that adaptation follows environment-specific genomic strategies constrained by both the ancestral genotype and the architecture of gene regulation.

Across 1200 generations in single-carbon environments, adaptation of diploid *S. cerevisiae* was driven predominantly by copy number variations (CNVs) rather than point mutations. Glucose-evolved populations exhibited early telomeric deletions that imposed transient, history-dependent fitness costs following galactose exposure, costs that diminished as duplications later emerged. Galactose-evolved populations, in contrast, accumulated persistent telomeric and subtelomeric duplications, maintaining strong specialization. The same genomic regions were repeatedly targeted in opposing directions across environments, underscoring CNVs as the principal and predictable axis of evolutionary change.

The dominance of CNVs over SNPs reflects their contrasting mutation spectra and context dependence. While the single-nucleotide mutation rate in yeast is ~10^{-10} per site per generation (roughly 10^{-3} per genome)[36,37], structural variants arise orders of magnitude more frequently and are influenced by ploidy, sequence architecture, and replication dynamics[38,39]. Tandem-array rearrangements at loci such as *CUP1* occur near 10^{-7} per division, and loss-of-heterozygosity events often exceed base-substitution rates, particularly in diploids[40].

Although point mutations have long been viewed as the canonical drivers of adaptation, evidence increasingly points to CNVs as rapid, flexible, and often reversible mechanisms of evolutionary change[16,17,19,41–49]. In *S. cerevisiae*, gene amplifications confer metal and toxin resistance, and recurrent CNVs in transporter and metabolic genes underlie ecological specialization across fungi and pathogens[48,50–55]. Yet, how such events shape adaptive trajectories across contrasting environments remains poorly understood. Our findings reveal that CNVs, rather than SNPs, form the structural basis of predictable evolutionary outcomes, dictating whether populations evolve toward generalist or specialist states and defining the genomic logic of adaptation in stable environments[47,56].

More broadly, our findings position CNVs as first-order levers of evolvability in microbes[17,18]. Parallel yet contrasting CNV signatures emerged in the two environments, i.e., mixed deletions/duplications in glucose versus duplication-biased remodeling in galactose. These changes to the genome likely map onto opposite physiological strategies with environment-specific trade-offs. The concentration of changes in telomeric regions and enrichment for transport/metabolic modules suggest that dosage tuning of peripheral genomic neighborhoods provides a rapid route to adaptation when nutrient regimes differ[22,57–59].

Second, the predictability and constraints of evolution are evident in our data: a linear modeling framework shows that ancestral fitness is the strongest determinant of evolved fitness, while evolutionary time contributes modestly and preconditioning/recovery exert only subtle, time-dependent effects[5,12,60–62]. This indicates that structural remodeling proceeds within bounds set by the starting genotype and network architecture, yielding incremental but repeatable fitness gains.

Finally, these results motivate targeted tests and applications. Mechanistically, dissecting causality and reversibility of specific CNVs (engineered gain/loss, LOH dynamics) can clarify how populations exit specialist states[18,21,63–66]. Ecologically and industrially, controlled fluctuation schedules could be used to steer CNV trajectories toward desired phenotypes (e.g., fermentation robustness, drug tolerance control)[25,67,68]. Methodologically, combining barcoded lineage tracking with long-read assemblies should resolve CNV architectures over time and link structural changes to fitness under explicit history manipulations, including the observed loss of history dependence in glucose lines after ~600 generations.

Methods

Strains used and experimental evolution.

A diploid SK1 (*a/α*) strain was generated by mating two isogenic haploids:

- *SK1 MATa ARS314::TRP1 ho::LYS2 lys2 ura3 leu2::hisG his3::hisG trp1::hisG*
- *SK1 MATα ARS314::LEU2 ho::LYS2 lys2 ura3 leu2::hisG his3::hisG trp1::hisG*

For each evolution experiment, the diploid strain was streaked from frozen stock onto YPD agar plates. A single colony was inoculated into glycerol–lactate medium (3 mL 100% glycerol, 2 mL 40% lactate, 50 mg complete amino acid mixture, 660 mg yeast nitrogen base per 100 mL) and cultured for 48 h.

From this starter culture, 1% (v/v) inoculum was transferred daily into fresh medium to establish six independent replicate populations in each of two carbon environments:

- Glucose (0.5% w/v) supplemented with 50 mg complete amino acid mixture and 0.66 mg yeast nitrogen base
- Galactose (0.5% w/v) supplemented with the same additives

All populations were propagated by serial 1:100 dilutions every 24 h and maintained at 30°C, 250 rpm shaking conditions for a total of ~1200 generations.

Growth Rate Measurements as a Proxy for Adaptation.

Growth rates were quantified at four evolutionary time points: 300, 600, 900, and 1200 generations. At each point, populations were revived from frozen stocks and inoculated into the same carbon source in which they had been evolved. Optical density (OD) measurements were then recorded to estimate growth rate, which served as a proxy for fitness.

We denote these assays as follows:

- ggg: populations evolved in glucose, passaged in glucose, and assayed in glucose.
- GGG: populations evolved in galactose, passaged in galactose, and assayed in galactose.
- gGg: populations evolved in glucose, passaged in galactose, and assayed in glucose.
- GgG: populations evolved in galactose, passaged in glucose, and assayed in galactose.

This design ensured that the measured growth rates reflected adaptation to the *native selection environment* of each evolving population.

Reciprocal Environment Assays.

