Norwegian University of Life Sciences
Faculty of Biosciences
Department of Animal and Aquacultural Sciences

Philosophiae Doctor (PhD)
Thesis 2019:23

Epigenetic and genetic adaptation in experimental yeast populations

Epigenetisk og genetisk tilpasning i eksperimentelle gjærpopulasjoner

Simon Stenberg

Epigenetic and genetic adaptation in experimental yeast populations

Epigenetisk og genetisk tilpasning i eksperimentelle gjærpopulasjoner

Philosophiae Doctor (PhD) Thesis

Simon Stenberg

Norwegian University of Life Sciences
Faculty of Biosciences
Department of Animal and Aquacultural Sciences

Ås/Adamstuen (2019)

Thesis number 2019:23
ISSN 1894-6402
ISBN 978-82-575-1586-7

Simon Stenberg
Epigenetic and genetic adaptation in experimental yeast populations
February 18, 2019
Main supervisor: Dag Inge Våge
Co supervisors: Stig W Omholt, Jonas Warringer and Arne B Gjuvsland
ISSN: 1894-6402
ISBN: 978-82-575-1586-7
Thesis: 2019:23

Norwegian University of Life Sciences
CIGENE
Department of animal and aquacultural sciences
Universitetstunet 3
1430 Ås Norway

Abstract

Evolution is a highly complex process where life change over time by the influence of relatively simple forces. Bringing evolution into the lab is a valuable asset to answer questions about processes that shape evolution. What determines how populations adapt to changing environments? To what degree can we predict evolutionary outcomes? How fast can populations adapt to a novel environment? I have utilized and contributed to the development of an experimental evolution platform using a combination of modern tools to answer fundamental questions about adaptive evolution using the baker's yeast *Saccharomyces cerevisiae* as a model. In Paper I, we showed how the concerted improvement of genetically independent fitness components can be used as a signature of natural selection, disclosing selection pressures to which adaptation has occurred. Accordingly, we reported adaptation of natural yeast lineages to diverse nitrogen niches and pinpoint the mutations underlying this diversity. In Paper II, we let yeast evolve in the lab in order to study if fast adaptation can be explained without invoking epigenetics. By comparing our results to data-driven simulations we found that fast adaptation to arsenic could be explained by pleiotropy of fitness components and slightly elevated mutation rate without invoking epigenetic explanations. In paper III we performed massively parallell experimental evolution in order to probe the stability of adaptation and identified reversible adaption to oxidative stress, through mitochondrial DNA erosion. Chronic stress led to genetic assimilation of adaptation by complete degradation of mitochondrial DNA. Overall, these findings illustrate the power of experimental evolution as a tool to understand evolution.

Sammendrag

Evolusjon er en svært kompleks prosess der levende organismer endres over tid som følge av relativt enkle krefter. Evolusjon inn i laboratoriet er et et verdifullt verktøy i arbeidt med å gi svar på noen grunnleggende spørsmål rundt evolusjon. Hva bestemmer hvordan populasjoner tilpasser seg endrede miljøbetingelser? I hvilken grad kan vi forutsi resultatet av evolusjon? Hvor fort kan populasjonener tilpasse seg et helt nytt miljø? Jeg har anvendt og videreutviklet en plattform for eksperimentell evolusjon som utnytter moderne labmetoder for å svare på grunnleggende spørsmål om adaptiv evolusjon i modellorganismen gjær (Saccharomyces cerevisiae). I artikkel I viste vi hvordan samvariasjon i genetisk uavhengige fitnesskomponenter kan brukes som en signatur på naturlig seleksjon for å avdekke hvilke seleksjonstrykk som har ført til adaptasjon. Vi avdekket adatasjon av naturlig forekommende gjærstammer til ulike nitrogenkilder og finkartla de kausale mutasjonene som har gitt opphav til den diversiteten vi ser. I artikkel II brukte vi labevolusjon av gjær til å studere om hurtig tilpassning kan forklares uten å ty til epigenetiske mekanisker. Vi sammenholdt eksperimentelle resultater med datadrevne simuleringer og fant at rask adaptasjon til arsenikk kunne forklares uten epigenetikk, dersom man forutsetter pleiotropi for fitnesskomponetner og noe forhøyet mutasjonsrate. I artikkell III gjorde vi stor-skala paralell eksperimentell evolusjon for undersøke hvor stabile tilpasningene var og avdekket reversibel adaptasjon til oksidativt stress, gjennom delvis nedbryting av mitokondrielt DNA. Vedvarendende stress førte til genetisk assimilering ved fullstending nedbryting av mitokondrielt DNA. Disse resultatene illustrere at eksperimentell evolusjon er et kraftfullt verktøy for å forstå evolusjon.

Abbreviations

CNV Copy number variation.

DFE Distribution of fitness effects.

ETC Electron transport chain.

INDEL Insertions and deletions.

LOH Loss of heterozygosity.

LTEE Long-term evolution experiment.

mtDNA Mitochondrial DNA.

QTL Quantitative trait loci.

ROS Reactive oxygen species.

SNP Single nucleotide polymorphism.

Contents

Thesis

1

1.1 Publications

This thesis includes the following three papers:

I Sebastian Ibstedt*, Simon Stenberg*, Sara Bagés, Arne B Gjuvsland, Francisco Salinas, Olga Kourtchenko, Jeevan K.A Samy, Anders Blomberg, Stig W Omholt, Gianni Liti, Gemma Beltran & Jonas Warringer

*Equal contributions

Concerted evolution of life stage performances signals recent selection on yeast nitrogen use

Molecular Biology and Evolution, 2015,32 (1): 153–161

DOI:10.1093/molbev/msu285

II Arne B Gjuvsland, Enikö Zörgö, Jeevan K.A Samy, Simon Stenberg, Ibrahim H Demirsoy, Francisco Roque, Ewa Maciaszczyk-Dziubinska, Magdalena Migocka, Elisa Alonso-Perez, Martin Zackrisson, Robert Wysocki, Markus J Tamás, Inge Jonassen, Stig W Omholt & Jonas Warringer

Disentangling genetic and epigenetic determinants of ultrafast adaptation

Molecular Systems Biology, 2016, 12(12): 892

DOI:10.15252/msb.20166951

III Simon Stenberg, Jing Li, Arne B. Gjuvsland, Karl Persson, Timmy Forsberg, Jia-Xing Yue, Payam Ghiaci, Ciaran Gilchrist, Martin Zackrisson, Mikael Molin, Gianni Liti, Stig W. Omholt & Jonas Warringer

Superoxide induces adaptive editing of mitochondrial DNA

[Manuscript in preparation], 2019

I have contributed to the following papers that are not included as they fall outside the aim of this thesis:

- Hanna Alalam, Fabrice E. Graf, Martin Palm, Marie Abadikhah, Martin Zackrisson, Matilda Mattsson, Chris Hadjineophytou, Linnéa Persson, Simon Stenberg, Payam Ghiaci, Per Sunnerhagen, Jonas Warringer & Anne Farewell

  Conjugation factors controlling F-plasmid antibiotic resistance transmission Hanna

  BioRχiv, 2018

  DOI:10.1101/271254

- Ben P. Harvey, Balsam Al-Janabi, Stefanie Broszeit, Rebekah Cioffi, Amit Kumar, Maria Aranguren-Gassis, Allison Bailey, Leon Green, Carina M. Gsottbauer, Emilie F. Hall, Maria Lechler, Francesco P. Mancuso, Camila O. Pereira, Elena Ricevuto, Julie B. Schram, Laura S. Stapp, Simon Stenberg & Lindzai T. Santa Rosa

  Evolution of marine organisms under climate change at different levels of biological organisation

  Water, 2014, 6:3545-3574

  DOI:10.3390/w6113545

- Martin Zackrisson, Johan Hallin, Lars-Göran Ottosson, Peter Dahl, Esteban Fernandez-Parada, Erik Ländström, Luciano Fernandez-Ricaud, Petra Kaferle, Andreas Skyman, Simon Stenberg, Stig W Omholt, Uroš Petrovič, Jonas Warringer & Anders Blomberg

  Scan-o-matic: High-Resolution Microbial Phenomics at a Massive Scale

  G3: Genes|Genomes|Genetics, 2016, 6:3003-3014

  DOI:10.1534/g3.116.032342

- Melania D'Angiolo, Matteo De Chiara, Jia-Xing Yue, Agurtzane Irizar, Simon Stenberg, Karl Persson, Agnes Llored, Benjamin Barre, Sophie Loeillet, Joseph Schacherer, Roberto Marangoni, Eric Gilson, Alain Nicolas, Jonas Warringer & Gianni Liti

  A yeast living fossil reveals the origin of genome introgressions

  [Manuscript in preparation], 2019

1.2 Aim

My PhD aimed to develop a massively parallelized experimental evolution framework that will allow establishing a causally cohesive understanding of the genotype-phenotype relation on a scale otherwise not easily achieved.

- To develop platform for capturing genotype-phenotype causality during adaptive evolution (**Paper I**, **Paper II**, **Paper III**).
- We sought to develop a simple framework for identifying selection pressures driving phenotype evolution in natural populations, based on unidirectional change in independent fitness components (**Paper I**).
- We sought to explore whether the fastest adaption trajectories could be explained by purely Neo-Darwinistic processes acting on genetic variation (**Paper II**).
- We sought to identify and understand non-genetic contributions to adaptation and their potential subsequent assimilation as permanent genetic solutions (**Paper III**).

1.3 Motivation

Our understanding of evolution has major implications for our understanding of all branches of life sciences as it is the very foundation of modern biology. Sequencing genomes is becoming inexpensive, offering a crucial tool to study genetic changes in genomes on a large scale. However, to connect genetic evolution with phenotypic evolution, a systematic high-throughput approach to phenotyping is necessary to utilize advancements in genomics fully. Recent developments in phenomics have opened new opportunities for studying phenotypic evolution and thereby synthesizing a framework for experimental evolution. Developing such a platform would allow us to study experimental evolution on a scale that has not been possible before. The ability to phenotype, genotype and store thousands of populations in parallel allows approaching questions about variation in repeated evolution, relative contributions of different genetic and epigenetic solutions to phenotypic changes and capturing rare evolutionary steps. To ask questions about pervasiveness and variation, large biological replication is necessary. Massively parallelizing experimental evolution is the next step for a deeper understanding of evolution.

Introduction 2

2.1 Experimental evolution

Studying contemporary and past evolution in the wild is the classical way to understand adaptation. Organisms vary in ecology, behavior and all layers down to intracellular molecular mechanisms resulting in a immense complexity of organisms. Populations constitute and modulate the environmental factors for other populations in a complex ecological network (Odling-Smee, Erwin, Palkovacs, Feldman, & Laland, 2013). Wild populations in natural habitats occupy diverse and fluctuating environments, where most environmental variables are unknown or difficult to control for. When the environmental variables coincide with genetic variation it can lead to false associations between environment and phenotypic evolution. Wild populations are often structured where subpopulations experience different environmental conditions which also complicates the study of evolution in the wild. The environmental variation also varies in temporal and spatial dimensions, resulting in selection pressures to be different based on an organism's position in time and space (Sork, Stowe, & Hochwender, 1993; Wadgymar et al., 2017).

Interaction between multiple genetic variants (epistasis), the tendency of genetic variants to influence many phentotypes (pleiotropy) and the tendency for non-adaptive variants to increase in frequency due to physical proximity to adaptive variants (genetic hitchhiking) further complicate the task to link genetic changes to phenotypic changes (Barton, 2000; Breen, Kemena, Vlasov, Notredame, & Kondrashov, 2012; Gorter, Aarts, Zwaan, & De Visser, 2018; Jerison, Kryazhimskiy, Mitchell, & Bloom, 2017; G I Lang et al., 2013; Smith & Haigh, 1974; Solovieff, Cotsapas, Lee, Purcell, & Smoller, 2013).

The approach to study evolution in natural environments is investigating covariance between phenotypic and genotypic dimensions as well as covariance between phenotypes (Lande & Arnold, 1983). As discussed above, there are numerous layers of complexity that all could contribute to this covariance, which makes it difficult to establish true causal relationships without risking spurious correlations between confounding variables (Lande & Arnold, 1983; Scheiner, Donohue, Dorn, Mazer, & Wolfe, 2002). Furthermore, environments and conditions which led to existing genetic adaptation may no longer

be present and thus, repeating studies with perfect replication of genetic and environmental conditions may not be possible.

The alternative approach is to study evolution in the lab in controlled environments. In a laboratory environment, virtually all environmental variables and genetic variation can be held under near perfect control. By using clonal microorganisms or inbred lines, the initial genetic variation can be removed completely or set to a particular known level. As the laboratory is an artificial environment, results from experimental laboratory evolution should be interpreted with caution as traits and evolutionary changes might not be relevant to wild populations. However, using model organisms in the lab has proven to be fruitful to elucidate processes that are relevant from a broad biological perspective such as sexual reproduction, cooperation, multicellularity, genetic hitchhiking and clonal interference (Barrick & Lenski, 2013; J. C. Gray & Goddard, 2012; Kawecki et al., 2012; G I Lang et al., 2013; Levy et al., 2015; Momeni, Waite, & Shou, 2013; Ratcliff, Denison, Borrello, & Travisano, 2012).

Since natural evolution taking place outside a laboratory environment, is one realized evolutionary trajectory without any true replication, the variance around this trajectory is virtually always unknown. This leads to the question, how different outcomes would we see if the evolution in a particular environmental context were repeated over and over? Technical challenges still remain to achieve a quantitative assessment of variability of adaptive trajectories, which points to the need of a platform to systematically measure adaptation across a variety of conditions (Lobkovsky & Koonin, 2012). The laboratory experimental evolution approach allows the study of massively replicated evolution, following independently evolving populations with identical environmental and genetic conditions. This allows us to answer questions about the degree of stochasticity of evolution. In **Paper III** we followed adaptive trajectories of 1152 evolving, initially genetically identical, parallel populations. Populations were followed in 8 different controlled environments, totalling to 9216 adaptive trajectories.

Experimental evolution has been carried out in a number of different organisms such as phages, bacteria, viruses, yeast, insects and mammals (Buckling, Craig MacLean, Brockhurst, & Colegrave, 2009; M K Burke & Rose, 2009; Girard, McAleer, Rhodes, & Garland Jr., 2001; Holder & Bull, 2001; Kawecki et al., 2012; R E Lenski, Rose, Simpson, & Tadler, 1991; Meyer et al., 2012). The organisms used as a model in experimental evolution have several im-

plications. The type of traits that are available naturally depend on the type of organism, but the genetic complexity of a trait can be large even in unicellular organisms with relatively small genomes such as *S.cerevisiae* (W. C. Ho & Zhang, 2014). The generation time is arguably the most critical parameter of the model organism as it limits the practical feasibility of an experiment. A thousand generations could mean vastly different amounts of experiment time. One advantage of using yeast as a model is that it has a relatively short generation time compared to other eukaryotic model organisms, which allows tracking of a relatively long perspective of generations. Using *S.cerevisiae*, we were able to track evolution for hundreds of generations in experimental conditions.

In microorganisms, a number of approaches have been developed to study evolution in the lab (Van den Bergh, Swings, Fauvart, & Michiels, 2018). The most straightforward approach is serial transfers, where a population is subsampled, presumably randomly, upon reaching a certain population size resulting in a number of bottlenecks that are transferred to a culture with fresh media (Barrick & Lenski, 2013; Van den Bergh et al., 2018). This can be done in both liquid and solid cultures. As the population expands and external and internal nutrients are depleted the physiological state of the cells changes (Gasch & Werner-Washburne, 2002; Saldanha, Brauer, & Botstein, 2004). Chemostats and turbidostats are alternatives of experimental setup to serial transfers (Barrick & Lenski, 2013). A chemostat maintains a constant flow of nutrients, continually replenishing the medium with nutrients. The turbidostat, keeps the population size, or more accurately, the turbidity of the culture constant (Dunham, 2010). Typically the population size is much (>100 fold) larger in chemostats and turbidostats than in serial transfer experiments and cells are generally regarded as being maintained in the same physiological state throughout the experiment, typically exponential growth. Serial transfers are arguably more scalable i.e. they are more suitable for miniaturization which allows increased parallelization. However, recently a platform has been developed that attempts to demonstrate a more scalable microfluidics platform with turbidostat capabilities (Wong, Mancuso, Kiriakov, Bashor, & Khalil, 2018). The choice of the platform has implications on the evolutionary outcomes in experimental evolution since essential evolutionary parameters such as population size, effective population size, and ecological, such as oxygen and nutrient availability, are dictated by the choice.

Since *S.cerevisiae* tolerates well to be frozen and revived again, an experimental lineage that is evolving can be frozen at each serial transfer to create a 'frozen fossil record'. Frozen stocks allow going back and revive samples from earlier time points in the evolving lineage and repeat, validate and further explore evolutionary scenarios. We constructed an extensive fossil record of the evolution experiments in **Paper III** which allowed us to revisit and revive samples for phenotyping and sequencing. This makes *S.cerevisiae* a powerful tool for experimental evolution studies (Dunham, 2010).

Whole genome sequencing of evolved populations allows detection of genetic changes occurring and fixing during experimental evolution (see 3.4.2), ideally time resolved to follow the dynamics of genetic changes (Araya, Payen, Dunham, & Fields, 2010; Payen et al., 2016). Tracking genetic variants in evolving experimental populations has shown that clonal interference is common in these experiments. Where lineages carrying different beneficial variants compete and in the absence of meiotic recombination, these mutations cannot be combined into the same haplotype while existing simultaneously in the population (Hegreness, 2006; G I Lang et al., 2013; Van den Bergh et al., 2018). Genetic hitchhiking is also common in experimental evolution, where variants increase in frequency because of linkage to a beneficial variant increasing in frequency by selection (G I Lang et al., 2013; Van den Bergh et al., 2018).

Experimental evolution has been used to survey the diversity of genetic solutions during adaptation, which includes SNPs, indels, segemental genome duplications, transposon activity (Gresham et al., 2008; Van den Bergh et al., 2018). The interaction between genetic variants that occur in the same genome known as epistasis is wide-spread, to the point where variants with no epistatic interaction appears to be uncommon. Diminishing returns of beneficial mutations where the combination of beneficial effects is lower than the sum is thought to be one explanation why fitness gains tend to decrease over time during experimental evolution in constant environments (Khan, Dinh, Schneider, Lenski, & Cooper, 2011; Kryazhimskiy, Rice, Jerison, & Desai, 2014; Van den Bergh et al., 2018).

The perhaps most ambitious example of an experimental evolution experiment is the *Escherichia coli* long-term evolution experiment (LTEE). Twelve flasks of initially genetically identical *E.coli* populations in a very nutrient-limited media was started in 1988. Samples have been frozen along the experiment to be able to track lineages back in time. As of 2017, the pop-

ulations have undergone over 60 000 generations, which in comparison to human evolution would correspond to approximately 1 million years (Good, McDonald, Barrick, Lenski, & Desai, 2017; Van Hofwegen, Hovde, & Minnich, 2016). After approximately 30 000 generations a significant adaptive event occurred which increased fitness dramatically. A genomic rearrangement allowed expression of a gene *citT* during aerobic growth, allowing the bacteria to metabolically utilize citrate, a carbon source present in the media, which wild-type *E.coli* is unable to do (Blount, Barrick, Davidson, & Lenski, 2012).

To observe this phenomenon in one type of organism that has among the shorter generation times of all organisms, took 14 years. It raises the question that if we are to study rare evolutionary events such as phenotypic innovation, can we design experiments that can capture this without having to spend that amount of time? Indeed, the mutations observed in the LTEE could be replicated in a much shorter time (Van Hofwegen et al., 2016). However, different phenotypic traits show a broad distribution of time to evolve, and it is almost certain that we miss rare events with the limited time perspective we have in experimentally evolving populations. One potential strategy to mitigate the time variable is to scale up experiments horizontally to sample populations in high-throughput instead of few populations at a longer timescale.

Utilizing the power of experimental evolution has given insight into the dynamics of selection on genetic variants, what types of genetic variants that underlie adaptations, the effect of mutations that precede substantial fitness increase by allowing access to novel adaptive trajectories and the effects of asexual reproduction (Colegrave & Collins, 2008; Elena & Lenski, 2003; Van den Bergh et al., 2018).

2.2 *S.cerevisiae* as a model

Humans have domesticated the yeast *Saccharomyces cerevisiae* in the utility of fermentation and later as a vehicle for production of pharmaceuticals and other products (Martínez, Liu, Petranovic, & Nielsen, 2012; Sicard & Legras, 2011). It is widely used for fermenting beverages, food and production of bioethanol. *S.cerevisiae* is one of the most well-studied organisms on the planet. It was the first eukaryotic genome to be sequenced and it has been and continues to be a very productive model system in biology (Mewes et al., 1997). The reasons for this are numerous but some examples are the short

generation time, ability to be genetically manipulated with relative ease and its life cycle is highly convenient for genetic studies. It is also suitable for freezing and thawing, which allows systematic storage of strains.

The haploid reference yeast genome contains ≈ 6000 genes, of which ≈17-19% are essential for survival in rich media (Giaever et al., 2002; Winzeler et al., 1999). The remaining 81-83% of genes are thereby not lethal in optimal growth conditions when individually deleted and have been systematically knocked out in *S.cerevisiae* to form the yeast deletion collections. The deletions has been done in the lab strain backgrounds S288C and Sigma 127B but also in wine yeast (Giaever & Nislow, 2014). The exceptions are genes that are technically hard to knock out because they have paralogs that are closely located or related, lie in repetitive or subtelomeric regions. The yeast deletion collection is a tremendous resource for yeast functional genomics and genetics and has been used in many research applications (Winzeler et al., 1999).

The use of domesticated yeast and studies on mosaic laboratory strains of yeast has given general insights into biology with significant impact on the understanding of cell biology, such as cell cycle control, secretion, telomeres and transcription (Hartwell, 1974; Kornberg, 2007; Mellman & Emr, 2013; Varela & Blasco, 2010). However, we still know relatively little about wild yeasts and their ecology and the lab yeasts are fundamentally different from wild, e.g. in terms of mitotic growth and sporulation (Warringer et al., 2011). Recently efforts have been made to sequence genomes of numerous wild isolates of yeast. These efforts have led to the conclusion that *S.cerevisiae* has its geographical origin in China (Duan et al., 2018; Liti et al., 2009; Peter et al., 2018). *S.cerevisiae* also display extensive reshuffling within subtelomeres (Yue et al., 2017).

The *S.cerevisiae* life cycle is distinguished by vegetative growth in both diploid and haploid states. Mating of two haploid cells forms diploids, and diploids may subsequently sporulate by going through meiosis to form four haploid spores germinating into haploid mitotically reproducing haploids. When cells are starving they enter a quiescent/G_0 resting state where the cell cycle is arrested until nutrients are accessible again (J. V. Gray et al., 2004; Jorgensen & Tyers, 2004). Wild-type yeast can switch mating type through a transposable mating type locus - *MAT*. This switch is dependent on an endonuclease *HO* which is deleted in most laboratory strains to allow stable heterothallic propagation of yeast strains of a single mating type.

As all papers included in this thesis uses *S.cerevisiae* as a model organism, this introduction will be focused, but not completely limited to, the yeast *S.cerevisiae*. The use of model organisms is thought to be translatable into general concepts about biology. However, detailed knowledge about the specific model system used to address biological questions is essential.

2.2.1 The *S.cerevisae* mitochondria

The mitochondrion is an essential endosymbiotic organelle with several functions including respiration and maintenance of iron-sulfur clusters, fatty acid synthesis, amino acid production, heme synthesis (Kispal et al., 2005; Malina, Larsson, & Nielsen, 2018). The mitochondria are dynamic structures that fuse and divide and form tubular branched networks that grow larger during respiratory growth (Westermann, 2010). Mitochondria has their own genomes in virtually all eukaryotic species, with varied mitochondrial genome (mtDNA) size and organization due to independent loss and transfer of segments to the nuclear genome (M. W. Gray, Lang, & Burger, 2004). Despite this tendency of transferring genetic material from the mtDNA to the nucleus, a set of retained genes involved in oxidative phosphorylation (OXPHOS) and respiration is remarkably conserved among eukaryotes (Wallace, 2007). The reason for why not all mitochondrially encoded genes are transferred to the nucleus and why this specific set of genes are retained remains unknown. However, one suggested explanation is that a set of core components of the electron transport chain can be under rapid transcriptional feedback control of their oxidized protein products, as a result of their ROS-generating activity, when encoded inside the mitochondrion (Malina et al., 2018; Wallace, 2007).

The *S.cerevisae* mitochondrial genome is approximately 75 kB with eight protein-coding genes encoded in the mitochondrial genome (see table 2.1), the human mitochondrial genome in comparison is 16 kB with 13 protein encoding genes (Penta, Johnson, Wachsman, & Copeland, 2001). The mitochondrial genome in yeast is composed of linear fragments of varying length, partly tandem repeats of the genome, and a minority of circular molecules (Westermann, 2014; Williamson, 2002). The vast majority of the ($\approx$1000) proteins with activity inside the mitochondria or in the mitochondrial membrane is encoded in the nucleus (Fox, 2012). Yeast is able to tolerate a complete loss of mtDNA (ρ^0), but it results in a small colony phenotype called *petite* when expanded from a single cell (Chen & Clark-Walker, 1999). *S.cerevisae* is crab-tree positive, meaning that they prefer fermentation even in the presence of oxygen, a feature yeast share with cancer cells (De Deken,

1966; Diaz-Ruiz, Rigoulet, & Devin, 2011). Electron transport during respiration occurs in the mitochondrial inner membrane and 7 subunits of the protein complexes of the electron transport chain, including parts of the main producers of ROS, are encoded in the mtDNA (Westermann, 2014). ρ^0 mutants are therefore unable to respire on non-fermentable carbon sources such as glycerol or ethanol (Lipinski, Kaniak-Golik, & Golik, 2010). *Petites* have been shown to rescue double mutants of both superoxide dismutases, indicating lower superoxide production, however, *petites* are more sensitve to H_2O_2 (Grant, MacIver, & Dawes, 1997; Longo, Gralla, & Valentine, 1996)

Tab. 2.1: Protein coding genes encoded on *S.cerevisiae* mitochondrial genome (Fox, 2012)

Gene	Function
COX1	Subunit I of cytochrome c oxidase (ETC complex **IV**)
COX2	Subunit II of cytochrome c oxidase (ETC complex **IV**)
COX3	Subunit III of cytochrome c oxidase (ETC complex **IV**)
ATP6	Subunit of the F0 component of ATP synthase
ATP8	Subunit of the F0 component of ATP synthase
OLI1(ATP9)	Subunit of the F0 component of ATP synthase
COB	Cytochrome b (ETC complex **III**)
VAR1	Mitochondrial small subunit ribosomal protein

Mitochondria are actively vertically transmitted to daughter cells, a process that involves a cytoskeleton associated protein machinery, mtDNA inheritance being partially disconnected and not well understood (Westermann, 2014). The mtDNA replication of circular templates is suggested to start by transcription of the three active origins of replication (ori) followed by rolling-circle replication initiated by homologous recombination producing linear concatemers of mtDNA (Maleszka, Skelly, & Clark-Walker, 1991; Malina et al., 2018; Westermann, 2014). Each cell contains ≈50 copies of the 75kB mtDNA genome per haploid genome and the mtDNA is organized in protein-DNA complexes called nucleoids where each nucleoid contains several mtDNA molecules (Westermann, 2014; Williamson, 2002). The rate of recombination is high in the mtDNA compared to nuclear DNA (Shannon, Rao, Douglass, & Criddle, 1972). During cell division yeast have a tendency towards homoplasmy of mtDNA during segregation meaning that within a cell, all mitochondrial genomes are eventually genetically identical (Alberio, Mineri, Tiranti, & Zeviani, 2007; Shibata & Ling, 2007). In human cancer cells, mtDNA mutations are commonly observed to be homoplasmic (Penta et al., 2001).

The electron transport chain is a primary site of production of reactive oxygen species (ROS) (Murphy, 2009; Turrens, 1997). However, even as the production of ROS is a product of the electron transport chain during respiration, disturbing parts of the electron transport chain can increase the production of ROS (Barros, 2003). Disrupting the assembly of the cytochrome C oxidase complex (ETC complex IV) increase ROS-production (Bode et al., 2013). Thus, removal of key components of the ETC may remove ROS production, whereas others do increase ROS production.

The maintenance of the mitochondrial genome is dependent on nuclear-encoded proteins, for example, the sole mitochondrial DNA polymerase *MIP1* (Genga, Bianchi, & Foury, 1986; Lipinski et al., 2010). The transcription at ori-sites are suggested to be conducted by *RPO41*. However, very small segments of mtDNA in ρ^- mutants, as small as 35 base pairs, can be maintained stably across generations independent of *RPO41* (W L Fangman, Henly, Churchill, & Brewer, 1989; Walton L Fangman, Henly, & Brewer, 1990). It has been suggested that cells can actively degrade mtDNA as a response to stress (Kang & Hamasaki, 2002; Koprowski et al., 2003; Litvinchuk et al., 2013). In line with the idea of nuclear-encoded mtDNA regulation, there are several nuclear-encoded proteins that regulate mtDNA copy number. Exonuclease activity has been shown to be linked to mtDNA copy number reduction (Fikus et al., 2000). Also the nucleoid may be involved in this, by promoting either replicaiton or degradation (Koprowski et al., 2003; Moraes, 2001; Zelenaya-Troitskaya, Newman, Okamoto, Perlman, & Butow, 1998). Environmental stimuli such as H_2O_2 has also been linked to increasing mtDNA copies in yeast (Hori, Yoshida, Shibata, & Ling, 2009)

In **Paper III** we investigate the mitochondrial genome by sequence analysis (see 3.4.2) to measure mtDNA copy numbers. In addition, we use growth phenotypes on non-fermentable carbon sources to use respiration as a marker for mitochondrial respiratory function.

2.3 Evolution

Evolution is the process of change in living organisms, responsible for the divergence of life into millions of species occupying a vast number of habitats across the planet. The question of how evolution works is not only crucial for understanding almost all of biology, but also for medicine, biotechnology, food science, human health and engineering. Globally, one in six human deaths is due to cancer, in which the process of clonal adaptation of somatic

cells to its tissue environment, is a fundamental aspect to carcinogenesis and development of tumor resistance to chemo- and radiotherapy (Greaves & Maley, 2012). A growing threat to human health is the development of resistance to antibiotics in bacteria - also through clonal adaptation (Munita & Arias, 2016; Willyard, 2017). To combat these threats to human health, a deep understanding of the processes driving evolution is crucial.

The constituent forces of evolution are natural selection, gene flow, mutation, recombination, non-random mating and genetic drift. To understand and predict evolution one needs to consider the potential influence of each of these evolutionary forces. The different evolutionary forces can all contribute to changes in genotype and phenotypes frequencies. In this thesis, I aim to deepen our understanding of adaptive evolution by developing and applying a massively parallelized experimental evolution platform that tracks phenotype change in yeast populations over time.

Box 1: Definitions

Adaptation Fitness increase due to natural selection or a trait that has evolved under natural selection

Phenotypic plasticity The capacity of a single genotype to produce multiple phenotypes

Evolution Heritable change in traits over time

Epigenetic Here used in the widest sense, to denote any transgenerational inheritance of non-genetic traits

Acclimatization Non-heritable plastic response in an individual to environmental shift

Genetic assimilation Process by which a phenotype originally produced non-genetically in response to an environmental later becomes genetically encoded via artificial- or natural selection.

Genetic accommodation Evolution as a response to novel traits by previous evolution of non-genetic response. Genetic assimilation is a special case of Genetic accommodation.

2.4 Natural Selection

As resources are limited in all environments, the number of individuals of a particular type that the environment can support is limited. Therefore, competition arises between genotypes seeking to exploit these resources for conversion into biomass and reproduction.

In the experimental setups in **Paper II** and **Paper III** the evolving populations are growing in a controlled environment where typically the carbon/energy source in the form of glucose, is limiting growth, whereas all other nutrients are in excess. This results in competition for glucose and decrease of growth at the end of the growth phase, and finally, starvation leading to a quiescent/G_0 state in stationary phase, until the batch cycle is ended and resources are replenished. To start growing exponentially again requires exiting the quiescent/G_0 stage, an exit triggered by detection and assimilation of nutrients, expression of proteins involved in uptake and utilization, which results in a lag phase (J. V. Gray et al., 2004). In some conditions, nitrogen is instead limiting growth and cells have to compete for rapid and, potentially, efficient (in terms of nitrogen use), uptake and capitalize nitrogen into biomass.

The competition is influenced by a genotype's absolute fitness - the ratio between its birth rate and its death rate. The absolute fitness depends on several variables such as how often a genotype mate, the time until reproductive age, the fraction of a genotype's offspring that survives to reproductive age. These fitness components are also often referred to as life history traits. The life cycle is relatively simple in unicellular organisms and the most important fitness component is the time it takes for the population to double in experimental conditions where nutrients is relatively abundant as the time spent in starvation is relatively short. As this reflects the genotype's rate to privatize resources and convert them into next generation cells before resources are depleted. Other fitness components are potentially under selection, such as the time to resume growth from stationary phase or the death rate when resources are depleted. Other fitness components are potentially under strong or very strong selection in nature, where nutrients are scarce and rapidly consumed and starvations phases long. An interesting question regarding whether or not the efficiency to capitalize resources into biomass rather than the rate of producing offspring could be under selection has been much debated (MacLean & Gudelj, 2006). In **Paper I** and **Paper II** we extract doubling times, lag times and efficiency as key components for the analysis. We extract only the doubling times in **Paper III** (see 3.2).

The fitness components also depend on phenotypes on many biological layers. From the highest level of conflict and interactions between cells to the lowest level of expression levels, localization and post-translational modifications of biomolecules. These phenotypes on different layers - are a product of the environment and the genotype. As genotypes are passed

to the next generation by mating or cell division, a part of the variation in fitness components and therefore fitness - are heritable. This means that the capacity to compete - reflected in fitness - have a heritable component. Given the heritable variation in fitness as a result of genetic variation in a population, genotypes associated with some traits that are associated with higher fitness will be more likely to be reproduced than others. The biased success of certain genotypes to reproduce and thereby increase within a population in a particular environment is natural selection.

In a strict neo-Darwinian paradigm, the benefit exerted from a genotype relative to other genotypes, with regards to the ability to pass on the genotype to successive generations, is classically measured as Darwinian (or relative) fitness (ω). As neo-Darwinian evolution means changes in genotype frequencies over time, the relative fitness describes the amount of bias imposed on a genotype's frequency change from one generation to the next due to selection. The fitness associated with a specific genotype needs to be higher than the mean fitness of a population to be favored by selection.

However, if a trait with selective advantage is produced, as a result of plastic or epigenetic responses, there is no genotype frequency change necessarily involved in the trait frequency changes in the population. If heritable, this also constitutes heritable variation in fitness being favored by selection. The neo-Darwinian paradigm would need inclusion of these types of non-genetic changes to explain or predict adaptive evolution as a result of non-genetic variation heritable fitness effects.

Absolute and relative fitness is often complicated to measure directly. The most widely used method to measure relative fitness or the selection coefficient is to do competition assays. Competing two or more genotypes allows quantification of relative fitness by estimating the relative abundance of genotypes before and after selection (Chevin, 2011). The selective advantage - expressed as the selection coefficient - can be quantified by measuring the slope of the log curve of the relative abundance of a particular genotype, as elegantly exemplified by Levy et al. (2015). This requires tagging of specific genotypes either by fluorescence or molecular barcodes that are difficult and costly to adapt to all experimental scales.

Genotypes associated with higher fitness than the mean fitness of the population reflects in a force of natural selection. The differential between the genotype fitness and the mean fitness of the population is expressed as the selection coefficient ($s = 1 - \omega$) - the pushing force of selection. The frequency

increase of genotypes associated with high fitness is referred to as positive selection sometimes expressed as positive selection coefficient. Selection coefficients have by definition no direction and only express the pushing force of selection. In contrast, the process of elimination of deleterious alleles by natural selection is called purifying or negative selection. Selection on beneficial alleles that increase fitness and removal of deleterious alleles in the population together results in adaptation and adaptations. Adaptation may temporary cease as the heritable adaptive variation may be depleted i.e. if no better hertiable biological states can be accesses from the current state, but the natural selection in the form of purifying selection will always be present as novel maladaptive variation is continuously generated by mutation.

It is also important to make the distinction between absolute and relative fitness. As a population adapt by the increase in frequency of genotypes with higher relative fitness and increase the mean fitness of the population, the population could still decrease and ultimately go to extinction if the fittest genotype has an absolute fitness that is negative i.e. we have adaptation on an allele frequency level but the population nevertheless spirals to extinction.

2.5 Non-adaptive evolution

Not all evolution is adaptive. Genotype and phenotype frequencies can change and mean fitness can decline or increase due to mutations, gene flow and genetic drift. Allele frequencies may drift in a population through random sampling effects, simply because sampling a small number of individuals has a larger probability of changing allele frequency than a large sample (Hartl & Clark, 2007, p.95).

The influence of mutations, gene flow and mating has to be considered when designing experiments to study evolution. In my experiments, we have removed or dramatically reduced gene flow and mating by eliminating their influences as evolutionary forces. Evolving populations in my setup have been spatially separated, leaving cross-contamination as the only potential source of gene flow. Cross-contamination can be detected by sequencing and were determined to be negligible. The experiments also exclusively involve yeast strains exclusively of one mating type and that have their mating type switching locus *HO* deleted (Cubillos, Louis, & Liti, 2009). This allows full control over the mating and will not occur in asexually haploid reproducing cultures. Nevertheless, we observed diploidization to occur in **Paper III** even if mating of haploids should not occur. The mechanism is not clear, but

likely corresponds to endoreplication followed by positive selection on the diploidized state, and we believe it to be a rare occurrence.

The elimination of gene flow and mating leaves only mutations and drift as remaining forces of non-adaptive evolution. In my evolution experiments in **Paper III**, the population sizes have been reasonably large. The populations are typically inoculated at $\approx 5 \times 10^4$ cells and doubling ≈ 4.6 times until $\approx 1 \times 10^6$ cells. In *S.cerevisiae* the rate of any type of mutations is 1.65×10^{-3} per genome per generation. Meaning that ≈ 1650 new mutations arise in every batch cycle. The chance of a particular mutation emerging multiple times, or reverting, during the 200 generations is low. Thus, mutation generates variation, altering copy numbers from 0 to 1, but contributes to negligible further change in frequencies after that, i.e. it is a negligible evolutionary force.

What about genetic drift? The influence of genetic drift is inversely proportional to the effective population size, $\frac{1}{N_e}$. As s is the strength of selection it follows that when $s = \frac{1}{N_e}$, drift and selection is equally strong. From this relationship, it follows that a genotype has to be more beneficial (higher s) the smaller the N_e is, to escape the influence of genetic drift. Census population size in our experiments is relatively high meaning that genetic drift should have a relatively weak influence on increasing genotype frequencies. However, drift dominates when genotype frequency is very low as new genotypes start at one copy and most of the neutral genotypes will rapidly go to extinction, but also many of the beneficial ones will be lost to drift at very low frequencies (Hartl & Clark, 2007).

The number of generations for an allele to fix in a population can be approximated to $t = 4N$. The probability of a neutral allele to fix is $\frac{1}{2N}$ meaning that the probability of extinction is $1 - \frac{1}{2N}$. This means that most neutral mutations will go extinct or fix in a significant number of generations in relatively large populations (Hartl & Clark, 2007, p.113).

The probability of a beneficial allele to survive extinction by genetic drift in constant population size is $2s$. To survive serial bottlenecks further decrease the probability of a beneficial allele to survive genetic drift. Population size is at its largest prior to a bottleneck event, meaning that the probability of beneficial mutations to occur is highest prior to the bottleneck event. A beneficial mutation that occurs later, and thus closer to a bottleneck event, has a smaller probability of surviving the bottleneck event than a mutation that occurs earlier. Most beneficial alleles will have a small probability of

surviving bottlenecks and genetic drift (Wahl & Gerrish, 2001). Using genetic drift as stress by evolving small populations of digital organisms *in silico*, populations would tend to evolve genome robustness to mitigate the adverse effects of drift (Franklin, LaBar, & Adami, 2018; LaBar & Adami, 2016; Richard E. Lenski, Ofria, Collier, & Adami, 1999; Wilke, Wang, Ofria, Lenski, & Adami, 2001).

There is an important distinction between the census and effective population size in the experimental evolution setup on solid media. Yeast cells cannot move, and the only expanding and mixing force is the budding growth and the physical force of pinning. Therefore, colonies are spatially structured without mixing resulting in a departure of effective population size from census population size as the colonies will genetically behave as if they were smaller than the census population size (Hallatschek, Hersen, Ramanathan, & Nelson, 2007; Kayser, Schreck, Yu, Gralka, & Hallatschek, 2018). Additionally, the bottlenecking of populations further reduce the $\frac{N_e}{N}$ ratio (Hartl & Clark, 2007, p.121-122). However, exactly quantifying effective population size is difficult.

2.6 Genetic variation and mutation

For genetic adaptation to occur, genetic variation is necessary, and the source of genetic variation is ultimately mutations. A population may have varying degrees of standing genetic variation before adaptation occurs. A completely isogenic population relies entirely on *de novo* mutations for selection to act on. A number of studies have experimentally investigated evolution in relation to standing genetic variation compared to *de novo* mutations and when standing genetic variation is high, selection primarily acts on standing genetic variation but *de novo* mutations can act together with standing genetic varation (Molly K. Burke, Liti, & Long, 2014; Vázquez-García et al., 2017).

Mutations include single nucleotide variation (SNP/SNV) and insertions and deletions (indels) but also more significant mutations such as gains or losses of genes or chromosomes. One type of larger mutations are aneuploidies where the resulting chromosome count is not a multiplicative of the haploid chromosome count. Another form of large mutations is chromosomal rearrangement such as translocations.

The rate of base substitution per site per generation in *S.cerevisiae* is estimated to be in the range of 1.67×10^{-10} to 0.33×10^{-9} (Lynch et al., 2008; Zhu,

Siegal, Hall, & Petrov, 2014). It is challenging to get precise estimates of the rate of mutation for each position across the genome because mutations are rare events and require a very large number of samples to get sufficient power. Estimation of mutation rates is usually performed using recessive reporter loci such as *CAN1*. However, mutation rates are known to vary along the chromosome (Gregory I Lang & Murray, 2011). A classic method to estimate Distribution of fitness effects (DFE) is mutation accumulation lines. In these experiments microorganisms are accumulating mutations during a given time and propagated with a very low effective population size - typically a single cell bottleneck. This allows a measure of the distribution of the effects of mutations when selection is dominated by genetic drift (Heilbron, Toll-Riera, Kojadinovic, & MacLean, 2014). In **Paper II** we estimated mutation rates in basal media and in the presence of As(III) using a fluctuation assay of *CAN1* (Gregory I. Lang & Murray, 2008; Luria & Delbruck, 1943)

SNPs, indels and larger mutations such as chromosomal copy number or gene copy number variations occur at different rates both between classes of mutations and within classes of mutations. For example is mutations in yeast strongly biased for C/G to T/A mutations (Zhu et al., 2014). By surveying a diverse set of yeast strains it has become evident that there is variation within *S.cerevisiae* not only in nucleotide content but also aneuploidy, ploidy and loss of heterozygosity (LOH) events (Bergström et al., 2014; Duan et al., 2018; Peter et al., 2018). Both ploidy changes and aneuploidy has been shown to occur and be adaptive in experimental evolution (Harari, Ram, Rappoport, Hadany, & Kupiec, 2018; Mulla, Zhu, & Li, 2014; Selmecki et al., 2015).

Studies have estimated the DFE of mutations for different species. For yeast, the DFE has been estimated, both for deleterious and beneficial mutations. These studies suggest that beneficial mutations are rare, but not as rare as previously thought. Of the deleterious mutations, lethal mutations are quite common - 30% of the mutations with fitness effects are lethal. The vast majority of spontaneous mutations will be deleterious or neutral (Eyre-Walker & Keightley, 2007; Joseph & Hall, 2004; Wloch, Szafraniec, Borts, & Korona, 2001). By tracking barcoded lineages of yeast, the distribution of fitness effects can be determined very precisely (Levy et al., 2015). The DFE is dependent on environment, but it has been suggested that some environments such as temperature universally increase the strength of selection genome-wide (Berger, Stangberg, & Walters, 2018).

In **Paper III** we extracted doubling time effects of each single gene knockout in the yeast deletion collection to estimate the distribution of effect sizes of loss of function mutations in each of the environments that the experimental populations were exposed to. Also, we constructed a library of most chromosomal duplications from which doubling time effects also were extracted to create a comprehensive effect size distribution of both loss-of-function mutations and aneuploidies. This dataset was also used in the individual based adaptation modeling conducted in **Paper III**.

Duan and colleagues could not find any difference in the frequency of aneuploidy between domesticated and wild isolates of *S.cerevisiae* (Duan et al., 2018). This suggests that aneuploidies, which seems like a drastic change for a cell, is not an artifact from domestication or laboratory cultivation. Aneuploidies are frequently detected in evolving yeast populations in the lab (Fisher, Buskirk, Vignogna, Marad, & Lang, 2018; Gerstein & Berman, 2015; Gresham et al., 2008; Mulla et al., 2014; Selmecki, Dulmage, Cowen, Anderson, & Berman, 2009; Vázquez-García et al., 2017). A study suggested that aneuploidies are a transient state that will be replaced by more fit solutions as adaptation continues, indicating that the aneuploidy is a first line of defense in response to stress (Yona et al., 2012). Another study showed that aneuploidies are commonly driven by a small number of genes responsible for aneuploidy fitness and that the fitness landscape explored is broader from aneuploidies than that of amplification of single genes (Sunshine et al., 2015).