297 To assess the stability and plasticity of adaptation, evolved populations were exposed to the non-home
298 carbon source for a short duration (~15 generations) and subsequently returned to their original
299 environment. For glucose-evolved populations, the non-home environment was 0.5% galactose
300 medium, while for galactose-evolved populations, the non-home environment was 0.5% glucose
301 medium.
302
303 Fitness was assayed in the home environment at three distinct stages:
304
305 - Immediately after growth in the non-home environment
306 - Following ~7 generations of regrowth in the home environment
307 - Following ~15 generations of regrowth in the home environment
308
309 At each assay point, 100–200 µL of culture (volume adjusted according to OD) was inoculated into 10
310 mL of the relevant assay medium. Cultures were incubated at 30 °C with shaking at 250 rpm, and optical
311 density (OD) measurements were collected at regular intervals. Growth dynamics were quantified by
312 calculating a composite growth rate, incorporating both the lag phase duration and the exponential
313 growth rate derived from the OD growth curves.
314
315 Each kinetic assay was performed in triplicate for every population at every time point. In parallel,
316 ancestral strains were revived and assayed under identical conditions, providing a consistent baseline
317 for estimating relative fitness gains during evolution.
318
319 **Whole-Genome Sequencing and Reference Genome Construction.**
320 Evolved populations at four evolutionary time points (300, 600, 900, and 1200 generations) were revived
321 in 5 mL of their respective evolution environments. Cells were harvested, and genomic DNA was
322 extracted using the phenol–chloroform method[70]. Sequencing libraries were prepared, and 150 bp
323 paired-end sequencing was performed on the DNBSEQ platform at ~100× coverage by Haystack
324 Analytics. The ancestral diploid SK1 strain was also sequenced at the same depth for comparison.
325
326 To construct a draft SK1 reference genome, three ancestral haploid SK1 strains from our laboratory
327 were independently sequenced at ~100× coverage. Each haploid genome was aligned to the S288C
328 reference genome using BWA-MEM, and variants were called with GATK v4.4.0.0. Common variants
329 across the three strains were identified and incorporated into the reference using
330 FastaAlternateReferenceMaker, generating a modified reference genome. This process of alignment,
331 variant calling, and incorporation of common polymorphisms was iterated until convergence, yielding a
332 high-confidence SK1 draft genome[42].
333
334 The diploid ancestral SK1 genome was then aligned to this draft reference, and the resulting BAM file
335 was used to polish the assembly with Pilon. Gene annotations were transferred from the S288C genome
336 to the SK1 draft genome using Liftoff, yielding a polished and annotated SK1 reference genome[69,70].

This polished SK1 genome was subsequently used as the reference for alignment of all evolved population sequencing data.

Variant Calling and Filtering.

For each evolved population, sequencing reads were aligned to the polished SK1 reference genome using BWA-MEM. Duplicate reads were identified and marked with Picard Tools. Variant calling was performed with GATK4, following the recommended best practices pipeline: HaplotypeCaller, GenomicsDBImport, and GenotypeGVCFs. Variants were jointly called across all time points for each replicate line to maintain consistency in variant representation. Variant annotation was carried out using SnpEff[71].

From the resulting VCF files, total read depth and allele-specific read depths (reference and alternate alleles) were extracted for each sample. Variants were filtered out based on the following conditions:

- the average coverage across all time points was <30×; and
- all populations across the four time points had <25× coverage; and
- the allele frequency remained unchanged across time.

Mutations were considered fixed when the alternate allele frequency reached thresholds of:

- 0.5–0.6 for heterozygous alternate alleles,
- 0.9–1.0 for homozygous alternate alleles, or
- 0–0.1 for homozygous reference alleles, relative to the reference genome.

All variants were visually validated using the Integrated Genome Viewer (IGV).

Because the number of fixed SNPs was low and many were shared across replicate populations, variant calls for one representative replicate line (across all time points in both environments) were independently verified using an alternative approach. Variants were re-called with bcftools mpileup, followed by filtering with the same thresholds as above. All the 5 SNPs called using GATK in glu1 population (glucose-evolved, replicate line1) were present in the variant list obtained from bcftools mpileup. There were no unique variants to GATK but 1 unique variant to bcftools mpileup. In the case of gal1 population (galactose-evolved, replicate line1), 3 SNPs were common to both the methods and 1 unique to GATK. This confirms the presence of few fixed variants in the population.

The raw sequencing data of all evolved populations is available at

375 http://www.ncbi.nlm.nih.gov/bioproject/1298216.
376
377 **Copy Number Variation (CNV) Analysis.**
378 To ensure comparable sequencing coverage across samples, the median coverage across the genome
379 was calculated for each evolved replicate at every time point and in both evolution environments
380 **(Supplementary Figure S7).**
381
382 A custom coverage-based pipeline was used to identify copy number variants (CNVs) in all evolved
383 populations. First, per-site read depth across the genome was calculated using samtools depth. To
384 detect contiguous genomic regions exhibiting CNVs, the genome was divided into 500 bp windows, and
385 the median coverage per window was determined. Each window's coverage was then normalized by
386 the genome-wide median coverage for that sample. To account for standing variation present in the
387 ancestor, the ratio of normalized coverage in evolved populations to that of the ancestral reference was
388 computed for each window, yielding a measure of relative coverage. All the windows that had a relative
389 coverage less than 0.25 were excluded from further analysis.
390
391 To classify CNV states, we implemented an unsupervised Hidden Markov Model (HMM) allowing
392 discrete states of 0, 0.5, 1, 1.5, 2, 3, and 4. A CNV state of 1 indicated equal copy number in evolved
393 and ancestral genomes, values <1 indicated deletions, and values >1 indicated duplications in the
394 evolved populations[72].
395
396 To validate the CNV calls, one representative replicate from each environment was independently
397 analyzed at a time point that exhibited maximum CNVs (300 generations for glucose and 1200
398 generations for galactose), along with the ancestor, using CNVpytor, a read depth-based tool for CNV
399 analysis[73]. A bin size of 500bp was used, and CNV calls with p-value<0.0001, and Q0<0.5 were filtered.
400 The copy number in each window of evolved populations was normalized with that in the ancestor to
401 account for only the changes that occurred during the experiment. The results were compared with the
402 coverage-based pipeline used in this study by calculating accuracy, precision (weighted average of true
403 positives to sum of true and false positives for each class), and recall (weighted average of true positives
404 to sum of true positives and false negatives for each class). In the glucose-evolved line, results obtained
405 from CNVpytor had more deletions than duplications. On comparing the two methods, the coverage-
406 based pipeline had an accuracy of 86.58%, precision of 86.94% and recall of 86.58%. In the case of
407 galactose-evolved line, duplications were higher than deletions, and the coverage-based pipeline could
408 predict this with accuracy of 79.58% (precision and recall of 79.61% and 79.58% respectively) when
409 compared to CNVpytor. This confirms that the trends of deletions in glucose and duplications in
410 galactose are consistently observed not just in our pipeline but also in another independent tool.
411
412 **CNV Annotation and Pathway Enrichment Analysis**
413 Using the CNV states obtained from coverage-based analysis, all windows with a state of 1 (no change
414 relative to ancestor) were excluded. The remaining windows were categorized as duplications (>1) or

415 deletions (<1). Each set of windows was annotated to identify the genes contained within the windows
416 for every replicate at each time point. Genes of unknown function were filtered out, and annotations
417 across time points were merged within each replicate to obtain the total number of unique genes
418 affected by CNVs. To visualize genomic variation, the average CNV state per window was calculated
419 across replicates and plotted for all time points in both environments.
420
421 For Pathway Enrichment Analysis (PEA), only CNVs shared across all replicates were considered. A t-
422 test was performed for each genomic window based on the relative coverage ratio across replicates.
423 Windows with $p < 0.05$ and a CNV state $\neq 1$ in at least four replicates were identified as high-confidence
424 CNVs. These were further separated into duplications and deletions, annotated, and filtered to exclude
425 genes of unknown function. For each time point and environment, a nonredundant list of unique genes
426 showing CNVs across replicates was compiled.
427
428 Because the resulting gene lists were large (300–1200 genes), they were randomly partitioned into bins
429 of 100 genes, and KEGG pathway enrichment analysis was performed on each bin using the
430 clusterProfiler package in R[57]. This random segregation and enrichment analysis was repeated 3–12
431 times, depending on gene list size. The adjusted p-values from all runs were combined using Fisher's
432 method, and pathways were ranked by gene count and adjusted p-value to identify the most significantly
433 affected pathways. This analysis was performed independently for each time point in both evolution
434 environments.
435

Acknowledgements

This work was supported by the DBT/Wellcome Trust India Alliance grant IA/S/19/2/504632 awarded to SS.

Figures

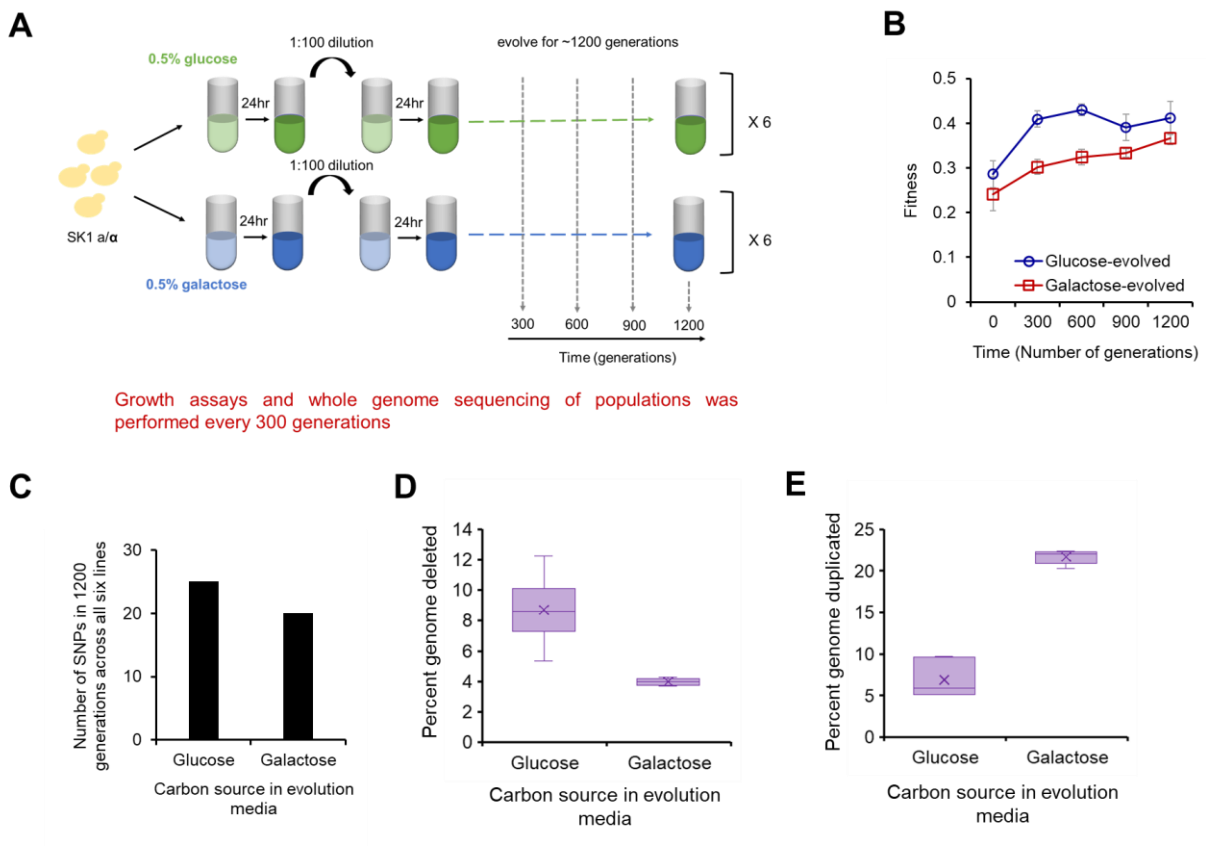


Figure 1. Experimental design, fitness gains, and a genomic overview of adaptation. (A) Schematic of the experimental evolution. Six independent diploid *S. cerevisiae* populations were propagated for 1200 generations in either glucose or galactose. Populations were sampled at 0, 300, 600, 900, and 1200 generations for phenotyping and whole-genome sequencing. **(B)** Home-environment fitness trajectories (mean relative growth rate ± s.e.m., n = 6 lines per environment) show steady improvement in both media; glucose-evolved lines (blue) rise rapidly and then plateau, while galactose-evolved lines (red) increase more gradually. **(C)** SNP accumulation is rare. Bars show the total number of fixed SNPs detected across all six replicate lines over 1200 generations in each environment (joint calling across time points; stringent depth and allele-frequency filters; visual confirmation in IGV). **(D)** and **(E)** Copy-number variation (CNV) dominates genome change and differs by environment. Box-and-whisker plots summarize CNV across 1200 generations. **(D)** deletions only and **(E)** duplications only. Boxes show interquartile range with median; whiskers span full range; × indicates mean. Glucose-evolved lines carry abundant deletions, whereas galactose-evolved lines show persistent, higher duplication loads, exhibiting environment-specific CNV strategies.