In an experimental evolution setting (see 2.1) with a clonal ancestral strain, the population relies completely on *de novo* mutations. When serially passaging populations, which is common practice in experimental evolution protocols, mutations are lost in the bottleneck event. Wahl, Gerrish, and Saika-Voivod (2002) calculated the optimal dilution ratio, $\sim 1:5-1:10$, to maximize the supply of beneficial mutations to survive bottlenecks. Additionally, Wahl et al. (2002) found that surviving mutations are almost equally likely to occur late, as they are to occur early during the exponential growth phase. This is explained by a balance of late and early mutations. Mutations occurring early are much more likely to survive, but late in the growth phase, a much larger number of mutations occur in the population.

When studying adaptive evolution, the mutations that are of interest are beneficial mutations or combination of mutations that are beneficial. As discussed earlier, beneficial mutations are rare. There are essentially three modes to

model the relationship between mutations and selection. Either beneficial mutations are significantly rare and their fitness effect is significantly large - this is referred to as strong selection, weak mutation (SSWM). An alternative is that beneficial mutations are not uncommon, but instead have a weak fitness effect, which is referred to as weak selection, strong mutation (WSSM). The third mode is when mutations with large fitness effects are common and clonal interference is present - strong selection, strong mutation (SSSM). The dynamics of beneficial mutations in relation to selection in asexual organisms are reviewed in Sniegowski and Gerrish (2010).

2.7 Recombination and genetic hitchhiking

Recombination of genetic material is one of the sources of genetic variation where the combination of alleles is shuffled between parental haplotypes in a highly regulated process during meiosis (Lam & Keeney, 2015). The probability of recombination between two loci expressed in centimorgans (cM), is highly dependent on the physical distance between loci. A consequence of this is that two loci that are in close proximity to each other are less likely to be separated from each other than increasingly distant loci during recombination. Two loci are genetically linked if their association in gametes show a higher correlation than would be expected by two independent entirely randomly segregating loci. This means that a locus that is increasing in frequency during evolution can have passenger loci that is neutral or deleterious to increase in frequency as well, beacuse of genetic linkage.

Genetic linkage poses a practical problem in addition to the biological consequences of this phenomenon that it is notoriously difficult to disentangle the driver loci from the passenger loci without testing each individual combination of genetic variants. In sexually reproducing populations the genetic linkage is correlated with, but not equal to, physical distance. However, in asexually reproducing populations meiosis does not occur and thus, meiotic recombination does not take place. Under such circumstances, two loci are linked irrespective of physical distance and the entire genome is acting as a single linkage group.

In my experiments in **Paper III**, there is no sexual reproduction during experimental evolution, meaning no recombination influencing evolution.

If sexual reproduction is allowed, beneficial alleles would be allowed to be recombined into the same haplotype without having to arise in the same lineage. In **Paper I** however, we use meiotic recombination as a tool to study evolution by recombining ancestral variants in a multitude of combination to map phenotypes to specific genetic variants (see 3.3). The influence of genetic hitchhiking is consequently large in my experiments in **Paper III** as genotypes under selection cannot be recombined to disperse neutral or deleterious passengers regardless of genetic distance.

2.8 Epigenetics and cellular memory

The principal carrier of heritability between generations is undoubtedly DNA. However, examples of heritable traits that are not causally linked with genetic sequence variations exist. The most studied form of epigenetic inheritance is transgenerational chromatin modifications and DNA methylation (Bird, 2002). For example, studies in humans have shown an epigenetic inheritance pattern of traits linked to starvation of parents in Sweden (Kaati, Bygren, & Edvinsson, 2002). However, DNA methylation is one of many examples of potential routes of epigenetic inheritance in a broader sense, such as prions, ncRNA, and mRNA transcripts (Henikoff & Greally, 2016).

2.8.1 Cellular memory

Cellular memory is defined as a maintenance, sometimes even across generations, of a state by for example continued expression triggered by some environmental factor (Henikoff & Greally, 2016). Transgenerational cellular memory is a type of transgenerational inheritance of traits induced by environmental factors and inherited to progeny that has never experienced the environment. Yeast cells experiencing stress imposed by elevated concentrations of $NaCl$ show higher tolerance to H_2O_2 in up to 5 generations after the initial stress (Guan, Haroon, Bravo, Will, & Gasch, 2012). The H_2O_2 tolerance persisted longer than would be expected of dilution of a tolerant subpopulation (Guan et al., 2012). A similar example has been observed in meiotic progeny of yeast where germinated spores of yeasts display a higher degree of stress tolerance in several environments. In mitotic haploid cells, a similar response could be observed induced by glucose starvation (Gutierrez, Taghizada, & Meneghini, 2018). Shifting carbon sources produce a transgenerational cellular memory in yeast where time spent in glucose determines

the lag phase in maltose, dependent on respiratory activity (Cerulus et al., 2018).

Cellular memory that is passed to daughter cells is, in the broad definition of epigenetics used in this introduction and elsewhere, indistinguishable from epigenetic memory (Henikoff & Greally, 2016; Klosin, Reis, et al., 2017). Therefore, I will make no distinction between cellular memory and epigenetics in this introduction as the examples discussed within 2.8 concerns transgenerational cellular memory.

2.8.2 Epigenetics

Recently, it was shown that environmentally induced epigenetic memory in the form of expression changes associated with chromatin modification can be inherited for at least 14 generations in *C.elegans* (Klosin, Casas, Hidalgo-Carcedo, Vavouri, & Lehner, 2017). In yeast, epigenetic inheritance based on methylated DNA is poorly understood since it lacks many of the proteins involved in methylation of DNA, namely the DNA methyltransferases *(Dnmt)* (Bulkowska et al., 2007; Suzuki & Bird, 2008). However, silent chromatin can be epigenetically inherited in yeast in a *SIR*-dependent manner, which has made yeast a model for studying variegated expression and chromatin gene silencing (Grunstein & Gasser, 2013). There are other forms of non-genetic inheritance of traits in yeast often included in the concept of epigenetics. Prions are generally associated with disease, but there are examples of prions resulting in adaptive heritable trait variation (Jarosz et al., 2014; True & Lindquist, 2000). Another example of non-genetic inheritance of traits was revealed by over-expression of non-prion proteins with intrinsically disordered domains, traits that remained heritable for ≈ 100 generations after stopped overexpression (Chakrabortee et al., 2016). Intermediate chromatin-based epigenetic gene silencing can in yeast facilitate adaptation during experimental evolution through mutational target expansion and increasing effective population size in an artificial experimental context where the site of the selected gene is altered such that its exposure to *SIR*-dependent gene silencing act on it to a varying degree (Stajic, Perfeito, & Jansen, 2019). In this thesis I use epigenetic in practical phenomenological definition where traits that are inherited *short-term* and not included in classical genetics, meaning that the term refers to the pattern of inheritance and not the molecular mechanism. Some epigenetic variation is stable far longer than the scope of my experiments, but short-term inheritance in my setup would signal non-genetic inheritance (Rando & Verstrepen, 2007).

Epigenetic changes may constitute a far faster route for organisms to cope with a novel environment than through genetic adaptation as discussed in **Paper II** and **Paper III**. The increased potential for a rapid response is due to that the epigenetic response may be activated potentially in many or all cells simultaneously. Genetic adaptation relies on random mutations to occur, avoid a loss to drift and then spread in the population.

That adaptive non-genetic trans-generational traits exist is hard to dispute. However, how non-genetic inheritance influences evolutionary changes remains a discussed topic. There are suggestions to extend the current neo-Darwinian paradigm to an 'extended modern synthesis' to include epigenetics and other sources of variation (Pigliucci, 2007). Examples of epigenetics playing a relatively important role in evolution may help facilitate an understanding of the role of epigenetics in evolution. Others argue that we can understand evolution without extending the neo-Darwinian paradigm to include epigenetics (Haig, 2007; Laland et al., 2014). The existence of non-genetic transgenerational traits is not in it self challenging the neo-Darwinian paradigm. It has been argued that epigenetic switches are a property of the genome itself and thus evolved by neo-Darwininan evolution (Haig, 2007). If epimutations (heritable change that does not result in change in DNA sequence) are directed towards cellular or individual needs or if epimutations result in evolution, by interaction with mutations, that cannot be accounted for by the current model - then it might be a need to evaluate the need of an extension of the modern synthesis. A foundational observation by Luria and Delbruck (1943) of a fundamental principle of neo-Darwinian evolution is the existing variation in fitness independent of the organisms needs - random mutation (Rando & Verstrepen, 2007). A unified model of genetic and non-genetic evolution would be needed to account for such interactions (Day & Bonduriansky, 2011).

Epigenetic switches may play a role in evolution even if the epimutations is not a non-random response to the environmental shift. For example, cryptic genetic variation may be uncovered by an environmental shift, that are otherwise 'protected' from selection through epigenetic silencing. The genetic variation in silenced loci is then random, but non-randomly exposed to produce phentoypic variation following an environmental shift (Gibson & Dworkin, 2004; Rando & Verstrepen, 2007).

In the evolution experiments included here, we test the contribution of non-genetic inheritance by relaxing selection pressure. In the absence of

counterselection as a result of deleterious effects of fixed genotypes in the relaxed selection, the fixed genotypes are not expected to change in frequency other than from mutation and drift - which are weak forces (see 2.5). Counterselection can be monitored by measuring growth defects in the absence of stress of adapted populations. By measuring the effect of counterselection and propagating adapted populations in the absence of stress and reimposing selection allows quantification of the contribution of non-genetic effects. If adaptation is only short-term heritable and subsequently lost in absence of counterselection, standard genetic adaptive solutions can be rejected to drive that adaptation (see Fig 4A in **Paper III**).

2.9 Phenotypic plasticity

All phenotyping changes in response to the environment are not evolution. Phenotypic plasticity is an organism's ability to respond to environmental changes. Plastic changes can be changes in development, physiology, morphology and behavior (Price, Qvarnström, & Irwin, 2003). As evolution is changes in allele frequencies, evolutionary changes cannot occur within a generation or individual. Phenotypic plasticity is that the genotype is able to produce several different phenotypes in different environments. This means that the phenotypic changes can occur in one individual that is exposed to a varying environment. Phenotypic plasticity can be conceptualized as a reaction norm as in fig 2.1, where genotype A and Genotype B show a different degree of plasticity when exposed by one environment to another.

Plasticity in itself can be considered a trait, as the degree of plasticity and ability to be plastic can be adaptive (Price et al., 2003). It has been argued that phenotypic plasticity is not the target of selection in itself but rather a by-product of selection in variable environments (Via, 1993). Plastic variation is not heritable, but the genetic factors determining the degree of plasticity are. However, being plastic is not only beneficial, but it is also a trait that potentially comes with costs (DeWitt, Sih, & Wilson, 1998).

If the environment is constant, there is arguably no benefit of being plastic, only to be well adapted to the constant environment. Therefore, one would expect that the evolution of the ability of phenotypic plasticity is driven by changing environments. Considering migration in a metapopulation can favor a plastic genotype over local adaptation. Not surprisingly this benefit is massively reduced when global costs of plasticity are increased (Sultan & Spencer, 2002).

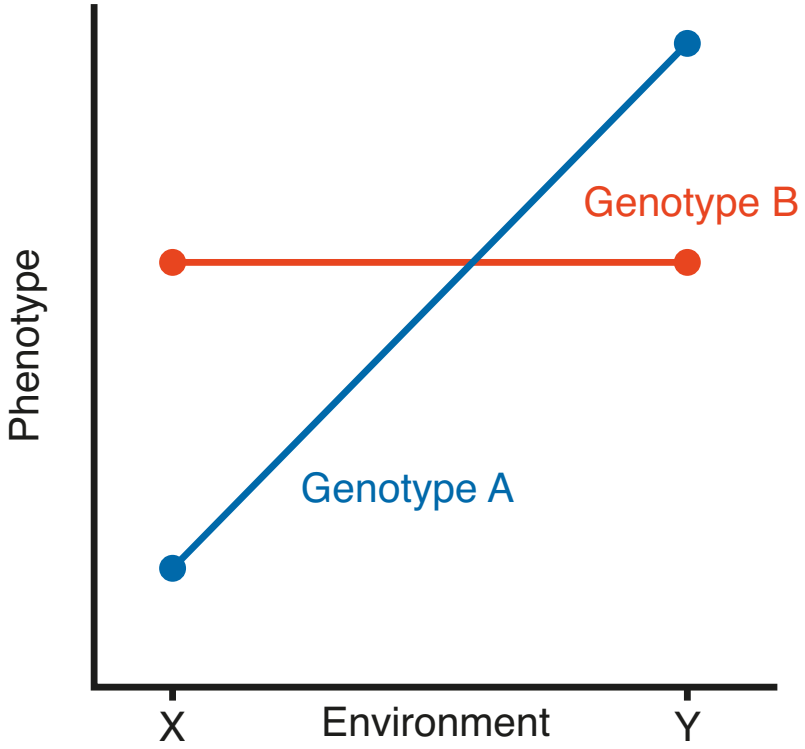


Fig. 2.1: Reaction norms of two genotypes A and B with varying plasticity, A:high and B:low, in two different environments X and Y.

Phenotypic plasticity has been thought to result in a slower rate of evolution by shielding genotypes from selection. However, plasticity may play different roles than merely buffering selection. Phenotypic plasticity might facilitate colonization into new environments followed by genetic adaptation that would not have been possible without plasticity (Ghalambor, McKay, Carroll, & Reznick, 2007). Organisms do however not seem to be constrained in phenotypic plasticity by costs of plasticity (Murren et al., 2015)

Plasticity that places an organism closer to the optimum on the fitness landscape is not unexpected to find. There are however examples of maladaptive plasticity where plastic responses appear to be of a disadvantage to the organism. One such potential example is the acclimatization to high-altitude (low pO_2) of low-altitude mammal populations. This plasticity seems to come

with a cost of disease of prolonged exposure (Cheviron, Bachman, & Storz, 2013; Storz, Scott, & Cheviron, 2010).

As discussed in 2.8 the contribution of non-genetic effects can be estimated by relaxing and reimposing selection. Phenotypic plasticity can be measured in a similar manner, but phenotypic plasticity should not be heritable at all in contrast to epigenetic effects that would be inherited on a short time scale.

2.10 Genetic assimilation

A controversial topic in relation to epigenetics is whether non-genetic traits can be 'assimilated' by facilitating genetic changes to irreversibly transform the trait into a genetically heritable trait. The term genetic assimilation was coined by Waddington after his observations in *D.melanogaster*. Waddington subjected pupae to heat shock which resulted in variation in an arbitrary trait - crossveinless. After successive positive artificial selection on this trait, a portion of the flies displayed this trait even without prior heat shock treatment (Waddington, 1953). An alternative perspective to describe this phenomenon is 'stabilizing selection' as defined by Schmalhausen, not to be confused by selection to maintain the mean trait of a population, discussed by Waddington (Waddington, 1953).

The role of genetic assimilation in evolutionary theory is still being debated since Waddingtons experiments (Crispo, 2007; Haig, 2007; Laland et al., 2014; Pigliucci, 2007; Price et al., 2003; Schlichting & Wund, 2014; Schneider & Meyer, 2017). The shortage of clear examples, the unanswered question of how common genetic assimilation is and if evolutionary theory already is sufficient to incorporate genetic assimilation are some of the questions that are raised against including genetic assimilation into standard evolutionary theory.

In the case of maladaptive plasticity as exemplified in 2.9, there may be selection to stabilize such a reaction norm to avoid maladaptive plasticity, as mentioned in Storz et al. (2010). If there is a cost of phenotypic plasticity or epigenetic switches and the environment stabilizes over time, selection may favor stabilization of reaction norms as well. A different mechanism of genetic assimilation may be by relaxation of stabilizing selection on plasticity, given a stabilization of environmental fluctuations. Genetic adaptation that increases fitness and as a byproduct reduce plasticity may be positively selected if the benefit of plasticity is removed. If selection on stabilization on reaction

norms can be explained, then the importance of genetic assimilation is not a mechanistic question as much as whether or not phenotypic plasticity and epigenetics shapes genetic evolution.

In 2.8 I discussed how to evaluate the relative contribution of genetics and non-standard genetic solutions to adaptation. Extending on this idea to relaxing and reimposing selection, one can evaluate the presence of genetic assimilation by relaxing and reimposing selection over time resolved adaptive trajectories. As genetic assimilation is defined in this context as traits that have an epigenetic explanation that gets replaced by genetic solutions, meaning that adaptation is initially epigenetic but becomes genetically encoded. This leads to the expectation of genetic assimilation to appear as short-term heritable adaptation early during adaptation that is genetically heritable at later stages of adaptation.

Potential routes for assimilation could be initial epigenetic adaptation followed by genetic adaptation with small effects. Genetic adaptations that, due to diminishing returns by strong negative epistasis with the epigenetic adaptations, has an indistinguishable large effect alone compared to the epigenetic adaptation alone. Diminishing returns of beneficial mutations are common and it is reasonable that epigenetic and genetic adaptations neither are completely additive (Wei & Zhang, 2018).

In this case, the genetic component of adaptation would be significantly slower as the epigenetic solution modifies the selection coefficient of the genetic solution by negative epistasis. An alternative route would be epigenetic adaptation followed by accommodating mutations that allow a stronger epigenetic response to the cost of loss of reversibility of epigenetic switching or selection on costly epigenetic switches in itself.

Recently diet-induced plasticity of carnivorous morphological traits was revealed in spadefoot toad tadpoles. The authors suggest an adaptive refinement of existing ancestral diet-induced plasticity into adaptive phenotypes in the form of a distinct carnivore morph (Levis, Isdaner, & Pfennig, 2018).

Approaches and methods 3

3.1 The *S.cerevisiae* genome

The *S.cerevisiae* genome was first published in 1996, being the first complete eukaryotic genome sequence (Mewes et al., 1997). The genome is distributed on 16 chromosomes with approximately 6000 genes (Engel et al., 2014; Mackiewicz et al., 2002; Peter et al., 2018). The genome is relatively compact with a protein-coding gene every 2 kB on average (Goffeau et al., 1996). The genes coding for rRNA is located on chromosome XII in repeats of hundreds of copies of 9 kB repeats, making this chromosome very hard to assemble completely. The rRNA cluster is the reason why the chromosome XII is split in two in the assembly discussed in 3.4.1 which is used as a reference genome for **Paper III**.

3.2 High-throughput phenotyping

To study evolution in the lab or the wild, a method to accurately measure genotypes and phenotypes are imperative. Compared to genetic data acquisition, phenotypic data is fundamentally more complicated. The dimensions of phenotype space are practically endless and we are not able to measure the entire 'phenome' of an organism. This means that regardless of how good analytical tools we have for genetic analysis, we will be limited in our understanding of the relationship between phenotypes and genotypes if phenotyping is lagging behind. We will undoubtedly miss hidden pleiotropy, genes that influence not only the phenotype we are measuring but potentially several phenotypes that are not being measured. Phenomics has been suggested to be the next step after the massive efforts of genomics (Houle, Govindaraju, & Omholt, 2010).

When studying evolution, fitness is the single most important parameter to measure. If fitness cannot be measured directly, the second best is to measure the traits which most contribute to a genotypes fitness. Fitness can be divided in fitness components that each contribute to a genotypes total fitness, but under different growth phases.

Micro-cultivation of microbial cultures has made it possible to measure phenotypes in microorganisms (Warringer, Ericson, Fernandez, Nerman,

& Blomberg, 2003). Bioscreen C is a platform that accurately measures the growth curves of microorganisms. However, the scalability for high-throughput is relatively low as each instrument is able to record growth of 200 wells. This platform is based on liquid cultivations which are laborious to handle in high-throughput compared to solid media. Recently an alternative on solid media has been published that is arguably more cost-efficiently scalable allowing accurate phenotyping of microorganism in high-throughput (Zackrisson et al., 2016).

Scan-o-matic is a software to analyze flatbed scanner images of plates containing growing yeast colonies. Robot-assisted pinning of yeast colonies between solid agar plates allows phenotyping of thousands of colonies simultaneously. A single plate contains 1536 colonies of which 384 are spatial controls interlaced between the experimental colonies. The plates are incubated in scanners inside incubators for the duration of the growth phase. Image acquisition occurs automatically at given intervals and the images are processed by Scan-o-matic to produce growth curves. Since each scanner will take four plates, 6144 growth curves can be simultaneously recorded with high precision. From the growth curves, several different phenotypic parameters are extracted such as maximal growth rate (Zackrisson et al., 2016).

As liquid and solid media constitute different environmental conditions - factors such as mixing, the access to nutrients, exposure to toxins, and population structure will differ between platforms. On solid media, yeast will grow in colonies as plaques which will have a structure of young cells growing in the edges of the colony, and older cells residing in the center of the colony. In contrast, in liquid cultures, the cultures are often grown with agitation that mix cells. The cells are potentially not equally exposed to additional stressful compounds added to the media in a colony compared to a liquid culture where each cell is completely submerged in the medium. Likewise, the access to nutrients is also limited for cells in a colony that are growing on top of other cells compared to cells in direct contact with the fresh agar surface.

The output from Bioscreen C is growth curves in the form of OD measurements from each micro-cultivation well over time. For increased precision and reproducibility in a high-throughput setting, this data is passed through an analytical software - PRECOG (Fernandez-Ricaud, Kourtchenko, Zackrisson, Warringer, & Blomberg, 2016). The software recalibrates the curves

using a function gathered from dilutions to attain OD values with a closer correspondence to the actual OD of the culture. The inherent issue with spectrophotometric growth assays is that when microorganisms reach a high enough density, they shield each other from the light giving a non-linear relationship of OD in relation to cell concentration. In addition to calibrating the OD, PRECOG extracts phenotypic parameters from the growth curves (Fernandez-Ricaud et al., 2016).

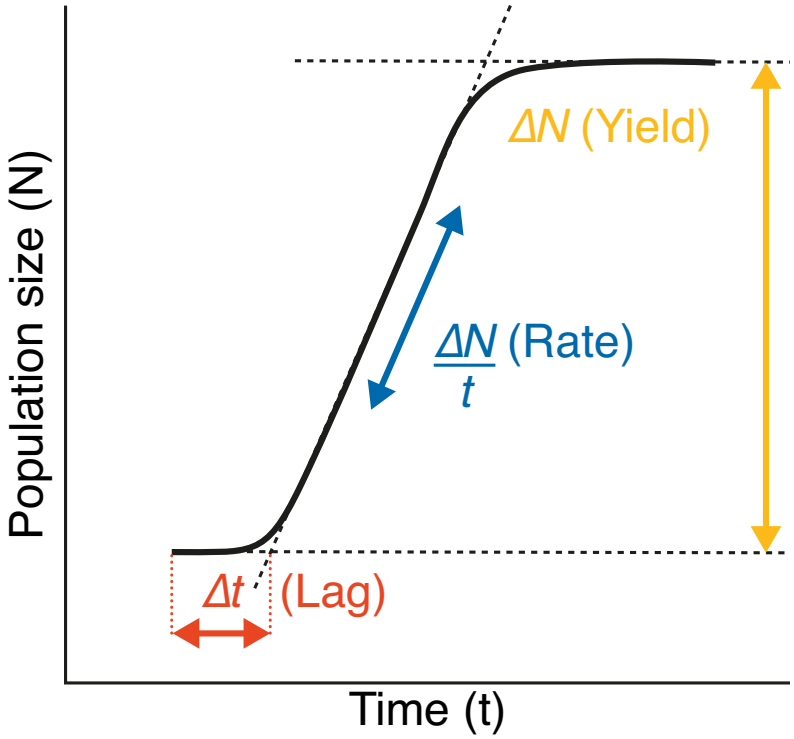


Fig. 3.1: Schematic growth curve with growth parameters/fitness components shown. Rate (blue) is the slope of the regression of the growth curve corresponding to doubling time. Yield (yellow) is the differential population size from inoculation to the end of the growth measurement. Lag (red) is the time until exponential growth which is the time difference between the start of the experiment and the intersection of the basal inoculation size and the regression line. Adapted from **Paper I**

The growth parameters extracted from the growth curves with PRECOG is yield, lag and rate as visualized in fig 3.1. For the Bioscreen C measurements, the output is OD which is not transformed to absolute cell numbers, but they are highly correlated. The yield is extracted by measuring the difference of

the population OD at the start and end of the measurement. The growth rate or doubling time is extracted by linear regression of the steepest slope of the exponential phase, and the slope of the curve is the doubling time (rate). The lag phase is calculated by the difference in time of the start of the experiment and the point of intersection by the regression line used to calculate the growth rate (Fernandez-Ricaud et al., 2016).

Scan-o-matic employes a similar growth parameter estimation as PRECOG but using actual population sizes. The pixel intensities recorded by the scanners in Scan-o-matic are calibrated to cell number using flow cytometry measurements. Otherwise, growth parameters are extracted in a similar way (Zackrisson et al., 2016). In both Scan-o-matic and data from Bioscreen C we condense the high-throughput growth curve data into single numerical parameters.

In **Paper I** we used Bioscreen C to phenotype the progeny of parental crosses for QTL analysis and in **Paper II**, we used Bioscreen C to phenotype strains in the presence of arsenic compounds. In **Paper III**, we did extensive phenotyping using Scan-o-matic resulting in well over 1 million growth curves to be analyzed.

3.3 Quantitative Trait Loci mapping

Linking phenotypes to genotypes is not a trivial task and have been approached by different methods. One way to statistically test the association between a genetic variation and traits is by genome-wide association studies. This method relies on the association between polymorphic SNPs in a population that is associated with a trait. Some issues with this approach are that due to the risk of spurious correlations due to population structure, enormous sample sizes are required to attain sufficient statistical power. In organisms such as yeast where we can conduct random crosses in the lab we can eliminate many of these problems.

Quantitative trait loci (QTL) mapping utilizes the fact that during meiosis loci will recombine and produce new haplotypes between two parental crosses (see2.7). By producing haploid offspring with different combinations of parental markers and phenotyping each offspring, phenotypic differences between parents can be associated with specific markers. The causal loci are in linkage disequilibrium with the marker or markers that show association with the trait. This can be done with relatively low sample sizes compared

to other association based methods. By increasing the number of rounds of meiosis and predicting the optimal amount of relatedness, the resolution and accuracy of QTL mapping can be significantly increased (Märtens, Hallin, Warringer, Liti, & Parts, 2016; She & Jarosz, 2018).

In **Paper I**, we had access to a previously genotyped collection of offspring of all possible pairwise crosses between geographically and genetically distinct natural populations of *S.cerevisiae*. The segregants were phenotyped using Bioscreen C (as discussed in 3.2), and QTL analysis was performed by using `r/qtl` package in `R`.

3.4 Genomics

To study genetic changes in evolving organisms, the genotypes of the evolving organisms needs to be determined. Advancements in sequencing technologies have made it feasible to sequence a large number of genomes simultaneously as the cost per sequenced nucleotide has dropped dramatically (Sboner, Mu, Greenbaum, Auerbach, & Gerstein, 2011). There are several different approaches to study genetic differences between individuals or populations. SNP-arrays is one method to genotype specific SNPs by pre-loaded probes on a chip. However, this is most appropriately used for studying germline genetic variation within a population as it requires a set of polymorphic SNPs decided beforehand. With this technology, *de novo* mutations cannot be detected. In organisms such as humans where a relatively large portion of the genome is not protein-coding compared to many microorganisms, the protein-coding parts of the genome can be specifically targeted. The exome is sequenced by enrichment by probes that pull down DNA from exons. The benefit is that this reduces the cost by limiting the sequencing to only the protein-coding parts of the genome. The downside is that in terms of studying evolution it would miss every genetic change that is not within the protein-coding part of the genome, such as ncRNA, regulatory sequences and potentially still unknown functional parts of the genome. For some applications, exome-seq is undoubtedly the appropriate approach, but in **Paper II** and **Paper III** we would risk missing causal genetic variation outside of the bounds of protein-coding parts of the genome. The significance of regulatory mutations in evolution is empirically supported (Wray, 2007). There are methods to detect regulatory mutations from exome-seq data, but it is still limited to adjacent regulatory sequences (Kim et al., 2016). Sequencing the entire genome makes no prior assumptions of interest. In *S.cerevisiae* the genome is

significantly smaller, thus cheaper, to sequence than humans, making it less of an economic and computational issue to sequence full genomes.

3.4.1 *De novo* sequencing and assembly

The enrichment of genetic variability of subtelomeric sequences in *S.cerevisiae* as discussed in 2.6, renders the subtelomeric regions of heightened interest. From the repetitive nature of subtelomeric and the homology between different chromosomes, subtelomeric regions are hard to assemble using short-read data (Louis, 1995). The existing assemblies, available at the time, of the reference strain used in **Paper III** had incomplete subtelomeres (Bergström et al., 2014). The PacBio platform outputs substantially longer reads than NGS platforms such as Illumina; the drawback is a substantially higher error rate (Rhoads & Au, 2015). As discussed in Rhoads and Au (2015) the issue of higher error rates can be mitigated by a combination of PacBio reads that are error-corrected using shorter reads with higher accuracy. The alternative, which is more viable for smaller genomes, is to achieve higher sequencing depths. The much longer PacBio reads can significantly help assembly of repetitive regions. In **Paper III** we sequenced the reference strain YPS128 using PacBio, assembled with the HGAP3 algorithm (Chin et al., 2013) to have a complete telomere-to-telomere assembly.

3.4.2 Variant calling

To find SNPs, Indels and other genetic variations in a population the re-sequenced reads needs to be compared to a reference sequence. This process has been reviewed in Sandmann et al. (2017) and Nielsen, Paul, Albrechtsen, and Song (2011). The raw reads need to be mapped to the reference genome and for each discrepancy between the reference and the reads, a decision needs to be made whether or not this is a variant or a sequencing error. For a single haploid individual, a minor disagreement of a minority of sequenced reads for a single position is probably a sequencing error as the alternative would be that the majority of reads have an error in the same position. When sequencing a pool of individual, or a population, the sampled DNA contains sequences from multiple individuals. Thus, the array of possibilities for each disagreement expands to contain alleles of a continuous distribution of frequencies. The quality of each base and mapping for each reads becomes more important when sequencing pools to evaluate the probability for each

candidate variation. The analytical pipeline we set up to call variants in **Paper II** and **Paper III** is summarized in Fig 3.2

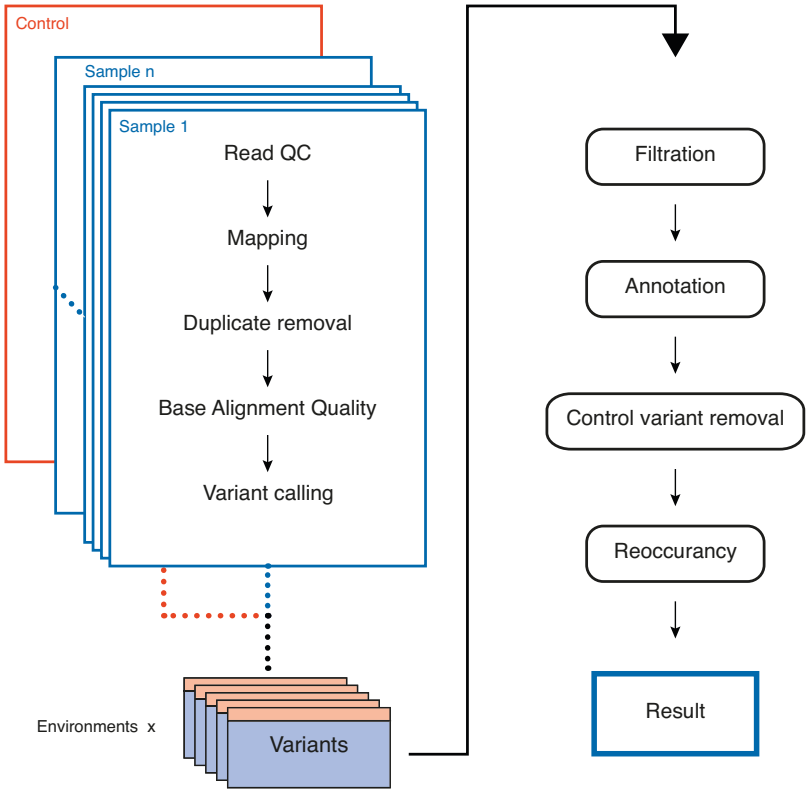


Fig. 3.2: Pipeline for processing re-sequencing reads for variant calling

The sequenced reads are quality controlled to detect low-quality base calls that subsequently might be trimmed or filtered out prior to mapping to the reference genome. As described in Li (2011) indels poses a problem of misalignment of bases adjacent to indels causing false positives. The approach we used to deal with this problem is the BAQ scoring method described in Li (2011). The alternative approach is to realign each genomic read with indels, which requires additional computational steps. The variants are called using Freebayes which is a Bayesian variant caller (Garrison & Marth, 2012). Freebayes assigns scores for the confidence of variant calls which is used to filter untrusted variants. The filtered variant set is matched against an annotation database to get effect estimations of each variant (Cingolani et al., 2012). In the case of experimental evolution when the ancestral strain also

is sequenced, the shared variants between the ancestor and evolved strains are removed. These variants are not of interest as they are likely to be fixed variants in the specific clone used for *de novo* assembly or the ancestral clone used as founder in the evolution experiments. The final set of variants are compared both on the gene level and nucleotide level to find variants that occur multiple times in the same gene among the populations in the same environment.

As SNPs and smaller indels are detected by differences between the reference sequence and sequence reads, larger genetic variations such as aneuploidies and copy number variations (CNVs) would be missed by this method. As these larger variations do not result in any sequence differences but the relative abundance of sequences. A locus that is duplicated in the re-sequenced sample would have exactly twice the relative abundance per genome relative to a re-sequenced control. There are several different methods to detect such signals, as reviewed in Zhao, Wang, Wang, Jia, and Zhao (2013). The method used in **Paper II** and **Paper III** is based on calculating the relative depth of coverage for a sliding window across the genome.

When the whole genome is sequenced, the detection of variants is still limited by several factors which will result in potential missed genetic variants. Larger structural rearrangements such as translocations will be missed as they result in neither sequence changes or changes in abundance. These require different methods such as *de novo* assembly (Tattini, D'Aurizio, & Magi, 2015). Genome annotation is limiting the evaluation and prediction of mutational effects, and especially regulatory sequences are relatively poorly annotated and not part of the SGD reference genome annotation. Besides, non-random genome fragmentation, PCR amplification bias, GC%-variation across the genome and repetitive regions further imposes risks of missing potential variants (Benjamini & Speed, 2012; Poptsova et al., 2014; Treangen & Salzberg, 2012). Genomic regions with systematically underrepresented among the sequenced reads will lack data to support calling of variants, and unambiguously mapping of short reads to repetitive regions is nearly impossible.

3.5 Genetic modifications

Detecting genetic changes in evolving populations can be done by sequencing as described in 3.4.2. However, discerning genetic changes that are responsible for adaptation requires empirical support. Reoccurrence of *de novo*

mutations in the same gene in parallel evolving populations is indicative of adaptation, but the variant needs to be shown to be responsible for fitness changes to be deemed adaptive. For every given genetic variant, the alternative, other than adaptive mutations, are passengers by genetic hitchhiking (see 2.7) or neutral or close to neutral mutations (see 2.5). A way to test the hypothesis that a given mutation is adaptive is to reconstruct it in isolation in the ancestral strain. Phenotypic variation between the mutant strain compared to the ancestral strain can thus be attributed to the mutation tested as it is the only difference between the strains.

Mackiewicz2002

3.5.1 Gene deletions and insertions

Yeast incorporates DNA into its chromosomes by homologous recombination during transformation of foreign DNA (Hinnen, Hicks, & Fink, 1978). This property of yeast allows genetic manipulation such as gene disruption or knock-outs by PCR based methods without much effort (Johnston, Riles, & Hegemann, 2002). Genes can be completely replaced by auxotrophic or antibiotic/fungicide resistance markers. In **Paper I** and **Paper III** we performed gene deletions using selective markers to compare complete loss-of-function to a given single nucleotide mutation. As mentioned in 2.2, all practically possible viable gene deletions exist in a yeast deletion collection. In **Paper III** we phenotyped all existing gene deletions in the deletion collection in several different conditions.

In addition to the deletion collection, a collection of all *S.cerevisiae* ORFs with native promoters on centromeric plasmids has been constructed (C. H. Ho et al., 2009). We transformed plasmids from this collection in **Paper I** to rescue reconstructed mutations as they simulate a hemizygosity or a single gene duplication.

3.5.2 *In vivo* site-directed mutagenesis

To avoid marker-effects and unnecessarily large genetic modifications we used the method *delitto perfetto* to accomplish single nucleotide mutations without leaving any genetic markers (Storici, Lewis, & Resnick, 2001; Storici & Resnick, 2006; Stuckey, Mukherjee, & Storici, 2011). To make these modifications, a cassette of *KanMX* and *URA3* markers are inserted into a ura^- strain at the site of modification. These two markers render the cells

ura^+ and resistant to geneticin (G418). *URA3* is a counter-selectable marker meaning that a strain with *URA3* will grow on media lacking uracil whereas strains without *URA3* will not, and if the strain is exposed to 5-fluoroorotic acid (5-FOA), the strain with *URA3* will not be able to grow, but the strain without the marker gene will be resistant.

This relationship allows initial selection of successful marker insertions on G418 and media lacking uracil. The next step is to insert a synthetic oligonucleotide harboring the genetic change in the center - otherwise identical to the site. This oligonucleotide will when the target strain is transformed, in some cases through homologous recombination replace the markers leaving the site identical to the initial strain except for the desired mutation. As this transformation with the oligonucleotide replaces the *URA3* and G418 marker, the positive colonies can be selected on 5-FOA and screened for geneticin sensitivity.

In **Paper I** we used this method to construct SNP that varied between parental strains to confirm the causal SNP to the QTL in the area around the SNP.

3.5.3 Chromosomal modifications

As with single nucleotide mutations, to quantify the phenotypic effects of duplications of chromosomes (aneuploidies), the chromosomal duplication has to be constructed in isolation without passenger mutations. The most straightforward approach to this is to isolate evolved aneuploid strains and do rounds of backcrossing to the ancestral strain to remove other mutations. This relies on that the aneuploid strain is viable in normal media where the backcrossing occurs. If selective media for which the evolution was carried out, then there is a risk of adaptation through other genetic solutions might occur.

To not rely on backcrossing and spontaneous occurrence methods have been developed to select for aneuploid strains that are disomic for a specific chromosome ($N + 1$). One method inserts a conditional centromere and markers to detect and select for chromosomal duplications. The modified centromere can be conditionally blocked to promote nondisjunction. A *HIS3* marker is inserted inside a *URA3* marker that when spontaneously excised forms a functional copy of *URA3*. Thus selecting for both *HIS3* and *URA3* cells select for disomic strains (Anders et al., 2009).

A different method relies on heterozygous gene deletion of an essential gene in a diploid strain. Also, the strain is transformed with the *can1::STE2pr-HIS3* marker which allows selection for only haploid *MATa* cells. Then, by bulk sporulation, only disomic strains will survive by selection on the essential gene deletion marker as both the selective marker and the essential gene is needed for survival (Zebrowski & Kaback, 2008). This method was employed in **Paper III** where we used it to construct an, as complete as possible, panel of disomic strains.

Main findings

4

4.1 Paper I

In **Paper I** we addressed the problem of how to identify adaptive phenotype changes in natural populations. This is challenging because not all phenotypic change, even life history traits, is adaptive. Orr proposed a method based on non-random directionality of QTLs (see 3.3), that if genetic variants present in a species that affect a trait consistently point in the same direction, then selection is the only reasonable explanation (Orr, 1998). We developed an approach that share features with Orr's proposal, considering the direction and strength of change on different components of a phenotype. The phenotype we consider in the proof of principle is fitness, or rather growth, in a particular environment, and the components are the components of growth - the lag, rate and yield of growth. Yeast inhabits sugar-rich environments that only contains a few nitrogen sources in limiting concentrations. Therefore, we reasoned that nitrogen utilization might be a candidate for environmental variation imposing selection on yeast. The nitrogen utilization environments consisted of a single source of nitrogen such as amino acids as the sole source of nitrogen. We found substantial variation between isolates of natural yeast populations. We could also detect correlations between fitness components of strains phenotyped in a large amount of nitrogen utilization environments.

The signal for concerted selection on fitness components is dependent on evidence that the variation in fitness components are driven by independently segregating loci. We performed a QTL mapping by phenotyping F1 offspring of pairwise crosses between all parental strains. These progeny had previously already been genotyped for parental markers. The vast majority of QTLs were private for a single fitness component, indicating a low degree of pleiotropy between fitness components in nitrogen utilization. We utilized a reciprocal hemizygosity screen to identify candidate genes comprising QTLs. In proline, we were able to associate *RIM15* to the fitness component efficiency and *PUT4* to the fitness component rate. In the case of allantoin, we were able to confirm individual SNPs in *DAL4* and *DAL1* between the parental strains responsible for variation in allantoin utilization.

This paper exemplifies detection of selection in the wild. We were able to detect concerted evolution on fitness components in nitrogen utilization

and identifying single SNPs responsible for phenotypic variation in certain environments.

4.2 Paper II

In **Paper II** we were interested in adaptation that is very fast, we wanted to explore whether the fastest adaptive trajectories could be explained without invoking non-genetic influences, in the form of epigenetics. We combined a modeling framework with phenomics and genomics. After exposing yeast to a number of environments, we could observe exceedingly fast adaptation to arsenic, at near-deterministic adaptive trajectories. The adapted populations were released from the selection pressure which showed heritable adaptation. Sequencing of the populations revealed diverse but previously observed drivers of arsenic adaptation, *FPS1*, *ASK10* and a gene duplication of *ACR3*.

By implementing a data-driven population genetics model of adaptation, we simulated adaptation given realistic parameters of mutation rates, DFE and epistasis. The model failed to produce fast adaptation similar to the empirically observed adaptation speed, even by increasing mutation rates and the number of available positive mutation mutations 5-fold. Only when implementing pleiotropy between fitness components, the empirical adaptation rates could be achieved by the model.

For ultra-fast adaptation to occur, the mutations have to arise early. To exclude directed epigenetic elevation of mutation rates we simulated of the adaptation of mutations driving ultra-fast adaptation given elevated mutation rates. The simulations revealed that slightly increasing the mutation rate, 1-3x of basal mutation rates, produced results matching the empirical results best. A fluctuation assay indicated a mutation rate of of a reporter gene to be 1.5-1.6x of basal mutation rate in arsenic.

4.3 Paper III

As stated both in 5.1 and 2.2.1, mitochondrial dysfunction is related to a number of human disease. One proposed explanation of mechanism is that ROS production by the the mithochondria leads to mtDNA damage, which in turn leads to damaged ETC protein malfunction, leading to increased production of ROS in a positive feedback. We hypothesized that if yeast were to be exposed to elevated mitochondrial ROS in the form of superoxide for

extensive evolutionary time we could elucidate a link between of prolonged mtDNA damage by elevated ROS levels and mitochondrial malfunction.

We developed a experimental evolution platform to combine high-throughput phenotyping and genomics to answer this question. By exposing 1152 populations of yeast to elevated ROS levels we observed exceedingly fast adaptation. By relaxing the selection pressure for ≈ 84 generations, we found that adaptation is only heritable for a few mitotic divisions and subsequently abruptly lost completely. However, the adaptation was found to be assimilated later during adaptation as later populations during adaptation retain adaptation upon release from superoxide selection.

We sequenced superoxide adapting populations and observed extensive mtDNA degradation correlating with adaptation to superoxide with a corresponding expected loss of respiratory capacity. Degradation was either complete or partial where a segment was retained targeting key components of the electron transport chain. In addition, we found pervasive duplications of chromosome II, II and V not coinciding with major growth rate increases during adaptation. However, chromosome V duplications exclusively accompanied complete mtDNA degradation.

Superoxide adaptation was rapidly lost and perfect respiratory capacity was restored when superoxide was reduced to basal levels in adapted populations despite no counterselection, indicating active regulation of mtDNA copy number. Adaptation to superoxide was fast, invariable, lost quickly following perfect inheritance of adaptation for several generations indicating epigenetic adaptation rather than Neo-Darwinian genetic adaptation. However, prolonged adaptation to superoxide failed to restore respiratory capacity and adaptation to elevated ROS level was irreversible for ≈ 84 generations indicating mtDNA erosion progress to complete mtDNA loss resulting in permanent transition to fermentation.

Future Perspectives

5

5.1 Future Perspectives

There are many extensions of this work that could be done to elucidate the pervasiveness and detailed mechanisms of the findings I report here by using a combination of methods used in **Paper I**, **Paper II** and **Paper III**. Specific research questions of interest with these papers as background include:

- As there are few if any examples of genetic assimilation in the wild, further investigating an example of genetic assimilation found in **Paper III** is very attractive. Is this common for yeasts in general or specific to YPS128? Unraveling natural variation of superoxide ultra-fast adaptation and genetic assimilation by exploring a panel of wild isolated *S.cerevisiae* would offer a route to answer that question. Is this a phenomenon that can be accounted for by standard neo-darwinian models? Given natural variation of the ability to adapt through mtDNA degradation and genetic assimilation, I could use the approach in **Paper II** to link natural genetic variants to trait variation by QTL-mapping to gain a mechanistic understanding for one example of genetic assimilation of a trait.