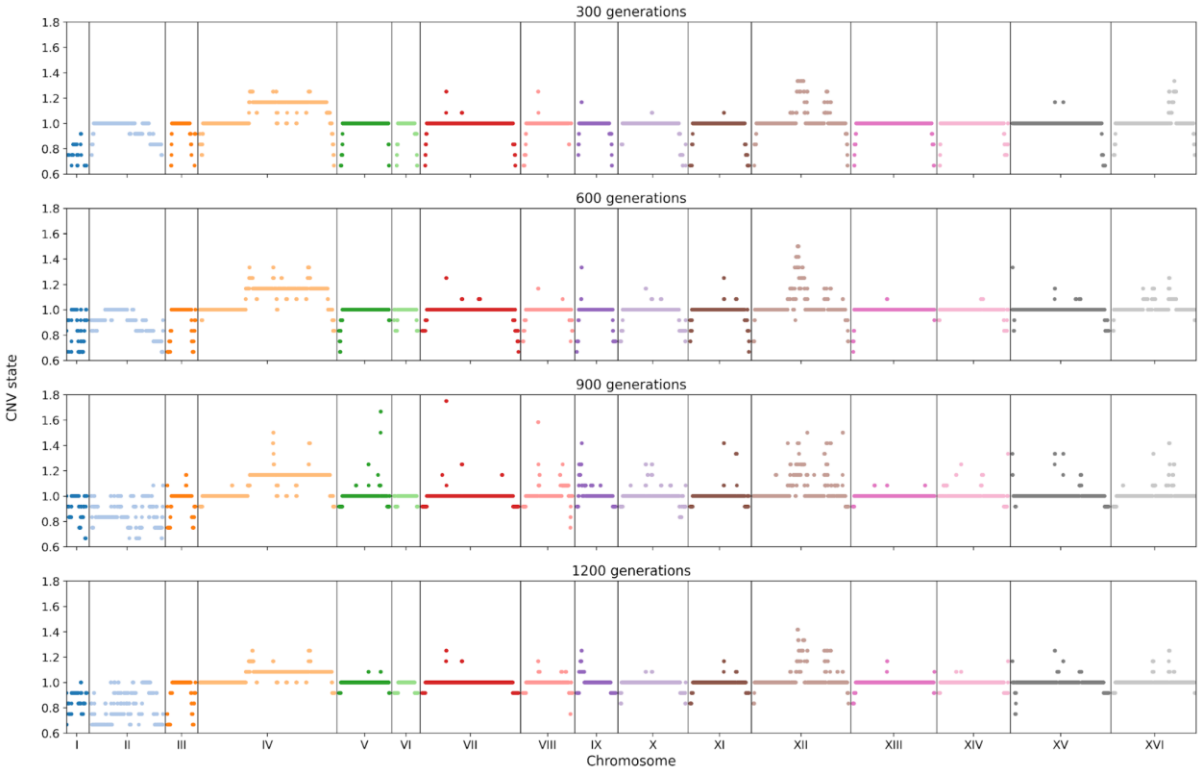


Figure 2. CNV dynamics in glucose-evolved populations across 1200 generations. CNV states for the six glucose-evolved lines at 300, 600, 900, and 1200 generations. The *x*-axis lists the 16 *S. cerevisiae* chromosomes. Each point is a 500-bp genomic window; values are normalized to the ancestor such that state = 1 indicates no change, <1 indicates deletion, and >1 indicates duplication. The colored horizontal bars denote the mean CNV state for that chromosome at that time point. At 300 generations, windows with deletion states (≈0.6–0.95) are widespread and concentrated near chromosome ends, revealing early, extensive telomeric loss. Through 600–900 generations, the deletion burden decreases (mean states trend toward 1). By 1200 generations, most chromosomes sit close to state 1 with fewer and weaker deletions, consistent with refinement/compensation of early structural changes. Overall, the trajectory indicates that early telomeric deletions dominate adaptation in glucose, followed by later remodeling that reduces deletion extent and introduces targeted duplications.

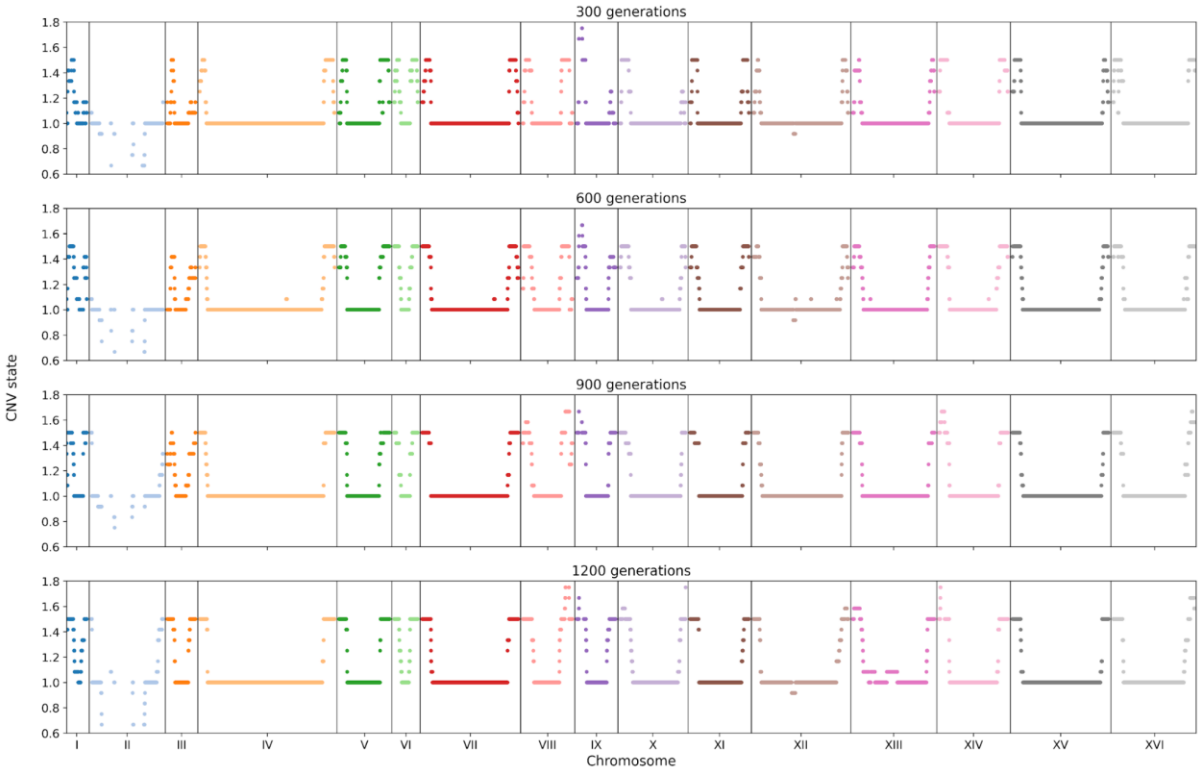
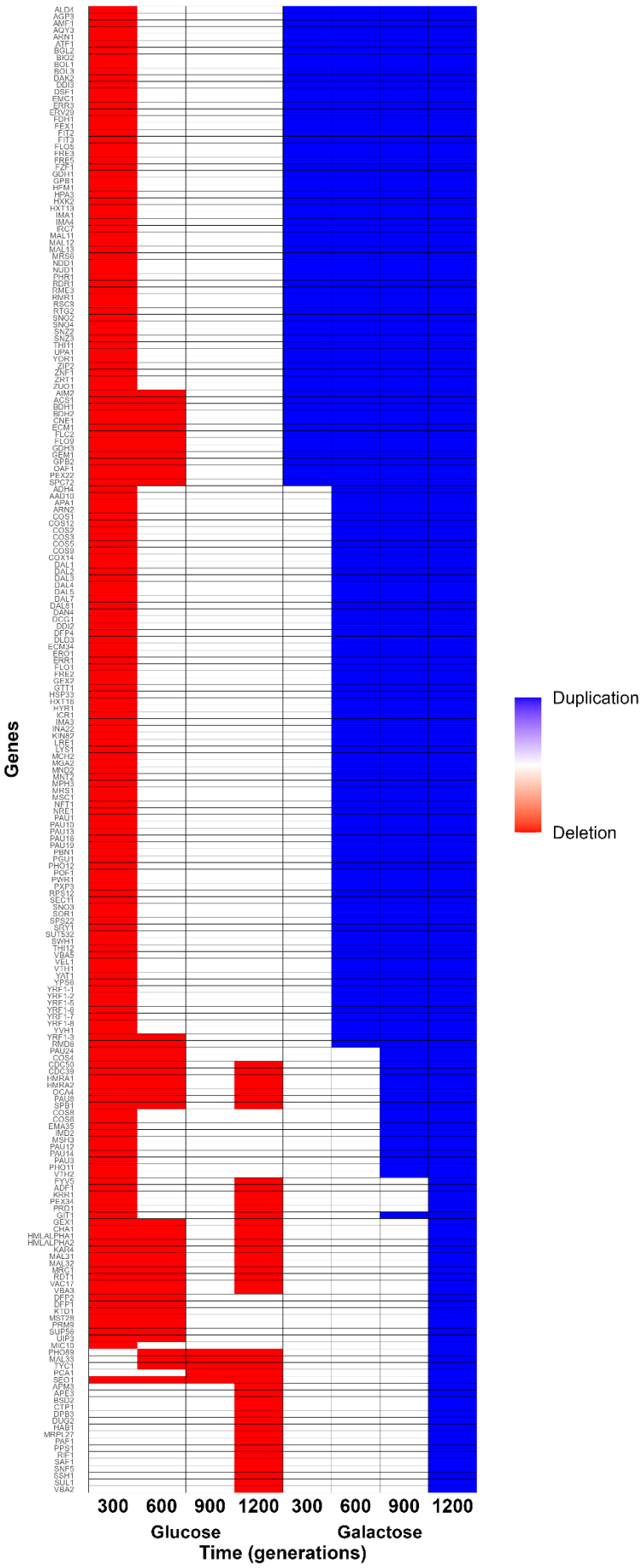


Figure 3. Persistent subtelomeric duplications in galactose-evolved populations. CNV states for the six galactose-evolved lines at 300, 600, 900, and 1200 generations. The *x*-axis lists the 16 *S. cerevisiae* chromosomes. Each point represents a 500-bp window; state = 1 indicates no change, <1 deletion, >1 duplication. The colored horizontal bar marks the mean CNV state for that chromosome at that time point. Across all time points, windows with duplication states (>1.0–1.7) are abundant and cluster near chromosome ends, revealing early-onset, persistent subtelomeric duplications under galactose selection except in chromosome II. Unlike the glucose-evolved trajectories (Figure 2), deletions are rare and mean chromosome states remain consistently >1 from 300 to 1200 generations, indicating sustained duplication loads rather than deletion-first remodeling. Together, these patterns demonstrate a duplication-biased CNV program in galactose that is stable over evolutionary time and concentrated in telomeric neighborhoods enriched for transport/metabolic modules.