- Disentangle the detailed molecular machinery involved in mtDNA degradation as a response superoxide stress. Guided by candidate genes where natural variation underlies trait variation of adaptive mtDNA I would be able to use knock-out mutants to confirm the involvement of specific proteins. These specific proteins could be studied in detail to untangle the molecular mechanism of rapid mtDNA degradation. I would also like to disrupt the known key proteins in mtDNA maintanance such as *MIP1, DIN7* and *ABF2* to see if mtDNA degradation can be inhibited.

- Further utilize the aneuploid strains we constructed in **Paper III** to study the potential epistatic relationship with mtDNA loss and chromosomal gains to elucidate the individual adaptive value of combinations of chromosome gains both in between chromosomes and in relation to mtDNA loss. Using the method described in Sunshine et al. (2015) to

map genes responsible for driving aneuploidies by truncation of the extra copies of chrosomomes, I could isolate individual genes responsible for fitness gains on duplicated chromosomes.

- Using the parallel experimental evolution platform developed in **Paper III** I could extend the perspective further to include more environments and specifically look for adaptation and loss of adaptations followed relaxation of selection. Specifically, environments that elevate ROS levels by other means than paraquat, for example H_2O_2, but also antioxidants. It would also be interesting to explore diauxic shifts in relation to mtDNA maintanance and the role of mtDNA degradation as a switch from respiration to fermentation as a model for the metabolic switch from respiratory OXSPHOS in cancer cells. Can mtDNA degradation leading to fermentation be induced in crab-tree negative yeasts?
- The mitochondrial dysfunction is connected to a number of diseases in humans, such as cancer, Alzheimer's disease and mitochondrial depletion syndromes (MDS), but also more controversially - ageing (Alberio et al., 2007; Carew et al., 2003; El-Hattab & Scaglia, 2013; Hsu, Tseng, & Lee, 2016; Krishnan, Greaves, Reeve, & Turnbull, 2007; Onyango, Dennis, & Khan, 2016; Penta et al., 2001; Shokolenko, 2014). A survey of human orthologs of associated mtDNA proteins with a background of molecular and evolutionary understanding of mtDNA degradation from our studies could be done. As yeast is able to function without mtDNA, it is a suitable model to study mitochondrial diseases. I would like to take what I learned from **Paper III** and explore the effects on markers of ageing in yeast.

Acknowledgement

6

There are so many people that deserve a thank you, that have shaped this journey from the start. First and foremost, a big thanks to my team of supervisors, Dag-Inge, Stig, Arne and Jonas. I have had the best supervision and support imaginable, that have helped me develop my thinking, writing and as a scientist. A big thanks to administrations of both GU and NMBU and other people who have helped me with the sometimes complicated enterprise to handle the bureaucracy of doing a PhD in Norway, while in Sweden.

Thanks for all the lunch discussions about every weird subject imaginable. There are too many people in this category to name, but you know who you are.

Thanks to Magnus, Mats and Thomas for answering my questions about bioinformatics and using large computers.

Thanks to Karl for bringing some artistry into the building, most appreciated artworks seen in publications, beer bottles and Christmas cards.

Thanks to Ulrika, Peter, Olga and Annabelle without whom I don't know how I would manage to develop my laboratory skills.

Thanks to Martin who developed and taught me the secrets of Scan-o-matic.

A special big thanks to the people I've forgot to mention

Sist men inte minst, tack till Mamma, Pappa, Erik och Emelie, tack för allt.

Bibliography

Alberio, S., Mineri, R., Tiranti, V., & Zeviani, M. (2007). Depletion of mtDNA: Syndromes and genes. *Mitochondrion*, *7*(1-2), 6–12.

Anders, K. R., Kudrna, J. R., Keller, K. E., Kinghorn, B. A. A., Miller, E. M., Pauw, D., … Strong, I. J. (2009). A strategy for constructing aneuploid yeast strains by transient nondisjunction of a target chromosome. *BMC Genetics*, *10*.

Araya, C. L., Payen, C., Dunham, M. J., & Fields, S. (2010). Whole-genome sequencing of a laboratory-evolved yeast strain. *BMC genomics*, *11*(1), 88.

Barrick, J. E., & Lenski, R. E. [R. E.]. (2013). Genome dynamics during experimental evolution. *Nature Rev Genet*, *advance on*.

Barros, M. (2003). H2O2 generation in Saccharomyces cerevisiae respiratory pet mutants: effect of cytochrome c. *Free Radical Biology and Medicine*, *35*(2), 179–188.

Barton, N. H. (2000). Genetic hitchhiking. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *355*(1403), 1553–1562.

Benjamini, Y., & Speed, T. P. (2012). Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Research*, *40*(10). arXiv: NIHMS150003

Berger, D., Stangberg, J., & Walters, R. J. (2018). A Universal Temperature-Dependence of Mutational Fitness Effects. *bioRxiv*, 268011. arXiv: 268011

Bergström, A., Simpson, J. T., Salinas, F., Barré, B., Parts, L., Zia, A., … Liti, G. (2014). A High-Definition View of Functional Genetic Variation from Natural Yeast Genomes. *Molecular Biology and Evolution*, *31*(4), 872–888.

Bird, A. [A.]. (2002). DNA methylation patterns and epigenetic memory. *Genes & Development*, *16*(1), 6–21.

Blount, Z. D., Barrick, J. E., Davidson, C. J., & Lenski, R. E. (2012). Genomic analysis of a key innovation in an experimental Escherichia coli population. *Nature*, *489*(7417), 513–518.

Bode, M., Longen, S., Morgan, B., Peleh, V., Dick, T. P., Bihlmaier, K., & Herrmann, J. M. (2013). Inaccurately Assembled Cytochrome c Oxidase Can Lead to Oxidative Stress-Induced Growth Arrest. *Antioxidants & Redox Signaling*, *18*(13), 1597–1612.

Breen, M. S., Kemena, C., Vlasov, P. K., Notredame, C., & Kondrashov, F. A. (2012). Epistasis as the primary factor in molecular evolution. *Nature*, *490*(7421), 535–538.

Buckling, A., Craig MacLean, R., Brockhurst, M. A., & Colegrave, N. (2009). The Beagle in a bottle. arXiv: 0208024 [gr-qc]

Bulkowska, U., Ishikawa, T., Kurlandzka, A., Trzcińska-Danielewicz, J., Derlacz, R., & Fronk, J. (2007). Expression of murine DNA methyltransferases Dnmt1 and Dnmt3a in the yeastSaccharomyces cerevisiae. *Yeast*, *24*(10), 871–882. arXiv: NIHMS150003

Burke, M. K. [M K], & Rose, M. R. (2009). Experimental evolution with Drosophila. *AJP: Regulatory, Integrative and Comparative Physiology*, *296*(6), R1847–R1854.

Burke, M. K. [Molly K.], Liti, G., & Long, A. D. (2014). Standing Genetic Variation Drives Repeatable Experimental Evolution in Outcrossing Populations of Saccharomyces cerevisiae. *Molecular Biology and Evolution*, *31*(12), 3228–3239.

Carew, J. S., Zhou, Y., Albitar, M., Carew, J. D., Keating, M. J., & Huang, P. (2003). Mitochondrial DNA mutations in primary leukemia cells after chemotherapy: clinical significance and therapeutic implications. *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, *17*(8), 1437–1447.

Cerulus, B., Jariani, A., Perez-Samper, G., Vermeersch, L., Pietsch, J. M., Crane, M. M., … Verstrepen, K. J. (2018). Transition between fermentation and respiration determines history-dependent behavior in fluctuating carbon sources. *eLife*, *7*.

Chakrabortee, S., Byers, J. S., Jones, S., Garcia, D. M., Bhullar, B., Chang, A., … Jarosz, D. F. (2016). Intrinsically Disordered Proteins Drive Emergence and Inheritance of Biological Traits. *Cell*, *167*(2), 369–381.e12.

Chen, X. J., & Clark-Walker, G. D. (1999). The petite mutation in yeasts: 50 years on. *International Review of Cytology*, *194*(November), 197–238.

Chevin, L. M. (2011). On measuring selection in experimental evolution. *Biology Letters*, *7*(2), 210–213.

Cheviron, Z. A., Bachman, G. C., & Storz, J. F. (2013). Contributions of phenotypic plasticity to differences in thermogenic performance between highland and lowland deer mice. *Journal of Experimental Biology*, *216*(7), 1160–1166.

Chin, C. S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., … Korlach, J. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature Methods*, *10*(6), 563–569. arXiv: nbt.2023 [doi:10.1038]

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., … Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*, *6*(2), 80–92.

Colegrave, N., & Collins, S. (2008). Experimental evolution: experimental evolution and evolvability. *Heredity*, *100*, 464.

Crispo, E. (2007). The Baldwin effect and genetic assimilation: Revisiting two mechanisms of evolutionary change mediated by phenotypic plasticity. *Evolution*, *61*(11), 2469–2479.

Cubillos, F. A., Louis, E. J., & Liti, G. (2009). Generation of a large set of genetically tractable haploid and diploid Saccharomyces strains. *FEMS yeast research*, *9*(8), 1217–1225.

Day, T., & Bonduriansky, R. (2011). A Unified Approach to the Evolutionary Consequences of Genetic and Nongenetic Inheritance. *The American Naturalist*, *178*(2), E18–E36.

De Deken, R. H. (1966). The Crabtree Effect: A Regulatory System in Yeast. *Journal of General Microbiology*, *44*(2), 149–156.

DeWitt, T. J., Sih, A., & Wilson, D. S. (1998). Costs and limits of phenotypic plasticity. *Trends in Ecology & Evolution*, *13*(2), 77–81.

Diaz-Ruiz, R., Rigoulet, M., & Devin, A. (2011). The Warburg and Crabtree effects: On the origin of cancer cell energy metabolism and of yeast glucose repression.

Duan, S.-F., Han, P.-J., Wang, Q.-M., Liu, W.-Q., Shi, J.-Y., Li, K., … Bai, F.-Y. (2018). The origin and adaptive evolution of domesticated populations of yeast from Far East Asia. *Nature Communications*, *9*(1), 2690.

Dunham, M. J. (2010). Chapter 19 - Experimental Evolution in Yeast: A Practical Guide. In W. Jonathan, G. Christine, & R. F. Gerald (Eds.), *Methods in enzymology* (Vol. Volume 470, pp. 487–507).

Elena, S. F., & Lenski, R. E. [R E]. (2003). Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat Rev Genet*, *4*(6), 457–469.

Engel, S. R., Dietrich, F. S., Fisk, D. G., Binkley, G., Balakrishnan, R., Costanzo, M. C., … Cherry, J. M. (2014). The Reference Genome Sequence of Saccharomyces cerevisiae : Then and Now. *G3 Genes|Genomes|Genetics*, *4*(3), 389–398.

Eyre-Walker, A., & Keightley, P. D. (2007). The distribution of fitness effects of new mutations. *Nature reviews. Genetics*, *8*(8), 610–8.

Fangman, W. L. [W L], Henly, J. W., Churchill, G., & Brewer, B. J. (1989). Stable maintenance of a 35-base-pair yeast mitochondrial genome. *Molecular and Cellular Biology*, *9*(5), 1917–1921.

Fangman, W. L. [Walton L], Henly, J. W., & Brewer, B. J. (1990). RP041-Independent Maintenance of [rho-] Mitochondrial DNA in Saccharomyces cerevisiae. *Molecular and Cellular Biology*, *10*(1), 10–15.

Fernandez-Ricaud, L., Kourtchenko, O., Zackrisson, M., Warringer, J., & Blomberg, A. (2016). PRECOG: A tool for automated extraction and visualization of fitness components in microbial growth phenomics. *BMC Bioinformatics*, *17*(1).

Fikus, M. U., Mieczkowski, P. A., Koprowski, P., Rytka, J., Śledziewska-Gójska, E., & Cieśla, Z. (2000). The product of the DNA damage-inducible gene of Saccharomyces cerevisiae, DIN7, specifically functions in mitochondria. Genetics Society of America.

Fisher, K. J., Buskirk, S. W., Vignogna, R. C., Marad, D. A., & Lang, G. I. (2018). Adaptive genome duplication affects patterns of molecular evolution in Saccharomyces cerevisiae. *PLoS genetics*, *14*(5), e1007396.

Fox, T. D. (2012). Mitochondrial protein synthesis, import, and assembly. *Genetics*, *192*(4), 1203–1234.

Franklin, J., LaBar, T., & Adami, C. (2018). Mapping the Peaks: Fitness Landscapes of the Fittest and the Flattest. *bioRxiv*, 298125.

Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *ArXiv*. arXiv: 1207.3907

Gasch, A. P., & Werner-Washburne, M. (2002). The genomics of yeast responses to environmental stress and starvation. *Functional and Integrative Genomics*, *2*(4-5), 181–192.

Genga, A., Bianchi, L., & Foury, F. (1986). A nuclear mutant of Saccharomyces cerevisiae deficient in mitochondrial DNA replication and polymerase activity. *Journal of Biological Chemistry*, *261*(20), 9328–9332.

Gerstein, A. C., & Berman, J. (2015). Shift and adapt: the costs and benefits of karyotype variations. *Current opinion in microbiology*, *26*, 130–6.

Ghalambor, C. K., McKay, J. K., Carroll, S. P., & Reznick, D. N. (2007). Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Functional Ecology*, *21*(3), 394–407.

Giaever, G., Chu, A. M., Ni, L., Connelly, C., Riles, L., Véronneau, S., ... Johnston, M. (2002). Functional profiling of the Saccharomyces cerevisiae genome. *Nature*, *418*(6896), 387–391. arXiv: 15334406

Giaever, G., & Nislow, C. (2014). The yeast deletion collection: A decade of functional genomics. *Genetics*, *197*(2), 451–465.

Gibson, G., & Dworkin, I. (2004). Uncovering cryptic genetic variation.

Girard, I., McAleer, M. W., Rhodes, J. S., & Garland Jr., T. (2001). Selection for high voluntary wheel-running increases speed and intermittency in house mice (Mus domesticus). *J Exp Biol*, *204*(Pt 24), 4311–4320.

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., ... Oliver, S. G. (1996). Life with 6000 Genes. *Science*, *274*(5287), 546–567.

Good, B. H., McDonald, M. J., Barrick, J. E., Lenski, R. E., & Desai, M. M. (2017). The dynamics of molecular evolution over 60,000 generations. *Nature*, *551*(7678), 45–50. arXiv: NIHMS150003

Gorter, F. A., Aarts, M. G., Zwaan, B. J., & De Visser, J. A. G. (2018). Local fitness landscapes predict yeast evolutionary dynamics in directionally changing environments. *Genetics*, *208*(1), 307–322.

Grant, C. M., MacIver, F. H., & Dawes, I. W. (1997). Mitochondrial function is required for resistance to oxidative stress in the yeast Saccharomyces cerevisiae. *FEBS Letters*, *410*(2-3), 219–222.

Gray, J. C., & Goddard, M. R. (2012). Sex enhances adaptation by unlinking beneficial from detrimental mutations in experimental yeast populations. *BMC Evol Biol*, *12*, 43.

Gray, J. V., Petsko, G. A., Johnston, G. C., Ringe, D., Singer, R. A., & Werner-Washburne, M. (2004). "Sleeping Beauty": Quiescence in Saccharomyces cerevisiae. *Microbiology and Molecular Biology Reviews*, *68*(2), 187–206. arXiv: arXiv:1011.1669v3

Gray, M. W., Lang, B. F., & Burger, G. (2004). Mitochondria of Protists. *Annual Review of Genetics*, *38*(1), 477–524.

Greaves, M., & Maley, C. C. (2012). Clonal evolution in cancer. *Nature*, *481*(7381), 306–313.

Gresham, D., Desai, M. M., Tucker, C. M., Jenq, H. T., Pai, D. A., Ward, A., ... Dunham, M. J. (2008). The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. *PLoS Genetics*, *4*(12), e1000303.

Grunstein, M., & Gasser, S. M. (2013). Epigenetics in Saccharomyces cerevisiae. *Cold Spring Harbor Perspectives in Biology*, *5*(7), 17491–17492.

Guan, Q., Haroon, S., Bravo, D. G., Will, J. L., & Gasch, A. P. (2012). Cellular memory of acquired stress resistance in Saccharomyces cerevisiae. *Genetics*, *192*(2), 495–505.

Gutierrez, H., Taghizada, B., & Meneghini, M. D. (2018). Nutritional and meiotic induction of heritable stress resistance states in budding yeast. *bioRxiv*.

Haig, D. (2007). Weismann Rules! OK? Epigenetics and the Lamarckian temptation. *Biology and Philosophy*, *22*(3), 415–428.

Hallatschek, O., Hersen, P., Ramanathan, S., & Nelson, D. R. (2007). Genetic drift at expanding frontiers promotes gene segregation. *Proceedings of the National Academy of Sciences*, *104*(50), 19926–19930. arXiv: 0812.2345

Harari, Y., Ram, Y., Rappoport, N., Hadany, L., & Kupiec, M. (2018). Spontaneous Changes in Ploidy Are Common in Yeast. *Current Biology*, *28*(6), 825–835.e4.

Hartl, L. D., & Clark, G. A. (2007). *Principles of population genetics* (4th Editio).

Hartwell, L. H. (1974). Saccharomyces cerevisiae cell cycle. *Bacteriological reviews*, *38*(2), 164–98.

El-Hattab, A. W., & Scaglia, F. (2013). Mitochondrial DNA Depletion Syndromes: Review and Updates of Genetic Basis, Manifestations, and Therapeutic Options. *Neurotherapeutics*, *10*(2), 186–198.

Hegreness, M. (2006). An Equivalence Principle for the Incorporation of Favorable Mutations in Asexual Populations. *Science*, *311*(5767), 1615–1617.

Heilbron, K., Toll-Riera, M., Kojadinovic, M., & MacLean, R. C. (2014). Fitness is strongly influenced by rare mutations of large effect in a microbial mutation accumulation experiment. *Genetics*, *197*(3), 981–990.

Henikoff, S., & Greally, J. M. (2016). Epigenetics, cellular memory and gene regulation.

Hinnen, A., Hicks, J. B., & Fink, G. R. (1978). Transformation of yeast. *Proceedings of the National Academy of Sciences*, *75*(4), 1929–1933.

Ho, C. H., Magtanong, L., Barker, S. L., Gresham, D., Nishimura, S., Natarajan, P., … Boone, C. (2009). A molecular barcoded yeast ORF library enables mode-of-action analysis of bioactive compounds. *Nature Biotechnology*, *27*(4), 369–377. arXiv: NIHMS150003

Ho, W. C., & Zhang, J. (2014). The genotype-phenotype map of yeast complex traits: Basic parameters and the role of natural selection. *Molecular Biology and Evolution*, *31*(6), 1568–1580.

Holder, K. K., & Bull, J. J. (2001). Profiles of adaptation in two similar viruses. *Genetics*, *159*(4), 1393–1404.

Hori, A., Yoshida, M., Shibata, T., & Ling, F. (2009). Reactive oxygen species regulate DNA copy number in isolated yeast mitochondria by triggering recombination-mediated replication. *Nucleic Acids Research*, *37*(3), 749–761.

Houle, D., Govindaraju, D. R., & Omholt, S. (2010). Phenomics: the next challenge. *Nat Rev Genet*, *11*(12), 855–866.

Hsu, C. C., Tseng, L. M., & Lee, H. C. (2016). Role of mitochondrial dysfunction in cancer progression.

Jarosz, D. F. [Daniel F], Brown, J. C., Walker, G. A., Datta, M. S., Ung, W. L., Lancaster, A. K., ... Lindquist, S. (2014). Cross-kingdom chemical communication drives a heritable, mutually beneficial prion-based transformation of metabolism. *Cell*, *158*(5), 1083–1093. arXiv: NIHMS150003

Jerison, E. R., Kryazhimskiy, S., Mitchell, J., & Bloom, J. S. (2017). Genetic Variation in Adaptability and Pleiotropy in Budding Yeast, 1–38.

Johnston, M., Riles, L., & Hegemann, J. H. (2002). Gene disruption. In G. Christine & R. F. Gerald (Eds.), *Methods in enzymology* (Vol. Volume 350, pp. 290–315).

Jorgensen, P., & Tyers, M. (2004). How cells coordinate growth and division.

Joseph, S. B., & Hall, D. W. (2004). Spontaneous Mutations in Diploid Saccharomyces cerevisiae. *Genetics*, *168*(4), 1817–1825.

Kaati, G., Bygren, L. O., & Edvinsson, S. (2002). Cardiovascular and diabetes mortality determined by nutrition during parents' and grandparents' slow growth period. *European Journal of Human Genetics*, *10*(11), 682–688.

Kang, D., & Hamasaki, N. (2002). Maintenance of mitochondrial DNA integrity: Repair and degradation. *Current Genetics*, *41*(5), 311–322.

Kawecki, T. J., Lenski, R. E., Ebert, D., Hollis, B., Olivieri, I., & Whitlock, M. C. (2012). Experimental evolution. arXiv: arXiv:1011.1669v3

Kayser, J., Schreck, C. F., Yu, Q., Gralka, M., & Hallatschek, O. (2018). Emergence of evolutionary driving forces in pattern-forming microbial populations.

Khan, A. I., Dinh, D. M., Schneider, D., Lenski, R. E., & Cooper, T. F. (2011). Negative epistasis between beneficial mutations in an evolving bacterial population. *Science*, *332*(6034), 1193–1196.

Kim, Y. C., Cui, J., Luo, J., Xiao, F., Downs, B., & Wang, S. M. (2016). Exome-based Variant Detection in Core Promoters. *Scientific Reports*, *6*(1), 30716.

Kispal, G., Sipos, K., Lange, H., Fekete, Z., Bedekovics, T., Janáky, T., ... Lill, R. (2005). Biogenesis of cytosolic ribosomes requires the essential iron-sulphur protein Rli1p and mitochondria. *EMBO Journal*, *24*(3), 589–598.

Klosin, A., Casas, E., Hidalgo-Carcedo, C., Vavouri, T., & Lehner, B. (2017). Transgenerational transmission of environmental information in C. elegans. *Science*, *356*(6335), 320–323.

Klosin, A., Reis, K., Hidalgo-Carcedo, C., Casas, E., Vavouri, T., & Lehner, B. (2017). Impaired DNA replication derepresses chromatin and generates a transgenerationally inherited epigenetic memory. *Science Advances*, *3*(8), e1701143.

Koprowski, P., Fikus, M. U., Dzierzbicki, P., Mieczkowski, P., Lazowska, J., & Ciesla, Z. (2003). Enhanced expression of the DNA damage-inducible gene DIN7 results in increased mutagenesis of mitochondrial DNA in Saccharomyces cerevisiae. *Molecular Genetics and Genomics*, *269*(5), 632–639.

Kornberg, R. D. (2007). The molecular basis of eucaryotic transcription.

Krishnan, K. J., Greaves, L. C., Reeve, A. K., & Turnbull, D. (2007). The ageing mitochondrial genome.

Kryazhimskiy, S., Rice, D. P., Jerison, E. R., & Desai, M. M. (2014). Microbial evolution. Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science*, *344*(6191), 1519–1522.

LaBar, T., & Adami, C. (2016). Different Evolutionary Paths to Complexity for Small and Large Populations of Digital Organisms. *PLoS Computational Biology*, *12*(12). arXiv: 1604.06299

Laland, K., Uller, T., Feldman, M., Sterelny, K., Müller, G. B., Moczek, A., … Strassmann, J. E. (2014). Does evolutionary theory need a rethink? *Nature*, *514*(7521), 161–4. arXiv: GGPU9I

Lam, I., & Keeney, S. (2015). Mechanism and control of meiotic recombination initiation. *Cold Spring Harb Perspet Biol*, *52*, 1–53. arXiv: 15334406

Lande, R., & Arnold, S. J. (1983). The Measurement of Selection on Correlated Characters. *Evolution*, *37*(6), 1210. arXiv: arXiv:1112.3516v1

Lang, G. I. [G I], Rice, D. P., Hickman, M. J., Sodergren, E., Weinstock, G. M., Botstein, D., & Desai, M. M. (2013). Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature*.

Lang, G. I. [Gregory I.], & Murray, A. W. [Andrew W.]. (2008). Estimating the per-base-pair mutation rate in the yeast Saccharomyces cerevisiae. *Genetics*, *178*(1), 67–82.

Lang, G. I. [Gregory I], & Murray, A. W. [Andrew W]. (2011). Mutation rates across budding yeast chromosome VI Are correlated with replication timing. *Genome Biology and Evolution*, *3*(1), 799–811.

Lenski, R. E. [R E], Rose, M. R., Simpson, S. C., & Tadler, S. C. (1991). Long-Term Experimental Evolution in Escherichia-Coli .1. Adaptation and Divergence during 2,000 Generations. *American Naturalist*, *138*(6), 1315–1341.

Lenski, R. E. [Richard E.], Ofria, C., Collier, T. C., & Adami, C. (1999). Genome complexity, robustness and genetic interactions in digital organisms. *Nature*, *400*(6745), 661–664.

Levis, N. A., Isdaner, A. J., & Pfennig, D. W. (2018). Morphological novelty emerges from pre-existing phenotypic plasticity. *Nature Ecology & Evolution*, *2*(8), 1289–1297.

Levy, S. F., Blundell, J. R., Venkataram, S., Petrov, D. A., Fisher, D. S., & Sherlock, G. (2015). Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature*, *advance on*.

Li, H. (2011). Improving SNP discovery by base alignment quality. *Bioinformatics*, *27*(8), 1157–1158.

Lipinski, K. A., Kaniak-Golik, A., & Golik, P. (2010). Maintenance and expression of the S. Cerevisiae mitochondrial genome-From genetics to evolution and systems biology.

Liti, G., Carter, D. M., Moses, A. M., Warringer, J., Parts, L., James, S. A., ... Koufopanou, V. (2009). Population genomics of domestic and wild yeasts. *Nature*, *458*(7236), 337–341.

Litvinchuk, A. V., Sokolov, S. S., Rogov, A. G., Markova, O. V., Knorre, D. A., & Severin, F. F. (2013). Mitochondrially-encoded protein Var1 promotes loss of respiratory function in Saccharomyces cerevisiae under stressful conditions. *European Journal of Cell Biology*, *92*(4-5), 169–174.

Lobkovsky, A. E., & Koonin, E. V. (2012). Replaying the tape of life: Quantification of the predictability of evolution. *Frontiers in Genetics*, *3*(NOV).

Longo, V. D., Gralla, E. B., & Valentine, J. S. (1996). Superoxide dismutase activity is essential for stationary phase survival in Saccharomyces cerevisiae: Mitochondrial production of toxic oxygen species in vivo. *Journal of Biological Chemistry*, *271*(21), 12275–12280.

Louis, E. J. (1995). The chromosome ends of Saccharomyces cerevisiae. *Yeast*, *11*(16), 1553–1573.

Luria, S. E., & Delbruck, M. (1943). Mutations of bacteria from virus sensitivity to virus resistance. *Genetics*, *28*(6), 491–511. arXiv: 1304.4330

Lynch, M., Sung, W., Morris, K., Coffey, N., Landry, C. R., Dopman, E. B., ... Thomas, W. K. (2008). A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proceedings of the National Academy of Sciences*, *105*(27), 9272–9277. arXiv: arXiv:1408.1149

Mackiewicz, P., Kowalczuk, M., Mackiewicz, D., Nowicka, A., Dudkiewicz, M., Laszkiewicz, A., ... Cebrat, S. (2002). How many protein-coding genes are there in the Saccharomyces cerevisiae genome? *Yeast*, *19*(7), 619–629.

MacLean, R. C., & Gudelj, I. (2006). Resource competition and social conflict in experimental populations of yeast. *Nature*, *441*(7092), 498–501.

Maleszka, R., Skelly, P., & Clark-Walker, G. (1991). Rolling circle replication of DNA in yeast mitochondria. *The EMBO Journal*, *10*(12), 3923–3929.

Malina, C., Larsson, C., & Nielsen, J. (2018). Yeast mitochondria: An overview of mitochondrial biology and the potential of mitochondrial systems biology.

Märtens, K., Hallin, J., Warringer, J., Liti, G., & Parts, L. (2016). Predicting quantitative traits from genome and phenome with near perfect accuracy. *Nature Communications*, *7*.

Martínez, J. L., Liu, L., Petranovic, D., & Nielsen, J. (2012). Pharmaceutical protein production by yeast: Towards production of human blood proteins by microbial fermentation. arXiv: j.copbio.2012.03.011 [10.1016]

Mellman, I., & Emr, S. D. (2013). A Nobel Prize for membrane traffic: Vesicles find their journey's end. *Journal of Cell Biology*, *203*(4), 559–561.

Mewes, H. W., Albermann, K., Bahr, M., Frishman, D., Gleissner, A., Hani, J., … Zollner, A. (1997). Overview of the yeast genome. *Nature*, *387*(6632 Suppl), 7–65.

Meyer, J. R., Quick, R. T., Lenski, R. E., Dobias, D. T., Weitz, J. S., & Barrick, J. E. (2012). Repeatability and Contingency in the Evolution of a Key Innovation in Phage Lambda. *Science*, *335*(6067), 428–432.

Momeni, B., Waite, A. J., & Shou, W. (2013). Spatial self-organization favors heterotypic cooperation over cheating. *eLife*, *2013*(2), 960.

Moraes, C. T. (2001). What regulates mitochondrial DNA copy number in animal cells?

Mulla, W., Zhu, J., & Li, R. (2014). Yeast: A simple model system to study complex phenomena of aneuploidy. arXiv: NIHMS150003

Munita, J. M., & Arias, C. A. (2016). Mechanisms of Antibiotic Resistance. *Microbiology Spectrum*, *4*(2), VMBF–0016–2015.

Murphy, M. P. (2009). How mitochondria produce reactive oxygen species. *Biochemical Journal*, *417*(1), 1–13. arXiv: NIHMS150003

Murren, C. J., Auld, J. R., Callahan, H., Ghalambor, C. K., Handelsman, C. A., Heskel, M. A., … Schlichting, C. D. (2015). Constraints on the evolution of phenotypic plasticity: Limits and costs of phenotype and plasticity. *Heredity*, *115*(4), 293–301.

Nielsen, R., Paul, J. S., Albrechtsen, A., & Song, Y. S. (2011). Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics*, *12*(6), 443–451. arXiv: NIHMS150003

Odling-Smee, J., Erwin, D. H., Palkovacs, E. P., Feldman, M. W., & Laland, K. N. (2013). Niche Construction Theory: A Practical Guide for Ecologists. *The Quarterly Review of Biology*, *88*(1), 3–28.

Onyango, I. G., Dennis, J., & Khan, S. M. (2016). Mitochondrial Dysfunction in Alzheimer's Disease and the Rationale for Bioenergetics Based Therapies. *Aging and Disease*, *7*(2), 201.

Orr, H. A. (1998). Testing natural selection vs. genetic drift in phenotypic evolution using quantitative trait locus data. *Genetics*, *149*(4), 2099–2104.

Payen, C., Sunshine, A. B., Ong, G. T., Pogachar, J. L., Zhao, W., & Dunham, M. J. (2016). High-Throughput Identification of Adaptive Mutations in Experimentally Evolved Yeast Populations. *PLOS Genetics*, *12*(10), e1006339.

Penta, J. S., Johnson, F. M., Wachsman, J. T., & Copeland, W. C. (2001). Mitochondrial DNA in human malignancy.

Peter, J., De Chiara, M., Friedrich, A., Yue, J. X., Pflieger, D., Bergström, A., … Schacherer, J. (2018). Genome evolution across 1,011 Saccharomyces cerevisiae isolates. *Nature*, *556*(7701), 339–344.

Pigliucci, M. (2007). Do we need an extended evolutionary synthesis? *Evolution*, *61*(12), 2743–2749. arXiv: arXiv:1011.1669v3

Poptsova, M. S., Il'Icheva, I. A., Nechipurenko, D. Y., Panchenko, L. A., Khodikov, M. V., Oparina, N. Y., … Grokhovsky, S. L. (2014). Non-random DNA fragmentation in next-generation sequencing. *Scientific Reports*, *4*(1), 1–6.

Price, T. D., Qvarnström, A., & Irwin, D. E. (2003). The role of phenotypic plasticity in driving genetic evolution. *Proceedings of the Royal Society B: Biological Sciences*, *270*(1523), 1433–1440.

Rando, O. J., & Verstrepen, K. J. (2007). Timescales of Genetic and Epigenetic Inheritance.

Ratcliff, W. C., Denison, R. F., Borrello, M., & Travisano, M. (2012). Experimental evolution of multicellularity. *Proceedings of the National Academy of Sciences*, *109*(5), 1595–1600.

Rhoads, A., & Au, K. F. (2015). PacBio Sequencing and Its Applications.

Saldanha, A., Brauer, M., & Botstein, D. (2004). Nutritional Homeostasis in Batch and Steady-State Culture of Yeast. *Molecular biology of the cell*, *15*(9), 4089–4104. arXiv: NIHMS150003

Sandmann, S., De Graaf, A. O., Karimi, M., Van Der Reijden, B. A., Hellström-Lindberg, E., Jansen, J. H., & Dugas, M. (2017). Evaluating Variant Calling Tools for Non-Matched Next-Generation Sequencing Data. *Scientific Reports*, *7*, 43169. arXiv: 1512.00567

Sboner, A., Mu, X. J., Greenbaum, D., Auerbach, R. K., & Gerstein, M. B. (2011). The real cost of sequencing: Higher than you think! *Genome Biology*, *12*(8).

Scheiner, S. M., Donohue, K., Dorn, L. A., Mazer, S. J., & Wolfe, L. M. (2002). Reducing environmental bias when measuring natural selection. *Evolution*, *56*(11), 2156–2167.

Schlichting, C. D., & Wund, M. A. (2014). Phenotypic plasticity and epigenetic marking: An assessment of evidence for genetic accommodation. *Evolution*, *68*(3), 656–672. arXiv: 9208212v1 [arXiv:hep-ph]

Schneider, R. F., & Meyer, A. (2017). How plasticity, genetic assimilation and cryptic genetic variation may contribute to adaptive radiations. *Molecular Ecology*, *26*(1), 330–350.

Selmecki, A. M., Dulmage, K., Cowen, L. E., Anderson, J. B., & Berman, J. (2009). Acquisition of aneuploidy provides increased fitness during the evolution of antifungal drug resistance. *PLoS Genetics*, *5*(10), 1000705.

Selmecki, A. M., Maruvka, Y. E., Richmond, P. A., Guillet, M., Shoresh, N., Sorenson, A. L., . . . Pellman, D. (2015). Polyploidy can drive rapid adaptation in yeast. *Nature*, *advance on*.

Shannon, C., Rao, A., Douglass, S., & Criddle, R. S. (1972). Recombination in yeast mitochondrial DNA. *Journal of Supramolecular Structure*, *1*(2), 145–152.

She, R., & Jarosz, D. F. [Daniel F.]. (2018). Mapping Causal Variants with Single-Nucleotide Resolution Reveals Biochemical Drivers of Phenotypic Change. *Cell*, *172*(3), 478–490.e15.

Shibata, T., & Ling, F. (2007). DNA recombination protein-dependent mechanism of homoplasmy and its proposed functions.

Shokolenko, I. N. (2014). Aging: A mitochondrial DNA perspective, critical analysis and an update. *World Journal of Experimental Medicine*, *4*(4), 46.

Sicard, D., & Legras, J. L. (2011). Bread, beer and wine: Yeast domestication in the Saccharomyces sensu stricto complex.

Smith, J. M., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetical research*, *23*(1), 23–35.

Sniegowski, P. D., & Gerrish, P. J. [P. J.]. (2010). Beneficial mutations and the dynamics of adaptation in asexual populations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1544), 1255–1263.

Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M., & Smoller, J. W. (2013). Pleiotropy in complex traits: Challenges and strategies. arXiv: NIHMS150003

Sork, V. L., Stowe, K. A., & Hochwender, C. (1993). Evidence for local adaptation in closely adjacent subpopulations for northern red oak (Quercus rubra L.) expressed as resistance to leaf herbivores. *The American Naturalist*, *142*(6), 928–936.

Stajic, D., Perfeito, L., & Jansen, L. E. T. (2019). Epigenetic gene silencing alters the mechanisms and rate of evolutionary adaptation. *Nature Ecology & Evolution*.

Storici, F., Lewis, L. K., & Resnick, M. A. (2001). In vivo site-directed mutagenesis using oligonucleotides. *Nature biotechnology*, *19*(8), 773–776.

Storici, F., & Resnick, M. A. (2006). The Delitto Perfetto Approach to In Vivo Site-Directed Mutagenesis and Chromosome Rearrangements with Synthetic Oligonucleotides in Yeast. *Methods in enzymology*, *409*, 329–345.

Storz, J. F., Scott, G. R., & Cheviron, Z. A. (2010). Phenotypic plasticity and genetic adaptation to high-altitude hypoxia in vertebrates. *Journal of Experimental Biology*, *213*(24), 4125–4136.

Stuckey, S., Mukherjee, K., & Storici, F. (2011). In Vivo Site-Specific Mutagenesis and Gene Collage Using the Delitto Perfetto System in Yeast Saccharomyces cerevisiae. In H. Tsubouchi (Ed.), *Dna recombination* (Chap. 11, Vol. 745, pp. 173–191).

Sultan, S. E., & Spencer, H. G. (2002). Metapopulation Structure Favors Plasticity over Local Adaptation. *The American Naturalist*, *160*(2), 271–283.

Sunshine, A. B., Payen, C., Ong, G. T., Liachko, I., Tan, K. M., & Dunham, M. J. (2015). The fitness consequences of aneuploidy are driven by condition-dependent gene effects. *PLoS Biol*, *13*(5), e1002155.

Suzuki, M. M., & Bird, A. [Adrian]. (2008). DNA methylation landscapes: Provocative insights from epigenomics.

Tattini, L., D'Aurizio, R., & Magi, A. (2015). Detection of Genomic Structural Variants from Next-Generation Sequencing Data. *Frontiers in Bioengineering and Biotechnology*, *3*(June), 1–8. arXiv: 15334406

Treangen, T. J., & Salzberg, S. L. (2012). Repetitive DNA and next-generation sequencing: Computational challenges and solutions. *Nature Reviews Genetics*, *13*(1), 36–46. arXiv: NIHMS150003

True, H. L., & Lindquist, S. L. (2000). A yeast prion provides a mechanism for genetic variation and genetic diversity. *Nature*, *407*, 477–483.

Turrens, J. F. (1997). Superoxide production by the mitochondrial respiratory chain.

Van den Bergh, B., Swings, T., Fauvart, M., & Michiels, J. (2018). Experimental Design, Population Dynamics, and Diversity in Microbial Experimental Evolution. *Microbiology and molecular biology reviews : MMBR*, *82*(3), 1–54.

Van Hofwegen, D. J., Hovde, C. J., & Minnich, S. A. (2016). Rapid Evolution of Citrate Utilization by Escherichia coli by Direct Selection Requires citT and dctA. *Journal of Bacteriology*, *198*(7), 1022–1034. arXiv: NIHMS150003

Varela, E., & Blasco, M. A. (2010). 2009 Nobel Prize in Physiology or Medicine: Telomeres and telomerase. *Oncogene*, *29*(11), 1561–1565.

Vázquez-García, I., Salinas, F., Li, J., Fischer, A., Barré, B., Hallin, J., … Liti, G. (2017). Clonal Heterogeneity Influences the Fate of New Adaptive Mutations. *Cell reports*, *21*(3), 732–744.

Via, S. (1993). Adaptive Phenotypic Plasticity: Target or By-Product of Selection in a Variable Environment? *The American Naturalist*, *142*(2), 352–365.

Waddington, C. (1953). Genetic assimilation of an acquired character. *Evolution*, *7*(2), 118–126.

Wadgymar, S. M., Lowry, D. B., Gould, B. A., Byron, C. N., Mactavish, R. M., & Anderson, J. T. (2017). Identifying targets and agents of selection: innovative methods to evaluate the processes that contribute to local adaptation. *Methods in Ecology and Evolution*, *8*(6), 738–749.

Wahl, L. M., & Gerrish, P. J. [Philip J]. (2001). The probability that beneficial mutations are lost in populations with periodic bottlenecks. *Evolution*, *55*(12), 2606–2610.

Wahl, L. M., Gerrish, P. J., & Saika-Voivod, I. (2002). Evaluating the impact of population bottlenecks in experimental evolution. *Genetics*, *162*(2), 961–971.

Wallace, D. C. (2007). Why Do We Still Have a Maternally Inherited Mitochondrial DNA? Insights from Evolutionary Medicine. *Annual Review of Biochemistry*, *76*(1), 781–821.

Warringer, J., Ericson, E., Fernandez, L., Nerman, O., & Blomberg, A. (2003). High-resolution yeast phenomics resolves different physiological features in the saline response. *Proceedings of the National Academy of Sciences*, *100*(26), 15724–15729.

Warringer, J., Zörgö, E., Cubillos, F. A., Zia, A., Gjuvsland, A., Simpson, J. T., ... Blomberg, A. (2011). Trait variation in yeast is defined by population history. *PLoS Genetics*, *7*(6), e1002111.

Wei, X., & Zhang, J. (2018). Patterns and mechanisms of diminishing returns from beneficial mutations. *bioRxiv*. arXiv: 467944

Westermann, B. (2010). Mitochondrial dynamics in model organisms: What yeasts, worms and flies have taught us about fusion and fission of mitochondria.

Westermann, B. (2014). Mitochondrial inheritance in yeast.

Wilke, C. O., Wang, J. L., Ofria, C., Lenski, R. E., & Adami, C. (2001). Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature*, *412*(6844), 331–333.

Williamson, D. (2002). The curious history of yeast mitochondrial DNA. *Nature reviews. Genetics*, *3*(6), 475–81.

Willyard, C. (2017). The drug-resistant bacteria that pose the greatest health threats.

Winzeler, E. A., Shoemaker, D. D., Astromoff, A., Liang, H., Anderson, K., Andre, B., ... Davis, R. W. (1999). Functional characterization of the Saccharomyces cerevisiae genome by gene deletion and parallel analysis. *Science*, *285*(5429), 901–906. arXiv: arXiv:1011.1669v3

Wloch, D. M., Szafraniec, K., Borts, R. H., & Korona, R. (2001). Direct estimate of the mutation rate and the distribution of fitness effects in the yeast Saccharomyces cerevisiae. *Genetics*, *159*(2), 441–52.

Wong, B. G., Mancuso, C. P., Kiriakov, S., Bashor, C. J., & Khalil, A. S. (2018). Precise, automated control of conditions for high-throughput growth of yeast and bacteria with eVOLVER. *Nature Biotechnology*, *36*, 614.

Wray, G. A. (2007). The evolutionary significance of cis-regulatory mutations.

Yona, A. H., Manor, Y. S., Herbst, R. H., Romano, G. H., Mitchell, A., Kupiec, M., ... Dahan, O. (2012). Chromosomal duplication is a transient evolutionary solution to stress. *Proceedings of the National Academy of Sciences*, *109*(51), 21010–21015. arXiv: arXiv:1408.1149

Yue, J. X., Li, J., Aigrain, L., Hallin, J., Persson, K., Oliver, K., ... Liti, G. (2017). Contrasting evolutionary genome dynamics between domesticated and wild yeasts. *Nature Genetics*, *49*(6), 913–924.

Zackrisson, M., Hallin, J., Ottosson, L.-G., Dahl, P., Fernandez-Parada, E., Ländström, E., ... Blomberg, A. (2016). Scan-o-matic: High-Resolution Microbial Phenomics at a Massive Scale. *G3: Genes|Genomes|Genetics*, *6*(September), 3003–3014.

Zebrowski, D. C., & Kaback, D. B. (2008). A simple method for isolating disomic strains of Saccharomyces cerevisiae. *Yeast*, *25*(5), 321–326.

Zelenaya-Troitskaya, O., Newman, S. M., Okamoto, K., Perlman, P. S., & Butow, R. A. (1998). Functions of the high mobility group protein, Abf2p, in mitochondrial DNA segregation, recombination and copy number in Saccharomyces cerevisiae. *Genetics*, *148*(4), 1763–76.

Zhao, M., Wang, Q., Wang, Q., Jia, P., & Zhao, Z. (2013). Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives - Springer. *BMC bioinformatics*, *14 Suppl 1*(Suppl 11), S1.

Zhu, Y. O., Siegal, M. L., Hall, D. W., & Petrov, D. A. (2014). Precise estimates of mutation rate and spectrum in yeast. *Proceedings of the National Academy of Sciences*, *111*(22), E2310–E2318.

List of Figures

List of Tables

Concerted Evolution of Life Stage Performances Signals Recent Selection on Yeast Nitrogen Use

Sebastian Ibstedt,[†,1] Simon Stenberg,[†,2] Sara Bagés,[1] Arne B. Gjuvsland,[2] Francisco Salinas,[3] Olga Kourtchenko,[1] Jeevan K.A. Samy,[2] Anders Blomberg,[1] Stig W. Omholt,[4] Gianni Liti,[3] Gemma Beltran,[5] and Jonas Warringer*[,1,2]

[1]Department of Chemistry and Molecular Biology, University of Gothenburg, Gothenburg, Sweden

[2]Centre for Integrative Genetics (CIGENE), Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences (UMB), Ås, Norway

[3]IRCAN, CNRS UMR 6267, INSERM U998, University of Nice, Nice, France

[4]Department of Biotechnology, Faculty of Natural Sciences and Technology, Norwegian University of Science and Technology, Trondheim, Norway

[5]Department of Biochemistry and Biotechnology, Universitat Rovira i Virgili, Tarragona, Spain

[†]These authors contributed equally to this work.

***Corresponding author:** E-mail: jonas.warringer@cmb.gu.se.

Associate editor: Joshua Akey

Abstract

Exposing natural selection driving phenotypic and genotypic adaptive differentiation is an extraordinary challenge. Given that an organism's life stages are exposed to the same environmental variations, we reasoned that fitness components, such as the lag, rate, and efficiency of growth, directly reflecting performance in these life stages, should often be selected in concert. We therefore conjectured that correlations between fitness components over natural isolates, in a particular environmental context, would constitute a robust signal of recent selection. Critically, this test for selection requires fitness components to be determined by different genetic loci. To explore our conjecture, we exhaustively evaluated the lag, rate, and efficiency of asexual population growth of natural isolates of the model yeast *Saccharomyces cerevisiae* in a large variety of nitrogen-limited environments. Overall, fitness components were well correlated under nitrogen restriction. Yeast isolates were further crossed in all pairwise combinations and coinheritance of each fitness component and genetic markers were traced. Trait variations tended to map to quantitative trait loci (QTL) that were private to a single fitness component. We further traced QTLs down to single-nucleotide resolution and uncovered loss-of-function mutations in *RIM15*, *PUT4*, *DAL1*, and *DAL4* as the genetic basis for nitrogen source use variations. Effects of SNPs were unique for a single fitness component, strongly arguing against pleiotropy between lag, rate, and efficiency of reproduction under nitrogen restriction. The strong correlations between life stage performances that cannot be explained by pleiotropy compellingly support adaptive differentiation of yeast nitrogen source use and suggest a generic approach for detecting selection.

Key words: **selection, adaptation, life history, nitrogen, yeast, metabolism.**

Introduction

Exposing natural selection and environmental factors driving adaptive differentiation of phenotypes and genotypes is an extraordinary challenge. The classical approach aims to unveil covariations in phenotype, genotype, or environmental factors with a reasonable proxy for fitness. Unfortunately, spurious associations emerge due to genetic drift and population structure. Genetic hitchhiking, whereby nonadaptive alleles and their phenotypes piggyback with adaptive variants at nearby loci, compounds the problem (Barton 2000), as do pleiotropy, whereby nonadaptive phenotypes hitchhike with adaptive by association to the same gene variant (Stearns 2010). Consequently, only a fraction of environmental factors, genotypes, and phenotypes that covary with a fitness proxy is directly linked to selection. An elegant approach toward exposing selection on a phenotype was taken by Orr (1998). If alleles (quantitative trait loci [QTLs]) enhancing a trait value are consistently found in one as compared with another lineage of a species, then positive selection is likely to have acted on that trait. Nevertheless, for most traits in most species, too few QTLs are known for the test to constitute a powerful alternative. Grouping traits into aggregates circumvents this lack of power (Fraser et al. 2010). However, it is rarely clear how a functionally relevant trait grouping is to be achieved.

To be exposed to selection in a particular environment, alleles have to alter net population growth, i.e., birth or death, in that environment. All organisms pass through life cycles composed of distinct life stages (Stearns 1992). Success in any of these life stages, either in the form of increased birth relative death or faster progression to a subsequent life stage in which reproduction occurs, directly affects net population growth (Roff 1992). Performance in these life stages is

Article

therefore the most immediate components of fitness and lower level phenotypes affect fitness via them. In many organisms, life stages are intimately linked such that they consistently are subject to the same environmental variations and the same selective pressures. We therefore reasoned that if an environment exposes a population to a long-term selection, optimization of performance in multiple life stages would often emerge. In absence of selection however, such fitness components would tend to fluctuate independently of each other, as dictated by the specifics of genetic drift, genetic hitchhiking, and pleiotropy. This conjecture directly suggested a test for adaptive differentiation in the form of robust correlations between fitness components. To be valid, this test requires fitness components to be genetically independent, as confounding correlations due pleiotropy and genetic hitchhiking otherwise could emerge.

To explore the power of this test, we studied natural isolates of the model yeast *Saccharomyces cerevisiae*, which has a well-understood life history with fitness components that are straight forward to define, measure and dissect genetically. Sex in natural *S. cerevisiae* is rare and heavily oriented toward self-fertilization (Ruderfer et al. 2006; Tsai et al. 2008). Selection is therefore expected to act primarily on the asexual life cycle (Warringer and Blomberg 2014). Accordingly, the time it takes for an asexual population to pass through the lag phase to the reproductive stage (lag time), its net rate of growth in the reproductive stage (doubling time), and the total population density achieved before the reproductive stage ends (efficiency) are key fitness components. Population growth of chemoheterotrophic microbes, such as yeasts, is limited by access to nutrients, primarily energy, carbon, reducing equivalents, and nitrogen. *Saccharomyces cerevisiae* obtains energy, carbon, and reducing equivalents from sugar, while using a wide diversity of nitrogen sources to satisfy its nitrogen needs (Cooper 1982). Decomposing or damaged fruit, flowering plant nectar, and tree saps (exudates) are primary microhabitats occupied by yeast (Landry et al. 2006; Hittinger 2013). These are rich in sugar but poor in nitrogen, with one or a few nitrogen sources dominating each microhabitat (supplementary fig. S1*A* and *B*, Supplementary Material online) (Gardener and Gillman 2001). Selection for optimal use of single nitrogen sources can therefore be expected to have been strong in natural yeasts and to have driven recent adaptive differentiation. To evaluate the validity of the proposed test for selection, we therefore measured asexual population growth of natural yeasts in a vast array of nitrogen-restricted, single nitrogen-source environments.

Results

Natural Yeasts Are Highly Differentiated with Regard to Nitrogen Source Use

To evaluate fitness components of natural yeasts we considered four *S. cerevisiae* strains, each representing one of the nonreproductively isolated yeast populations discovered outside China: the West African DBVPG6044 (WA), the North American YPS128 (NA), the European DBVPG6765 (E), and the Sake Y12 (S). Lineages differed genetically by 0.3–0.7%, corresponding to several million generations of evolution, and encompassed >50% of the known SNP and phenotypic variation outside China (Liti et al. 2009; Warringer et al. 2011; Bergstrom et al. 2014) (supplementary fig. S1*C*, Supplementary Material online). Strains were clonally propagated in 28 nitrogen environments corresponding to all low-complexity nitrogen sources used by yeast. These were present at equal and yield-limiting nitrogen concentrations (supplementary fig. S1*D* and *E*, Supplementary Material online). From high-resolution growth curves, it was abundantly clear that natural strains vary greatly in their capacity to use different nitrogen sources (fig. 1*A*). Quantifying these variations, we resolved the growth supported by the various nitrogen sources into its components and extracted the rate (population doubling time), lag (time to initiate growth), and efficiency of population expansion (total change in population density) (fig. 1*B*). Variations among natural lineages were pervasive with significant strain differences (false discovery rate [FDR], $q \leq 5\%$) in 45% of the 84 fitness component measures (fig. 1*C* and supplementary fig. S2, Supplementary Material online). The lower the mean capacity to convert nitrogen into growth, the larger was the variation between strains, with the West African standing out as a generally poor performer (fig. 1*D* and *E*). Thus, natural *S. cerevisiae* strains are highly differentiated with regards to nitrogen source use and the differentiation is strongest for poor nitrogen sources. The latter is in agreement with recent wine yeast observations (Gutierrez et al. 2013). Given that entry into stationary phase could occur for reasons other than nitrogen depletion, we investigated whether total population densities achieved truly reflected exhaustion of nitrogen. Surveying the nitrogen remaining in the medium as a function of time in strains WA and NA, we found nitrogen depletion to neatly coincide with exit from the exponential growth phase (fig. 1*F*). In fact, 10–35% of the total population expansion occurred after this time point, presumably by mobilization of stored intracellular nitrogen. Thus, stationary phase levels reflected depletion of not only external but also internally stored nitrogen.

Concerted Selection on Lag, Rate, and Efficiency of Yeast Nitrogen Use

To test whether the differentiation of nitrogen source use was caused by adaptation and selection, we compared measures of lag, rate, and efficiency of population growth over all nitrogen-restricted environments and strains. Correlations between all pairs of fitness components were excellent (fig. 2*A*). Thus, the efficiency of population growth covaried extensively (Pearson, $r = 0.85$ and 0.61) with both the rate and lag. This correlation between fitness components has two components: covariation that is due to genetic variants that are shared between strains and covariation that is due to genetic variants that differ between strains. The former corresponds to concerted evolution of fitness components in the ancestral lineage. The latter corresponds to concerted evolution of fitness components in lineage(s) after their separation. As

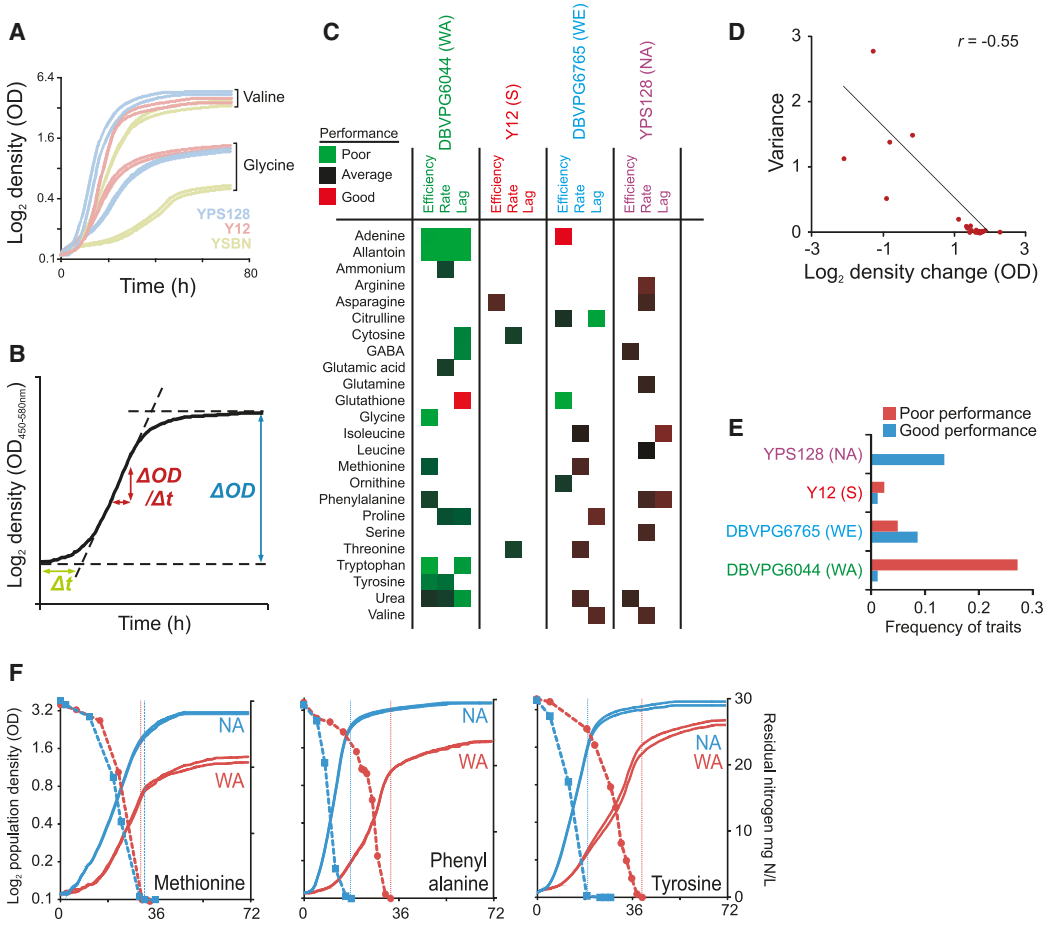


FIG. 1. Natural yeasts are highly differentiated for nitrogen source use. (*A*) Population density as a function of time for Y12 (S), YPS128 (NA), and the S288C derivative YSBN in sample single nitrogen-source environments ($n = 2$). (*B*) Extraction of lag (time to initiate proliferation), rate (population doubling time), and efficiency (total change in population density) of population growth from high-density growth curves. (*C*) Natural yeasts vary in capacity to convert different nitrogen sources into population growth. Nitrogen sources significantly (FDR, $q < 5\%$) better (red) or worse (green) for one natural yeast lineage ($n = 6$) than others ($n = 18$) are shown. (*D*) Variation in nitrogen source utilization between natural yeast lineages is largest for poor nitrogen sources. Variance over the four natural strains is plotted as a function of mean. Population growth efficiency is displayed. Line = linear regression. (*E*) The North American is superior and the West African inferior at utilizing single-nitrogen sources. Fraction of fitness component measures in which each strain ($n = 6$) is significantly (FDR, $q < 5\%$) better (blue) or worse (red) than other natural yeasts ($n = 18$) is shown. (*F*) Yeast variation in total population density reached reflects variations in nitrogen use efficiency. Population size (left *y*-axis, continuous bold lines) and nitrogen remaining in the medium (right *y*-axis, broken bold lines) was quantified as a function of time in West African and North American strains. Broken lines = time of external nitrogen depletion.

>98% of nucleotides are shared between strains, we expect much of the covariation between fitness components to arise from these invariant positions. We can estimate this ancestral coevolution by considering trait averages, over the four strains, as a proxy for ancestral traits. Comparing trait averages for different fitness components over the different environments, we find these to be strongly correlated (supplementary fig. S3A, Supplementary Material online). Thus, fitness components have largely coevolved with regards to nitrogen source use in the ancestor of current yeasts. Fitness component coevolution after the separation of lineages would be reflected in a correlation over the four strains. Given the low sample size (four), which adds substantial randomness that hides any existing covariation, we cannot provide an accurate estimate of recent fitness component coevolution. The average correlation ($r = 0.30$, considering each environment and pair of fitness component independently) is likely to be a substantial underestimate and should be regarded as lower bound for coevolution. Under an assumption of variations in fitness components being due to

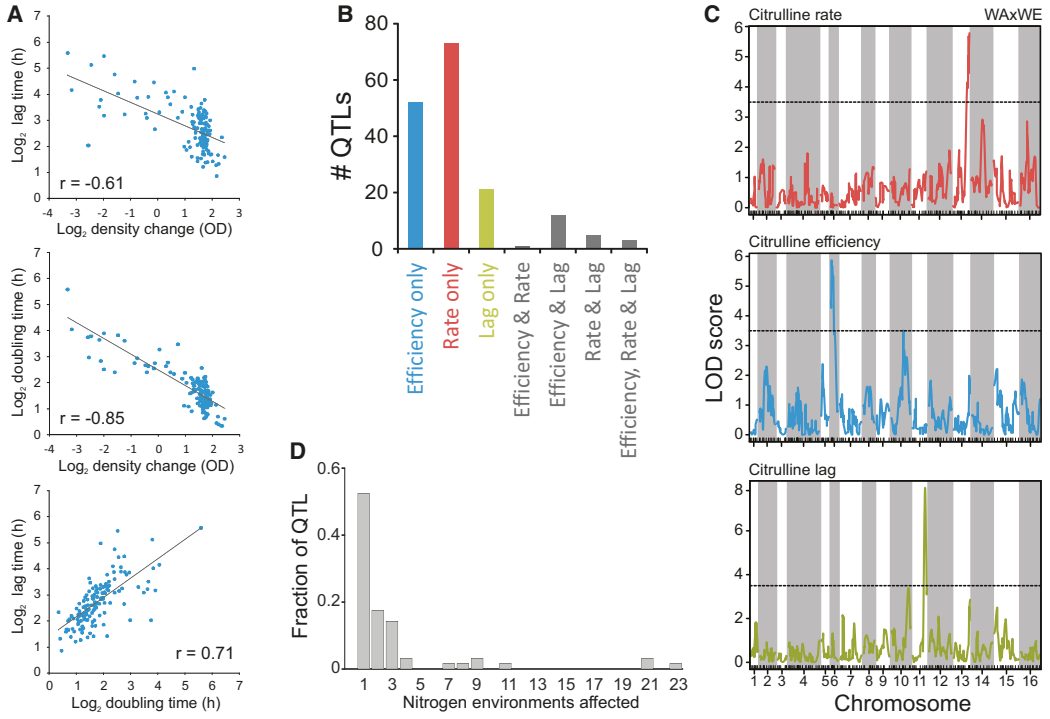


FIG. 2. Concerted selection on lag, rate, and efficiency of yeast nitrogen use. (*A*) The efficiency, rate, and lag of population growth are strongly correlated in natural yeast strains. The $\log_2$ of population growth efficiency, rate, and lag were pairwise compared over all natural yeasts and nitrogen environments. Means ($n = 6$), Pearson correlations (r), and linear regressions are shown. (*B*) Efficiency, rate, and lag of nitrogen source use are genetically independent. QTL (permutation test, $\alpha = 0.05$) private to a fitness component or shared between components were summed over six yeast crosses and 28 nitrogen environments. (*C*) QTL private to a single fitness component control variation in the efficiency, rate, and lag of citrulline use in the WE × WA cross. Logarithm of the odds (LOD) scores represent cosegregation of growth and genetic markers in 92 F1 recombinants. Bands represent chromosome boundaries. Dotted lines show threshold for significance (permutation test, $\alpha = 0.05$). (*D*) Nitrogen use QTL is private to a single nitrogen source. Fraction of total QTL detected for one or more than one nitrogen source is displayed. Each cross and fitness component was considered separately. There is a weak support for more environment pleiotropy than expected by chance (Mann–Whitney–Wilcoxon, $P = 0.016$).

different alleles, these correlations are challenging to explain without invoking selection. However, if underlying alleles are pleiotropic and consistently affect multiple fitness components in the same direction, selection need not necessarily be involved. Rejecting pleiotropy is not straight forward for alleles that are shared between strains. However, it can be done for alleles that differ between strains because following mating between two strains, we expect variations in two fitness components to cosegregate in offspring recombinants only if they are explained by the same, or genetically linked, alleles. We therefore mated the four natural lineages in all six pairwise combinations. Resulting hybrids were sporulated to obtain 92 F1 meiotic progenies from each cross. F1 segregants were genotyped at nucleotide sites for which parents were known from genome sequencing to be polymorphic (Cubillos et al. 2011). We then precisely followed the net population growth of these 552 recombinants in all nitrogen environments, extracting >42,000 fitness component measures. Finally, we evaluated the cosegregation of each SNP marker and each life history trait by QTL linkage mapping, for each cross and nitrogen environment separately. Overall, we detected 230 robust (permutation test, $\alpha = 0.05$) QTL. 87.4% of all QTL were unique for a single fitness component (fig. 2*B*), rejecting the null hypothesis of genetic variants with pleiotropic effects on fitness components. For example, QTLs private to lag, rate, or efficiency determined population growth in the West African × Wine/European cross in citrulline (fig. 2*C*), leucine, and isoleucine (supplementary fig. S3*B*, Supplementary Material online). In agreement with this fitness component specificity of QTLs, only weak correlation ($r = 0.15$) between fitness components remained in the average environment and cross. Together with the strong correlations between natural isolates, this supports that natural variation in yeast nitrogen source use indeed reflects adaptive differentiation.

QTLs were largely unique to a single nitrogen source (fig. 2*D* and supplementary fig. S4*A*, Supplementary Material online), meaning that different alleles tended to

control variation in the use of different nitrogen sources and are acted upon independently by selection. Nevertheless, a few extremely pleiotropic QTL, affecting the same fitness component in a large number of nitrogen-restricted environments, were observed. For example, a single QTL on chromosome VI accounted for poor growth efficiency of the Wine/European strain in almost all nitrogen-restricted niches, suggesting a general relaxation of selection for efficient nitrogen use in this strain. Predominantly, QTL also emerged only in a single of the three crosses in which a particular genetic background was represented (supplementary fig. S4*B*, Supplementary Material online). Therefore, the penetrance of the underlying alleles was highly dependent on genetic context, suggesting widespread epistasis. Although the results reported above provide strong indications, the limitations of QTL data should be acknowledged. Due to lack of power, detected QTLs do not explain all of the heritable variation in traits (Bloom et al. 2013). Furthermore, the breakup of parental allele structures during meiosis and the emergence of novel allele combinations can both disrupt and promote epistasis, affecting trait values. Finally, QTL represents the combined effect of all alleles in a region. These effects call for some caution and means that stringent conclusions can only be reached by identifying and measuring the effects of individual mutations.

Alleles Private to a Single Fitness Component Control Natural Variation in Yeast Nitrogen Source Use

To address the shortcoming of the QTL analysis, we traced the genetic basis of three QTLs down to single alleles. All nonessential genes within these three QTL regions were identified. We then crossed reference strain (BY4741) gene knockouts for each of these genes to both well and poor-performing natural strains, resulting in pairs of diploid hemizygotes. A trait difference within such hemizygotic pairs was taken to imply that the two alleles at the hemizygotic locus affect the trait differently. To exclude confounding effects from haploinsufficiency, we next mated a poor- and well-performing natural strain. In two of the resulting diploid hybrids, we reciprocally deleted either of the two alleles. A trait difference between these was considered final confirmation of the causative loci. We first focused on the chromosome VI QTL controlling variation exclusively in population growth efficiency in all crosses involving the Wine/European strain, and in almost all nitrogen environments (fig. 3*A* and supplementary fig. S5*A*, Supplementary Material online). Hemizygotic diploids only containing the WE allele of *RIM15* mimicked the poor growth efficiency of the WE parent (fig. 3*B*). Furthermore, the WE *RIM15* consistently imposed poor efficiency on hemizygotes from crosses with other natural yeasts (fig. 3*C* and supplementary fig. S5*B*, Supplementary Material online). In all cases, *RIM15* fully explained the defects of the WE parent. The WE *RIM15* contains an early two base pair insertion, *rim15c.459_460insCA*, shifting the reading frame to cause an early stop codon (supplementary fig. S5*C*, Supplementary Material online). Thus, WE *RIM15* is a null allele. No rate or lag QTL were found in the *RIM15* region in the relevant crosses and nitrogen sources and diploids only containing the WE *RIM15* allele showed no impairments of the rate or lag (supplementary fig. S5*D* and *F*, Supplementary Material online). *RIM15* encodes a poorly understood protein kinase known to control stationary phase entry (Wanke et al. 2005) and sporulation efficiency (Bergstrom et al. 2014). As sporulation is induced by nitrogen depletion, it is tempting to speculate that *rim15* sporulation defects arise as a consequence of nitrogen use impairment.

We next considered the strong chromosome XV QTL affecting population growth rate on proline in all crosses involving the West African DBVPG6044 (fig. 3*D*). Hemizygotic diploids from crosses with reference strain gene knockouts identified the West African *PUT4*, encoding a high-affinity proline permease (Andreasson et al. 2004), as the only allele in the region contributing to the trait variation (fig. 3*E*). Reciprocal *PUT4* hemizygotes from crosses between the West African and other natural strains confirmed that the West African *PUT4* impairs proline growth rate, accounting for 54–96%, depending on cross, of the West African defect (fig. 3*F*). No efficiency or lag proline QTL was found in the *PUT4* region and no diploids hemizygotic for the WA *PUT4* allele showed any impairments of population growth efficiency or lag (supplementary fig. S6*A*–*C*, Supplementary Material online). Thus, the *PUT4* defect is private to growth rate. WA *PUT4* allele harbors no nonsynonymous but several synonymous and promoter mutations, implying that expression differences cause the proline growth variation (supplementary fig. S7*A*, Supplementary Material online). WA *PUT4* is inherited by the SK1 and Y55 lab strains, but the proline defect is completely buffered in Y55, emphasizing that its exposure to selection depends on genetic context (supplementary fig. S7*B* and *C*, Supplementary Material online).

Finally, we considered the exceedingly slow growth of the West African when utilizing allantoin (fig. 3*G*), the nitrogen secretion product of mammals other than apes (Young et al. 1944). Allantoin lag time is unaffected, but the enormous growth rate effect precluded growth efficiency estimation. The defect affected close to 50% of WA offspring, regardless of cross, suggesting a monogenic effect (supplementary fig. S8*A*, Supplementary Material online). Arguing against monogenicity however, results from the lab strain cross implied that neither *DAL4* nor *DAL1 WA* alleles were able to support allantoin growth (fig. 3*H*). *DAL4* and *DAL1* encode enzymes catalyzing the first two steps in allantoin uptake and degradation and are arranged back-to-back in an allantoin use gene cluster (supplementary fig. S8*B* and *C*, Supplementary Material online). This arrangement fuelled concern that deletion of one *DAL* gene impairs expression of its neighbor, as suggested (Naseeb and Delneri 2012). Inspecting WA *DAL1* and *DAL4*, we uncovered a single nucleotide frameshifting insertion in *DAL4* and an early proline to serine nonsynonymous mutation, predicted to be strongly detrimental, in *DAL1* (supplementary fig. S8*D*, Supplementary Material online). Repairing either of these mutations with S288C variants failed to restore allantoin growth (fig. 3*I*), implying that they independently disrupt allantoin utilization. To confirm this, we repaired either of the two mutations while also

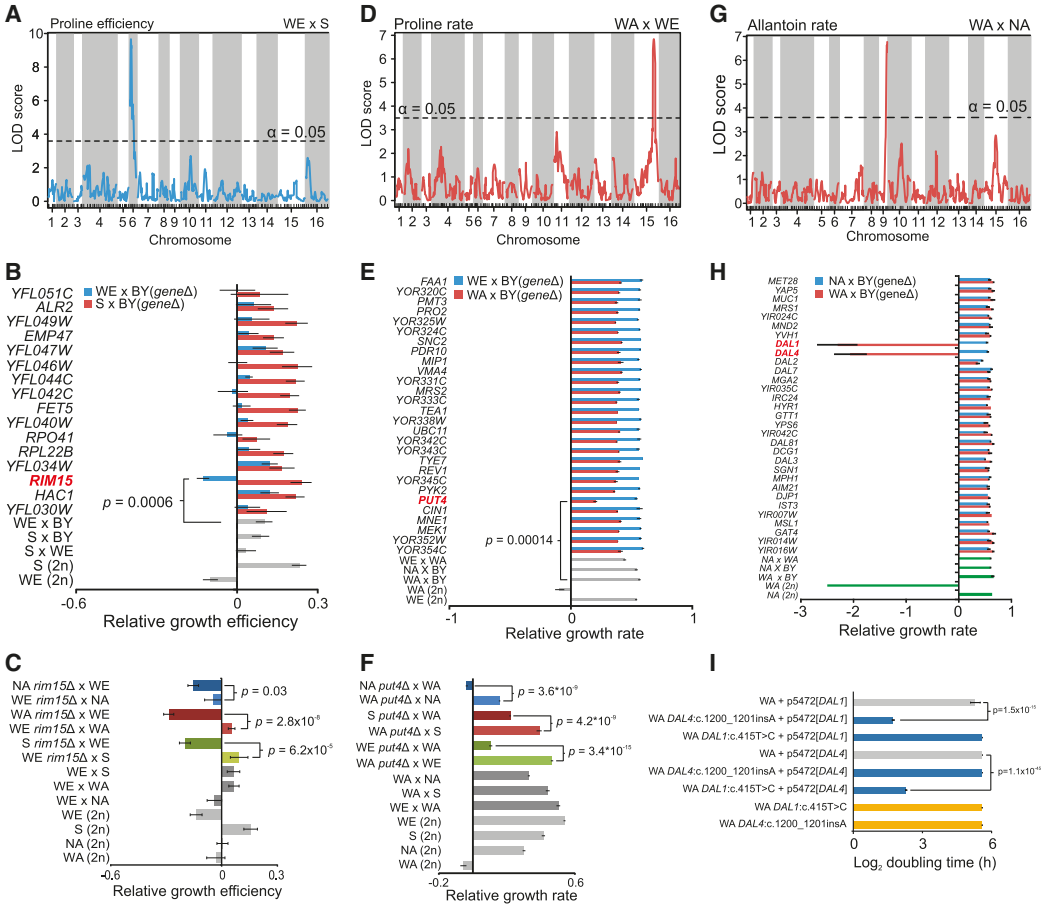


FIG. 3. Alleles private to single fitness components control variation in yeast nitrogen source use. (*A*–*C*) *RIM15* accounts for poor population growth efficiency of the WE lineage during nitrogen limitation. (*A*) LOD score plots of the cosegregation of growth efficiency and genetic markers among 92 F1 recombinants from the WE × S cross in proline. Bands represent chromosome boundaries and dotted line shows threshold for significance (permutation test, $\alpha = 0.05$). (*B*) Proline growth efficiency of diploid hemizygotes retaining only the WE (blue) or S (red) allele in crosses between WE or S and single BY4741 gene deletions. Gene deletions correspond to nonessential genes in the chromosome VI QTL region. Mean ($n = 2$) $\log_2$ values were normalized to those of the YSBN control ($n = 4$). Error bars = SEM. (*C*) Proline growth efficiency of reciprocal diploid hemizygotes retaining only the WE or only the alternate *RIM15* allele in crosses between WE and other natural lineages. Mean ($n = 8$) $\log_2$ values were normalized to those of the YSBN control ($n = 5$). Error bars = SEM, *P*-values = Student's *t*-test. (*D*–*F*) *PUT4* allele accounts for poor proline growth rate in the WA lineage. (*D*) LOD score plot of cosegregation of proline growth rate and genetic markers among 92 F1 recombinants from the WA × WE cross. Bands represent chromosome boundaries. Dotted line indicates threshold for significance (permutation test, $\alpha = 0.05$). (*E*) Proline growth rates of diploid hemizygotes retaining only the WE (blue) or WA (red) allele in crosses between WE or WA and single BY4741 gene deletions. Gene deletions correspond to nonessential genes in the chromosome XV QTL region. Mean ($n = 2$) $\log_2$ values were normalized to those of the YSBN control ($n = 4$). Error bars = SEM, *P*-values = Student's *t*-test. (*F*) Proline growth rates of reciprocal diploid hemizygotes retaining only the WA or only the alternate *PUT4* allele in crosses between WE and other natural lineages. Mean ($n = 8$) $\log_2$ values were normalized to those of the YSBN control ($n = 5$). Error bars = SEM ($n = 2$), *P*-values = Student's *t*-test. *PUT4* explained 97 ± 6% of the WA–NA, 67 ± 4% of the WA–WE, and 54 ± 3% of the WA–S variation. (*G*–*I*) A frameshift mutation in *DAL4* and a nonsynonymous mutation in *DAL1* both impair WA growth rate on allantoin. (*G*) LOD score plot of cosegregation of allantoin growth rate and each genetic marker among 92 F1 recombinants from the WA × WE cross. Bands represent chromosome boundaries. Dotted line shows threshold for significance (permutation test, $\alpha = 0.05$). (*H*) Allantoin growth rates of diploid hemizygotes retaining only the WE (blue) or WA (red) allele in crosses between WE or WA and single BY4741 gene deletions. Gene deletions correspond to nonessential genes in the chromosome IX QTL region. Mean ($n = 8$) $\log_2$ values were normalized to those of the YSBN control ($n = 5$). Error bars = SEM. (*K*) Addition of the S288C *DAL1* or *DAL4* to a haploid WA via a centromeric plasmid and replacement of the candidate SNP in the other *DAL* gene with the S288C nucleotide variant identified *dal4c.1201delA* and *dal1c.415C>T* as independently impairing the allantoin growth rate of WA. Mean ($n = 6$–12) $\log_2$ population growth rates are displayed. Error bars = SEM.

introducing a centromeric plasmid with the functional S288C version of the other allele. Perfect allantoin growth was restored in both constructs. Thus, the *DAL1* and *DAL4* are null alleles arising from *dal4c.1201delA* and *dal1c.415C > T* mutations, respectively. Each mutation disrupts growth to a degree that is comparable to the effect of both mutations together, reflecting positive epistasis between loss-of-function mutations in components of a linear pathway (Lehner 2011). Curiously the lab strain Y55, which has inherited both WA null alleles, achieved reasonably fast allantoin growth, suggesting the existence of unknown routes for allantoin use (supplementary fig. S8*E* and *F*, Supplementary Material online).

Discussion

From a genetics perspective, it is not surprising that distinct alleles control variation in lag, rate, and efficiency of population expansion. Effects on any single one of these direct components of fitness, both by environmental factors (Warringer et al. 2008) and gene deletion (Warringer et al. 2003), are common. Biochemically, a positive coupling between the rate and efficiency of population expansion is controversial because the rate of metabolic reactions is negatively correlated with the energy remaining after reactions (Westerhoff et al. 1983; Heinrich et al. 1997; Pfeiffer et al. 2001). Under energy restriction, this is expected to force a trade-off between population growth rate and efficiency and there is ample experimental support for this (Postma et al. 1989; Spor et al. 2008, 2009). However, under nitrogen restriction, accelerated burning of energy by fast metabolic reactions does not necessarily reduce biomass yields and increase in biomass yield does not require slow metabolism. Energy is available in excess. Consequently, the observed correlation between rate and efficiency does not violate thermodynamics. Assimilated nitrogen can either be stored or used inside cells or channeled into reproduction. Reflecting this distinction, the nitrogen content per cell can vary over orders of magnitude. Enhanced population growth efficiency under nitrogen restriction represents the prioritized channeling of nitrogen to the next generation. This is important, because it suggests an evolutionary mechanism whereby efficiency enhancing mutations may be selected. In absence of internal storages, efficiency enhancing mutations cannot increase frequency in well-mixed populations because nonmutants enjoy equal access to the resource and benefit equally from efficiency enhancements (Hardin 1968; MacLean 2008). Privatization of nitrogen by internal accumulation means that only genomes carrying efficiency enhancing mutations benefit from them, enabling selection on population growth efficiency.

The fundamental assumption of this article is that a strong correlation between fitness components that reflect performance in different life stages, but are nonpleiotropic, only can emerge through concerted selection. Concerted selection follows from the intrinsic link between life stages underlying these fitness components, which are exposed to the same environmental variations and selective pressures. However, the conjecture does not necessarily predict such a strong correlation as observed here. It is tempting to speculate that diminishing return of consecutive mutations on a trait (Chou et al. 2011; Khan et al. 2011; Kryazhimskiy et al. 2014) may constitute part of the explanation for the strength of detected correlations. First, it provides a powerful incentive for parallel accumulation of mutations enhancing different fitness components. Two mutations enhancing different fitness components will simply tend to have a larger aggregate fitness contribution than two mutations enhancing the same fitness component. Second, the diminishing return of consecutive mutations means that even if different numbers of beneficial mutations have been accumulated in different fitness components, their aggregate trait contributions will tend to be rather similar.

Although there are few intrinsic limitations to the applicability of the proposed test for selection, caveats need to be recognized. First, it is prudent to point out that our evidence for coevolution of fitness components is strongest with regard to coevolution in the ancestral lineage. In contrast, our evidence for nonpleiotropy between fitness components relates to variants that differ between strains, i.e., more recent evolutionary events. It is hard to envision why fixed variants would be largely pleiotropic, whereas polymorphisms would be largely nonpleiotropic with regard to effects on fitness components. Nevertheless, we can currently not completely rule out this possibility. Second, the test is straight forward to apply only to traits that directly reflect performance in different life stages. Great care should be taken before extending the test to organisms with hard to define, hard to measure, or hard to genetically dissect fitness components. Third, absence of correlation between fitness components is not a sufficient ground for rejecting selection. Over shorter time periods, optimization of a single-fitness component may occur, even if multiple fitness components are under selection, as few mutations are accumulated in each genome. This has been made abundantly clear in artificial laboratory selections (MacLean and Gudelj 2006; Novak et al. 2006; Bachmann et al. 2013). Optimization of single fitness components may also occur in organisms where life stages are less intimately linked and occur in different environmental contexts, as selection pressures then may be radically different. Fourth, the test is unable to distinguish between positive and recently relaxed selection. Here, the *RIM15*, *PUT4*, *DAL1*, and *DAL4* alleles were lineage-specific loss-of-function mutations, implying recent relaxation of selection in that lineage (Zorgo et al. 2012). Nevertheless, extension-of-function mutations in *RIM15*, *DAL4*, and *PUT4* emerge rapidly in nitrogen-restricted artificial laboratory selections (Hong and Gresham 2014). Hence, it cannot be excluded that loss-of-function mutations are ancestral and that independent repair recently have occurred in well-functioning lineages. Fifth, all fitness components should ideally be accounted for. We could not quantify effects on sexual recombination, which although exceedingly rare may be under selection or on diploid growth traits, which often differs from haploid growth traits (Zorgo et al. 2013). Furthermore, we could not measure performance in the yeast spore state, which may be under selection in nature (Neiman 2011), or in the net death phase, when growth has ceased. From a general perspective, not accounting for all possible

phenotypic and genetic relationships between fitness components leaves a margin for error. Assigning selection may still be valid, but the interpretation of which fitness components that are under selection may be confounded due to undetected pleiotropy between them.

Materials and Methods

Strains

Saccharomyces cerevisiae natural isolates were collected and haploidized as described (Cubillos et al. 2009; Liti et al. 2009). Haploids were crossed in all pairwise combinations. Twenty-three tetrads from each cross were sporulated to obtain 92 F1 recombinants. These were genotyped at 164–180 polymorphic positions (Cubillos et al. 2011). Diploid hemizygotes were obtained by mating haploids to single gene deletion BY4741 gene deletions using robotics. *RIM15* and *PUT4* reciprocal hemizygotes of crosses between natural isolates were obtained by one-step PCR and manual crossing. WA *dal4c.1201delA* and *dal1c.415C* > *T* mutations were replaced by the S288C variants using site-specific in vivo mutagenesis (Stuckey et al. 2011). WT and versions with either the *dal4c.1201delA* or *dal1c.415C* > *T* repaired were transformed with centromeric p5472 MoBY plasmids (Ho et al. 2009). A haploid S288C derivative, YSBN10, was used as control.

Medium and Cultivation

Microcultivation was performed in synthetic defined (SD) medium as described (Warringer and Blomberg 2003; Warringer et al. 2003), using a single nitrogen source present at 30 mg N/l. Experiments were run over 72–144 h until all isolates had entered stationary phase (Warringer and Blomberg 2003). Where a stationary phase had not been reached at the end of the experiment, experiments were discarded to avoid confounding effects. Population growth lag, rate, and efficiency were extracted as described (Warringer et al. 2008). Fitness components were $\log_2$ transformed and normalized to those of at least four controls. Means of replicates ($n = 2$) were used for linkage analysis. Other experiments were performed with larger numbers of repeats, as indicated.

Nitrogen Uptake

Microcultivation of yeast cells was performed in SD medium as described above. At each sampling time point, 20 biological replicates of each microcultivated sample were pooled. Supernatants were collected and analyzed by diethyl ethoxymethylenemalonate derivatization and high-performance liquid chromatography analysis as described (Gomez-Alonso et al. 2007). Concentrations of the relevant nitrogen source were calculated using internal and external standards.

QTL Analysis

QTL analysis was performed using a nonparametric model in R/QTL at 2-cM density (Broman et al. 2003). Significance was estimated by permutation tests ($\alpha = 0.05$). QTL positions were counted as identical if positions were within one average marker distance.

Sequence and SNP Analysis

Sequence data were taken from Bergstrom et al. (2014). Negative SNP consequences were predicted by Sorting Intolerant From Tolerant analysis (Kumar et al. 2009).

Statistics

Two group comparisons were performed using a homoscedastic two-tailed Student's *t*-test and FDRs at $q \leq 5\%$. Significance of pleiotropy was tested using randomized permutations and a Mann–Whitney–Wilcoxon test.

Supplementary Material

Supplementary figures S1–S8 and Supplementary Materials and Methods are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was supported by the Research Council of Norway under the eVITA program (grant numbers 178901/V30 and 222364/F20) and from the Carl Trygger foundation (grant number 08-400).

References

Andreasson C, Neve EP, Ljungdahl PO. 2004. Four permeases import proline and the toxic proline analogue azetidine-2-carboxylate into yeast. *Yeast* 21:193–199.

Bachmann H, Fischlechner M, Rabbers I, Barfa N, Branco Dos Santos F, Molenaar D, Teusink B. 2013. Availability of public goods shapes the evolution of competing metabolic strategies. *Proc Natl Acad Sci U S A.* 110:14302–14307.

Barton NH. 2000. Genetic hitchhiking. *Philos Trans R Soc Lond B Biol Sci.* 355:1553–1562.

Bergstrom A, Simpson JT, Salinas F, Barre B, Parts L, Zia A, Nguyen Ba AN, Moses AM, Louis EJ, Mustonen V, et al. 2014. A high-definition view of functional genetic variation from natural yeast genomes. *Mol Biol Evol.* 31(4):872–888.

Bloom JS, Ehrenreich IM, Loo WT, Lite TL, Kruglyak L. 2013. Finding the sources of missing heritability in a yeast cross. *Nature* 494:234–237.

Broman KW, Wu H, Sen S, Churchill GA. 2003. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19:889–890.

Chou HH, Chiu HC, Delaney NF, Segre D, Marx CJ. 2011. Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science* 332:1190–1192.

Cooper TG. 1982. Nitrogen metabolism in *Saccharomyces cerevisiae*. In: Strathern JN, Jones EW, Broach JR, editors. The molecular biology of the yeast *Saccharomyces cerevisiae*: metabolism and gene expression. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press. p. 39–99.

Cubillos FA, Billi E, Zorgo E, Parts L, Fargier P, Omholt S, Blomberg A, Warringer J, Louis EJ, Liti G. 2011. Assessing the complex architecture of polygenic traits in diverged yeast populations. *Mol Ecol.* 20:1401–1413.

Cubillos FA, Louis EJ, Liti G. 2009. Generation of a large set of genetically tractable haploid and diploid Saccharomyces strains. *FEMS Yeast Res.* 9:1217–1225.

Fraser HB, Moses AM, Schadt EE. 2010. Evidence for widespread adaptive evolution of gene expression in budding yeast. *Proc Natl Acad Sci U S A.* 107:2977–2982.

Gardener MC, Gillman MP. 2001. Analyzing variability in nectar amino acids: composition is less variable than concentration. *J Chem Ecol.* 27:2545–2558.

Gomez-Alonso S, Hermosin-Gutierrez I, Garcia-Romero E. 2007. Simultaneous HPLC analysis of biogenic amines, amino acids, and ammonium ion as aminoenone derivatives in wine and beer samples. *J Agric Food Chem.* 55:608–613.

Gutierrez A, Beltran G, Warringer J, Guillamon JM. 2013. Genetic basis of variations in nitrogen source utilization in four wine commercial yeast strains. *PLoS One* 8:e67166.

Hardin G. 1968. The tragedy of the commons. The population problem has no technical solution; it requires a fundamental extension in morality. *Science* 162:1243–1248.

Heinrich R, Montero F, Klipp E, Waddell TG, Melendez-Hevia E. 1997. Theoretical approaches to the evolutionary optimization of glycolysis: thermodynamic and kinetic constraints. *Eur J Biochem.* 243: 191–201.

Hittinger CT. 2013. *Saccharomyces* diversity and evolution: a budding model genus. *Trends Genet.* 29:309–317.

Ho CH, Magtanong L, Barker SL, Gresham D, Nishimura S, Natarajan P, Koh JL, Porter J, Gray CA, Andersen RJ, et al. 2009. A molecular barcoded yeast ORF library enables mode-of-action analysis of bioactive compounds. *Nat Biotechnol.* 27:369–377.

Hong J, Gresham D. 2014. Molecular specificity, convergence and constraint shape adaptive evolution in nutrient-poor environments. *PLoS Genet.* 10:e1004041.

Khan AI, Dinh DM, Schneider D, Lenski RE, Cooper TF. 2011. Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* 332:1193–1196.

Kryazhimskiy S, Rice DP, Jerison ER, Desai MM. 2014. Microbial evolution. Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science* 344:1519–1522.

Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc.* 4:1073–1081.

Landry CR, Townsend JP, Hartl DL, Cavalieri D. 2006. Ecological and evolutionary genomics of *Saccharomyces cerevisiae*. *Mol Ecol.* 15: 575–591.

Lehner B. 2011. Molecular mechanisms of epistasis within and between genes. *Trends Genet.* 27:323–331.

Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA, Davey RP, Roberts IN, Burt A, Koufopanou V, et al. 2009. Population genomics of domestic and wild yeasts. *Nature* 458:337–341.

MacLean RC. 2008. The tragedy of the commons in microbial populations: insights from theoretical, comparative and experimental studies. *Heredity* 100:471–477.

MacLean RC, Gudelj I. 2006. Resource competition and social conflict in experimental populations of yeast. *Nature* 441:498–501.

Naseeb S, Delneri D. 2012. Impact of chromosomal inversions on the yeast DAL cluster. *PLoS One* 7:e42022.

Neiman AM. 2011. Sporulation in the budding yeast *Saccharomyces cerevisiae*. *Genetics* 189:737–765.

Novak M, Pfeiffer T, Lenski RE, Sauer U, Bonhoeffer S. 2006. Experimental tests for an evolutionary trade-off between growth rate and yield in *E. coli*. *Am Nat.* 168:242–251.