Genes
Duplication
Deletion
300
600
900
1200
300
600
900
1200
Glucose
Galactose
Time (generations)

486

Figure 4. **Divergent copy number evolution in glucose versus galactose.** Many genes deleted during adaptation in glucose appear duplicated during adaptation in galactose, visible as rows that are red in the left block and blue in the right block. This highlights environment specific resolution of selection on the same loci over evolutionary time. Each row is a gene. Columns show four time points in glucose (300, 600, 900, 1200 generations) followed by four time points in galactose (300, 600, 900, 1200 generations). Color encodes the direction of copy number change relative to the ancestor: red indicates deletions, blue indicates duplications, white indicates no detected change (scale bar at right).

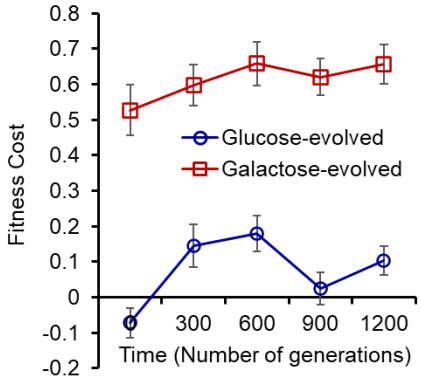


Figure 5. History-dependent fitness costs after cross-environment preculture. For each evolved background, fitness cost was computed as the relative decrease in composite growth rate in the home environment when cells were precultured in the non-home carbon source. Composite growth rate integrates lag and exponential growth from OD time courses (Methods). Lines show the mean cost across six replicate populations at each evolutionary time point (300, 600, 900, 1200 generations); blue, glucose-evolved populations precultured in galactose and assayed in glucose; red, galactose-evolved populations precultured in glucose and assayed in galactose. Glucose-evolved populations display a transient specialization: a pronounced cost early (peaking by ~600 generations) that collapses by 900 generations and remains low thereafter, indicating reduced sensitivity to preculture history. In contrast, galactose-evolved populations show a persistent, increasing cost across the experiment, consistent with deepening specialization to galactose and strong history dependence following glucose exposure.

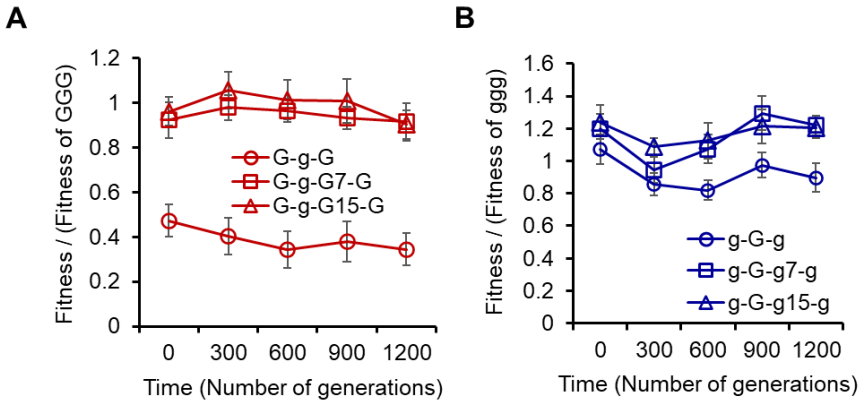


Figure 6. Short-term history effects and recovery in home environment. For each evolutionary background, we measured growth in the home environment after a brief exposure to the non-home carbon source, with or without a recovery phase back in the home medium. Fitness is shown as a ratio to the corresponding home-only control (for (A): divided by fitness of ggg; for (B): divided by fitness of GGG), where "g/G" denote glucose/galactose and the string indicates evolution → preculture → (optional recovery duration) → assay. Thus, g-G-g = glucose-evolved, precultured in galactose, assayed in glucose; g-G-g7-g / g-G-g15-g include ~7 or ~15 generations of recovery in glucose before assay. Analogously, G-g-G, G-g-G7-G, G-g-G15-G are for galactose-evolved lines. Lines show the mean across six replicate populations at each evolutionary time point (0, 300, 600, 900, 1200 generations). Composite fitness integrates lag and exponential growth from OD time courses (Methods). **(A)** Galactose-evolved populations (right) show a strong penalty when assayed immediately after glucose preculture (G-g-G), whereas 7–15 generations of re-exposure to galactose (G-g-G7-G, G-g-G15-G) restore fitness toward the home-only baseline across all evolutionary time points. **(B)** Glucose-evolved populations exhibit history dependence early: performance in glucose after galactose preculture depends on recovery duration (g-G-g < g-G-g7-g ≈ g-G-g15-g) through ~600 generations, after which the curves converge, indicating a loss of history dependence.