Orr HA. 1998. Testing natural selection vs. genetic drift in phenotypic evolution using quantitative trait locus data. *Genetics* 149: 2099–2104.

Pfeiffer T, Schuster S, Bonhoeffer S. 2001. Cooperation and competition in the evolution of ATP-producing pathways. *Science* 292:504–507.

Postma E, Verduyn C, Scheffers WA, Van Dijken JP. 1989. Enzymic analysis of the crabtree effect in glucose-limited chemostat cultures of *Saccharomyces cerevisiae*. *Appl Environ Microbiol.* 55:468–477.

Roff DA. 1992. The evolution of life histories: theory and analysis. New York: Chapman and Hall.

Ruderfer DM, Pratt SC, Seidel HS, Kruglyak L. 2006. Population genomic analysis of outcrossing and recombination in yeast. *Nat Genet.* 38: 1077–1081.

Spor A, Nidelet T, Simon J, Bourgais A, de Vienne D, Sicard D. 2009. Niche-driven evolution of metabolic and life-history strategies in natural and domesticated populations of *Saccharomyces cerevisiae*. *BMC Evol Biol.* 9:296.

Spor A, Wang S, Dillmann C, de Vienne D, Sicard D. 2008. "Ant" and "grasshopper" life-history strategies in *Saccharomyces cerevisiae*. *PLoS One* 3:e1579.

Stearns FW. 2010. One hundred years of pleiotropy: a retrospective. *Genetics* 186:767–773.

Stearns SC. 1992. The evolution of life histories. Oxford: Oxford University Press.

Stuckey S, Mukherjee K, Storici F. 2011. In vivo site-specific mutagenesis and gene collage using the delitto perfetto system in yeast *Saccharomyces cerevisiae*. *Methods Mol Biol.* 745:173–191.

Tsai IJ, Bensasson D, Burt A, Koufopanou V. 2008. Population genomics of the wild yeast *Saccharomyces paradoxus*: quantifying the life cycle. *Proc Natl Acad Sci U S A.* 105:4957–4962.

Wanke V, Pedruzzi I, Cameroni E, Dubouloz F, De Virgilio C. 2005. Regulation of G0 entry by the Pho80-Pho85 cyclin-CDK complex. *EMBO J.* 24:4271–4278.

Warringer J, Anevski D, Liu B, Blomberg A. 2008. Chemogenetic fingerprinting by analysis of cellular growth dynamics. *BMC Chem Biol.* 8:3.

Warringer J, Blomberg A. 2003. Automated screening in environmental arrays allows analysis of quantitative phenotypic profiles in *Saccharomyces cerevisiae*. *Yeast* 20:53–67.

Warringer J, Blomberg A. 2014. Yeast phenomics-large scale mapping of the genetic basis for organismal traits. In: Hancock JM, editor. Phenomics. Boca Raton: CRC Press. p. 172–207.

Warringer J, Ericson E, Fernandez L, Nerman O, Blomberg A. 2003. High-resolution yeast phenomics resolves different physiological features in the saline response. *Proc Natl Acad Sci U S A.* 100:15724–15729.

Warringer J, Zorgo E, Cubillos FA, Zia A, Gjuvsland A, Simpson JT, Forsmark A, Durbin R, Omholt SW, Louis EJ, et al. 2011. Trait variation in yeast is defined by population history. *PLoS Genet.* 7: e1002111.

Westerhoff HV, Hellingwerf KJ, Van Dam K. 1983. Thermodynamic efficiency of microbial growth is low but optimal for maximal growth rate. *Proc Natl Acad Sci U S A.* 80:305–309.

Young EG, Wentworth HP, Hawkins WW. 1944. The absorption and excretion of allantoin in mammals. *J Pharmacol Exp Therapeut.* 81: 1–9.

Zorgo E, Chwialkowska K, Gjuvsland AB, Garre E, Sunnerhagen P, Liti G, Blomberg A, Omholt SW, Warringer J. 2013. Ancient evolutionary trade-offs between yeast ploidy states. *PLoS Genet.* 9:e1003388.

Zorgo E, Gjuvsland A, Cubillos FA, Louis EJ, Liti G, Blomberg A, Omholt SW, Warringer J. 2012. Life history shapes trait heredity by accumulation of loss-of-function alleles in yeast. *Mol Biol Evol.* 29: 1781–1789.

II

Published online: December 15, 2016

Report

SOURCE DATA | TRANSPARENT PROCESS | OPEN ACCESS

molecular systems biology

Disentangling genetic and epigenetic determinants of ultrafast adaptation

Arne B Gjuvsland[1,*], Enikö Zörgö[1,2], Jeevan KA Samy[1], Simon Stenberg[1], Ibrahim H Demirsoy[2], Francisco Roque[3], Ewa Maciaszczyk-Dziubinska[4], Magdalena Migocka[4], Elisa Alonso-Perez[2], Martin Zackrisson[2], Robert Wysocki[4], Markus J Tamás[2], Inge Jonassen[3], Stig W Omholt[5] & Jonas Warringer[1,2,**]

Abstract

A major rationale for the advocacy of epigenetically mediated adaptive responses is that they facilitate faster adaptation to environmental challenges. This motivated us to develop a theoretical–experimental framework for disclosing the presence of such adaptation-speeding mechanisms in an experimental evolution setting circumventing the need for pursuing costly mutation–accumulation experiments. To this end, we exposed clonal populations of budding yeast to a whole range of stressors. By growth phenotyping, we found that almost complete adaptation to arsenic emerged after a few mitotic cell divisions without involving any phenotypic plasticity. Causative mutations were identified by deep sequencing of the arsenic-adapted populations and reconstructed for validation. Mutation effects on growth phenotypes, and the associated mutational target sizes were quantified and embedded in data-driven individual-based evolutionary population models. We found that the experimentally observed homogeneity of adaptation speed and heterogeneity of molecular solutions could only be accounted for if the mutation rate had been near estimates of the basal mutation rate. The ultrafast adaptation could be fully explained by extensive positive pleiotropy such that all beneficial mutations dramatically enhanced multiple fitness components in concert. As our approach can be exploited across a range of model organisms exposed to a variety of environmental challenges, it may be used for determining the importance of epigenetic adaptation-speeding mechanisms in general.

Keywords adaptation; epigenetics; evolution; modelling; population genetics
Subject Categories Evolution; Genome-Scale & Integrative Biology
DOI 10.15252/msb.20166951 | Received 15 March 2016 | Revised 11 November 2016 | Accepted 16 November 2016
Mol Syst Biol. (2016) 12: 892

Introduction

The need for an extended evolutionary theory where epigenetic mechanisms have a more prominent explanatory position is a much-debated issue (Laland *et al*, 2014). This discussion has arisen due to a deeper understanding of the epigenetic mechanisms underlying phenotypic plasticity and parental influence (Rando & Verstrepen, 2007; Carone *et al*, 2010; Halfmann & Lindquist, 2010; Daxinger & Whitelaw, 2012). A major rationale for advocating an important role for environmentally guided DNA, RNA, protein and metabolite alterations mediated by epigenetic adaptive mechanisms is that such alterations provide an evolutionary advantage by facilitating faster adaptation (Richards, 2006) to new or recurrent environmental changes. To assess the adaptive importance of epigenetic mechanisms relative to a pure mutation–selection regime for a variety of adaptations in a wide range of organisms is a challenging undertaking, however. Even in those cases where we have identified causative genetic variation underlying a specific adaptation and thus may be tempted to promote a gene-centric explanation, we have to show that epigenetic mechanisms have not acted transiently during the adaptation process to guide gene-based solutions by allowing silenced variation to take effect (Masel & Siegal, 2009; Halfmann *et al*, 2010), altering mutation effect sizes (Laland *et al*, 1999; Plucain *et al*, 2014), or enhancing mutation rates either locally (Molinier *et al*, 2006; MacLean *et al*, 2013) or globally (Roth *et al*, 2006; Zhang & Saier, 2009; Martincorena & Luscombe, 2013) through elevated DNA damage (Ruden *et al*, 2008) or impaired DNA repair (Tu *et al*, 1996; Hoege *et al*, 2002; Moore *et al*, 2014; Supek & Lehner, 2015). Such documentation is arguably beyond reach through studies of natural adaptations, but could conceivably be addressed by artificial selection experiments in the laboratory (Conrad *et al*, 2011; Dettman *et al*, 2012).

State-of-the-art experimental evolution methodology can verify the long-term stability of an adaptation after removal of the

1 Centre for Integrative Genetics (CIGENE), Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, Ås, Norway
2 Department of Chemistry and Molecular Biology, University of Gothenburg, Gothenburg, Sweden
3 Computational Biology Unit, University of Bergen, Bergen, Norway
4 Institute of Experimental Biology, University of Wroclaw, Wroclaw, Poland
5 Centre for Biodiversity Dynamics, Department of Biology, NTNU – Norwegian University of Science and Technology, Trondheim, Norway
*Corresponding author. Tel: +47 67232713; E-mail: arne.gjuvsland@nmbu.no
**Corresponding author. Tel: +46 317863961; E-mail: jonas.warringer@cmb.gu.se

selection regime creating it, and reversion of candidate mutations by gene editing can validate their adaptive effect. However, refuting epigenetic adaptive mechanisms reviving silenced mutations or influencing mutation rate demands costly mutation–accumulation experiments that precisely mimic the adaptive regime and can currently only address the issue of general changes in mutation rates (Zhu *et al*, 2014). Thus, unless one can remedy these obstacles, even an experimental evolution framework appears impracticable for determining the adaptive importance of epigenetics.

We hypothesized that a possible route to overcome these obstacles would be to make use of a theoretical–experimental approach involving a data-driven evolutionary population model capable of explaining experimental results as a function of mutation effect sizes, mutation target sizes and mutation rate changes. As very fast adaptations are arguably a natural point of departure to search for evolutionarily important epigenetic adaptive mechanisms, we tested this hypothesis by precisely tracking the adaptation dynamics of budding yeast populations adapting to a panel of environmental challenges and by making use of an individual-based evolutionary population model to explain the fastest adaptive trajectories. We found that ultrafast arsenic adaptation could be fully accounted for by gene-based solutions causing extensive positive pleiotropy between fitness components. And we could only account for the experimentally observed homogeneity of adaptive speed and heterogeneity of molecular solutions if mutation rates were close to empirical estimates of the basal mutation rate. As the introduced theoretical–experimental approach can be exploited across a wide range of model organisms and environments (Long *et al*, 2015), it can tentatively become an instrumental generic tool for illuminating the influence of epigenetics on adaptation.

Results

Arsenic adaptation is ultrafast and heritable

To identify growth challenges eliciting ultrafast adaptation, we exposed $n = 4$ independent haploid yeast populations, derived from a single clone, to each of 18 energy-constrained environments over 250 generations (Appendix Table S1). The iterative batch experimental evolution (Appendix Fig S1A) forced adapting populations to cycle through a lag phase, an exponential growth phase and a stationary growth phase. We tracked the associated fitness components—length of lag phase, population doubling time and efficiency of growth—at high accuracy (Fig 1A). The four populations, hereafter termed P1–P4, exposed to arsenic in its most toxic form, As(III), adapted faster than populations exposed to other challenges and went from poor to optimal performance for all three fitness components within just a few mitotic divisions (Fig 1B and C, and Appendix Fig S1B). In the absence of arsenic, the adapted strain performed on par with the founder; thus, fitness component increases were adaptive responses to arsenic and not to other selective pressures (Fig 1B). We found estimates of the number of viable cells (colony-forming units; CFU) to match estimates of the total number of cells in populations and to be unaffected by the presence of As(III) (Appendix Fig S1C). The three estimated fitness components therefore reflected the time to the first cell division, cell division time and the energy efficiency of cells, and together they should capture total fitness well.

The extraordinarily fast As(III) adaptations could conceivably be due to a single rare adaptive mutation standing at substantial frequency in the shared founder population rather than *de novo* mutations. To account for this possibility, we repeated the arsenic adaptations in four new populations, hereafter termed P5–P8, that were initiated from distinct founder populations derived by clonal expansion from four different single cells (see Materials and Methods). As the original As(III) adapting populations (P1–P4), populations P5–P8 showed the ultrafast and near-deterministic adaptive leaps. With the higher sampling density these were detectable already after 10 generations (Fig EV1). The probability of adaptive variants standing in all P5–P8 populations was estimated to be 3.9×10^{-5} (see Materials and Methods).

We tested the possibility of the ultrafast adaptation being directly due to non-genetic mechanisms by releasing each of the four adapted populations P1–P4 from selection for 75 generations. All populations retained their extreme As(III) tolerance (Fig 1D). This excludes a plain phenotypic plasticity mechanism at the level of the individual cell. Moreover, we are not aware of any reports of transgenerational epigenetic inheritance of fitness over 75 generations, making it a quite unlikely explanation.

Ultrafast arsenic adaptation is driven by *FPS1*, *ASK10* and *ACR3* mutations

The genetic nature of As(III) adaptation motivated us to sequence the end point populations P1–P4 by SoLID sequencing to identify *de novo* single nucleotide polymorphisms (SNPs) and copy number variations (CNVs) rising to high frequencies (Appendix Tables S2 and S3). To identify and validate causative mutations, 35 of the top candidates were individually reconstructed in founder cells and fitness components were recorded (Fig 2A). About 75% of final adaptive gains in each of the four populations could be explained by a single population specific mutation. Adaptation in P2, P3 and P4 was predominantly due to a premature stop codon in *FPS1* (P4; encoding the aquaglyceroporin through which As(III) enters the cell; Wysocki *et al*, 2001), a duplication of *ACR3* (P2; encoding the As(III) exporter; Wysocki *et al*, 1997) and a loss-of-function SNP in *ASK10* (P3; a positive regulator of Fps1; Lee *et al*, 2013), respectively (Fig EV2). The mutations were neutral (Acr3, Ask10) or negative (Fps1) in absence of As(III), excluding that they were driven to high frequencies by other selection pressures (Appendix Fig S2). The P1 population harboured two medium-frequency *FPS1* nonsynonymous mutations (A410S and F413L) that resisted reconstruction, but were predicted to impair Fps1 function by amino acid conservation over Fps1 orthologs. The P1 mutations occurred in different haplotypes (reads), affecting close to the entire population (Appendix Table S2).

Conceivably, early As(III) adaptation could have been epigenetic in origin and only later assimilated as *FPS1*, *ASK10* and *ACR3* mutations into the genome (Pigliucci *et al*, 2006). We therefore sequenced the P1–P4 populations throughout their adaptive trajectories by Illumina sequencing, tracking the frequency change of these causative mutations over time (Fig 2B). *FPS1*, *ASK10* and *ACR3* mutations were all at undetectable frequencies in the founder population, emerged and rose in frequency early and were practically

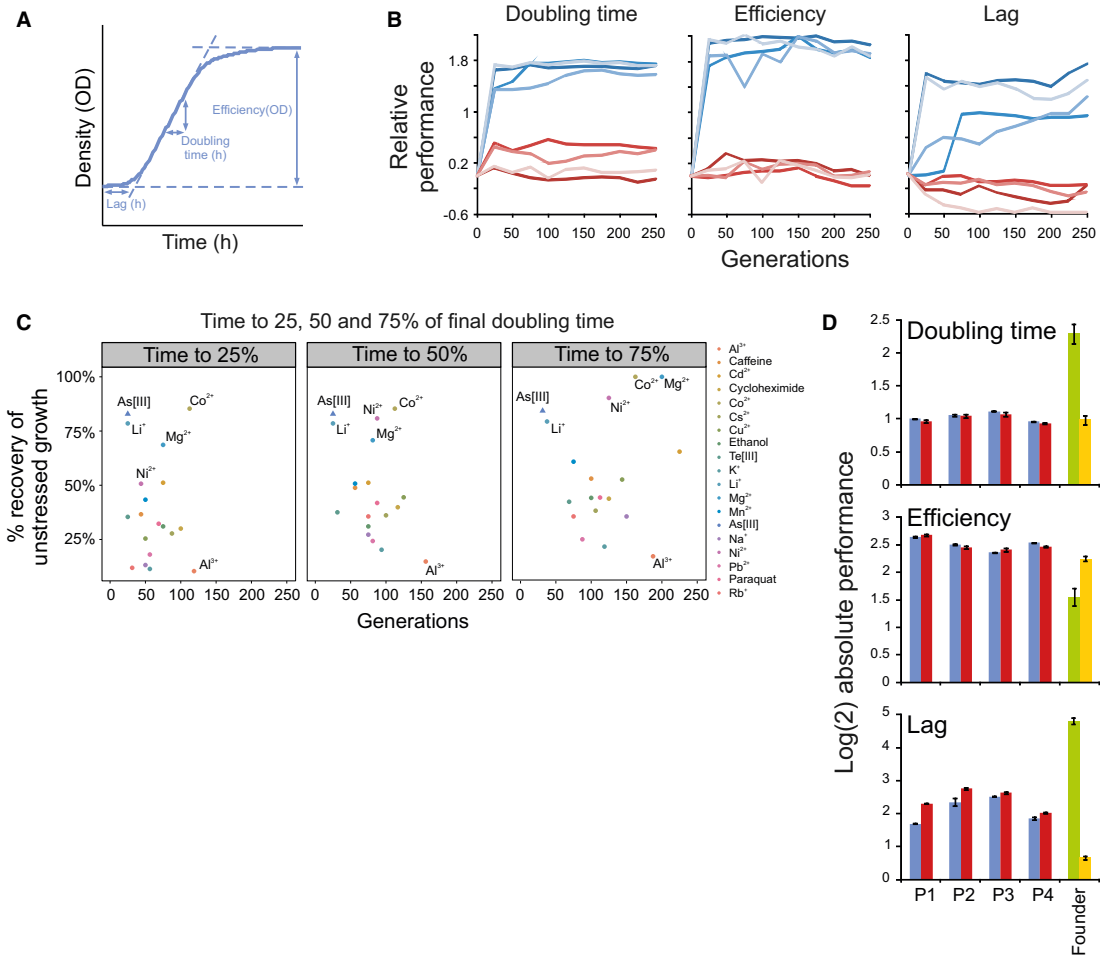


Figure 1. Arsenic adaptation is ultrafast and heritable.

A Schematic illustration of the extraction of the fitness components: length of lag phase (h), growth rate (doubling time, h) and growth efficiency (total change in population size, OD). Absolute fitness components were log(2)-transformed. When comparisons across experimental plates were performed, absolute log(2) values were first normalized to the corresponding mean value of many founder populations in randomized positions on the same plate, producing relative fitness components (see Materials and Methods). A positive relative performance equals good growth.

B Fitness components of As(III) adapting populations (n = 2) relative the founder (n = 4). Blue = 5 mM As(III). Red = no As(III). Colour = populations P1–P4.

C Mean adaptation speed under 18 selection pressures (n = 4 independent populations) for doubling time. A monotone function was fitted to each adaptation curve using least squares and the function *isoreg* in *R* (version 2.15.3). Two measures of adaptive speed were extracted from the function: (x-axis) the number of generations required to reach 25, 50 and 75% of the final doubling time (t = 250 generations) and (y-axis) the fraction (%) of the initial gap to the founder doubling time in optimal environments (no stress added) that was recovered at these time points. Colour indicates challenge.

D Absolute log(2) fitness components of arsenic-adapted (t = 250 generations) populations in 5 mM As(III), before (blue bars) and after (red bars) a 75 generations release from selection. Green bars: founder in 5 mM As(III). Yellow bars: founder without arsenic. Error bars represent SEM (n = 2).

Source data are available online for this figure.

fixated before 100 generations. The rise in frequency of the *ACR3* duplication was slightly delayed relative the adaptive progression in P2 (compare: Figs 1B and 2B), which may be due to sequencing or negative selection against the large duplication during sequencing preparations. No other SNPs from the original end point sequencing were confidently called at earlier time points, and only a few previously undetected SNPs were discovered (Appendix Fig S3). Of these, only three were predicted to affect protein function (see Materials

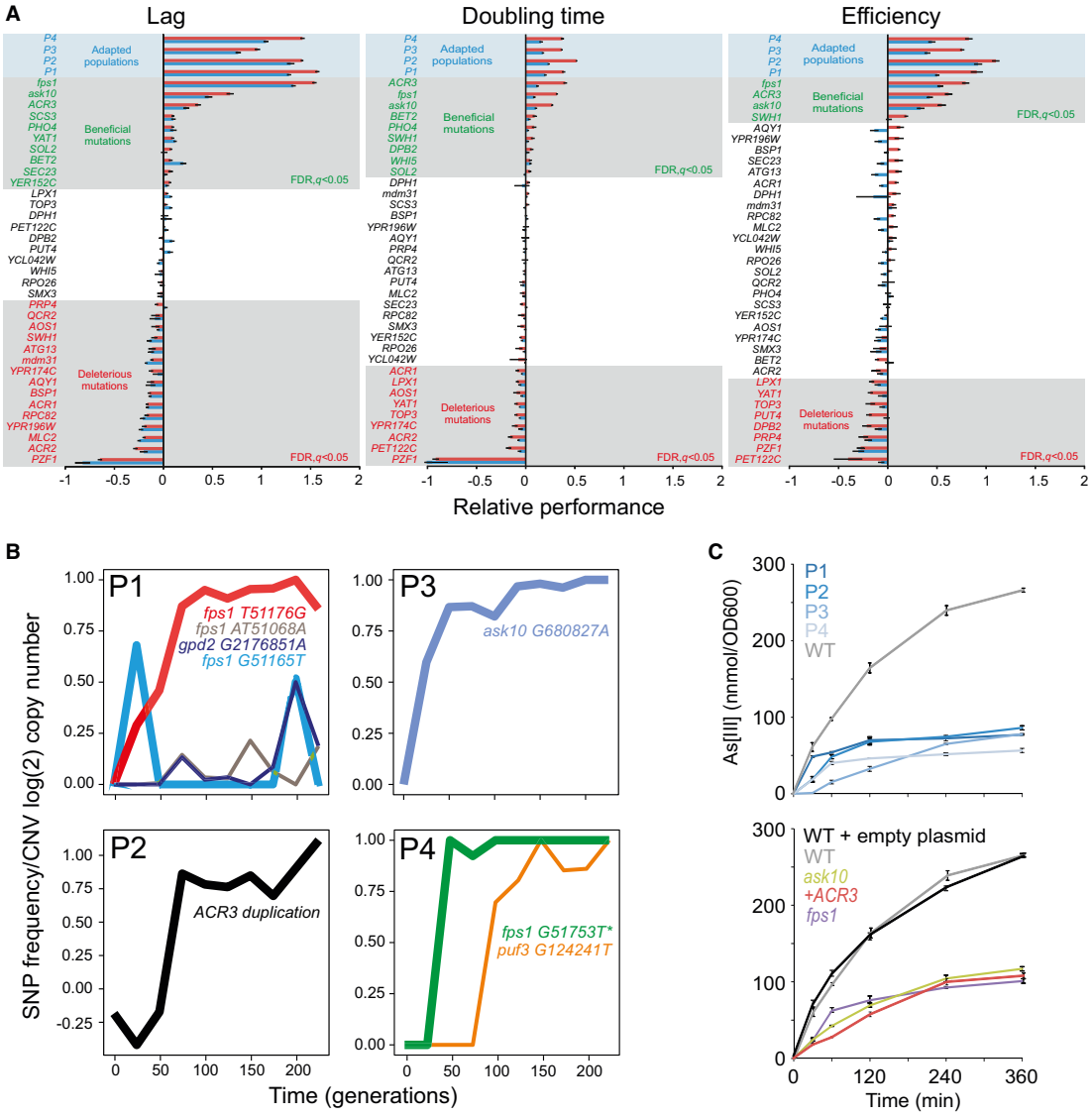


Figure 2. Ultrafast arsenic adaptation is due to rapid fixation of positively pleiotropic *FPS1*, *ASK10* and *ACR3* mutations.

A Candidate driver mutations (gene duplications or SNP; see Materials and Methods) were reconstructed individually in a WT background and their growth in presence of As(III) evaluated. Mean (n = 2) fitness components in 3 and 5 mM As(III) (blue and red bars, respectively) relative to the founder (n = 20) are shown. Grey field = significant (FDR, q < 0.05) effects at 5 mM As(III). Blue field = As(III) adapted populations. Error bars represent SEM.

B Adapting P1–P4 populations were sampled at every 25th generation and deep sequenced. The frequency of confidently called mutations (total read depth of > 100, frequency of > 10% in ≥ 2 time points and snpEff effect = "Moderate" or "High") predicted to affect protein function (SIFT < 0.05) is shown. Fps1 G51165T failed to pass the quality filter in the re-sequencing but is shown for completeness. For the *ACR3* containing duplicated region in P2, the grand copy number mean of all the segments within the duplicated region (chr. XVI 880799–944600) is shown. Colours indicate mutations. Bold line = causative mutations. *reconstructed *FPS1* mutation.

C Arsenic accumulation inside cells. Top panel: Arsenic-adapted populations (t = 250 generations). Bottom panel: As(III) causative mutations individually reconstructed in WT backgrounds. Error bars = SD (n = 2).

Source data are available online for this figure.

and Methods): a previously undetected, low-frequency frameshift in *FPS1* (P1), a late emerging and low-frequency *GPD2* mutation (P1) and a late emerging mutation in *PUF3* (P4) (Fig 2B). Neither Gpd2, the minor isoform of the glycerol dehydrogenase, nor Puf3, involved in mitochondrial function and mRNA stability, is known to be linked to As(III) metabolism. Overall, although we cannot completely exclude very small transient contributions from epigenetics or from variants in other genes, loss-of-function mutations in *FPS1* and *ASK10* and duplications of *ACR3* emerged as the dominant proximal causes of ultrafast As(III) adaptation.

As the identified causative mutations implied adaptation to be mediated by As(III) exclusion, we followed the accumulation of arsenic inside cells. We found it to be delayed and stabilized at a lower final level in all adapted populations (Fig 2C, top panel). The reduced intracellular arsenic levels were almost completely accounted for by the reconstructed *fps1*, *ask10* and *ACR3* mutations (Fig 2C, bottom panel). Thus, by reducing the intracellular concentration of arsenic, these mutations affected all three fitness components through the same mechanism: As(III) exclusion.

Ultrafast arsenic adaptation is due to positive pleiotropy between fitness components

Given these experimental results, we assessed quantitatively to what extent the observed adaptive trajectories could be accounted for by plain neo-Darwinian mechanisms. We employed an evolutionary population model based on individual cells that combined population genetics and population dynamics through genotype–phenotype maps with parameters describing mutation rates and effect sizes. As each of the three reconstructed critical mutations had a large positive effect on all three fitness components, we first evaluated whether positive pleiotropy was needed to account for the ultrafast adaptation. To this end, we simulated the experimental setup while varying mutation parameters and using three types of genotype–phenotype maps: mutations affecting only cell division time (population doubling time), mutations affecting both cell division time and the time to the first cell division (population lag time) in the same direction (positive pleiotropy) and mutations affecting both doubling time and lag but with independently sampled effects (random assignment of positive and negative pleiotropy). Efficiency was not taken into account because it may not confer fitness benefits in an energy-restricted regime (MacLean, 2008) and would require a much more complex model with weak empirical backing.

The parameter set giving the fastest adaptation for the doubling time-only model clearly failed to approach the ultrafast adaptation of arsenic adaptations (Fig 3A) despite: (i) having overall mutation rates 5× those reported in yeast (Lynch *et al*, 2008), (ii) 65% of mutations being non-neutral and 17% of mutations being beneficial (both values 5× the numbers reported by (Hall *et al*, 2008) and (iii) a mean selection coefficient of 0.15 (2.5× the level reported by Joseph & Hall, 2004). Overall, populations adapted dramatically faster in the models with pleiotropy between fitness components than in the models without fitness component pleiotropy (Figs 3B and EV3A and B). Populations exclusively experiencing positive pleiotropy adapted only moderately faster than populations experiencing both positive and negative pleiotropy. Thus, positive pleiotropy between fitness components can indeed accelerate adaptation drastically, and the presence of negative pleiotropy only moderately limits this acceleration. The benefits of positive pleiotropy were similar for slow- and fast-adapting populations (Appendix Fig S4). With positive pleiotropy included, the fastest scenarios approached the observed arsenic adaptations (Fig 3B).

Next, we used the empirical lag and doubling time values for reconstructed mutations and simulated the fate of the *de novo FPS1, ASK10* and *ACR3* mutations in competition assays, starting with a single-mutant cell at the start of the first batch cycle in an otherwise homogenous founder population (Fig 3C). We simulated mutants with only the doubling time effect of the reconstructed mutation, only the lag effect and both the doubling time and the lag effect. Factoring in both the doubling time and lag effect strongly reduced the risk of losing the *FPS1* mutation due to chance (loss in 14 of 25 doubling time scenarios vs. 0 of 25 doubling time and lag scenarios) and accelerated fixation of remaining mutations. The positively pleiotropic doubling time and lag effects combined into a very large selection coefficient ($s = 0.64$; Appendix Fig S5), driving the mutation to fixation in ~25 generations. Thus, the predicted adaptive performance approached that of the experimentally ultrafast arsenic adaptations, without even taking efficiency into account. The selection coefficients for *ASK10* and *ACR3* were somewhat lower ($s = 0.41$ and $s = 0.36$ respectively, Appendix Fig S5), causing a slightly longer fixation time (Fig 3C). The additive effects of doubling time and lag changes on fitness are shown analytically in the Materials and Methods section.

Ultrafast arsenic adaptation occurs at near-basal mutation rates

Even though the competition simulations show that the focal causal mutations can account for the ultrafast adaptation, these results rest on that mutations emerge early. Given that epigenetic mechanisms can direct DNA damage and DNA repair (Molinier *et al*, 2006; Roth *et al*, 2006; Zhang & Saier, 2009; MacLean *et al*, 2013; Martincorena & Luscombe, 2013) to elevate mutation rates, this early emergence may in principle be epigenetically facilitated. To resolve this issue, we defined the beneficial mutation target set to contain *ACR3* duplications and loss-of-function mutations in *FPS1* and *ASK10*, the mutational targets of the latter being premature stop codons and changes in strongly conserved amino acids (SIFT < 0.06; Appendix Fig S6A and B). We simulated populations experiencing basal as well as elevated (3×, 5× and 10×) point mutation and duplication rates and tracked the frequencies of all beneficial mutations using lag and rate values equalling those empirically observed. At basal mutation rates, the founder genotype went extinct within 25 generations in the fastest evolving populations (Fig 3D—upper left panel). While the predicted large variations in adaptive speed between populations at a basal mutation rate (Fig 3D—upper left panel; compare time to vertical black line) are a possible scenario, it produces a distribution of simulated adaptations from which we would be unlikely to draw the four nearly deterministic ultrafast adaptations observed in the experimental data (Fig 1B and Appendix Fig S2A). A mutation rate closer to the upper bound of empirical estimates of the basal mutation rate (3×) (Lynch, 2006; Lang & Murray, 2008; Lynch *et al*, 2008) increased the homogeneity in adaptive speed considerably, while allowing heterogeneity in adaptive solutions. In this case, the founder genotype went extinct in 35 generations in the median population (Fig 3D, upper right panel). Mutation rates (> 5×) above empirical estimates of the basal

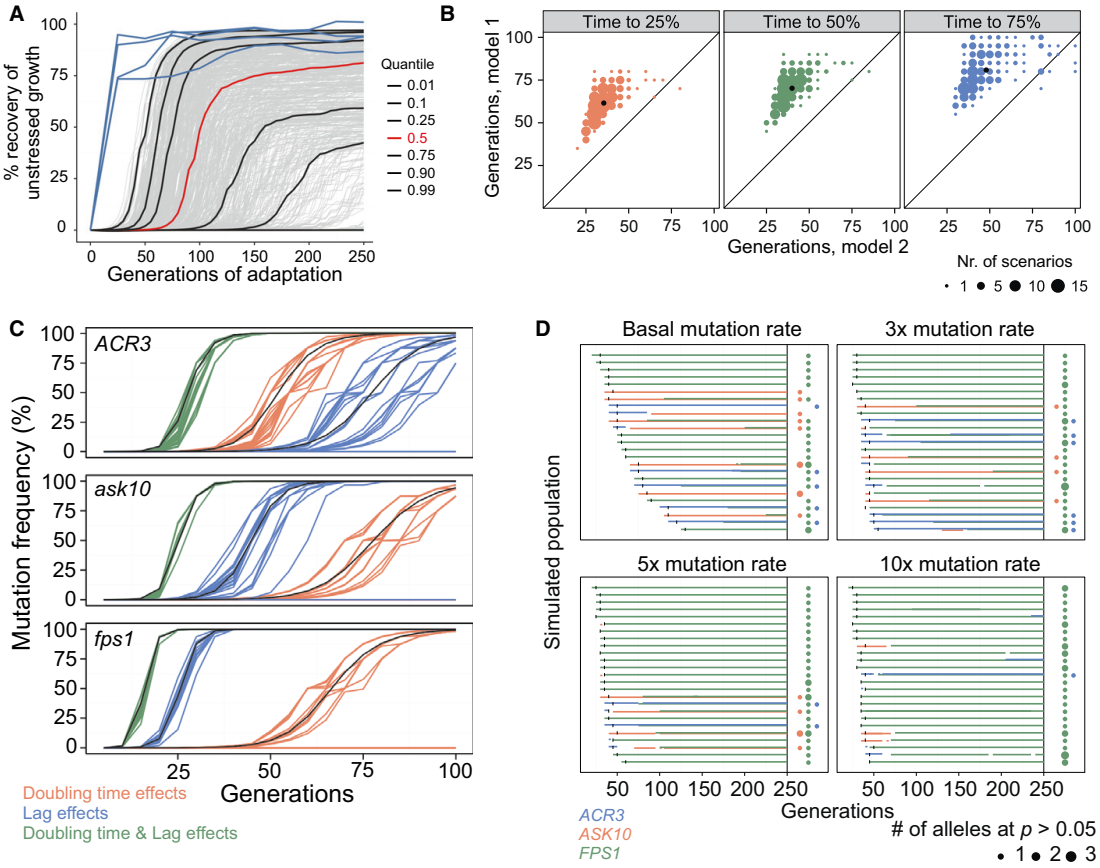


Figure 3. Ultrafast arsenic adaptation is explained by plain neo-Darwinian mechanisms.

A, B Simulating populations (n = 500) with different mutation parameters adapting to arsenic, in an individual cell-based model. Each cell has a cell division time (population doubling time) and a time to the first cell division (lag). Cell division times were recorded every 5th cell division, and population means were expressed as the fraction of founder growth w/o arsenic recovered. (A) Mutations affect only cell division time. Grey lines = 500 adapting populations. Black lines = quantiles corresponding to the fastest 1, 10, 25, 75, 90 and 99% of populations. Red line = median population. Blue line = empirical arsenic populations, P1–P4. (B) Contrasting simulations with mutations only affecting cell division time (model 1, M1, y-axis) or both cell division time and the time to the first cell division with the same effect size and direction (model 2, M2, x-axis). The number of cell divisions required to recover 25% (left panel), 50% (middle panel) and 75% (right panel) of founder growth w/o arsenic is shown. The 227 fastest scenarios, with 75% recovery in ≤ 100 cell divisions, are displayed. Dot size = number of populations. Black dot = median population.

C Simulated competition in arsenic between WT and *ACR3* (top panel), *ASK10* (middle panel), and *FPS1* (lower panel) mutation-carrying cells, respectively. Mutations emerge in a single cell at t = 0 in WT population and have empirical effects on doubling time and/or growth lag. Coloured lines = model with individual cells and genetic drift. Black lines = deterministic model based on subpopulations of cells.

D Arsenic adaptation of 25 simulated populations in an individual cell-based model, given basal, 3×, 5× and 10× mutation rates. Beneficial mutations in *FPS1* (green), *ASK10* (red) and *ACR3* (blue) occur randomly at estimated frequencies and have empirical lag and doubling time effects. Left hand side: Before t = 250 generations. Horizontal bars = mutated alleles exist at (sum) P > 50%. Vertical black lines = WT allele goes extinct. Right hand side: At t = 250 generations. Dot presence = WT allele extinct. Dot colour = gene name. Dot size = number of mutated alleles at P > 5%.

mutation rate gave results that were incompatible with the experimental data, as the superior *FPS1* mutations were then consistently fixed, leaving no room for *ACR3*- and *ASK10*-based solutions (Fig 3D, lower panels). The simulations therefore suggested a mutation rate between 1× and 3× at loci under selection and excluded substantially higher mutation rates.

The confinement of the mutation rate needed to explain the co-appearance of homogeneity in adaptation speed and heterogeneity in adaptive solutions is in line with existing theory of adaptation dynamics in asexual populations (Sniegowski & Gerrish, 2010) (Appendix Fig S7). With a mutation rate of ~1–3×, our dataset falls in the *strong selection, strong mutation* regime, where a handful of

mutations compete for fixation. Higher mutation rate brings us into the *strong selection, weak mutation* regime, where *FPS1* with its high selection coefficient is the preferred solution. At lower mutation rates, we move towards the *weak selection, strong mutation* regime, where rare mutation events lead to large variation in adaptation speed.

Notably, in contrast to our experimental observations, the model predicted that *FPS1* mutations would eventually emerge and fixate in populations where *ACR3* or *ASK10* were already fixated (Fig 3D). The reconstructed *FPS1, ASK10* and *ACR3* mutations explain most of the fitness gains seen in P1–P4, but not all of it. Unidentified causal mutations accounted for ~25% of total fitness gains, and this fraction is larger in *ASK10* and *ACR3* populations than in *FPS1* populations. We may therefore have underestimated the late fitness of *ACR3* and *ASK10* containing clones relative to *FPS1* containing clones. Indeed, when we mimicked this possibility by letting the adaptive gains of *ASK10* and *ACR3* mutations approach that of *FPS1*, *FPS1* was less prone to fixate in *ASK10/ACR3* backgrounds (Appendix Fig S8).

There are no robust experimental means to estimate local mutation rates at loci under selection. However, loss-of-function mutation rates can be estimated for marker genes (Lang & Murray, 2008) and used as somewhat crude proxies. We therefore tested the model prediction that *FPS1*, *ASK10* and *ACR3* mutation rates are at the most moderately elevated in cells adapting to As(III) by measuring the loss-of-function mutation rate for *CAN1* at its native locus (Fig EV4). Both in non-adapted (WT) and adapted (*FPS1* mutated) genotypes, we found a slightly elevated (1.5–1.6×) loss-of-function mutation rate at the *CAN1* locus during As(III) exposure. These estimates do not allow precise conclusions on loss-of-function mutation rates at *FPS1* and *ASK10* loci, and emphatically not for the duplication rate at the *ACR3* locus. However, they exclude a dramatic elevation of the general point mutation rate during arsenic exposure, in agreement with model predictions.

Discussion

As all large effect mutations driving As(III) adaptation enhanced multiple fitness components, positive pleiotropy appeared as the main reason for the observed ultrafast adaptation, a conclusion confirmed by the theoretical modelling. The underlying reason being that the large effect mutations all contributed to fast exclusion of As(III), preventing As(III) from accumulating inside cells and thereby from delaying the time to the first cell division, delaying cell division and increasing the energy costs of homeostasis maintenance.

Mutations inactivated Fps1, the main entrance pathway for As(III) or its critical activator, Ask10, or duplicated the major extrusion system, Acr3 (Fig EV2), in agreement with that Fps1 and Acr3 are the most critical contributors to intracellular As(III) accumulation and toxicity (Wysocki *et al*, 1997, 2001; Talemi *et al*, 2014). The ultrafast adaptation observed for As(III) was not shared by the other selective pressures evaluated. This may be explained by the nature of the elements used. First, several toxic metals share properties and structures with critical metals. For instance, toxic Na^+, Li^+, Rb^+ and Cs^+ are similar to K^+, which plays many important physiological roles and exert their toxicity partially by substituting K^+ in the cell (Cyert & Philpott, 2013). Likewise, the toxicity of Cd^{2+} can in part be attributed to its similarity with essential elements, such as Zn^{2+} and Ca^{2+} (Wysocki & Tamas, 2010). These toxic metals often use the same transport systems to enter or leave the cell as the essential metal and a simple exclusion strategy, as for As(III), will not be possible. Second, some elements are critical at low concentrations but toxic at elevated levels, for example Cu^{2+}. For these elements, solutions in the form of exclusion are likely to be severely constrained by their essentiality (Wysocki & Tamas, 2010; Cyert & Philpott, 2013). In contrast, As(III) is always toxic and has no known function in cells. Third, an element may "hijack" an important cellular process for mediating its toxicity (e.g. the sulphate assimilation pathway is central to Te(IV) toxicity) (Ottosson *et al*, 2010). Such a mode of toxic action is likely to prevent fast adaptation. Fourth, exclusion of toxic elements may be mediated by multiple pathways. In these cases, exclusion may be constrained by redundancy such that no single mutation has a large effect. Finally, even when mutations at a locus have large effects, the locus may be too small for mutations to be encountered frequently. Adaptation to rapamycin is, for example, dominated by large effect mutations in the rapamycin binding protein Fpr1 (Lorenz & Heitman, 1995), but the gene is a tiny 345 bp and the expected waiting time for loss-of-function mutations is therefore long.

There are currently two approaches to measure and model fitness in experimental populations (Barrick & Lenski, 2013). The standard approach measures the fitness of individual genotypes as their frequency change over time in competition assays (Gresham *et al*, 2008; Lang *et al*, 2011). This is simplified if each genome in a population is barcoded before the onset of selection with a unique sequence tag (Levy *et al*, 2015; Venkataram *et al*, 2016), allowing very accurate estimation of the fitness distribution of standing and de novo mutations for use as a model input. Given that change in fitness of the population is also exactly measured and a suitable modelling framework in place, such approaches are certainly useful for understanding the speed of adaptation. So far, however, these approaches have focused on steady-state adaptation where a population has evolved in a constant environment for a long time, with selection acting only on doubling time (Kosheleva & Desai, 2013; Rice *et al*, 2015).

We employed the alternative approach: break fitness in batch-to-batch experiments down into its components, both experimentally and theoretically. This approach certainly comes with caveats attached. It is not always clear that the estimated and modelled fitness components—here cell division time and time to the first cell division—fully capture fitness. In experimental microbial populations, death rates may not be negligible and it is debatable whether efficient use of resources, as reflected in the final growth yield of a population, is a selectable trait or not (MacLean, 2008; Ibstedt *et al*, 2015). Furthermore, to estimate fitness components, mutations must be reconstructed or reversed and the fitness component of individual genotypes must be estimated. This is laborious, in particular if interactions between mutations and between individuals (Moore *et al*, 2013) are to be measured. Here, we considered evolutionary scenarios of very fast adaptation, where single mutations drive adaptation and rapidly rise to fixation, without measurable death occurring. In such scenarios, the caveats above are lesser concerns. Under slow, absent or negative adaptation, clonal interference, positive epistasis, cell–cell interactions and death may all be substantial.

In such evolutionary scenarios, more complex models may be needed.

A marked benefit of breaking fitness down, and connecting it to genotypes via the intervening phenotypic layers, is the possibility to identify the causal factors underlying particular patterns of adaptation. This is illustrated by our discovery that positive pleiotropy between fitness components is the driving force of the observed ultrafast adaptation. To understand adaptation dynamics at an even deeper level, both experimentation and modelling must be extended to molecular phenotypes. For example, by connecting the time to the first cell division and the cell division time to the biochemical and network properties of As(III) metabolism (Talemi *et al*, 2014), a complete and formalized understanding of the causes of ultrafast As (III) adaptation could potentially be obtained.

In conclusion, our results show that even ultrafast adaptation can be achieved based on purely genetic, de novo solutions, without invoking either direct or indirect action of epigenetics (Lenski & Mittler, 1993; Brisson, 2003; Galhardo *et al*, 2007; Ram & Hadany, 2012). Proof by example provides no grounds for rejecting the hypothesis that transgenerational epigenetic mechanisms mediating fast organismal adaptation can be evolutionarily relevant. However, as adaptation speed is a frequent argument for why adaptation by transgenerational epigenetic mechanisms would be favoured by selection and become widespread in nature, our result is a reminder of the forcefulness of plain neo-Darwinian adaptation mechanisms. But the major instrumental value of our study is that it provides a framework of generic worth across a range of experimentally evolvable organisms and environments to systematically assess how important epigenetic mechanisms are for achieving fast adaptation.

Materials and Methods

Strains and medium

Haploid, asexual BY4741 cells (*MATa*; *his3*Δ1; *leu2*Δ0; *met15*Δ0; *ura3*Δ0) (Brachmann *et al*, 1998), stored at −80°C in 20% glycerol, were used as WT and to initiate founder populations. Gene duplication events were mimicked by transforming WT cells with centromeric *URA3* and *KANMX4* plasmids from the MoBY collection (Ho *et al*, 2009). Each plasmid contained a single gene (BY4741 alleles). Point mutations were individually reconstructed in BY4741 backgrounds using *in vivo* site-specific mutagenesis, as described (Stuckey *et al*, 2011). Strains were cultivated in synthetically complete (SC) medium containing: 0.14% yeast nitrogen base (YNB, CYN2210, ForMedium), 0.50% ammonium sulphate, 0.077% complete supplement mixture (CSM, DCS0019, ForMedium), 2.0% (w/v) glucose and pH buffered to 5.8 with 1.0% (w/v) succinic acid and 0.6% (w/v) NaOH. Except for glucose, all required nutrients were present in excess. Where indicated, the medium was supplemented with 3 or 5 mM $NaAsO_2$ (As(III), Sigma-Aldrich) or other environmental challenges as described in Appendix Table S1.

Experimental evolution

Except for the four follow-up arsenic adapting populations, P5–P8, reported in Fig EV1, all experimental evolutions were initiated from a single founder population. The founder population was constructed by clonal colony expansion up to an estimated 1 million cells, from a single cell, on SC agar medium (as above, +2% (w/v) agar) with no added stress. The colony was dissolved in liquid SC medium to create the founder population, the optical density was measured, and an average of 10^5 cells were randomly drawn by pipetting of 5 µl of cell suspension into experimental wells to initiate each adaptation. The follow-up experiment of arsenic adapting populations, P5–P8, was initiated identically, except that each of the four populations was initiated from four different founder populations. These were clonally expanded from four distinct cells, up to a population size of $\sim 3 \times 10^7$. Assuming that adaptive mutations during the clonal expansion from a single cell are Poisson distributed, with normal mutation rates and the mutation target sizes reported in Appendix Fig S6, the probability that a single P5–P8 population housed one or more standing adaptive variants is ~0.08. The probability that all of P5–P8 housed one or more standing adaptive variant at experiment start is $\sim 3.9 \times 10^{-5}$. Experimental evolutions were performed in a batch-to-batch mode in flat-bottom 96-well micro-titre plates containing SC complete medium supplemented with stress factors (Appendix Table S1). To reduce the risk of cross-contamination, every second well was left empty, such that all pairs of populations were separated by empty wells. No indication of cross-contamination between As(III) populations was found in the sequence data. Except for the follow-up As(III) adapting populations P5–P8, populations were propagated over 50 batch-to-batch cycles as 175 µl, non-shaken cultures maintained at 30°C. The follow-up As(III) adapting populations P5–P8 were propagated over 20 cycles. In all cycles, populations were cultivated well into stationary phase. The cultivation length corresponding to ~120-h cultures over the first 10 cycles, ~96 h in cycles 10–30 and ~72 h in cycles 30–50. Stationary phase population sizes corresponded to on average $N = 3.5 \times 10^6$ cells, with the largest deviation corresponding to half that size. To initiate each new cycle, 5 µl of re-suspended and randomly drawn stationary phase cell cultures, corresponding to an average of $N = 10^5$ cells, was multi-pipetted into fresh medium. Each batch cycle corresponded to ~5 population size doublings. The adaptation schema thus progressed over ~250 population doublings (~100 doublings for follow-up As(III) adapting populations P5–P8). Except for the follow-up arsenic adaptations, P5–P8, 50 µl of each population was sampled at the end of every 5^{th} cycle, pipetted into 100 µl of 30% glycerol and stored as a frozen fossil record at −80°C. For populations P5–P8, sampling was instead performed at every batch cycle.

Fitness component extraction

To estimate fitness components, frozen samples were first thawed and re-suspended. 10 µl was pipetted into random wells in 100-well honeycomb plates, each well containing 350 µl of liquid SC medium. Populations were pre-cultivated without shaking at 30°C for 72 h until well into stationary phase. Following re-suspension, 10 µl of each pre-culture was randomly sampled and transferred to 100-well honeycomb plates, containing 350 µl of liquid SC medium supplemented by relevant stress factors. Populations were cultivated in Bioscreen C (Growth Curves Oy, Finland) instruments for 72 h at 30°C and at maximum horizontal shaking for 60 s every other minute (Warringer & Blomberg, 2003; Warringer *et al*, 2003).

Optical density (turbidity) was recorded every 20 min using a wide-band (420–580 nm) filter. Stochastic noise was removed by light smoothing of the raw data, the background light scattering was subtracted, and optical densities were transformed into population size estimates using an empirically based calibration (Fernandez-Ricaud *et al*, 2016). From population size growth curves, population doubling times, length of the lag phase and population growth efficiency (total gain in population size) were extracted (Fernandez-Ricaud *et al*, 2016). Population growth parameters were log(2)-transformed to better adhere to normal distribution assumptions. When comparisons across plates were made, log(2) estimates were first normalized to the corresponding mean of 4–20 WT (founder) controls distributed in fixed but randomized positions. For doubling time and lag, the relative growth measures equalled: mean of log(2) WT estimates – log(2) experimental estimate. For the population growth efficiency, the relative growth measure equalled: log(2) experimental estimate – mean of log(2) WT estimates. Positive values thus always indicate adaptation. The normalization accounts for systematic bias between plates, instruments and batches. Finally, a mean was formed across replicates.

Viable and total cell counts

We cultivated populations of founder (BY4741) and adapted (the reconstructed *FPS1* mutation) genotypes in 100-well honeycomb plates in 350 µl of SC medium (as above) with and without 5 mM As(III) until OD = 1.00 (n = 12). To count viable cells, we plated diluted cells on solid (1.5% agar) medium with no added arsenic and counted the number of colony-forming units (CFU) and multiplied by the dilution factor. No significant difference in CFU was observed between presence and absence of 5 mM As(III) for either founder or *FPS1* cells. We counted total cells in two ways, both by passing sonicated (30 s; to dissolve cell aggregates) and diluted samples through a flow cytometer (BD FACSaria, BD Biosciences, US) at a known flow rate, counting passage events and multiplying by the dilution factor and by direct counting of cells in a hemocytometer (Bürker counting chamber, Knittel Gläser).

As(III) accumulation

Exponentially growing cells (in 150 ml of complete SD medium, with or without uracil) were exposed to 1 mM As(III) and sampled at indicated time points. Cells were washed (2×) in ice-cold water and centrifuged. Cell pellets were re-suspended in water, boiled (10 min) and centrifuged to collect the supernatant. The As(III) content of each sample was measured (n = 2) using a flame atomic absorption spectrometer (3300, Perkin Elmer).

Fluctuation assay mutation rate estimation

The *CAN1* fluctuation assay was performed as described (Lang & Murray, 2008). Single streaked WT and reconstructed *FPS1* colonies were isolated on solid SC medium and inoculated and cultivated overnight in SC medium. Cultures were diluted to a fixed cell density and distributed into the wells of four (WT and *FPS1*, with and without As(III) 96-well plates, each containing 25 µl of SC medium. Wells were sealed to prevent evaporation and cross-contamination and cultivated for 3 days at 30°C, without shaking. We counted cells in three random wells for each plate with a hemocytometer (as above), using the mean as an estimate of cell count per wells in each sample. Discarding wells in the outer frame, we plated the remaining 57 independent cultures on SC agar medium lacking arginine but containing 0.6 g/l of *L*-canavanine sulphate (Sigma-Aldrich). After 3 days at 30°C, we estimated the fraction, P_0, of plated patches without any canavanine-resistant colonies (i.e. without colonies carrying *CAN1* loss-of-function mutations). The *CAN1* loss-of-function mutation rate, μ, (mutations per *CAN1* locus per cell division) was then estimated as: $\mu = -\ln(P_0)/N$.

Sequencing and sequence analysis

To sequence founder and As(III) adapting populations, frozen samples were thawed and re-suspended and 10 µl was pipetted into 100 µl of SC medium with weak (2 or 3 mM) As(III) selection, minimizing allele frequency change. Populations were cultivated into stationary phase. DNA from end point populations was extracted, prepared and sequenced using ABI 5,500 × l SOLiD™ sequencing, completely according to industrial standards. Quality controlled reads were aligned to the yeast reference genome (sacCer3) using ABI's BioScope v1.3. SNPs and indels were called using SAMtools mpileup (Li *et al*, 2009), filtering for regions > 5× mean coverage. CNVs were called using a sliding window approach, with window size: 300 bp and step length: 300 bp. A CNV was conservatively called using CNV seq (Xie & Tammi, 2009) if the log(2) coverage ratio founder/adapted population) exceeded 0.5 or fell short of −0.5. For re-sequencing of earlier time points, DNA was extracted using Epicentre MasterPure Yeast DNA Purification kit, with an added lyticase digestion step. Libraries were prepared using the Illumina Nextera XT enzymatic kit. Paired-end sequencing was performed on a HiSeq 2500, according to industrial standards. Reads were quality-trimmed (Phred score cut-off of 25). Nextera transposase sequences were removed using Trim Galore (v.0.3.8). Reads were mapped to the S288C reference genome (R64-1-1_20110203) using BWA MEM (v.0.7.7-r441). PCR duplicates were removed post-mapping using Picard tools (v.1.109). Samples were sequenced 4–6×, and libraries from the same sample were merged, again using Picard tools. To avoid false variant calls as a result of misalignment around indels, base alignment quality scores were calculated using SAMtools (v.0.1.18 [r982:295]) (Li, 2011). Variants were called using FreeBayes (v0.9.14-8-g1618f7e), treating all sequenced time points from each population as a cohort and filtered in three steps. First, we removed variants with a quality score < 20. Second, we removed variants standing in our founder population (P > 10% in at least two libraries of founder cultivated in absence of As(III)) relative the reference genome. Third, we removed variants standing in the founder after DNA preparative cultivation (3 mM As(III)) (P > 0.2). Variants were annotated using snpEff (v.3.6c). To only retain mutations likely to affect protein function (Fig 2B), we filtered for "Moderate" and "High" snpEff calls and further for non-synonymous SNPs with a SIFT (Sorting Intolerant From Tolerant; Ng & Henikoff, 2003) score of < 0.05. Appendix Fig S3 shows the results without the protein function filtering. Nextera enzyme digestion is biased, making CNV calling inaccurate. We therefore only estimated the copy number of the *ACR3* region, calling the duplicated segment using a sliding window with a size and step

length = 100 bp. Reads with mapping quality < 1 were discarded. Segments were calculated with the circular binary segmentation algorithm using DNAcopy (v.1.40.0) in R.

Deterministic modelling of competition assays

As a basis for the individual-based stochastic model, we developed a deterministic model of competition assays using simple subpopulation growth curves. This deterministic model was used to (i) analytically solve (black lines) fates of novel mutations (Fig 3C) and (ii) to analytically convert mutation effects on doubling time and lag into a selection coefficient (Appendix Fig S5). The model describes a competition assay between two subpopulations, WT (P_1) and mutated (P_2), in a batch cycle set-up mimicking the experimental framework (Appendix Fig S1A). Within each subpopulation (index $i = 1, 2$) individuals share genotype and genetically determined fitness component values. We let $N_i(t)$ be the population size of population P_i at time t. Batch cycles start with $N_i(0)$ individuals in P_i. The total bottleneck population size is $N = N_1(0) + N_2(0)$. No net growth occurs in the time period until time λ_i (lag period). Thereafter, growth is exponential with doubling time τ_i, giving:

$$N_i(t) = N_i(0)2^{\frac{t-\lambda_i}{\tau_i}} \tag{1}$$

A batch cycle ends at time t_{end} after M population doublings such that: $2^M N = N_1(t_{end}) + N_2(t_{end})$.

Simulations underlying black lines in Fig 3C

Parameter settings: $N = 10^5$, $M = 5$ and lag (λ_{WT} = 804.8, λ_{fps1} = 276.6, λ_{ask10} = 503.3 λ_{acr3} = 630.4 min) and doubling times (τ_{WT} = 162.3, τ_{fps1} = 130.2, τ_{ask10} = 134.6 τ_{acr3} = 122.8 min) based on mean empirical values. We assumed *ACR3* duplication effects and the marginal effect of the plasmid to be multiplicative. We simulated competition assays over 20 cycles starting from a single-mutant cell ($N_2(0) = 1$) in the first cycle (Fig 3C black lines). We recorded the frequency ratio, $r(t) = N_2(t)/N_1(t)$ of mutant over WT genotypes. We computed mutation selection coefficients, s, by regressing $\ln(r(t))$ on the number of (P1) generations (Appendix Fig S5A). Finally, we estimated the sensitivity of these selection coefficient estimates to potential measurement error in lag and doubling time (Appendix Fig S5B).

Analytic results on selection coefficients

With a simplifying assumption on t_{end}, we also derived analytical expressions for the joint contribution of lag and doubling time effects on selection coefficients. We assumed $\lambda_2 \leq \lambda_1$ and computed $r(t)$ for a single cycle starting at $t = 0$ and ending at $t_{end} = \lambda_1 + M\tau_1$ where P1 has doubled M times.

$$\begin{aligned} r(t) &= r_0, && 0 \leq t \leq \lambda_2 \\ r(t) &= r_0 2^{\frac{t-\lambda_2}{\tau_2}}, && \lambda_2 < t \leq \lambda_1 \\ r(t) &= r_0 2^{\frac{t-\lambda_2}{\tau_2} \frac{t-\lambda_1}{\tau_1}}, && \lambda_1 < t \end{aligned} \tag{2}$$

The selection coefficient s expressed as the per generation slope of $ln(r(t))$ becomes:

$$\begin{aligned} s &= [ln(r(M\tau_1 - \lambda_1)) - \ln(r(0))]/\mathrm{M} \\ s &= \left[ln\left(r_0 2^{\frac{(M\tau_1+\lambda_1)-\lambda_2}{\tau_2} - \frac{(M\tau_1+\lambda_1)-\lambda_1}{\tau_1}} - \ln(r_0)\right)\right]/M \\ s &= \left[ln\left(2^{\frac{M\tau_1+\lambda_1-\lambda_2}{\tau_2} - M}\right)\right]/M \\ s &= \ln(2)\left[\frac{M\tau_1 + \lambda_1 - \lambda_2}{\tau_2} - M\right]/M \\ s &= \ln(2)\left[\left(\frac{\tau_1}{\tau_2} - 1\right) + \frac{\lambda_1 - \lambda_2}{M\tau_2}\right] \end{aligned} \tag{3}$$

When mutations only affect doubling time ($\lambda_2 = \lambda_1$), equation 3 simplifies to

$$s = ln(2)\left(\frac{\tau_1}{\tau_2} - 1\right) \tag{4}$$

This is equivalent to equation 3.2 in (Chevin, 2011). When mutations only affect lag time ($\tau_1 = \tau_2$), equation 3 simplifies to

$$s = ln(2)\left(\frac{\lambda_1 - \lambda_2}{M\tau_2}\right) \tag{5}$$

Thus, the selection coefficient due to a difference in lag time is reduced when the number of mitotic divisions between bottlenecks or the doubling time increases. Furthermore, for the parameter values for our reconstructed mutation, the lag and doubling time effects on selection coefficients are close to additive.

Individual-based model of batch experimental evolution

To account for the combined effects of random mutation events, random subsampling of cells, clonal interference and epistasis, we mimicked the experimental framework (Appendix Fig S1A) in an individual-based model. Each cell has its individual genotype that determines the time to the first cell division and its cell division time. Cells with identical genotypes divide at the same time. Parameters, similar to those for the deterministic model above, describe population size at the start of each batch cycle (N), average number of mitotic divisions before serial transfer (M) and total number of batch cycles. When the total population size reaches $2^M N$ cells, N cells are subsampled randomly to found the next cycle. The model assumed no cell death, that all effects on cell division rate and lag are genetic, no meiosis or ploidy change and no interactions between cells. We found no evidence of cell death, no evidence of ploidy change in sequence data and meiosis is inactivated by deletion of the mating type-switching gene. The model was used for three sets of simulations of increasing complexity:

Individual-based competition assays

We simulated the competition assays studied with subpopulation growth curves ($N = 10^5$, $M = 5$, 20 batch cycles, genotypic lag and doubling time parameters as above) in Fig 3C (coloured lines). Competition assays were initiated with a single-mutant (*fps1*, *ask10* or *ACR3*) cell in a founder population and no other mutations emerging. Frequency trajectories are stochastic due to the random sampling of individuals. In the extreme case, random sampling leads to loss-of-mutant lines.

Simulating experimental evolution with mutations in FPS1, ASK10, ACR3

We simulated experimental evolution starting from a founder population accumulating mutations in *FPS1* (loss-of-function), *ASK10* (loss-of-function) and *ACR3* (duplication). The basal duplication rate of *ACR3* was set to 3×10^{-7} duplications per division. Beneficial mutational target sizes for *FPS1* and *ASK10* were computed by downloading the SIFT yeast database (http://sift-db.bii.a-star.edu.sg/public/Saccharomyces_cerevisiae/EF4.74/) and extracting all possible stop gain base changes and non-synonymous mutations with attached SIFT scores, for *FPS1* and *ASK10* (Appendix Fig S6). In Fig 3D, beneficial mutation was considered as all mutations with a SIFT score < 0.06, corresponding to the observed driver mutation with the highest SIFT score. Mutational target sizes were multiplied with a global mutation rate estimate (Lynch *et al*, 2008) of 0.33×10^{-9} mutations/bp/division. Empirical lag and doubling time values were used (above), assuming complete negative epistasis. We ran four sets of simulations ($n = 25$) over 50 cycles with $N = 10^5$ and $M = 5$, varying the mutation rate from basal to 3×, 5× and 10× the basal rate.

Simulating adaptation across the realistic range of mutation parameter values

We simulated the experimental framework ($N = 10^5$, $M = 5$, 50 batch cycles) starting from a clonal wild-type (WT) population with empirical arsenic WT lag and doubling time values ($L_{wt} = 805$ min, $T_{wt} = 162$ min) adapting towards wild-type performance in absence of arsenic ($L_{wt,N} = 271$ min, $T_{wt,N} = 126$ min). Mutation events were sampled after each cell division with parameters being the overall mutation rate, the proportion of mutations affecting fitness, the proportion of fitness-affecting mutations that are beneficial, and the distribution of selection coefficients for non-neutral mutations. Following (Joseph & Hall, 2004), the selection coefficient s_r for a given mutation m was sampled from a gamma distribution with shape α and scale β. The selection coefficients in Joseph and Hall (2004) are given by the ratio of exponential growth rates of the mutant and wild-type strain, respectively, and following (Chevin, 2011) we compensated for the overestimation (factor $ln(2)$) of the per generation selection coefficients in equations 3–5. Sampled selection coefficients, $s = s_r\, ln(2)$, were inserted into equations 3–5 and rearranged to provide the mutation induced change in doubling or lag time from the WT. For doubling time, the equation becomes

$$\Delta T_{wt}^{m} = \frac{T_{wt} s_r}{s_r + 1} \tag{6}$$

ΔT_{wt}^{m} is the change in doubling time when mutation m emerges as the first mutation in a WT cell with doubling time T_{wt}. In the very rare cases where ΔT_{wt}^{m} exceeded $T_{wt} - T_{wt,N}$, the value was truncated. Negative epistasis was implemented in the form of diminishing return of positive mutations. If the mutation m emerges in a cell with genotype G and doubling time T_G, the resulting change in doubling time was modelled as

$$\Delta T_{G}^{m} = \frac{T_G - T_{wt,N}}{T_{wt} - T_{wt,N}} \Delta T_{wt}^{m} \tag{7}$$

The capping of extreme effects and diminishing return of consecutive positive mutations means that adapting cells asymptotically approach WT growth in absence of arsenic. All empirical populations followed this behaviour (Figs 1A and EV1), and there is strong experimental support for a diminishing return of positive mutations (Chou *et al*, 2011; Khan *et al*, 2011). We simulated models where mutations affected only cell division time (population doubling time, M1), cell division time (population doubling time), and time to the first cell division (population lag time) with the same effect sign and size (M2) and cell division time (population doubling time), and time to the first cell division (population lag time) with random effect sign and magnitudes (M3) using 500 mutation parameter sets based on empirical values. These corresponded to full-factorial combinations of the mutation rate ($\frac{\mu}{5}, \frac{\mu}{3}, \mu, 3\mu, 5\mu$, where the base mutation rate $\mu = 0.33 \times 10^{-9}$ mutations/bp/division; Lynch *et al*, 2008), the fraction of fitness-affecting mutations ($\frac{y}{5}, \frac{y}{3}, y, 3y, 5y$, where $y = 0.034$ (Hall *et al*, 2008), the fraction of fitness-affecting mutations that are beneficial ($\frac{z}{5}, \frac{z}{3}, z, 3z, 5z$, where $z = 0.13$; Hall *et al*, 2008) and the scale ($\beta = 13.3$, 20, 27.35, 33, 40, 47 min) of the effect size distribution. Thirty-three minutes corresponded to the reported empirical value (Joseph & Hall, 2004). The shape of the effect size distribution was kept constant at $\alpha = 2$. We recorded population averages for doubling times at the end of each batch cycle.

Data availability

Models are available as Code EV1 and can also be downloaded from https://bitbucket.org/ajkarloss/yeast_sim. The SOLiD sequencing data are accessible at EBI (http://www.ebi.ac.uk/ena/data/view/PRJEB17740) with accession number PRJEB17740. The Illumina sequencing data are accessible at NCBI (https://www.ncbi.nlm.nih.gov/sra?term=SRP092403) with accession number SRP092403.

Expanded View for this article is available online.

Acknowledgements
We thank Payam Ghiaci for help with DNA extraction. Financial support by the Polish National Science Centre (grant number 2012/07/B/NZ1/02804) to RW, from the Swedish Research Council (grant numbers 325-2014-6547 and 621-2014-4605) and from the Carl Tryggers Foundation (grant number CTS 12:521) to JW, and from the Research Council of Norway (grant numbers 178901/V30 and 222364/F20) is acknowledged.

Author contributions
ABG, SWO and JW conceived, designed and coordinated the study. EZ, IHD, SS and MZ performed and analysed evolution experiments. EA-P performed mutation rate experiments. EM-D, MM, RW and MJT designed, performed and analysed arsenic accumulation. SS, FR and IJ analysed sequence data. ABG and JKAS designed, performed and analysed simulations. ABG, SWO and JW wrote the paper, with input from all other authors.

Conflict of interest
The authors declare that they have no conflict of interest.

References

Barrick JE, Lenski RE (2013) Genome dynamics during experimental evolution. *Nat Rev Genet* 14: 827 – 839

Beese SE, Negishi T, Levin DE (2009) Identification of positive regulators of the yeast Fps1 glycerol channel. *PLoS Genet* 5: e1000738

Bienert GP, Thorsen M, Schussler MD, Nilsson HR, Wagner A, Tamas MJ, Jahn TP (2008) A subgroup of plant aquaporins facilitate the bi-directional diffusion of As(OH)3 and Sb(OH)3 across membranes. *BMC Biol* 6: 26

Brachmann CB, Davies A, Cost GJ, Caputo E, Li J, Hieter P, Boeke JD (1998) Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* 14: 115 – 132

Brisson D (2003) The directed mutation controversy in an evolutionary context. *Crit Rev Microbiol* 29: 25 – 35

Bun-ya M, Shikata K, Nakade S, Yompakdee C, Harashima S, Oshima Y (1996) Two new genes, PHO86 and PHO87, involved in inorganic phosphate uptake in *Saccharomyces cerevisiae*. *Curr Genet* 29: 344 – 351

Carone BR, Fauquier L, Habib N, Shea JM, Hart CE, Li R, Bock C, Li C, Gu H, Zamore PD, Meissner A, Weng Z, Hofmann HA, Friedman N, Rando OJ (2010) Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals. *Cell* 143: 1084 – 1096

Chevin LM (2011) On measuring selection in experimental evolution. *Biol Lett* 7: 210 – 213

Chou HH, Chiu HC, Delaney NF, Segre D, Marx CJ (2011) Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science* 332: 1190 – 1192

Conrad TM, Lewis NE, Palsson BO (2011) Microbial laboratory evolution in the era of genome-scale science. *Mol Syst Biol* 7: 509

Cyert MS, Philpott CC (2013) Regulation of cation balance in *Saccharomyces cerevisiae*. *Genetics* 193: 677 – 713

Daxinger L, Whitelaw E (2012) Understanding transgenerational epigenetic inheritance via the gametes in mammals. *Nat Rev Genet* 13: 153 – 162

Dettman JR, Rodrigue N, Melnyk AH, Wong A, Bailey SF, Kassen R (2012) Evolutionary insight from whole-genome sequencing of experimentally evolved microbes. *Mol Ecol* 21: 2058 – 2077

Fernandez-Ricaud L, Kourtchenko O, Zackrisson M, Warringer J, Blomberg A (2016) PRECOG: a tool for automated extraction and visualization of fitness components in microbial growth phenomics. *BMC Bioinformatics* 17: 249

Galhardo RS, Hastings PJ, Rosenberg SM (2007) Mutation as a stress response and the regulation of evolvability. *Crit Rev Biochem Mol Biol* 42: 399 – 435

Ghosh M, Shen J, Rosen BP (1999) Pathways of As(III) detoxification in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 96: 5001 – 5006

Gresham D, Desai MM, Tucker CM, Jenq HT, Pai DA, Ward A, DeSevo CG, Botstein D, Dunham MJ (2008) The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. *PLoS Genet* 4: e1000303

Halfmann R, Alberti S, Lindquist S (2010) Prions, protein homeostasis, and phenotypic diversity. *Trends Cell Biol* 20: 125 – 133

Halfmann R, Lindquist S (2010) Epigenetics in the extreme: prions and the inheritance of environmentally acquired traits. *Science* 330: 629 – 632

Hall DW, Mahmoudizad R, Hurd AW, Joseph SB (2008) Spontaneous mutations in diploid *Saccharomyces cerevisiae*: another thousand cell generations. *Genet Res* 90: 229 – 241

Ho CH, Magtanong L, Barker SL, Gresham D, Nishimura S, Natarajan P, Koh JL, Porter J, Gray CA, Andersen RJ, Giaever G, Nislow C, Andrews B, Botstein D, Graham TR, Yoshida M, Boone C (2009) A molecular barcoded yeast ORF library enables mode-of-action analysis of bioactive compounds. *Nat Biotechnol* 27: 369 – 377

Hoege C, Pfander B, Moldovan GL, Pyrowolakis G, Jentsch S (2002) RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature* 419: 135 – 141

Ibstedt S, Stenberg S, Bages S, Gjuvsland AB, Salinas F, Kourtchenko O, Samy JK, Blomberg A, Omholt SW, Liti G, Beltran G, Warringer J (2015) Concerted evolution of life stage performances signals recent selection on yeast nitrogen use. *Mol Biol Evol* 32: 153 – 161

Joseph SB, Hall DW (2004) Spontaneous mutations in diploid *Saccharomyces cerevisiae*: more beneficial than expected. *Genetics* 168: 1817 – 1825

Khan AI, Dinh DM, Schneider D, Lenski RE, Cooper TF (2011) Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* 332: 1193 – 1196

Kosheleva K, Desai MM (2013) The dynamics of genetic draft in rapidly adapting populations. *Genetics* 195: 1007 – 1025

Laland K, Uller T, Feldman M, Sterelny K, Muller GB, Moczek A, Jablonka E, Odling-Smee J, Wray GA, Hoekstra HE, Futuyma DJ, Lenski RE, Mackay TF, Schluter D, Strassmann JE (2014) Does evolutionary theory need a rethink? *Nature* 514: 161 – 164

Laland KN, Odling-Smee FJ, Feldman MW (1999) Evolutionary consequences of niche construction and their implications for ecology. *Proc Natl Acad Sci USA* 96: 10242 – 10247

Lang GI, Botstein D, Desai MM (2011) Genetic variation and the fate of beneficial mutations in asexual populations. *Genetics* 188: 647 – 661

Lang GI, Murray AW (2008) Estimating the per-base-pair mutation rate in the yeast *Saccharomyces cerevisiae*. *Genetics* 178: 67 – 82

Lee J, Reiter W, Dohnal I, Gregori C, Beese-Sims S, Kuchler K, Ammerer G, Levin DE (2013) MAPK Hog1 closes the *S. cerevisiae* glycerol channel Fps1 by phosphorylating and displacing its positive regulators. *Genes Dev* 27: 2590 – 2601

Lenski RE, Mittler JE (1993) The directed mutation controversy and neo-Darwinism. *Science* 259: 188 – 194

Levy SF, Blundell JR, Venkataram S, Petrov DA, Fisher DS, Sherlock G (2015) Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature* 519: 181 – 186

Li H (2011) Improving SNP discovery by base alignment quality. *Bioinformatics* 27: 1157 – 1158

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25: 2078 – 2079

Liu Z, Boles E, Rosen BP (2004) Arsenic trioxide uptake by hexose permeases in *Saccharomyces cerevisiae*. *J Biol Chem* 279: 17312 – 17318

Long A, Liti G, Luptak A, Tenaillon O (2015) Elucidating the molecular architecture of adaptation via evolve and resequence experiments. *Nat Rev Genet* 16: 567 – 582

Lorenz MC, Heitman J (1995) TOR mutations confer rapamycin resistance by preventing interaction with FKBP12-rapamycin. *J Biol Chem* 270: 27531 – 27537

Lynch M (2006) The origins of eukaryotic gene structure. *Mol Biol Evol* 23: 450 – 468

Lynch M, Sung W, Morris K, Coffey N, Landry CR, Dopman EB, Dickinson WJ, Okamoto K, Kulkarni S, Hartl DL, Thomas WK (2008) A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc Natl Acad Sci USA* 105: 9272 – 9277

Maciaszczyk-Dziubinska E, Migdal I, Migocka M, Bocer T, Wysocki R (2010) The yeast aquaglyceroporin Fps1p is a bidirectional arsenite channel. *FEBS Lett* 584: 726 – 732

MacLean RC (2008) The tragedy of the commons in microbial populations: insights from theoretical, comparative and experimental studies. *Heredity* 100: 471 – 477

MacLean RC, Torres-Barcelo C, Moxon R (2013) Evaluating evolutionary models of stress-induced mutagenesis in bacteria. *Nat Rev Genet* 14: 221 – 227

Martincorena I, Luscombe NM (2013) Non-random mutation: the evolution of targeted hypermutation and hypomutation. *BioEssays* 35: 123 – 130

Masel J, Siegal ML (2009) Robustness: mechanisms and consequences. *Trends Genet* 25: 395 – 403

Molinier J, Ries G, Zipfel C, Hohn B (2006) Transgeneration memory of stress in plants. *Nature* 442: 1046 – 1049

Moore LS, Stolovicki E, Braun E (2013) Population dynamics of metastable growth-rate phenotypes. *PloS One* 8: e81671

Moore LS, Wei W, Stolovicki E, Benbenishty T, Wilkening S, Steinmetz LM, Braun E, David L (2014) Induced mutations in yeast cell populations adapting to an unforeseen challenge. *PloS One* 9: e111133

Mukhopadhyay R, Rosen BP (1998) *Saccharomyces cerevisiae* ACR2 gene encodes an arsenate reductase. *FEMS Microbiol Lett* 168: 127 – 136

Mukhopadhyay R, Shi J, Rosen BP (2000) Purification and characterization of ACR2p, the *Saccharomyces cerevisiae* arsenate reductase. *J Biol Chem* 275: 21149 – 21157

Ng PC, Henikoff S (2003) SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res* 31: 3812 – 3814

Ottosson LG, Logg K, Ibstedt S, Sunnerhagen P, Kall M, Blomberg A, Warringer J (2010) Sulfate assimilation mediates tellurite reduction and toxicity in *Saccharomyces cerevisiae*. *Eukaryotic Cell* 9: 1635 – 1647

Pigliucci M, Murren CJ, Schlichting CD (2006) Phenotypic plasticity and evolution by genetic assimilation. *J Exp Biol* 209: 2362 – 2367

Plucain J, Hindre T, Le Gac M, Tenaillon O, Cruveiller S, Medigue C, Leiby N, Harcombe WR, Marx CJ, Lenski RE, Schneider D (2014) Epistasis and allele specificity in the emergence of a stable polymorphism in *Escherichia coli*. *Science* 343: 1366 – 1369

Ram Y, Hadany L (2012) The evolution of stress-induced hypermutation in asexual populations. *Evolution* 66: 2315 – 2328

Rando OJ, Verstrepen KJ (2007) Timescales of genetic and epigenetic inheritance. *Cell* 128: 655 – 668

Rice DP, Good BH, Desai MM (2015) The evolutionarily stable distribution of fitness effects. *Genetics* 200: 321 – 329

Richards EJ (2006) Inherited epigenetic variation–revisiting soft inheritance. *Nat Rev Genet* 7: 395 – 401

Roth JR, Kugelberg E, Reams AB, Kofoid E, Andersson DI (2006) Origin of mutations under selection: the adaptive mutation controversy. *Annu Rev Microbiol* 60: 477 – 501

Ruden DM, Jamison DC, Zeeberg BR, Garfinkel MD, Weinstein JN, Rasouli P, Lu X (2008) The EDGE hypothesis: epigenetically directed genetic errors in repeat-containing proteins (RCPs) involved in evolution, neuroendocrine signaling, and cancer. *Front Neuroendocrinol* 29: 428 – 444

Sniegowski PD, Gerrish PJ (2010) Beneficial mutations and the dynamics of adaptation in asexual populations. *Philos Trans R Soc Lond B Biol Sci* 365: 1255 – 1263

Stuckey S, Mukherjee K, Storici F (2011) In vivo site-specific mutagenesis and gene collage using the delitto perfetto system in yeast *Saccharomyces cerevisiae*. *Methods Mol Biol* 745: 173 – 191

Supek F, Lehner B (2015) Differential DNA mismatch repair underlies mutation rate variation across the human genome. *Nature* 521: 81 – 84

Talemi SR, Jacobson T, Garla V, Navarrete C, Wagner A, Tamas MJ, Schaber J (2014) Mathematical modelling of arsenic transport, distribution and detoxification processes in yeast. *Mol Microbiol* 92: 1343 – 1356

Thorsen M, Jacobson T, Vooijs R, Navarrete C, Bliek T, Schat H, Tamas MJ (2012) Glutathione serves an extracellular defence function to decrease arsenite accumulation and toxicity in yeast. *Mol Microbiol* 84: 1177 – 1188

Thorsen M, Lagniel G, Kristiansson E, Junot C, Nerman O, Labarre J, Tamas MJ (2007) Quantitative transcriptome, proteome, and sulfur metabolite profiling of the *Saccharomyces cerevisiae* response to arsenite. *Physiol Genomics* 30: 35 – 43

Tu Y, Tornaletti S, Pfeifer GP (1996) DNA repair domains within a human gene: selective repair of sequences near the transcription initiation site. *EMBO J* 15: 675 – 683

Venkataram S, Dunn B, Li Y, Agarwala A, Chang J, Ebel ER, Geiler-Samerotte K, Herissant L, Blundell JR, Levy SF, Fisher DS, Sherlock G, Petrov DA (2016) Development of a comprehensive genotype-to-fitness map of adaptation-driving mutations in yeast. *Cell* 166: 1585 – 1596 e1522

Warringer J, Blomberg A (2003) Automated screening in environmental arrays allows analysis of quantitative phenotypic profiles in *Saccharomyces cerevisiae*. *Yeast* 20: 53 – 67

Warringer J, Ericson E, Fernandez L, Nerman O, Blomberg A (2003) High-resolution yeast phenomics resolves different physiological features in the saline response. *Proc Natl Acad Sci USA* 100: 15724 – 15729

Wysocki R, Bobrowicz P, Ulaszewski S (1997) The *Saccharomyces cerevisiae* ACR3 gene encodes a putative membrane protein involved in arsenite transport. *J Biol Chem* 272: 30061 – 30066

Wysocki R, Chery CC, Wawrzycka D, Van Hulle M, Cornelis R, Thevelein JM, Tamas MJ (2001) The glycerol channel Fps1p mediates the uptake of arsenite and antimonite in *Saccharomyces cerevisiae*. *Mol Microbiol* 40: 1391 – 1401

Wysocki R, Fortier PK, Maciaszczyk E, Thorsen M, Leduc A, Odhagen A, Owsianik G, Ulaszewski S, Ramotar D, Tamas MJ (2004) Transcriptional activation of metalloid tolerance genes in *Saccharomyces cerevisiae* requires the AP-1-like proteins Yap1p and Yap8p. *Mol Biol Cell* 15: 2049 – 2060

Wysocki R, Tamas MJ (2010) How *Saccharomyces cerevisiae* copes with toxic metals and metalloids. *FEMS Microbiol Rev* 34: 925 – 951

Wysocki R, Tamás MJ (2011) *Saccharomyces cerevisiae* as a model organism for elucidating arsenic tolerance mechanisms. In *Cellular effects of heavy metal*, Bánfalvi G (ed.), pp 87 – 112. Heidelberg: Springer Verlag

Xie C, Tammi MT (2009) CNV-seq, a new method to detect copy number variation using high-throughput sequencing. *BMC Bioinformatics* 10: 80

Yompakdee C, Ogawa N, Harashima S, Oshima Y (1996) A putative membrane protein, Pho88p, involved in inorganic phosphate transport in *Saccharomyces cerevisiae*. *Mol Gen Genet* 251: 580 – 590

Zhang Z, Saier MH Jr (2009) A mechanism of transposon-mediated directed mutation. *Mol Microbiol* 74: 29 – 43

Zhu YO, Siegal ML, Hall DW, Petrov DA (2014) Precise estimates of mutation rate and spectrum in yeast. *Proc Natl Acad Sci USA* 111: E2310 – E2318

Superoxide induces adaptive editing of mitochondrial DNA

Simon Stenberg[1], Jing Li[2], Arne B. Gjuvsland[1], Karl Persson[2], Timmy Forsberg[2], Payam Ghiaci[2], Ciaran Gilchrist[2], Martin Zackrisson[2], Mikael Molin[2], Gianni Liti[3], Stig W. Omholt[4], Jonas Warringer[1,2]*

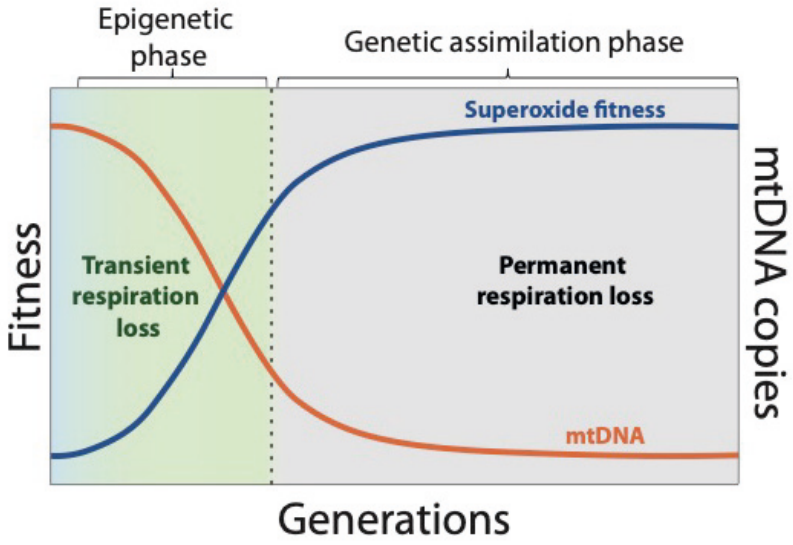
[1] Centre for Integrative Genetics (CIGENE), Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, PO Box 5003, 1432 Ås, Norway

[2] Department of Chemistry and Molecular Biology, University of Gothenburg, PO Box 462, 40530 Gothenburg, Sweden

[3] Université Côte d'Azur, CNRS, INSERM, IRCAN, Nice, France;

[4] Department of Circulation and Medical Imaging, Cardiac Exercise Research Group, Norwegian University of Science and Technology (NTNU), 7491, Trondheim, Norway

* Corresponding author: jonas.warringer@cmb.gu.se



Abstract

The importance of oxidative stress as a driver of mitochondrial impairment has become a contentious issue. Because age-related disease is intimately linked to mitochondrial function, and oxidative stress increases with age, this issue sorely needs to be resolved. Here we show that budding yeast adapts to superoxide (O_2^-) distress by almost immediately reducing the copy number of key mitochondrial respiratory genes, and completely restores mitochondrial function by replenishing the lost genes very abruptly eight to sixteen cell divisions after stress removal. Chronic exposure to O_2^- distress after the first ultrafast adaptation prevents the mitochondria from being functionally restored after stress removal and the adaptation becomes genetically fixed. The results bring chronic oxidative stress back to the fore as a key inducer of mitochondrial dysfunction without invoking random mitochondrial DNA mutations, and offer intriguing possibilities for better understanding as well as treatment of a range of age-related diseases.

Mitochondria are intracellular organelles possessed by all eukaryotes. Although involved in many aspects of anabolism and cellular homeostatic regulation (Palikaras & Tavernarakis, 2014; Westermann, 2010), their primary role is to produce adenosine triphosphate (ATP) by the mitochondrial respiratory chain. The vast majority of mitochondrial proteins are encoded by the nuclear genome in all eukaryotes (Lionaki, Gkikas, & Tavernarakis, 2016; Palikaras & Tavernarakis, 2014) and the mitochondrial genome contains only a small but well conserved set of genes encoding mostly key respiratory chain proteins (Wallace, 2007).

Mitochondrial dysfunction associates with the pathogenesis of a wide range of age-related human diseases, including Alzheimer's disease, Parkinson's disease (Ammal Kaidery & Thomas, 2018), the deterioration of skeletal and cardiac muscle (Hepple, 2016), and macular degeneration (Hyttinen, Viiri, Kaarniranta, & Błasiak, 2018). A prominent explanation for this association is the mitochondrial free radical theory of aging (MFRTA) proposed by Harman almost 50 years ago (Harman, 1972). As mitochondria are the principle source of intracellular reactive oxygen species (ROS), the hypothesis postulated that the accumulation of mitochondrial DNA (mtDNA) mutations due to ROS exposure lead to abnormal mitochondrial respiratory chain proteins. This leads to further increase of ROS production and more mtDNA mutations through positive feedback. However, the lack of convincing experimental support for a predicted positive feedback loop has led to austere criticism of the MFRTA (Kujoth et al., 2005; Payne & Chinnery, 2015; Zsurka, Peeva, Kotlyar, & Kunz, 2018). Moreover, the etiological importance of mtDNA mutations has also become a contentious issue (Khrapko & Vijg, 2009). As mitochondria retain a key role in the complex web of processes leading to cellular and organismal ageing (Payne & Chinnery, 2015), a pressing question is therefore whether there is a causative link between ROS generation, mitochondrial impairment and age-related pathogenesis.

Here we report that the primary adaptive response of yeast cell populations to a moderate superoxide (O_2^-) distress, generated by paraquat (N,N-dimethyl-4-4′-bipiridinium dichloride),

depends directly on a regulatory system reducing the copy number of key mtDNA genes underlying the electron transport chain (ETC). Intriguingly, the regulatory system possesses a cellular memory of the oxidative distress for 8-16 cell divisions before it abruptly completely restores mitochondrial mtDNA copy number and respiratory chain function. And most importantly, under chronic O_2^- distress, the capacity for restoring mitochondrial function breaks down due to erosion of critical mtDNA segments, leading to permanent mtDNA impairment and irrevocable loss of respiratory capacity. This novel link between chronic oxidative distress and mitochondrial dysfunction, caused by the breakdown of a regulatory system apparently designed for handling temporary 0_2^- distress, suggests a new avenue for research concerning the connection between mitochondrial impairment and age-related pathogenesis.

Ultrafast adaptation to O_2^- distress

Even though very fast adaptation in haploid budding yeast may be achieved by normal Darwinian random mutation and selection (Gjuvsland et al., 2016), consistently reproducible ultrafast adaptation is a putative signature of a regulatory response not involving novel genetic variation. We therefore first compared the temporal adaptation profiles to paraquat-mediated oxidative distress with those of seven other types of stress not directly challenging mitochondrial function (Table S1). Paraquat induces predominantly O_2^--production in mammalian and yeast mitochondria (Cochemé & Murphy, 2008a) (Fig 1A). In yeast, ETC complex III and mitochondrial NADPH dehydrogenases donate electrons to paraquat, which passes these on to O_2 (Castello, Drechsel, & Patel, 2007; Cochemé & Murphy, 2008b). This mode of O_2^- generation closely resembles the main natural process, where electrons leak from the mitochondrial ETC directly to O_2, primarily from ETC complex I and III (Fang & Beattie, 2003; Turrens, 1997).

We adapted 1152 isogenic yeast populations to mild doses of paraquat (Fig 1B, S1), and >8000 populations were in total exposed to seven non-mitochondrial challenges (Table S1) to distinguish between generic and specific responses. We continuously counted cells (Zackrisson et al., 2016), and extracted cell division times (doubling time, *D*) and number of cell divisions (population doublings, *G*) from each batch cycle. We preserved a frozen fossil record of 96 random populations per challenge that later were revived and re-analysed in a well-replicated (n=6), randomized design to provide near error-free adaptation profiles for these 768 populations.

All populations exposed to excess O_2^- adapted faster than all other populations (Fig 1C, S2A). The early adaptation was astoundingly fast with a 106 min (68% of the mean adaptation potential) reduction in mean cell doubling time the first 10 generations, during which there was a shift from multi-phase to single-phase growth (Fig S2B). Adaptation then slowed down and later levelled out at 75 generations (mean), at which time the populations had realised 72.9% (mean) of their adaptation potential.

O_2^- adaptation is strictly regulated

To resolve whether the ultrafast paraquat adaptation was due to Darwinian or regulatory processes (Fig 2A) we released 96 well-adapted (70-90% of endpoint adaptation) frozen fossil populations from paraquat exposure for 10 batch cycles, corresponding to a mean of 84 generations (Fig 2B). After each relaxation cycle, we re-exposed populations to O_2^- distress, tracking adaptation loss as a function of generations of relaxed selection. We found that 86 of the 96 populations (90%) at this stage of adaptation retained their adaptive gains over 1-2 batch cycles of relaxed selection (8-16 generations), before abruptly losing them over a single batch cycle (Fig 2C). The growth of the adapted populations was on par with the founder population during the relaxation, showing that their abrupt loss of adaptive gain occurred despite no or marginal Darwinian counterselection against the acquired adaptations (Fig S3A). In contrast, all populations from the two other environments showing fast adaptation (arsenic and glycine) only slowly and progressively lost their adaptive gains (Fig S3B), despite stronger (s=0.08, 0.06) Darwinian counterselection (Fig S3A).