References

1. Bell, G. Fluctuating selection: the perpetual renewal of adaptation in variable environments. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **365**, 87–97 (2010).
2. Botero, C. A., Weissing, F. J., Wright, J. & Rubenstein, D. R. Evolutionary tipping points in the capacity to adapt to environmental change. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 184–189 (2015).
3. Chevin, L.-M., Lande, R. & Mace, G. M. Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory. *PLoS Biol.* **8**, e1000357 (2010).
4. Levin, S. A. The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award Lecture. *Ecology* **73**, 1943–1967 (1992).
5. Lässig, M., Mustonen, V. & Walczak, A. M. Predicting evolution. *Nat. Ecol. Evol.* **1**, 77 (2017).
6. Stern, D. L. & Orgogozo, V. Is genetic evolution predictable? *Science* **323**, 746–751 (2009).
7. Orr, H. A. The genetic theory of adaptation: a brief history. *Nat. Rev. Genet.* **6**, 119–127 (2005).
8. Ahlawat, N., Mahilkar, A. & Saini, S. Resource presentation dictates genetic and phenotypic adaptation in yeast. *BMC Ecol. Evol.* **25**, 33 (2025).
9. Ahlawat, N., Venkataraman, P., Brajesh, R. G. & Saini, S. Effects of resource packaging on the adaptative and pleiotropic consequences of evolution. *NPJ Syst. Biol. Appl.* **11**, 78 (2025).
10. Nosil, P., Flaxman, S. M., Feder, J. L. & Gompert, Z. Increasing our ability to predict contemporary evolution. *Nat. Commun.* **11**, 5592 (2020).
11. Kryazhimskiy, S., Rice, D. P., Jerison, E. R. & Desai, M. M. Microbial evolution. Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science* **344**, 1519–1522 (2014).
12. Bakerlee, C. W., Nguyen Ba, A. N., Shulgina, Y., Rojas Echenique, J. I. & Desai, M. M. Idiosyncratic epistasis leads to global fitness-correlated trends. *Science* **376**, 630–635 (2022).
13. Doud, M. B., Lee, J. M. & Bloom, J. D. How single mutations affect viral escape from broad and narrow antibodies to H1 influenza hemagglutinin. *Nat. Commun.* **9**, 1386 (2018).
14. Tenaillon, O. *et al.* Tempo and mode of genome evolution in a 50,000-generation experiment. *Nature* **536**, 165–170 (2016).
15. Lang, G. I. *et al.* Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature* **500**, 571–574 (2013).
16. Barrick, J. E. & Lenski, R. E. Genome dynamics during experimental evolution. *Nat. Rev. Genet.* **14**, 827–839 (2013).
17. Yona, A. H. *et al.* Chromosomal duplication is a transient evolutionary solution to stress. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 21010–21015 (2012).
18. Hull, R. M., Cruz, C., Jack, C. V. & Houseley, J. Environmental change drives accelerated adaptation through stimulated copy number variation. *PLoS Biol.* **15**, e2001333 (2017).
19. Todd, R. T. & Selmecki, A. Expandable and reversible copy number amplification drives rapid adaptation to antifungal drugs. *eLife* **9**, e58349 (2020).
20. Selmecki, A. M. *et al.* Polyploidy can drive rapid adaptation in yeast. *Nature* **519**, 349–352 (2015).
21. James, T. Y. *et al.* Adaptation by Loss of Heterozygosity in Saccharomyces cerevisiae Clones Under Divergent Selection. *Genetics* **213**, 665–683 (2019).
22. O'Donnell, S. *et al.* Telomere-to-telomere assemblies of 142 strains characterize the genome structural landscape in Saccharomyces cerevisiae. *Nat. Genet.* **55**, 1390–1399 (2023).
23. Berman, J. Ploidy plasticity: a rapid and reversible strategy for adaptation to stress. *FEMS Yeast Res.* **16**, fow020 (2016).
24. Levy, S. F. *et al.* Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature* **519**, 181–186 (2015).
25. Nguyen Ba, A. N. *et al.* High-resolution lineage tracking reveals travelling wave of adaptation in laboratory yeast. *Nature* **575**, 494–499 (2019).
26. Elena, S. F. & Lenski, R. E. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat. Rev. Genet.* **4**, 457–469 (2003).
27. Kawecki, T. J. *et al.* Experimental evolution. *Trends Ecol. Evol.* **27**, 547–560 (2012).

28. Escalante-Chong, R. *et al.* Galactose metabolic genes in yeast respond to a ratio of galactose and glucose. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 1636–1641 (2015).
29. Rajeshkannan, null, Mahilkar, A. & Saini, S. GAL Regulon in the Yeast S. cerevisiae is Highly Evolvable via Acquisition in the Coding Regions of the Regulatory Elements of the Network. *Front. Mol. Biosci.* **9**, 801011 (2022).
30. Roop, J. I., Chang, K. C. & Brem, R. B. Polygenic evolution of a sugar specialization trade-off in yeast. *Nature* **530**, 336–339 (2016).
31. Trumbly, R. J. Glucose repression in the yeast Saccharomyces cerevisiae. *Mol. Microbiol.* **6**, 15–21 (1992).
32. Cerulus, B. *et al.* Transition between fermentation and respiration determines history-dependent behavior in fluctuating carbon sources. *eLife* **7**, e39234 (2018).
33. Mahilkar, A., Nagendra, P., Alugoju, P., E, R. & Saini, S. Public good-driven release of heterogeneous resources leads to genotypic diversification of an isogenic yeast population. *Evol. Int. J. Org. Evol.* **76**, 2811–2828 (2022).
34. Mahilkar, A., Nagendra, P., Venkataraman, P., Deshmukh, S. & Saini, S. Rapid evolution of pre-zygotic reproductive barriers in allopatric populations. *Microbiol. Spectr.* **11**, e0195023 (2023).
35. Stockwell, S. R., Landry, C. R. & Rifkin, S. A. The yeast galactose network as a quantitative model for cellular memory. *Mol. Biosyst.* **11**, 28–37 (2015).
36. Zhu, Y. O., Siegal, M. L., Hall, D. W. & Petrov, D. A. Precise estimates of mutation rate and spectrum in yeast. *Proc. Natl. Acad. Sci.* **111**, (2014).
37. Liu, H. & Zhang, J. Yeast Spontaneous Mutation Rate and Spectrum Vary with Environment. *Curr. Biol. CB* **29**, 1584-1591.e3 (2019).
38. Pankajam, A. V., Dash, S., Saifudeen, A., Dutta, A. & Nishant, K. T. Loss of Heterozygosity and Base Mutation Rates Vary Among Saccharomyces cerevisiae Hybrid Strains. *G3 Bethesda Md* **10**, 3309–3319 (2020).
39. Dutta, A., Dutreux, F. & Schacherer, J. Loss of Heterozygosity Spectrum Depends on Ploidy Level in Natural Yeast Populations. *Mol. Biol. Evol.* **39**, msac214 (2022).
40. Zhao, Y. *et al.* Structures of naturally evolved CUP1 tandem arrays in yeast indicate that these arrays are generated by unequal nonhomologous recombination. *G3 Bethesda Md* **4**, 2259–2269 (2014).
41. Dunham, M. J. *et al.* Characteristic genome rearrangements in experimental evolution of Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 16144–16149 (2002).
42. Lang, G. I., Parsons, L. & Gammie, A. E. Mutation rates, spectra, and genome-wide distribution of spontaneous mutations in mismatch repair deficient yeast. *G3 Bethesda Md* **3**, 1453–1465 (2013).
43. Payen, C. *et al.* High-Throughput Identification of Adaptive Mutations in Experimentally Evolved Yeast Populations. *PLoS Genet.* **12**, e1006339 (2016).
44. Brown, C. J., Todd, K. M. & Rosenzweig, R. F. Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment. *Mol. Biol. Evol.* **15**, 931–942 (1998).
45. Gresham, D. *et al.* The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. *PLoS Genet.* **4**, e1000303 (2008).
46. Gresham, D. *et al.* Adaptation to diverse nitrogen-limited environments by deletion or extrachromosomal element formation of the GAP1 locus. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 18551–18556 (2010).
47. Gorkovskiy, A. & Verstrepen, K. J. The Role of Structural Variation in Adaptation and Evolution of Yeast and Other Fungi. *Genes* **12**, 699 (2021).
48. Hu, G. *et al.* Variation in chromosome copy number influences the virulence of Cryptococcus neoformans and occurs in isolates from AIDS patients. *BMC Genomics* **12**, 526 (2011).
49. Andersson, D. I. & Hughes, D. Gene amplification and adaptive evolution in bacteria. *Annu. Rev. Genet.* **43**, 167–195 (2009).
50. Fogel, S., Welch, J. W., Cathala, G. & Karin, M. Gene amplification in yeast: CUP1 copy number regulates copper resistance. *Curr. Genet.* **7**, 347–355 (1983).
51. Onetto, C. A. *et al.* SO2 and copper tolerance exhibit an evolutionary trade-off in Saccharomyces cerevisiae. *PLoS Genet.* **19**, e1010692 (2023).

52. Ning, Y., Dai, R., Zhang, L., Xu, Y. & Xiao, M. Copy number variants of ERG11: mechanism of azole resistance in Candida parapsilosis. *Lancet Microbe* **5**, e10 (2024).
53. Vande Zande, P. *et al.* Step-wise evolution of azole resistance through copy number variation followed by KSR1 loss of heterozygosity in Candida albicans. *PLoS Pathog.* **20**, e1012497 (2024).
54. Triglia, T., Foote, S. J., Kemp, D. J. & Cowman, A. F. Amplification of the multidrug resistance gene pfmdr1 in Plasmodium falciparum has arisen as multiple independent events. *Mol. Cell. Biol.* **11**, 5244–5250 (1991).
55. Nair, S. *et al.* Adaptive copy number evolution in malaria parasites. *PLoS Genet.* **4**, e1000243 (2008).
56. Abdul-Rahman, F. & Gresham, D. Copy number variation facilitates rapid toggling between ecological strategies. *BioRxiv Prepr. Serv. Biol.* 2025.07.22.666191 (2025) doi:10.1101/2025.07.22.666191.
57. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics J. Integr. Biol.* **16**, 284–287 (2012).
58. Carlson, M., Celenza, J. L. & Eng, F. J. Evolution of the dispersed SUC gene family of Saccharomyces by rearrangements of chromosome telomeres. *Mol. Cell. Biol.* **5**, 2894–2902 (1985).
59. Charron, M. J., Read, E., Haut, S. R. & Michels, C. A. Molecular evolution of the telomere-associated MAL loci of Saccharomyces. *Genetics* **122**, 307–316 (1989).
60. Bakerlee, C. W., Phillips, A. M., Nguyen Ba, A. N. & Desai, M. M. Dynamics and variability in the pleiotropic effects of adaptation in laboratory budding yeast populations. *eLife* **10**, e70918 (2021).
61. Jerison, E. R. *et al.* Genetic variation in adaptability and pleiotropy in budding yeast. *eLife* **6**, e27167 (2017).
62. Lyons, D. M., Zou, Z., Xu, H. & Zhang, J. Idiosyncratic epistasis creates universals in mutational effects and evolutionary trajectories. *Nat. Ecol. Evol.* **4**, 1685–1693 (2020).
63. Hassan, N. *et al.* CRISPR-PCDup: a novel approach for simultaneous segmental chromosomal duplication in Saccharomyces cerevisiae. *AMB Express* **10**, 27 (2020).
64. Easmin, F. *et al.* CRISPR-PCD and CRISPR-PCRep: Two novel technologies for simultaneous multiple segmental chromosomal deletion/replacement in Saccharomyces cerevisiae. *J. Biosci. Bioeng.* **129**, 129–139 (2020).
65. Smukowski Heil, C. S. *et al.* Loss of Heterozygosity Drives Adaptation in Hybrid Yeast. *Mol. Biol. Evol.* **34**, 1596–1612 (2017).
66. Xu, H. *et al.* Chromosome drives via CRISPR-Cas9 in yeast. *Nat. Commun.* **11**, 4344 (2020).
67. Jiang, T., Liu, S., Cao, S. & Wang, Y. Structural Variant Detection from Long-Read Sequencing Data with cuteSV. *Methods Mol. Biol. Clifton NJ* **2493**, 137–151 (2022).
68. Chochinov, C. A. & Nguyen Ba, A. N. Bulk-Fitness Measurements Using Barcode Sequencing Analysis in Yeast. *Methods Mol. Biol. Clifton NJ* **2477**, 399–415 (2022).
69. Shumate, A. & Salzberg, S. L. Liftoff: accurate mapping of gene annotations. *Bioinforma. Oxf. Engl.* **37**, 1639–1643 (2021).
70. Walker, B. J. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS One* **9**, e112963 (2014).
71. Cingolani, P. Variant Annotation and Functional Prediction: SnpEff. *Methods Mol. Biol. Clifton NJ* **2493**, 289–314 (2022).
72. Johnson, M. S. *et al.* Phenotypic and molecular evolution across 10,000 generations in laboratory budding yeast populations. *eLife* **10**, e63910 (2021).
73. Suvakov, M., Panda, A., Diesh, C., Holmes, I. & Abyzov, A. CNVpytor: a tool for copy number variation detection and analysis from read depth and allele imbalance in whole-genome sequencing. *GigaScience* **10**, giab074 (2021).