If a Darwinian process were underlying the early paraquat adaptation, the above results would require the highly unlikely fast and synchronous loss of all novel adaptive genetic variation within the 86 populations. To quantify this unlikeliness more rigorously, we used data-driven individual-based evolutionary population modelling (Gjuvsland et al., 2016) to assess whether purely genetic processes could explain the early ultrafast O_2^- adaptation. We found that a Darwinian mutation-selection was incapable of explaining our results based on current knowledge of yeast biology (Fig 2D). In contrast, the maintenance of a constant adaptive gain under relaxed selection for a moderate length of time followed by an abrupt restoration of the founder population phenotype is expected if a transient cellular memory mechanism under regulatory control was involved.

Chronic O_2^- distress causes fixation

The observation that the remaining 10% of the paraquat-adapting populations retained their adapted state after 10 batch cycles, suggested that for these the adaptation was either completely Darwinian or had become genetically assimilated, i.e. the induced phenotypic variation had become constitutive by genetic fixation (Pigliucci, 2006). To resolve this issue we repeated the selection relaxation with frozen fossil populations from across the whole adaptation window (Fig 2B). This time we found early O_2^- adaptation to be inherited over 1-2 batch cycles for all 96 populations, before being abruptly lost (Fig 2E). Again, the abrupt restoration of the founder state occurred despite no or marginal counter-selection against adaptations after relaxation (Fig S3C, D). 46% of the populations had assimilated their adapted state (i.e. retaining it over ~84 generations of relaxed selection) after 24 generations, and all had done so after 242 generations (Fig 2E). The results imply that a deterministic regulatory process involving cellular memory caused the ultrafast adaptation to excess O_2^- in all 96 populations, and that chronic exposure ultimately leads to genetic assimilation (Fig 2E) and loss of the dynamic regulatory adaptive response.

Early O_2^- adaptation targets the ETC

To probe the link between mitochondrial function and O_2^- adaptation, we tracked mtDNA content through the early regulatory controlled adaptation phase in five random paraquat exposed populations. We found a clear inverse temporal association between mean coverage of mtDNA and adaptive gain (R^2 = 0.49-0.96) (Fig 3A). *COB*, coding for cytochrome b, the main source of ETC generated O_2^-, was lost early from all five populations, together with its gene neighbours *ATP6*, *ATP8*, *VAR1* and *COX1* (Fig S4A). Loss of respiratory capacity matched mtDNA loss in these 5 populations very well (R^2 = 0.49-0.80 at G_6 to G_{15}; Fig S4B) and all 96 fossil populations lost respiratory capacity in synchrony with early adaptive gains (R^2 = 0.59) (Fig 3B). This strongly suggests that prevention of O_2^- generation through loss of specific ETC components was the proximal reason for the observed early adaptation. To further corroborate this explanation, we repeated the selection relaxation experiment (Fig 2A) for the 5 focal populations. We found the respiration loss, assayed by growth on glycerol, to be consistently maintained for 8 to 16 generations after release from O_2^- distress, before being restored in synchrony with erasure of the cellular memory (Fig 3C). This demonstrates that the regulatory system causing the early adaptation response is capable of both degrading and restoring the mtDNA encoding the ETC in response to O_2^- levels. In contrast to growth on glycerol, the observed marginal counterselection on glucose during the relaxation phase shows that growth on glucose is almost independent of ETC function, while still being highly dependent on other mitochondrial functions (Kispal et al., 2005).

Genetic effects of chronic O_2^- distress

Our results show that chronic exposure to excess O_2^- causes the mtDNA regulatory system to lose its capacity to restore mtDNA homeostasis and respiratory growth (Fig 3C). We found little evidence for point mutations being responsible for this loss, neither in 44 random t_{50} endpoint populations (Fig S5A) nor in the 5 populations targeted by time-resolved sequencing (Fig S5B). However, all but 4 endpoint populations carried extra nuclear chromosome II (n=29), III (n=21) and/or V (n=16) copies (Fig 4A) at near fixation (mean p: 0.97). We backcrossed (3x) aneuploidy-carrying clones to founder cells and compared paraquat tolerance in meiotic progeny with and without extra chromosomes. Chromosome II or V duplication incurred moderate doubling time reductions (31 and 38 min; s = 0.11 and 0.12) (Fig 4B), while chromosome III gain conferred no such advantage (Fig 4B). Intriguingly, Chr V duplications were exclusive (p=1.1x10^{-5}) to ρ^0 populations (Fig S5D). The appearance of aneuploidies succeeded early O_2^- adaptation and mtDNA reductions (Fig 4C), and in two cases they succeeded complete mtDNA loss. Aneuploidies allowed respiratory growth, although gain of Chr II or V caused strong (147 and 98 min) doubling time defects. (Fig 4D). Taken together, the aneuploidy results suggest that adaptive aneuploidy is intimately linked to chronic O_2^- distress, while not being directly responsible for the mtDNA erosion causing the irrevocable loss of the capacity for homeostatic mitochondrial restoration.

The above genetic data are in concert with the conception that mtDNA erosion under chronic O_2^- distress is due to the same mechanism that is operative under temporary distress, but that its prolonged operation causes the mtDNA degradation to pass a threshold preventing regulatory restoration. To investigate this possibility we pool-sequenced 44 endpoint (t_{50}) adapted populations with short reads to estimate mitochondrial relative to euploid nuclear copy numbers. A single (A12) population retained the mitochondrial genome at only a slightly reduced level (mean reduction in coverage: 34%) (Fig 5A). 25 end-point populations had completely lost nearly their entire 77kb mitochondrial genome, becoming ρ^0 (rho nil, 1981, (Westermann, 2014)). Most of these (20) retained very small (1kb-3kb) mtDNA fragments at either of three (3(1), 18-28(11) and 51-60(8) kb) sites. The remaining 18 populations completely lost large (55.8 to 92.2% of mt genome) mtDNA sections, becoming ρ^- (rho negative). Retained segments of 6 to 34 kb remained at near founder levels (mean increase in coverage: 20 %), in some (5) cases despite lacking known origins of replications. These segments were clearly non-random (Fig 5B) and conferred moderate doubling time advantages compared to the ρ^0 populations (mean of 11 minutes; $p < 0.05$; $s = 0.05$) (Fig S6A). This variability in the mitochondrial genetic data is fully consistent with the operation of a single erosion mechanism. As yeast mtDNA at normal O_2^- levels consists of linear (~90%) and circular (~10%) arrangements (Bendich, 1993), and as linear mtDNA arrays can be replenished from circular mtDNA templates (Maleszka, Skelly, & Clark-Walker, 1991; Shibata & Ling, 2007), we hypothesised that the regulatory mechanism might be more prone to delete linear mtDNA than circular mtDNA. To test the immediate prediction from this hypothesis we determined the mtDNA arrangement in cells retaining segments of the original mtDNA genome by sequencing two ρ^- endpoint clones with reads much longer than the retained segments. We consistently found retained mtDNA segments to exist in circular arrangements, with joining of segment endpoints (Fig 5C, S5D). This is in line with the interpretation that the observed abrupt restoration of mtDNA content, once the cellular memory is erased, is due to replenishment of linear mtDNA arrays by rolling circle mtDNA amplification, and that the observed loss of restorative capacity under chronic O_2^- distress is due to degradation of critical circular mtDNA segments.

Discussion

The observed maintenance of fast growth on glucose despite extensive loss of ETC mtDNA, while the remaining mtDNA segments were retained at near founder population level, can hardly be reconciled with an extensive reduction in the number of mitochondria in response to the imposed temporary O_2^- distress. In addition to their role in energy production, mitochondria fulfil essential functions that do not require mtDNA, including fatty acid synthesis, amino acid production, heme synthesis and iron-sulphur cluster biogenesis, calcium buffering, and are a signalling hub for both innate immunity and cell death of mammalian cells (Pickles, Vigié, & Youle, 2018). Yeast mitochondria share most of these functions (Malina, Larsson, & Nielsen, 2018). A mitophagic response to moderate temporary O_2^- distress would thus be quite costly. Regulated mtDNA deletion therefore appears to be an adaptive response that is invoked when the control mechanisms operating under O_2^- eustress (Sies, 2018) fail to maintain O_2^- homeostasis,

which makes evolutionary sense. The fact that the regulatory system includes a cellular memory strongly supports that it is designed for handling temporary oxidative distress. Under chronic oxidative distress, it will be navigating outside its operational boundaries, likely leading to maladaptive physiological consequences. Our results concerning chronic O_2^- distress are in accordance with this conception.

Assuming our results transfer to mammals, a straightforward prediction is prevalence for loss of ETC genes in age-related disease. A 4977-bp Class 1 mtDNA deletion (mtDNAD4977) in humans has in particular been repeatedly associated with age-related disease states linked to oxidative stress (Phillips et al., 2017), and it includes several genes key for ETC function. The exact mechanism for the formation of this deletion is still not clear, but is thought to be due to faulty replication or repair of double-strand breaks (Phillips et al., 2017). The reason for the widely observed clonal expansion of mtDNA deletions is still enigmatic, but data from mice, rats, rhesus monkeys and humans strongly suggests that it requires to be explained in terms of the machinery for DNA replication (Kowald & Kirkwood, 2018). A feedback process involving a product inhibition mechanism downregulating mtDNA transcription as long as there exist sufficient components for the respiration chain, appears capable of explaining clonal expansion in the listed species (Kowald & Kirkwood, 2018). However, it does not provide a regulatory explanation for the mtDNA deletion as such, nor the restoration of normal mtDNA genomes we observe in yeast. Our results are fully consistent with the conception that the 4977-bp mtDNA excision is due to the operation of an epigenetic and highly specific regulatory system instead of random faulty repair or faulty replication. And they are consistent with the conception that clonal expansion has to be explained in terms mtDNA replication control, caused by regulatory maintenance of the eroded mt genomes until the oxidative distress comes back to eustress levels.

The above conceptions are strongly supported by experimental data on the action of doxorubicin, a cardiotoxic anthracycline antibiotic used in cancer chemotherapy. The major mechanism of doxorubicin cardiotoxicity in mice appears to be via damage or inhibition of the electron transport chain leading to loss of ATP and not general redox stress (Pointon et al., 2010), linking it to a previously reported doxorubicin induced ≈4 kb mtDNA deletion in mouse cardiomyocytes, where the incidence of the deletion increased with the dosage and with the duration of the doxorubicin administration (Adachi et al., 1993). And adult male rats receiving a single doxorubicin bolus restored mitochondrial function within 2 weeks, assayed by the concentration of 8OHdG adducts (Palmeira, Serrano, Kuehl, & Wallace, 1997). The temporal dilution of the adducts appeared not to be due to mitochondrial biogenesis but increase in mtDNA copy number, in line with a regulatory restoration response similar to what we observe in yeast.

An evolutionary rationale for the existence of an adaptive regulatory system in mammals similar to what is present in yeast is that temporary oxidative distress occurs in connection with several types of challenges in the natural environment like heat and ion stress, UV irradiation, pathogen exposure and nutrient deprivation (Filomeni, De Zio, & Cecconi, 2015; L. Li, Chen, & Gibson, 2013). Moreover, considering that the physiological consequences of chronic O_2^- distress

associated with high age have barely been exposed to natural selection, this suggests that mammalian mtDNA might be exposed to the same loss of capacity for mtDNA restoration.

Even though the above considerations do not provide conclusive evidence for the biomedical relevance of our findings, we think they warrant reassessment of huge amounts of existing data in light of the new findings as well as the design of new experimental work on mammalian systems. The presence of a regulatory system underlying specific mtDNA excision opens a completely new avenue for therapy compared to one that has to ameliorate pathogenic effects stemming from a random faulty repair or replication mechanism.

Materials and Methods

Strains: *Founder genotype:* We used a single streaked, domesticated (*MATα ura3*::*NatMX*-barcode *ho::HYGMX*) haploid clone of yeast strain YPS128 as founder genotype. YPS128 is a wild, oak isolate with a North American genome composition (Liti et al., 2009), growth typical of the species (Warringer et al., 2011), outstanding growth at normal O_2^- (avoiding adaptation to background medium) and no confounding transposon insertion in the respiratory Hap1 (in contrast to the reference strain, (Gaisne, Bécam, Verdière, & Herbert, 1999). *Aneuploidies:* We reconstructed individuals with single chromosome duplications (II, III, IV, V, VI, VIII, IX, X, XI, XII, XIII, XIV, XV, XVI – I and VII could not be obtained despite repeated tries and are likely to be inviable in YPS128) by two methods. First, we repeatedly (3x) backcrossed clones from endpoint (t_{50}) adapted populations carrying chromosome duplications to founder clones of the opposite mating type (*MATα ho::HYGMX*). Each backcross was performed on YPD (Yeast Peptone Dextrose) medium using haploids verified by qPCR to retain the chromosome duplication, diploid hybrids were selected after three days of growth on solid minimum media (0.675% Yeast Nitrogen Base (CYN2210, ForMedium), 2% (w/v) D-Glucose, pH=6-6.5 (NaOH), 2.5% agar) medium and sporulated overnight on solid 1% KAc sporulation medium to generate recombined haploids. These were genotyped at the *URA* and *ho* locus and *ura- MATα* haploids were passed on to the next round of backcrossing. After three rounds of backcrossing we selected *ura- MATα* haploids with (n=2 clones) and w/o (n=2 clones) the chromosome duplication of interest and estimated their respective fitnesses (see below) at high biological replication (n=6) and in a completely randomized design on the media of interest. We estimated the fitness (doubling time, D) effect of the chromosome duplication by comparing (subtraction) haploids with and w/o chromosome duplication(s). We assumed that additional mutations affecting doubling time in the adapted clones segregate independently of the chromosome duplication. Second, we reconstructed chromosome duplications that did not occur in adapted populations, or that resisted isolation by backcrossing in YPS128 founder backgrounds as in (Zebrowski & Kaback, 2008). This included genetic modifications of the founder clone genotype to match the *his3Δ* and *can1::STE2pr-HIS3* before essential gene deletion as described in (Zebrowski & Kaback, 2008).

Deletion collection: To estimate gene loss-of-function mutation effect sizes, we used the haploid BY4741 single gene deletion collection (*MAT***a**;*his3Δ1;leu2Δ0;met15Δ0;ura3Δ0; genex::kanMX*), for growth in normal and excess 0_2^-. Collection size: n=4580 (auxotrophic), each cultivated at n=6.

Environment composition: Yeast strains were consistently cultivated in a Synthetic Complete medium (SC; hereafter: "Background medium") composed of 0.14% Yeast Nitrogen Base (CYN2210, ForMedium), 0.50% NH_4SO_4, 0.077% Complete Supplement Mixture (CSM; DCS0019, ForMedium), 2.0% (w/w) glucose, pH set to 5.80 with 1.0% (w/v) succinic acid and 0.6% (w/v) NaOH. For all solid medium cultivations, 2.0% (w/v) agar was added. For pre-cultures to glycine, isoleucine, citrulline and tryptophan selection environments, the background medium was modified to avoid nitrogen storing and later growth on stored nitrogen: CSM was replaced by 20 mg/L uracil (not converted into usable nitrogen metabolites) and NH_4SO_4 was reduced to growth limiting concentrations (30 mg N/L). Selection environments represented simple modifications to the background medium: +0.8 µg/mL rapamycin, +400 µg/mL paraquat, +3 mM arsenic ([As III]; $NaAs_2O_3$), +62,5 mg/L citric acid. For the four nitrogen selection experiments environments, NH_4SO_4 in nitrogen background medium was replaced by 30 mg N/L of one of L-glycine, L-isoleucine, L-citrulline and L-tryptophan together with 20 mg/L uracil (not useable as nitrogen source). We cast all solid plates 10-15 hours prior to use, on an absolutely level surface, by pouring exactly 50mL of medium in the same upper right corner of each plate. We removed excess liquid by drying plates in a laminar air-flow in a sterile environment. All yeast populations were stored at -80°C in 20% glycerol and cultivated at 30.0°C. Populations were subsampled and transferred to and from evolution plates, storage plates and experimental plates using robotics (ROTOR HDA, Singer Instruments) at the indicated transfer format.

Adaptation regime: We single streaked and expanded a single, haploid founder clone to moderate colony size (~2 million cells). We subsampled the colony randomly (~50.000 cells), expanded it until stationary phase (~2 million cells; 36h) in 5mL of background medium, poured the culture on top of a solid plate (background medium; cast in PlusPlates, Singer Instruments, UK) and expanded the population again until stationary phase (72h), creating a lawn of founder cells. We subsampled the lawn and transferred 1152 subsamples (~150.000 cells) to each of eight fresh solid plates, using robotics (ROTOR HDA, Singer Instruments) and 384 short pin pads. These served as pre-cultures (t_{-1}) to the first selection cycle. Pre-cultures were propagated on background, or nitrogen background, medium. We expanded pre-cultured populations into stationary phase (~2 million cells; 72h). These were randomly subsampled (~50.000 cells) by robotics and 1536 short pin pads to create founder populations. The preparation procedure was designed to minimize genetic, epigenetic and plastic (environmental) variation in cellular states, between and within populations and plates. However, we note that none of these parameters will be exactly zero because of the need for population expansions and the structured genetics and

environments of colonies, all of which has small, unavoidable contributions. We initiated adaptation experiments by transferring (384 short pin pads; robotics) 1152 subsampled (~50.000 cells) founder populations to adaptation plates. We propagated this t_0 culture for 72h until stationary phase (no detectable population expansion). Each consecutive adaptation cycle (t_1 to t_{50}) was then initiated and terminated identically. Again, while the design minimized environmental and epigenetic variation over time and space, these parameters will not be exactly zero. The physical boundaries of plates, uneven medium solidification and water evaporation unavoidably creates environmental structure within and between plates that must be normalized away post-experiment. Furthermore, the freeze-thaw history is unavoidably shifted backwards in time and time spent in stationary phase before transfer does increase as populations adapt.

Cultivation for fitness estimation: We cultivated and estimated fitness (doubling time, *D*) for five types of samples with the following individual specifics:
Adapting populations (Data S1): We subsampled (1536 short pin pads) all 9216 adapting populations at the end of each batch-cycle and tracked their population size during a subsequent, directly connected but separated batch expansion at no replication (n=1) and moderate accuracy (no randomization). The separation of adaptation and phenotyping allowed the interleaving (384 short pin pads) of 384 non-evolved, fixed controls (founder genotypes; frozen) among evolving populations and the use of these as spatial controls to reduce bias from systematic variations in environmental parameters within and between plates. The separation systematically shifted the time vector of phenotype relative evolution plates and in the absence of doubling time estimates for the t_0 culture. This was accounted for post-experimentally. Per definition, we cannot establish a t_0 culture that is equivalent to later cultures in that it has passed through at least one preceding stress cultivation round. To avoid an often large, confounding effect from a difference in preculture, we therefore assumed t_1 estimates to be equivalent to t_0 while acknowledging that this, conservatively, do exclude any adaptation occurring in the 1st batch cycle (~2-2.5 doublings in 0_2^- excess). The data is shown in Fig S2A.
Frozen fossil populations (Data S2): We systematically subsampled (~50.000 cells) 96 random populations at the end of batch cycles 0, 1, 2, 3, 4, 5, 7, 9, 12, 15, 20, 25, 30, 35, 40, 45 and 50 in each selection regime. We expanded subsampled populations until stationary phase (72h) in liquid, selection medium micro-cultures (100 µL), added 100µL of glycerol (final concentration: 20% [w/w]) and stored the 768 adapting populations to create a frozen fossil record at (-80°C). The freezing process systematically shifted the time vector of phenotype relative evolution plates; this was handled post-experimentally as for the 9216 adapting populations. We thawed the frozen fossils at room temperature, re-suspended, subsampled (~50.000 cells), transferred subsamples to solid background or nitrogen background medium, and propagated populations until stationary phase (72h) to create no stress pre-cultures. We subsampled pre-cultures, transferred subsamples to solid medium selection plates in a well replicated (n=6) design with near complete randomization of replicates and adaptation stages, within and between plates and instruments. Randomization was achieved using the `randint` function in the python package

`NumPy (version 1.15.4)`. We interleaved 384 spatial controls (as for the 9216 adapting populations; not frozen in stress) among evolving populations. The data is shown in Figs 1C, 3A, 4C, S2B, S3D, S5B and S6A.

Selection relaxation experiments (Data S3, S4): To estimate the heredity of adaptive gains, we tracked the capacity of frozen fossil populations to retain adaptive gains over 1-11 batch cycles of relaxed selection. First (Data S3), we thawed the frozen fossils corresponding to (approximate) batch cycles at which 50-80% of endpoint adaptation had been achieved at room temperature, re-suspended, subsampled and transferred subsamples (~50.000 cells) to either solid background or nitrogen background medium (no stress). We propagated populations over ten consecutive batch cycles (~84 generations) of population expansion and contraction by subsample transfer at every 72h. This selection relaxation regime was performed identically to the adaptation regime, except that proliferation occurred in absence of the original selection pressure (no stress; background or nitrogen background medium). At the end of each relaxed selection batch cycle, all 768 populations were again preserved in glycerol at -80C (as above), creating a second, completely standardized record of frozen fossil populations. These were thawed, re-suspended, subsampled and transferred to 1536 pre-cultivation plates (no stress) with founder populations handled identically. All samples were randomized (as above) across plates, positions and instruments. At stationary phase (72h), pre-cultures were subsampled four times (n=4) and transferred to fresh selection plates, in a randomized design (as above). The pre-cultivation is absolutely required for standardization but systematically shifts the time vector of phenotype relative relaxed selection evolution plates. In the absence of directly comparable population doubling time and number of population doubling estimates for zero batch cycles of deselection, we conservatively assumed the same estimates as for 1st batch cycle of relaxed selection. In the second round (Data S4) selection relaxation follow-up, we thawed the frozen fossils corresponding to batch cycles 1, 2, 3, 4, 5, 7 and 50 of excess O_2^-, relaxed selection over 10 batch cycles of population expansion and contraction, pre-cultivated and cultivated these, exactly as above. The relaxation of selection experiment was repeated again for batch cycles 0, 1, 2, 3, 9 and 30 of five random 0_2^- adapting populations A7, A8, B12, B5 and B8, as above, except that selection relaxation was performed on three independent replicates of each population. The O_2^- adaptation retained in these was tracked at n=5. Data is shown in Fig 2E, 3C and S3B.

Deletion collection (Data S5): We thawed the frozen deletion collections at room temperature, re-suspended, subsampled and transferred subsamples (~50.000 cells) to either solid background (auxotrophic collection) or nitrogen background (prototrophic collection) medium. At stationary phase (72h), pre-cultures were subsampled six times (n=6) and transferred to experimental plates. n=3 replicates were distributed across each of the two experimental plates, in fixed positions, for a total of n=6 replicates. Data was used in Fig 2D.

Chromosome duplications (Data S6 and S7): Reconstructed chromosome duplications were handled as deletion collections, except that replication was higher. n=9. Data was used in Figs 2D, 4B and S5C.

Extracting population size: We tracked population size expansion for all cultures (Data S1 to S7) using the Scan-o-matic system (Zackrisson et al., 2016) version 1.7 (https://github.com/Scan-o-Matic/scanomatic.git). All data sets were acquired and analyzed identically. We deposited 1152 experimental populations (1536 short pin pads) on each experimental plate. To account for spatial variation across plates, without introducing bias from using the experimental values for normalization, we introduced (384 short pin pads) 384 controls (founder genotypes), in the 384 interleaved empty positions on each experimental plate. Controls were subsampled from a separate 384 pre-culture array, pre-cultivated in parallel to experimental cultures. The initial population size, is on average 1.55 fold larger for spatial controls and their history prior to the pre-culture differ from that of experiments; thus, they are not, and should not, be used as exact t_0 samples without further normalization, but are used to capture spatial variation across plates. Plates were maintained undisturbed and without lids for the duration of the experiment (72h) in high-quality desktop scanners (Epson Perfection V800 PHOTO scanners, Epson Corporation, UK) standing inside dark, temperature (30.0C) and moisture controlled thermostatic cabinets with intense air circulation. Scanners were connected via USB to standard desktop computer. Scanner power supplies were separately controlled by power managers that instantaneously shut down the scanner after scans, avoiding light stress. Images, capturing four plates per image, were acquired using SANE (Scanner Access Now Easy) and transmissive scanning at 600 dpi. Plates were fixed in place by custom-made acrylic glass fixtures. Orientation markers contained on each fixture ensured pixel-exact recognition of plate positioning. Each fixture was calibrated by scanner using a calibration model that provided positions for each feature of that fixture, relative to its orientation markers. Pixel intensities were normalized and standardized across instruments and tie using transmissive scale calibration targets (LaserSoft IT8 Calibration Target, LaserSoft Imaging, Germany), accounting for lamp aging and changes in light conditions across the whole pixel intensity spectrum. Each image stack was processed in a two-pass analysis. The first-pass was performed during image acquisition to set up the information needed for population size estimations. Positions in each image were matched to the fixed calibration model using the fixture orientation markers, allowing detection and annotation of plates and transmissive scale calibration strips. In the second-pass analysis, images were segmented to identify plate and transmissive scale calibration strip positions. The calibration strips were trimmed and the pixel intensities compared to the manufacturer's supplied values, such that normalized pixel values remained independent of shifts in light properties over time and space. Colonies were detected using a virtual grid across each plate based on pinning format, and the grid was adjusted such that intersections matched the center of the features detected. At every intersection, each colony and the surrounding area were segmented to determine the local background and pixel intensities. Differences in pixel intensity were converted to population size estimates by calibration to a pre-established, independent calibration function, obtained using cell number estimates from both spectrometer and FACS measurements. Based on these, we obtained population size growth curves. Raw measurements of population size were smoothed in a two-step procedure. First, a median filter identified and

removed local spikes in each curve. Second, a Gaussian filter reduced the influence of remaining local noise.

Extracting fitness, evolutionary time and adaptation potential:

Fitness (doubling time): We identified the steepest slope in each growth curve by local regression over five consecutive time points and converted slopes into population size doubling times. For quality control, the residuals of the regression model were used to determine goodness-of-fit and to flag growth curves suspected to be of poor quality. We manually inspected all flagged growth curves, discarding approximately 0.3% as erroneous. We extracted the population size doubling time, $D(h)$, for all retained populations, log transforming it, $\log_2(D)$. To minimize systematic errors caused by structured environmental variation within and between plates, we used the fixed spatial controls (founders) introduced at every fourth position. First, controls with extreme $\log_2(D)$ were removed and the remaining control positions were used to interpolate a normalization surface, across all 1536 positions on the plate; i.e. we estimated what $\log_2(D)$ value a control would have in each position. Second, the interpolated surface was smoothed, first with a kernel filter to exclude deviant positions, and then with a Gaussian smoothing ($\sigma = 1.5$) to soften the contours of the control landscape. Third, for each of the 1536 positions on plate, we estimated the expected control value, $\log_2(D)$, by calculating the distance-weighted (Gauss distribution, $\sigma = 1.5$) local mean of the smoothed control landscape. Fourth, we extracted the normalized, relative population size doubling time for each experiment, D_r, by subtracting the control value for that position as $D_r = \log_2(D_{experiment}) - \log_2(D_{control})$. D_r is reported as y-axis data for Data S1 and Fig S2A. For frozen fossil populations (Data S2), the control samples differ slightly in cultivation and storage history from other samples, creating a small but systematic error. We remove this error by estimating the adaptation, A, achieved up to each batch cycle as $A_t = D_{r,t} - D_{r,0}$. A is reported as y-axis data for Data S2 and Fig 1C.

Evolutionary time: We estimated the evolutionary time (generations, G) passed in each bath cycle for each population as the number of population size doublings from batch cycle start to end. We estimated evolutionary time for missing batch cycles by linear interpolation between values for adjacent measured batch cycles. Summing over all batch cycles up to a given time point, we obtain a measure, G, mean number of consecutive cell generations in each population since adaptation start, assuming that cell death is negligible. G is reported as x-axis data in Data S1, S2. We use a mean across relevant populations, G, as x-axis data in Fig 1C, 2C, 2E, 3, 4C, S2A, S3B, S4B, S5B and S6.

Adaptation potential and realized adaptation potential: We estimated the adaptation potential by postulating that the doubling time of YPS128 in absence of stress represents an absolute, lower boundary for the doubling time achievable by YPS128 in the relevant evolutionary time span, regardless of environment. To retain the spatial normalization, we converted the D_r to a normalized absolute doubling time D_{norm}, as $D_{norm} = 2^{Dr} D_{control,\ grand}$. $D_{control,\ grand}$ is the grand mean of all control doubling times run in the particular experimental series, removing plate and batch bias and allowing direct comparison across experiments. We estimated the absolute

adaptation achieved as $A_{abs} = D_{norm,t,k} - D_{norm,0,k}$. We fitted a Loess regression (span, α = 0.6; tricubic weighting) to each adaptation trajectory (A_{abs}, G) and extracted the Loess estimated $A_{abs,loess}$, at each cell generation, G_r , using function approx (linear method) in R. We estimated the adaptation potential at time t, as $P = D_{norm,0,k} - D_{norm,0,no\ stress}$. $D_{norm,0,k}$ is the normalized absolute doubling time of founder (t=0) population in environment k; $D_{norm,0,no\ stress}$ is the normalized absolute doubling time of founder (t=0) populations in absence of stress. We estimated the fraction of the adaptation potential that had been realized at each time point t, as $P_R = A_{abs,loess} / P$. We report mean P_R for the first 10 generations of superoxide adaptation as 72.9%.

Selection coefficients: We estimated selection coefficients for variants driving observed change in population doubling time as in (Gjuvsland et al., 2016).

We employed the simplifying assumptions that the change in cell division time imposed by a variant is the only parameter driving change in population doubling time and obtain:

$$s = ln\,(2)\left(\frac{\tau_{founder}}{\tau_1} - 1\right)$$

Where s = per generation selection coefficient for a genotype with a cell division time equal to the measured population size doubling time and $\tau_1$and $\tau_{founder}$ represent the population size doubling times, D_{norm}, of the population of interest, and the founder population.

Sequencing

Long read sequencing and de novo assembly of the YPS128 founder

Genomic DNA was extracted from an overnight founder culture using the phenol-chloroform protocol. Norwegian Sequencing Centre sequenced the genome on a PacBio RS II instrument using the P4-C2 chemistry. Additional PacBio sequencing data of the same YPS128 genotype was incorporated from (Yue et al., 2017). A total of 9 SMRT cells were used in the assembly that resulted in 1352628 reads, corresponding to approximately 205x genome coverage. We ran the *de novo* assembly using the hierarchical assembly protocol RS_HGAP_Assembly3.3 with an expected genome size of 12Mb.

Resequencing of adapted populations

DNA from O_2^- adapted populations was extracted using a modified protocol of the Epicentre MasterPure Yeast DNA Purification Kit. DNA was extracted from overnight O_2^- stressed cultures. Pool sequencing was performed at SciLife (Stockholm, Sweden), using Illumina HiSeq2500, 2x126bp. Libraries were prepared using the Nextera XT kit to accommodate the low DNA yield from small cultures. At least two founder controls were included in each flow cell.

Calling de novo point mutations: Sequenced reads were quality-trimmed and nextera transposase sequences were removed with TrimGalore(v.0.3.8). Reads were mapped to the YPS128 pacbio assembly using BWA MEM (v.0.7.7-r441). PCR and optical duplicates were flagged using Picard-tools (v.1.109[1716]). Base alignment quality scores were calculated using samtools calmd (v.0.1.18 [r982:295]) and variants were called using Freebayes (v0.9.14-8-g1618f7e). All alleles were reported regardless of frequency or genotype model since the populations are a pooled set of individuals. Variants were annotated using SnpEFF (v.3.6c). Variants below a quality score of 20 and variants present in the sequenced founder samples were filtered out

Calling aneuploidies: Aneuploidies were called using a sliding, non-overlapping 200bp window coverage of reads mapped. Reads with a MAPQ of <1 were not counted. Window coverage ratio was calculated as:

$$log_2\left(\frac{kw_i}{w_{founder,i}}\right)$$

where w_n is the depth of coverage of mapped reads in each 200bp window, i, $w_{founder}$ is the depth of coverage of each *i* in a founder sequenced in the same flow cell and

$$k = {\sum_{i=1}^{G} D_{founder}}\Big/{\sum_{i=1}^{G} D_{sample}}$$

where *G* is the YPS128 genome size and *D* the depth of coverage for each nucleotide. Aneuploidies were called determining the median log_2 window coverage as calculated above for each chromosome. Data is shown in Fig 4A, S5A and S5D.

Calling mtDNA copy number change: mtDNA copy number was calculated for each sample genome using a sliding, non-overlapping window of 1kB.The mtDNA copy number relative euploid nuclear genome ratio was calculated for each window as:

$$log_2\left(\frac{w_i}{w_{median,euploid}}\right)$$

where $w_{median,euploid}$is the median of all 1kB windows of the nuclear genome excluding chromosomes with detected aneuploidies. Data is shown in Fig 5A. We estimate the median (across all windows) absolute number of mtDNA molecules, assuming one copy of the nuclear genome and assuming no sequencing bias for mitochondrial DNA, as

$$2^{median\left(log_2\left(\frac{w_i}{w_{median,euploid}}\right)\right)}$$

This is shown in Fig 3A and S4B. We show the same number window-wise in Fig 5B and S4A.

Long read sequencing: DNA was extracted using Qiagen Genomic-tip 100/G DNA extraction kit. Libraries for Oxford Nanopore sequencing were prepared using 1D Native barcoding genomic DNA with the EXP-NBD104 and SQK-LSK108 kit. The flowcell version was FLO-MIN106. The raw nanopore reads were basecalled by guppy (v2.1.3) with a minimal quality score cutoff of 5 (options: --qscore_filtering --min_qscore 5). For all basecalled reads that passed the quality filter, demultiplexing was further performed by guppy with the help of the `guppy_reads_classifier.pl` from LRSDAY (v1.3.1) (Yue & Liti, 2018). The demultiplexed reads were processed by LRSDAY (v1.3.1) for adapter trimming, reads downsampling (downsampled to 50X coverage), *de novo* assembly, assembly polishing, assembly scaffolding, and dotplot visualization. This is shown in Fig 5C. In addition to de novo assembly, we also aligned the adapter trimmed reads to the reference genome of the founder strain YPS128 (Yue et al., 2017) by minimap2 (v2.13) (H. Li, 2018). IGV (v2.4.14) to visualize the resulting alignment. This is shown in Fig S6B.

Individual-based modelling of experimental populations

Population parameters: We simulated the experimental adaptations in an individual-based model, implemented in Python (https://github.com/HelstVadsom/GenomeAdaptation.git) (Gjuvsland et al., 2016). We repeated each simulation 1152x. We start from a haploid, isogenic founder population that is subsampled at the end of each batch phase to found the next cultivation cycle. Population parameters describe population size at the start of each batch cycle (N), number of cell divisions before subsampling in each batch cycle (M_t) and total number of batch cycles (n=50 cycles). When the total population size reaches $2^{Mt}N$ cells, N cells are sub-sampled randomly to found the next cycle. N was set to equal the approximate mean across all empirical sub-samplings. M_t was set to the equal the mean (across populations) empirical measure in each batch cycle t. Each cell divides 12 times before it dies. Mating, meiosis, sporulation or ploidy change do not occur. There is no population structure.

Mutation rate parameters: All cells begin as identical, haploid founder cells. Cells have 4947 nuclear encoded protein genes, and 16 chromosomes specified by the sequenced reference genome (R64-1-1.23). Cells have no mitochondrial genome and essential nuclear encoded genes are not included. Cells independently and randomly acquire nuclear genome mutations as chromosome duplications and point mutations in protein coding genes at the end of each cell division. Mutation rates are constant, equal for all genomes, for all chromosomes and for all nucleotide sites. Chromosomes and nucleotide sites mutate only once. Sites on new chromosomes do not mutate. Chromosome duplications occur at rate, $\mu = 4.85*10^{-5}$ duplications/cell division. Point mutations occur at rate $\mu = 0.33*10^{-9}$ point mutations/bp/division (Lynch et al., 2008; Zhu, Siegal, Hall, & Petrov, 2014).

Mutation effect size parameters: We track the mutations of each cell, its reproductive age, and its cell division time. Mutations and cell division time is passed on to daughter cells. Cells begin at a cell division equal to the founder population doubling time. Change in cell division time is affected by mutations only and mutations only affect cell division time. Because chromosome duplication and loss-of-gene function point mutations are the most common drivers of experimental adaptation (Chevereau et al., 2015) we approximated these as the only sources of change in cell division time. We estimated the cell division effect size of chromosome duplications by comparing the population doubling time, D_{norm}, of backcrossed spores, with and w/o each chromosome duplication. We estimated the cell division effect size of point mutations by downloading the SIFT yeast database (http://sift-db.bii.a-star.edu.sg/public/Saccharomyces_cerevisiae/EF4.74/) and extracting all possible stop gain base changes and nonsynonymous mutations, with attached SIFT scores (Vaser, Adusumalli, Leng, Sikic, & Ng, 2016). All stop gain base changes and all nonsynonymous mutations with a SIFT score <0.05 affect cell division time with an effect size equal to the population doubling time, D_r, effect of the corresponding gene deletion (see above). Aging, epigenetics, or mtDNA do not effect cell division time. There are no cell-cell interactions. We assumed that population doubling time in stress in the relevant evolutionary time span cannot become shorter than the measured mean founder population doubling time in absence of stress, $D_{founder,no\ stress}$. We

implement the well documented (Chou, Chiu, Delaney, Segrè, & Marx, 2011; Khan, Dinh, Schneider, Lenski, & Cooper, 2011) diminishing return of mutations with increasing fitness, by letting a mutation, *m*, define the cell division time in cells with the mutation, D_m , as:

$$D_m = k\left(D_G - D_{founder,no\ stress}\right) + D_{founder,no\ stress}$$

D_G is the cell division time of the genotype before the mutation occurred. *k* is defined as:

$$k = max\left(\frac{2^{D_r} D_{founder,stress} - D_{founder,no\ stress}}{D_{founder,stress} - D_{founder,no\ stress}}, 0\right)$$

$D_{founder,stress}$ is the measured mean doubling time of the founder in presence of the simulated stress. No other form of epistasis is included. We recorded population averages for cell division times at the end of each batch cycle, and plot it as a function of mean evolutionary time (total population doublings).

Simulating relaxed selection: We simulated loss of adaptive gains similarly, except for that we let mutation cell division effect sizes be defined by the measured aneuploidy and gene deletion effects in no stress.

Acknowledgements
The authors would like to acknowledge support from Science for Life Laboratory, the National Genomics Infrastructure, NGI, and Uppmax for providing assistance in massive parallel sequencing and computational infrastructure. The authors would also like to acknowledge Johan Hallin and Lars-Göran Ottosson for help and advice with strain construction and design of adaptation experiment, and Olga Kourtchenko for help with designing nitrogen utilization environments. Norwegian Sequencing Centre (NSC) is acknowledged for providing the PacBio sequencing

References

Adachi, K., Fujiura, Y., Mayumi, F., Nozuhara, A., Sugiu, Y., Sakanashi, T., … Toshima, H. (1993). A deletion of mitochondrial DNA in murine doxorubicin-induced cardiotoxicity. *Biochemical and Biophysical Research Communications*, *195*(2), 945–951. https://doi.org/10.1006/bbrc.1993.2135

Ammal Kaidery, N., & Thomas, B. (2018). Current perspective of mitochondrial biology in Parkinson's disease. *Neurochemistry International*, *117*, 91–113. https://doi.org/10.1016/j.neuint.2018.03.001

Bendich, A. J. (1993). Reaching for the ring: the study of mitochondrial genome structure. *Current Genetics*. https://doi.org/10.1007/BF00336777

Castello, P. R., Drechsel, D. A., & Patel, M. (2007). Mitochondria are a major source of paraquat-induced reactive oxygen species production in the brain. *Journal of Biological Chemistry*, *282*(19), 14186–14193. https://doi.org/10.1074/jbc.M700827200

Chevereau, G., Dravecká, M., Batur, T., Guvenek, A., Ayhan, D. H., Toprak, E., & Bollenbach, T. (2015). Quantifying the Determinants of Evolutionary Dynamics Leading to Drug Resistance. *PLoS Biology*, *13*(11), 1–18. https://doi.org/10.1371/journal.pbio.1002299

Chou, H.-H., Chiu, H.-C., Delaney, N. F., Segrè, D., & Marx, C. J. (2011). Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science (New York, N.Y.)*, *332*(6034), 1190–2. https://doi.org/10.1126/science.1203799

Cochemé, H. M., & Murphy, M. P. (2008a). Complex I is the major site of mitochondrial superoxide production by paraquat. *Journal of Biological Chemistry*, *283*(4), 1786–1798. https://doi.org/10.1074/jbc.M708597200

Cochemé, H. M., & Murphy, M. P. (2008b). Complex I is the major site of mitochondrial superoxide production by paraquat. *Journal of Biological Chemistry*, *283*(4), 1786–1798. https://doi.org/10.1074/jbc.M708597200

Fang, J., & Beattie, D. S. (2003). External alternative NADH dehydrogenase of Saccharomyces cerevisiae: a potential source of superoxide. *Free Radical Biology and Medicine*, *34*(4), 478–488. https://doi.org/10.1016/S0891-5849(02)01328-X

Filomeni, G., De Zio, D., & Cecconi, F. (2015). Oxidative stress and autophagy: The clash between damage and metabolic needs. *Cell Death and Differentiation*, *22*(3), 377–388. https://doi.org/10.1038/cdd.2014.150

Gaisne, M., Bécam, A. M., Verdière, J., & Herbert, C. J. (1999). A "natural" mutation in Saccharomyces cerevisiae strains derived from S288c affects the complex regulatory gene HAP1 (CYP1). *Current Genetics*, *36*(4), 195–200. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10541856

Gjuvsland, A. B., Zörgö, E., Samy, J. K., Stenberg, S., Demirsoy, I. H., Roque, F., … Warringer, J. (2016). Disentangling genetic and epigenetic determinants of ultrafast adaptation. *Molecular Systems Biology*, *12*(12), 892. https://doi.org/10.15252/msb.20166951

Harman, D. (1972). The biologic clock: the mitochondria? *Journal of the American Geriatrics Society*, *20*(4), 145–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/5016631

Hepple, R. T. (2016). Impact of aging on mitochondrial function in cardiac and skeletal muscle. *Free Radical Biology & Medicine*, *98*, 177–86. https://doi.org/10.1016/j.freeradbiomed.2016.03.017

Hyttinen, J. M. T., Viiri, J., Kaarniranta, K., & Błasiak, J. (2018). Mitochondrial quality control in AMD: does mitophagy play a pivotal role? *Cellular and Molecular Life Sciences*, *75*(16), 2991–3008. https://doi.org/10.1007/s00018-018-2843-7

Khan, A. I., Dinh, D. M., Schneider, D., Lenski, R. E., & Cooper, T. F. (2011). Negative epistasis between beneficial mutations in an evolving bacterial population. *Science*, *332*(6034), 1193–1196. https://doi.org/10.1126/science.1203801

Khrapko, K., & Vijg, J. (2009). Mitochondrial DNA mutations and aging: devils in the details? *Trends in Genetics*, *25*(2), 91–98. https://doi.org/10.1016/j.tig.2008.11.007

Kispal, G., Sipos, K., Lange, H., Fekete, Z., Bedekovics, T., Janáky, T., … Lill, R. (2005). Biogenesis of cytosolic ribosomes requires the essential iron-sulphur protein Rli1p and mitochondria. *EMBO Journal*, *24*(3), 589–598. https://doi.org/10.1038/sj.emboj.7600541

Kowald, A., & Kirkwood, T. B. L. (2018). Resolving the enigma of the clonal expansion of mtDNA deletions. *Genes*, *9*(3). https://doi.org/10.3390/genes9030126

Kujoth, G. C., Nash, N., Sherman, M., Hanna, A., Knight, J., Zehr, C., … Norton, D. (2005). Apoptosis in Mammalian Aging Mitochondrial DNA Mutations, Oxidative Stress, and. *Neurosci. Lett. Arch. Neurol. Ann. Neurol. Science Science Neuron Arch. Neurol. Trends*

Neurosci. Neuron Nat. Genet. J. Biol. Chem. Proc. Natl. Acad. Sci. U.S.A. J. Neurochem. Nature Proc. Natl. Acad. Sci. U.S.A. Science J. Neurosci. Nature Nature, *162*(805), 481–1491. https://doi.org/10.1126/science.1112125

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li, L., Chen, Y., & Gibson, S. B. (2013). Starvation-induced autophagy is regulated by mitochondrial reactive oxygen species leading to AMPK activation. *Cellular Signalling*, *25*(1), 50–65. https://doi.org/10.1016/j.cellsig.2012.09.020

Lionaki, E., Gkikas, I., & Tavernarakis, N. (2016). Differential protein distribution between the nucleus and mitochondria: Implications in aging. *Frontiers in Genetics*, *7*(SEP). https://doi.org/10.3389/fgene.2016.00162

Liti, G., Carter, D. M., Moses, A. M., Warringer, J., Parts, L., James, S. A., … Koufopanou, V. (2009). Population genomics of domestic and wild yeasts. *Nature*, *458*(7236), 337–341. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2659681/pdf/ukmss-4241.pdf

Lynch, M., Sung, W., Morris, K., Coffey, N., Landry, C. R., Dopman, E. B., … Thomas, W. K. (2008). A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proceedings of the National Academy of Sciences*, *105*(27), 9272–9277. https://doi.org/10.1073/pnas.0803466105

Maleszka, R., Skelly, P. J., & Clark-Walker, G. D. (1991). Rolling circle replication of DNA in yeast mitochondria. *The EMBO Journal*, *10*(12), 3923–3929. https://doi.org/10.1002/j.1460-2075.1991.tb04962.x

Malina, C., Larsson, C., & Nielsen, J. (2018). Yeast mitochondria: An overview of mitochondrial biology and the potential of mitochondrial systems biology. *FEMS Yeast Research*. https://doi.org/10.1093/femsyr/foy040

Palikaras, K., & Tavernarakis, N. (2014). Mitochondrial homeostasis: The interplay between mitophagy and mitochondrial biogenesis. *Experimental Gerontology*, *56*, 182–188. https://doi.org/10.1016/j.exger.2014.01.021

Palmeira, C. M., Serrano, J., Kuehl, D. W., & Wallace, K. B. (1997). Preferential oxidation of cardiac mitochondrial DNA following acute intoxication with doxorubicin. *Biochimica et Biophysica Acta - Bioenergetics*, *1321*(2), 101–106. https://doi.org/10.1016/S0005-2728(97)00055-8

Payne, B. A. I., & Chinnery, P. F. (2015). Mitochondrial dysfunction in aging : Much progress but many unresolved questions. *BBA - Bioenergetics*, *1847*(11), 1347–1353. https://doi.org/10.1016/j.bbabio.2015.05.022

Phillips, A. F., Tigano, M., Piganeau, M., Brunet, E., Sfeir, A., Millet, R., … Babin, L. (2017). Single-Molecule Analysis of mtDNA Replication Uncovers the Basis of the Common Deletion Article Single-Molecule Analysis of mtDNA Replication Uncovers the Basis of the Common Deletion. *Molecular Cell*, *65*, 527–538.

Pickles, S., Vigié, P., & Youle, R. J. (2018). Mitophagy and Quality Control Mechanisms in Mitochondrial Maintenance. *Current Biology*, *28*(4), R170–R185. https://doi.org/10.1016/j.cub.2018.01.004

Pigliucci, M. (2006). Phenotypic plasticity and evolution by genetic assimilation. *Journal of Experimental Biology*, *209*(12), 2362–2367. https://doi.org/10.1242/jeb.02070

Pointon, A. V., Walker, T. M., Phillips, K. M., Luo, J., Riley, J., Zhang, S. D., … Gant, T. W. (2010). Doxorubicin in vivo rapidly alters expression and translation of myocardial electron

transport chain genes, leads to ATP loss and caspase 3 activation. *PLoS ONE*, *5*(9), 1–17. https://doi.org/10.1371/journal.pone.0012733

Shibata, T., & Ling, F. (2007). DNA recombination protein-dependent mechanism of homoplasmy and its proposed functions. *Mitochondrion*. https://doi.org/10.1016/j.mito.2006.11.024

Sies, H. (2018). On the history of oxidative stress: Concept and some aspects of current development. *Current Opinion in Toxicology*, *7*, 122–126. https://doi.org/10.1016/j.cotox.2018.01.002

Turrens, J. F. (1997, February 1). Superoxide production by the mitochondrial respiratory chain. *Bioscience Reports*. Portland Press Limited. https://doi.org/10.1023/A:1027374931887

Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M., & Ng, P. C. (2016). SIFT missense predictions for genomes. *Nature Protocols*, *11*(1), 1–9. https://doi.org/10.1038/nprot.2015.123

Wallace, D. C. (2007). Why Do We Still Have a Maternally Inherited Mitochondrial DNA? Insights from Evolutionary Medicine. *Annual Review of Biochemistry*, *76*(1), 781–821. https://doi.org/10.1146/annurev.biochem.76.081205.150955

Warringer, J., Zörgö, E., Cubillos, F. A., Zia, A., Gjuvsland, A., Simpson, J. T., … Blomberg, A. (2011). Trait variation in yeast is defined by population history. *PLoS Genetics*, *7*(6), e1002111. https://doi.org/10.1371/journal.pgen.1002111

Westermann, B. (2010, August 1). Mitochondrial dynamics in model organisms: What yeasts, worms and flies have taught us about fusion and fission of mitochondria. *Seminars in Cell and Developmental Biology*. Academic Press. https://doi.org/10.1016/j.semcdb.2009.12.003

Westermann, B. (2014). Mitochondrial inheritance in yeast. *Biochimica et Biophysica Acta - Bioenergetics*. https://doi.org/10.1016/j.bbabio.2013.10.005

Yue, J.-X., Li, J., Aigrain, L., Hallin, J., Persson, K., Oliver, K., … Liti, G. (2017). Contrasting evolutionary genome dynamics between domesticated and wild yeasts. *Nature Genetics*, *49*(6), 913–924. https://doi.org/10.1038/ng.3847

Yue, J.-X., & Liti, G. (2018). Long-read sequencing data analysis for yeasts. *Nature Protocols*, *13*(6), 1213–1231. https://doi.org/10.1038/nprot.2018.025

Zackrisson, M., Hallin, J., Ottosson, L.-G., Dahl, P., Fernandez-Parada, E., Ländström, E., … Blomberg, A. (2016). Scan-o-matic: High-Resolution Microbial Phenomics at a Massive Scale. *G3: Genes|Genomes|Genetics* , *6*(September), 3003–3014. https://doi.org/10.1534/g3.116.032342

Zebrowski, D. C., & Kaback, D. B. (2008). A simple method for isolating disomic strains of Saccharomyces cerevisiae. *Yeast*, *25*(5), 321–326. https://doi.org/10.1002/yea.1590

Zhu, Y. O., Siegal, M. L., Hall, D. W., & Petrov, D. A. (2014). Precise estimates of mutation rate and spectrum in yeast. *Proceedings of the National Academy of Sciences*, *111*(22), E2310–E2318. https://doi.org/10.1073/pnas.1323011111

Zsurka, G., Peeva, V., Kotlyar, A., & Kunz, W. S. (2018). Is there still any role for oxidative stress in mitochondrial DNA-dependent aging? *Genes*, *9*(4). https://doi.org/10.3390/genes9040175

Tables

Table 1:

Environment	Compound	Concentration	Type
As(III)	Sodium meta arsenite ($NaAsO_2$)	3mM	Heavy metalloid
Rapamycin	Sirolimus	0.8μg/ml	TOR-inhibitor
Citric acid	Citric Acid	62.5mg/mL	Weak acid
Superoxide	Paraquat (Methyl viologen dichloride)	400μg/ml	Cyclic ROS-producer
Glycine	L-Glycine	160.86 mg/L (30mg/ml N)	Nitrogen utilization
Citrulline	L-Citrulline	125.14 mg/L (30mg/ml N)	Nitrogen utilization
Tryptophan	L-Tryptophan	218.82 mg/L (30mg/ml N)	Nitrogen utilization
Isoleucine	L-Isoleucine	281.1 mg/L (30mg/ml N)	Nitrogen utilization

Figure legends

Fig. 1 Ultrafast adaptation to O_2^- distress

A) Model of basal (top) and paraquat induced (bottom) O_2^- generation. **B)** Experiment design. We adapted 1152 homogeneous, haploid yeast populations to each of eight selection pressures over 50 cycles (t_1-t_{50}) of expansion and random subsampling. We tracked population size (n=1, no randomization), extracting population size doubling time (D) and population doublings (generations, G) from each batch cycle. We stored batch cycles 0-5, 7, 9, 12, 15, 20, 25, 30, 35, 40, 45 and 50 of 96 populations per environment and re-analyzed these (n=6, complete randomization) to capture near error free adaptation kinetics for 768 populations. **C)** Adaptation (y-axis, A) for frozen fossil populations (mean of n=96, each at n=6) adapting to each selection pressure (color) as a function of mean generations (x-axis, G). Shade: *S.E.M*

Fig. 2 Regulated O_2^- adaptation is genetically assimilated

A) Model for separating regulated and Darwinian (genetic) adaptation. Regulated adaptation is faster (left column) but not transmitted to offspring under relaxed selection (right column). Under genetically assimilation, regulatory solutions become genetically fixed and heredity emerges. **B)** Design of relaxed selection experiments. *First round:* We revived frozen populations having achieved 50-80% of endpoint (t_{50}) adaptation (blue area) and propagated and stored these over 10 batch cycles in absence of selection. We re-exposed stored populations (n=5, complete randomization) to the original selection pressure, extracting doubling times (D). *Second round:* We released populations at many stages of adaptation (arrows) from selection and tracked the kinetics of adaptation loss. **C)** O_2^- adaptation (y-axis, A) retained as a function of mean generations of relaxed selection (x-axis), first round experiments. Mean (n=96 populations, each measured at n=5) is shown. Shade: s.e.m. **D**) O_2^- adaptation as compared to in Darwinian models based on point mutations, chromosome duplication or both. **E)** O_2^- adaptation (y-axis, A) retained as a function generations of relaxed selection (x-axis, mean G) for n=96 populations (lines, each at n=5, randomization). Panels show adaptation loss after 6, 10, 15, 19, 24, 33 and 242 generations of adaptation.

Fig. 3 Early O_2^- adaptation targets the ETC

A) mtDNA content (red, left y-axis; median coverage relative euploid nuclear genome) as a function of generations of O_2^- evolution for five populations (panels). O_2^- adaptation (green, right y-axis; mean of n=6) is shown for comparison. Shaded area: s.e.m. **B)** Respiratory (glycerol) growth (red, right y-axis) as a function of generations of O_2^- adaptation (x-axis). O_2^- adaptation (green, left y-axis) is included for comparison. Mean doubling times (n=96, each at n=3, randomization) are shown. Shaded area: s.e.m. **C)** Respiratory (glycerol) growth (red, right y-axis) as a function of generations of relaxed selection (x-axis). Panels: generations (G) of O_2^-

adaptation before relaxation of selection. Retained O_2^- adaptation (green, left *y*-axis) is included for comparison. Mean doubling times (n=15 each at n=3, randomization) are shown. Shade: s.e.m.

Fig. 4 Chronic O_2^- distress causes aneuploidy

A) Coverage (color, $\log_2$ chromosome/genome median) for each chromosome (*y*-axis) in sequenced (n=44) O_2^- adapted endpoint (t_{50}) populations (*x*-axis). **B, D)** Fitness effect (*y*-axis, doubling time, D_{norm}) of re-current chromosome duplications (panels) in O_2^- (paraquat) (B) and glycerol (D). We backcrossed (3x) endpoint to founder cells and compared offspring (n=2 spores, each in n=6 replicates) w. and w/o duplicated chromosome II (B; red), III (B; purple) and V (B; brown). Error bars: s.e.m. **B)** O_2^- (paraquat) **D)** Respiration (glycerol) **C)** Kinetics of chromosome duplications in five (panels) O_2^- adapting populations. Left *y*-axis: coverage ($\log_2$ chromosome/genome median) of chromosome II (blue), III (red) and V (yellow) as a function of generations (G) of adaptation (*x*-axis). Dashed line: missing data. O_2^- adaptation is shown for comparison (right *y*-axis). Shade: s.e.m.

Fig. 5 Chronic O_2^- distress causes complete mtDNA loss

A) Chronic O_2^- distress ends in retention of mtDNA (ρ^+) at slightly reduced levels (top panel; 1 population, 2.3%), complete loss of mtDNA segments (ρ^-) encoding O_2^- generating functions (middle panel; 18/44, 41%) or complete loss of entire (ρ^0) mtDNA (bottom panel; 25/44, 57%). *Annotation box:* mtDNA encoded protein genes. **B)** Mean (n=44 endpoint populations) coverage relative the euploid nuclear genome, across the mt genome (1kb bins). Brown: tRNA genes, orange: protein genes, yellow: replication origins **C)** Pairwise comparison between the assembled mitochondrial contigs of two evolved clones (A7_2, D3_1) and the YPS128 reference mitochondrial genome. Red: forward sequence matches, blue: reverse sequence matches. The sequence repetitions are consistent with circularization of the retained mtDNA segments.

Fig. S1 Design of adaptation experiments, detailed design

A) Details of adaptation experiment design and workflow. We single streaked and expanded a single, haploid YPS128 founder clone, subsampled the colony randomly and expanded it in liquid background medium until stationary phase. We poured the culture on top of a single solid plate (red plate) expanding the population until stationary phase. We subsampled the solid stationary phase culture in 1152 dispersed positions, transferring and depositing subsamples on each of two fresh solid plates (blue plates) using robotics (ROTOR HDA, Singer Instruments) and 1536 short pin transfer pads to create pre-cultures (t_{-1}). The two pre-cultures were on background (rapamycin, arsenic, O_2^-, citric acid) or nitrogen background medium (tryptophan,

isoleucine, glycine, citrulline), the latter to deplete intracellular nitrogen storages. We expanded the 1152 x 2 precultures into stationary phase, repeatedly (x4) subsampled each pre-culture (as above), and transferred and deposited each subsamples onto one of the eight selection media (Table S1). We propagated the 1152 x 8 cultures, the time zero (t_0), cultures until stationary phase (green plates). We repeated the cycle of subsampling and population expansion over 50 consecutive batch cultivation cycles (orange plates, t_1 to t_{50}). We measured population size once (n=1) at 20 minute intervals and extracted doubling times, D (Data S1). We also subsampled every 12th population at the end of batch cycles 0-5, 7, 9, 12, 15, 20, 25, 30, 35, 40, 45 and 50, transferred subsamples to liquid selection medium (grey plates), expanded these to stationary phase and added glycerol to a final concentration of 20%. Plates were stored at -80C as a frozen fossil record of 768 populations. Frozen stocks were later subsampled, revived by transfer to liquid background or nitrogen background medium and expanded as pre-cultures to stationary phase (purple plates). Pre-cultures were repeatedly (n=5) subsampled, subsamples were transferred to selection medium in a randomized design, non-evolved founder subsamples were interleaved in every 4th position on each plate to serve as spatial controls and the population size expansion were tracked in 20 minute intervals. We extracted doubling times, D, normalized the doubling times of experiments to those of neighbouring spatial controls to generate relative doubling times, and then transformed these back to absolute doubling times without spatial bias (D_{norm}). Data S2.

Fig. S2 Ultrafast adaptation to O_2^- distress

A) Adaptation kinetics of 9216 adapting populations. Panels: Selection pressures. Lines: 1152 populations (n=1 replicate) per selection pressure. y-axis: adaptation. x-axis: generations, G. Inset: zoom-in on first cycles of O_2^- adaptation. **B)** Populations adapting to O_2^- transitions from growth over multiple phases to a single growth phase during early adaptation. Growth in excess O_2^- before adaptation (light green, t_0) and at early (intermediate green, t_2) and late (dark green, t_{50}) O_2^- adaptation is shown for an example population. *Left panel:* Population size (y-axis, on $\log_2$ scale, curves aligned to same start OD for visualization). *Right panel:* 1st derivative of population size (y-axis). x-axes: hours of growth in excess O_2^-. Arrow indicates second growth phase.

Fig. S3 Regulated O_2^- adaptation is genetically assimilated

A, C) No selection against O_2^- adaptation at normal O_2^- levels (no paraquat). Mean (left y-axis, n=96 populations, each at n=5 replicates) growth defect (D_t - D_{t0}) in absence of the selection agent. Translation into selection coefficients (right y-axis, s) is shown. **A)** First round experiment: Populations having achieved 70-90% of their endpoint adaptation to O_2^- distress. Selection against arsenic and glycine adaptations is shown for comparison. **C)** Second round

experiment: populations adapting to excess O_2^- captured at various stages of adaptation (*x*-axis, see Fig 2E). **B)** Arsenic and glycine adapting populations only slowly lose their adaptations under relaxed selection despite strong counter-selection against them. Arsenic and glycine adaptation retained (*y*-axis, *A*) as a function generations of relaxed selection (*x*-axis). First round experiments, populations having achieved 70-90% of their endpoint adaptation. Means (*n*=96 populations, each measured at *n*=5) are shown. Shade: s.e.m. **D)** Populations at early stage of O_2^- adaptation mostly grow normally at normal O_2^- levels. Growth curves of early O_2^- adapted (green, t_2) and founder (t_0 = red) populations at normal O_2^- (no paraquat). Median of *n*=96 populations, each at *n*=5 replicate) is shown; poor quality curves were excluded. Shade: s.d.

Fig. S4 Early O_2^- adaptation targets the ETC

A) Early O_2^- adaptation targets the O_2^- generating functions of the electron transport chain (ETC). mtDNA coverage (green bars, 1kb bins) relative the euploid nuclear genome in five O_2^- adapting populations (panels). Concentric circles: adaptation stages, from t_0 (outermost circle) to t_{15} (innermost circle). Brown: tRNA genes, orange: protein genes. **B)** mtDNA erosion cause loss of respiratory growth. Respiratory growth (doubling time, D_{norm}, in glycerol, left *y*-axis) as a function of generations of adaptation (*x*-axis) is shown in red in five O_2^- adapting populations (panels). Mean mtDNA coverage relative the euploid nuclear genome is shown in blue (right *y*-axis). Dotted line: interpolation between G_{70} and G_{242}. Shade: s.e.m.

Fig. S5 Chronic O_2^- distress causes aneuploidy

A) Chronic O_2^- distress causes chromosome II, III and V gains but not adaptive point mutations in specific genes. Number of sequenced (n_{tot} = 44, dotted line) O_2^- adapted endpoint populations (left *y*-axis) in which a genomic event occurred: Events classified were complete loss of entire or segments of mtDNA gene, chromosome amplifications and genes hit by point mutations. Bar color indicates variant type: sky blue = mtDNA loss, bright blue = chromosomal gain, black = frame shifting point mutation, orange = missense point mutation, light red = point mutation causing stop codon loss, red = point mutation causing stop codon gain, purple = point mutation upstream of ORF (<500bp), dark blue = synonymous point mutation, green = point mutation resulting in start codon lost. Right *y*-axis (gray line): mean allele frequencies. **B)** Point mutations do not coincide with early O_2^- adaptation in five (A7, A8, B5, B8, and B12) O_2^- adapting populations (panels). Frequencies (left *y*-axis) of SNPs and small INDELS as a function generations of adaptation (*x*-axis). Line color indicates variant type: orange = missense (orange), gray = intergenic, red = gain of stop codon, blue = synonymous, purple = upstream of ORF (<500bp). Variants pre-dating adaptation, supported by few (<10) reads or (<2) time points or shared across environments (>2) were filtered out. Right y-axis (green line): mean (*n*=6) adaptation. Shade: s.e.m. **C)** Gain of chromosome II and V but not III accelerates O_2^- growth. We

backcrossed (3x) clones carrying chromosome duplications to founder cells and isolated meiotic offspring w. and w/o each duplication. Mean (n=2 offspring, each at n=6 replicates) population size (y-axis, $\log_2$ scaled) as a function of time growing in excess O_2^- (h; x-axis). Panels and line color: chromosome duplication tested: chr II (red), III (green) or V (blue). Shade: s.e.m. **D)** Chronic O_2^- distress cause chromosome V duplication only under loss of the entire mtDNA (ρ^0). Number (n_{tot} =44) of O_2^- adapted endpoint populations (y-axis) carrying a beneficial chromosomal amplification (x-axis). Red bar: entire mtDNA lost (ρ^0), blue bar = segments of lost (ρ^-). Empty bars: ρ^0 and ρ^- populations w/o chromosome duplication. p-value: Fisher's exact test.

Fig. S6 Chronic O_2^- distress causes complete mtDNA loss

A) Loss of mtDNA segments (ρ^-) is more favorable than loss of entire mtDNA (ρ^0) under chronic O_2^- distress. Original experiment (left panel) was reproduced (right panel). Mean (n_{tot}=43, each measured at n=6) O_2^- adaptation (A) as a function of generations of adaptation, for populations with loss of mtDNA segments (green line, ρ^-, *n=18*) or the entire mtDNA (red line, ρ^0, *n=25*). Shade: s.e.m. B) Alignment between long (nanopore sequence) reads of two evolved clones (A7_2, D3_1) and the YPS128 reference mitochondrial genome. Genomic coordinates (top) are based on the YPS128 reference mitochondrial genome. Upper track: the YPS128 reference mitochondrial genome annotation. Protein-coding genes, tRNA, and other mitochondrial RNAs are colored in orange, blue, and purple respectively. Middle track: the relative read mapping coverage for the evolved clone A7_2. Lower track: the relative read mapping coverage for the evolved clone D3_1. Mismatches between the reads and the reference genome are denoted in colored vertical bars (middle, lower track). The step-like distribution of the mapping coverage together with the strong mismatch signal at the step boundaries (middle, lower tracks) may imply co-existence of both circular and linear mtDNA arrangements.

Visual abstract: Superoxide induces adaptive epigenetic editing of mitochondrial DNA

Cells edit the mtDNA copy number in response to O_2^- distress. The editing is epigenetically inherited but becomes genetically fixed when chronic O_2^- distress drives mtDNA erosion below a replenishment threshold (dotted line), leading to catastrophic, irreversible loss of respiration.

Figure 1

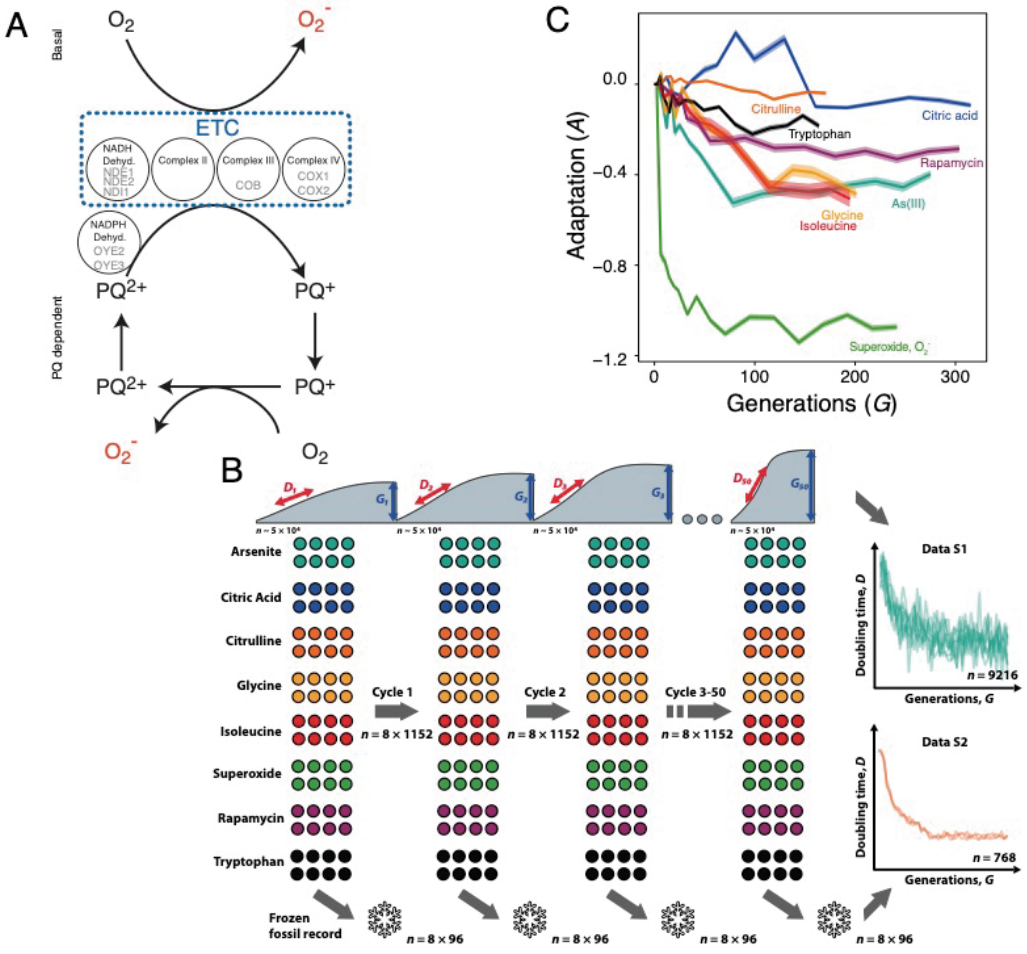

A
Basal
PQ dependent
ETC
NADH Dehyd.
NDE1
NDE2
NDI1
Complex II
Complex III
COB
Complex IV
COX1
COX2
NADPH Dehyd.
OYE2
OYE3
C
Adaptation (A)
Generations (G)
Citrulline
Tryptophan
Citric acid
Rapamycin
As(III)
Glycine
Isoleucine
Superoxide, O2-
B
Arsenite
Citric Acid
Citrulline
Glycine
Isoleucine
Superoxide
Rapamycin
Tryptophan
Cycle 1
Cycle 2
Cycle 3-50
n = 8 × 1152
Frozen fossil record
n = 8 × 96
Data S1
Doubling time, D
n = 9216
Generations, G
Data S2
n = 768

Figure 2

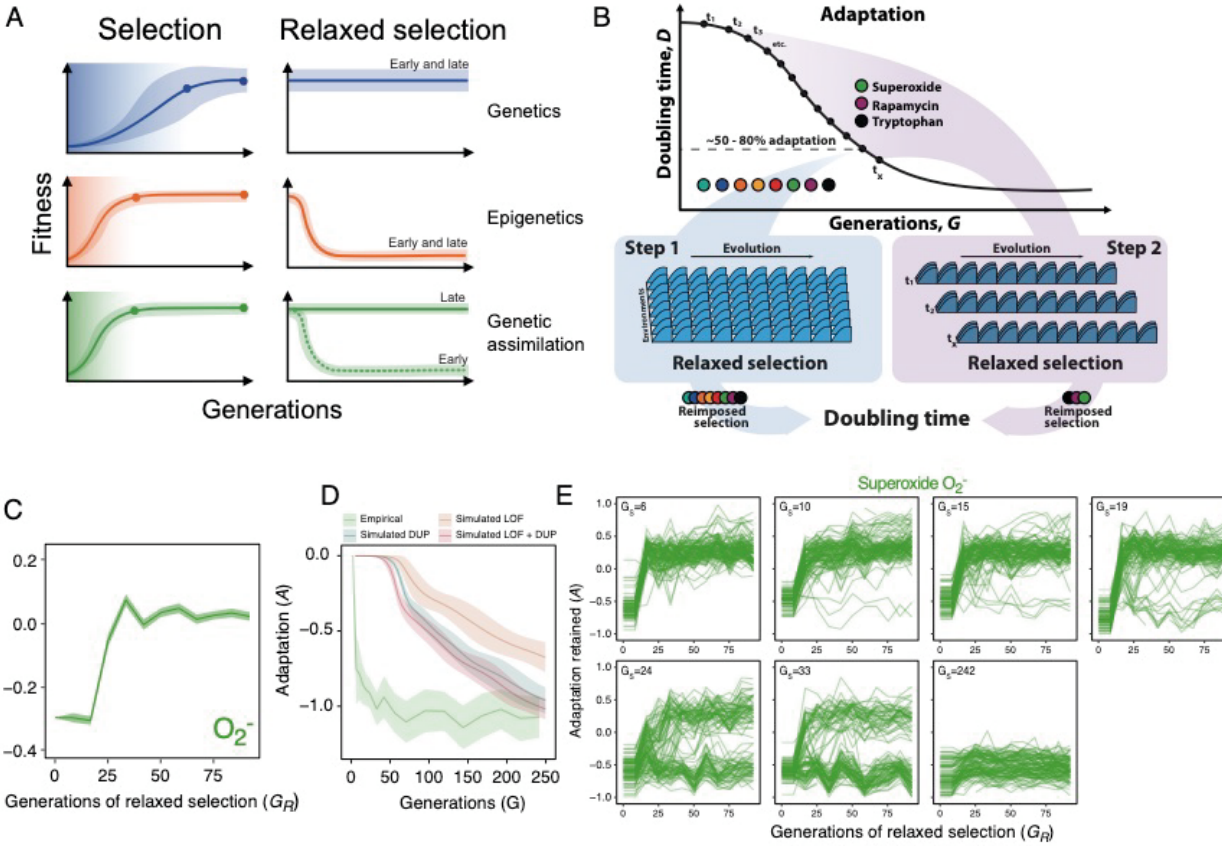

A
Selection
Relaxed selection
Early and late
Genetics
Epigenetics
Early and late
Late
Early
Genetic assimilation
Fitness
Generations
B
Adaptation
Doubling time, D
Superoxide
Rapamycin
Tryptophan
~50 - 80% adaptation
Generations, G
Step 1
Evolution
Relaxed selection
Step 2
Evolution
Relaxed selection
Reimposed selection
Doubling time
Reimposed selection
C
O2-
Generations of relaxed selection (GR)
D
Empirical
Simulated LOF
Simulated DUP
Simulated LOF + DUP
Adaptation (A)
Generations (G)
E
Superoxide O2-
Adaptation retained (A)
Generations of relaxed selection (GR)

Figure 3

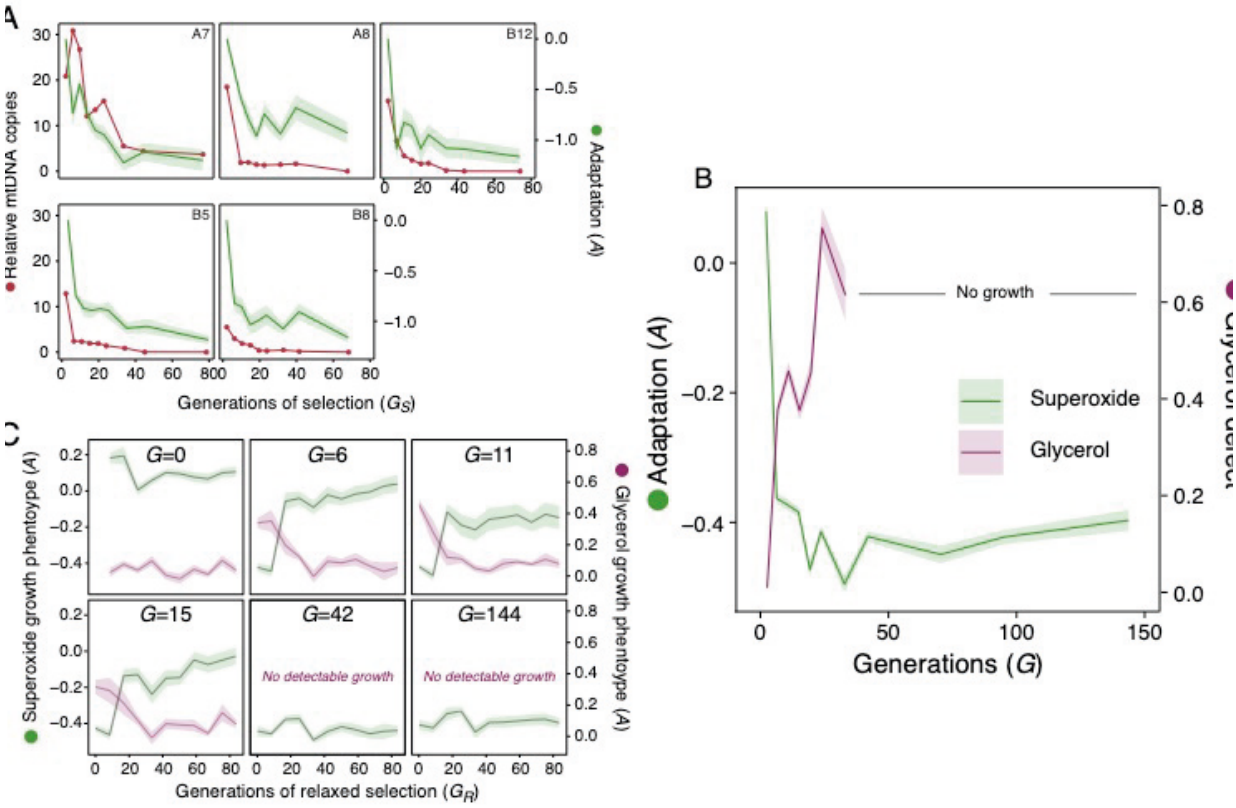

A
A7
A8
B12
B5
B8
Relative mtDNA copies
Adaptation (A)
Generations of selection (G_S)
G=0
G=6
G=11
G=15
G=42
G=144
No detectable growth
No detectable growth
Superoxide growth phentoype (A)
Glycerol growth phentoype (A)
Generations of relaxed selection (G_R)
B
No growth
Superoxide
Glycerol
Adaptation (A)
Generations (G)

Figure 4

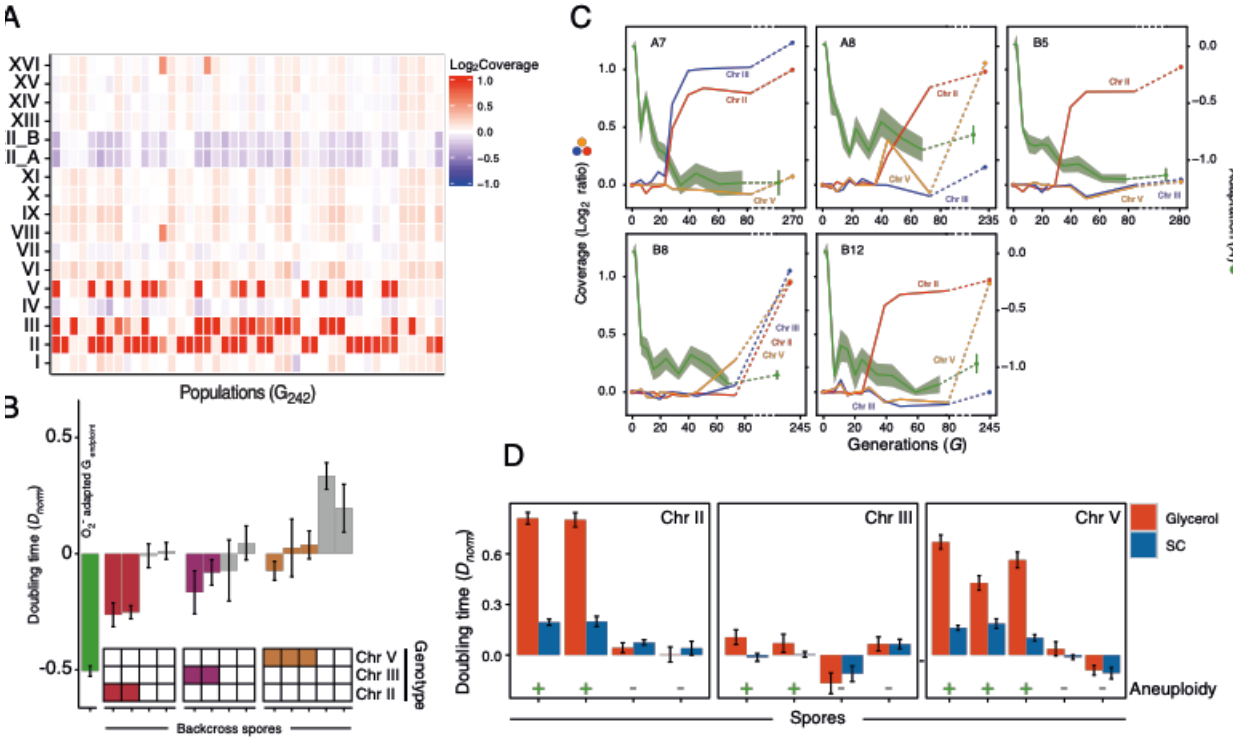

A
XVI
XV
XIV
XIII
II_B
II_A
XI
X
IX
VIII
VII
VI
V
IV
III
II
I
Log2Coverage
1.0
0.5
0.0
−0.5
−1.0
Populations (G242)
B
Doubling time (Dnorm)
0.5
0
-0.5
O2− adapted Gendpoint
Chr V
Chr III
Chr II
Genotype
Backcross spores
C
A7
A8
B5
B8
B12
Chr II
Chr V
Chr III
Coverage (Log2 ratio)
0 20 40 60 80 270
0 20 40 60 80 235
0 20 40 60 80 280
0 20 40 60 80 245
Generations (G)
D
Chr II
Chr III
Chr V
Doubling time (Dnorm)
0.6
0.3
0.0
Glycerol
SC
Aneuploidy
Spores

Figure 5

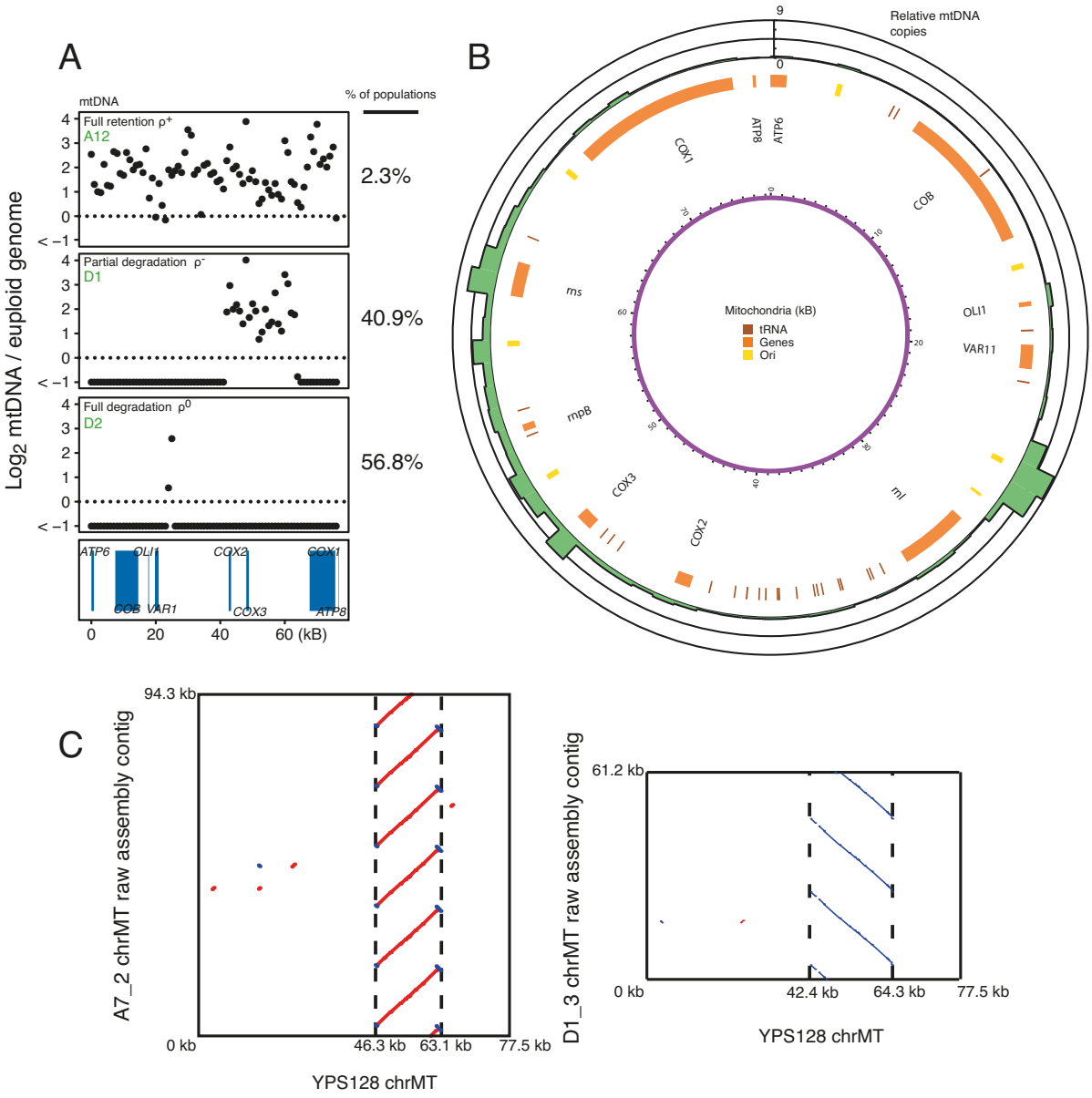
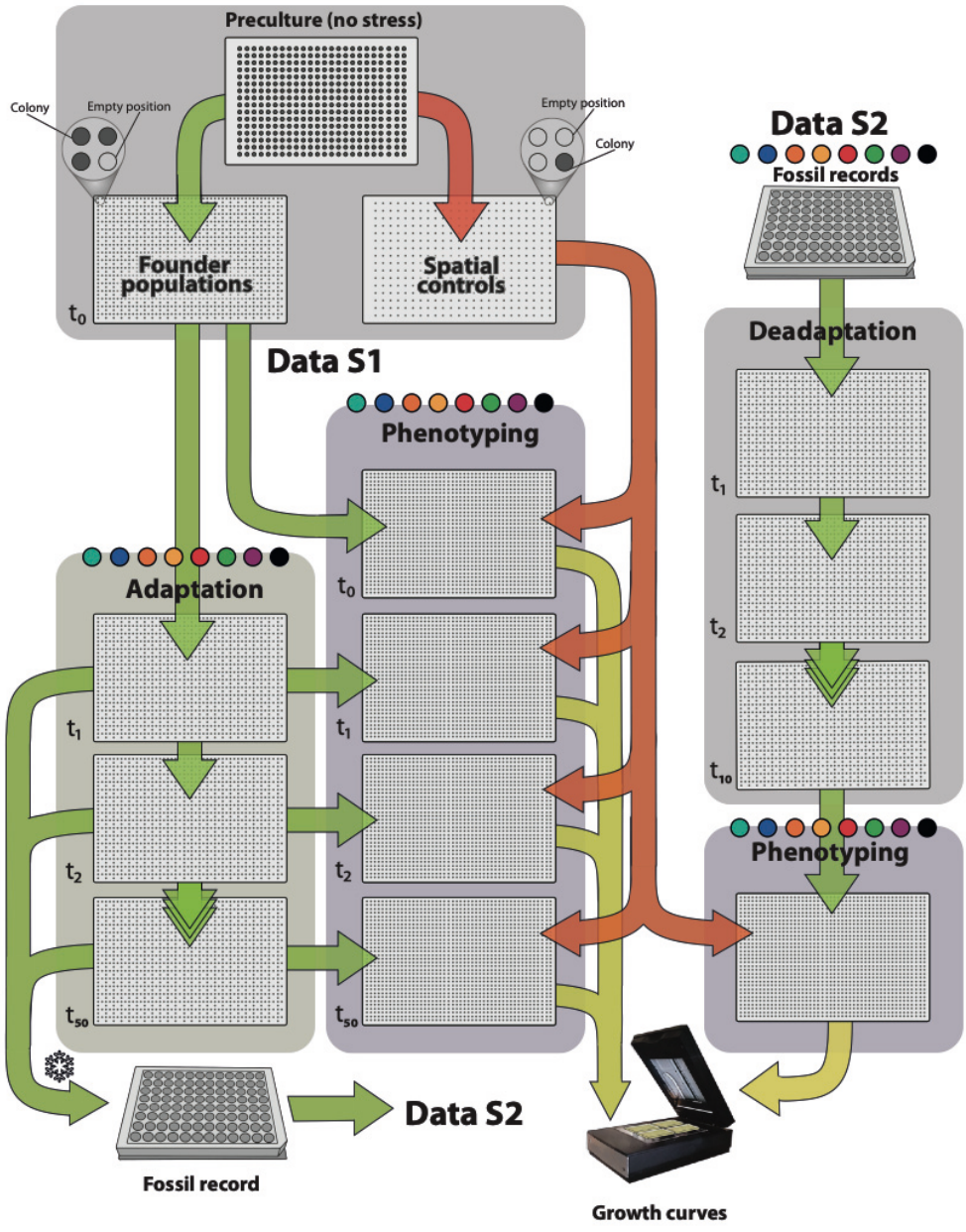

A
mtDNA
% of populations
Full retention ρ+
A12
2.3%
Partial degradation ρ-
D1
40.9%
Full degradation ρ0
D2
56.8%
Log2 mtDNA / euploid genome
ATP6
OLI1
COX2
COX1
COB
VAR1
COX3
ATP8
0
20
40
60 (kB)
B
Relative mtDNA copies
Mitochondria (kB)
tRNA
Genes
Ori
ATP6
ATP8
COX1
COB
OLI1
VAR11
rnl
COX2
COX3
rnpB
rns
C
94.3 kb
A7_2 chrMT raw assembly contig
0 kb
46.3 kb
63.1 kb
77.5 kb
YPS128 chrMT
61.2 kb
D1_3 chrMT raw assembly contig
0 kb
42.4 kb
64.3 kb
77.5 kb
YPS128 chrMT

Figure S1


Preculture (no stress)
Colony
Empty position
Empty position
Colony
Founder populations
Spatial controls
t_0
Data S1
Phenotyping
Adaptation
t_0
t_1
t_2
t_{50}
t_1
t_2
t_{50}
Data S2
Fossil records
Deadaptation
t_1
t_2
t_{10}
Phenotyping
Fossil record
Data S2
Growth curves

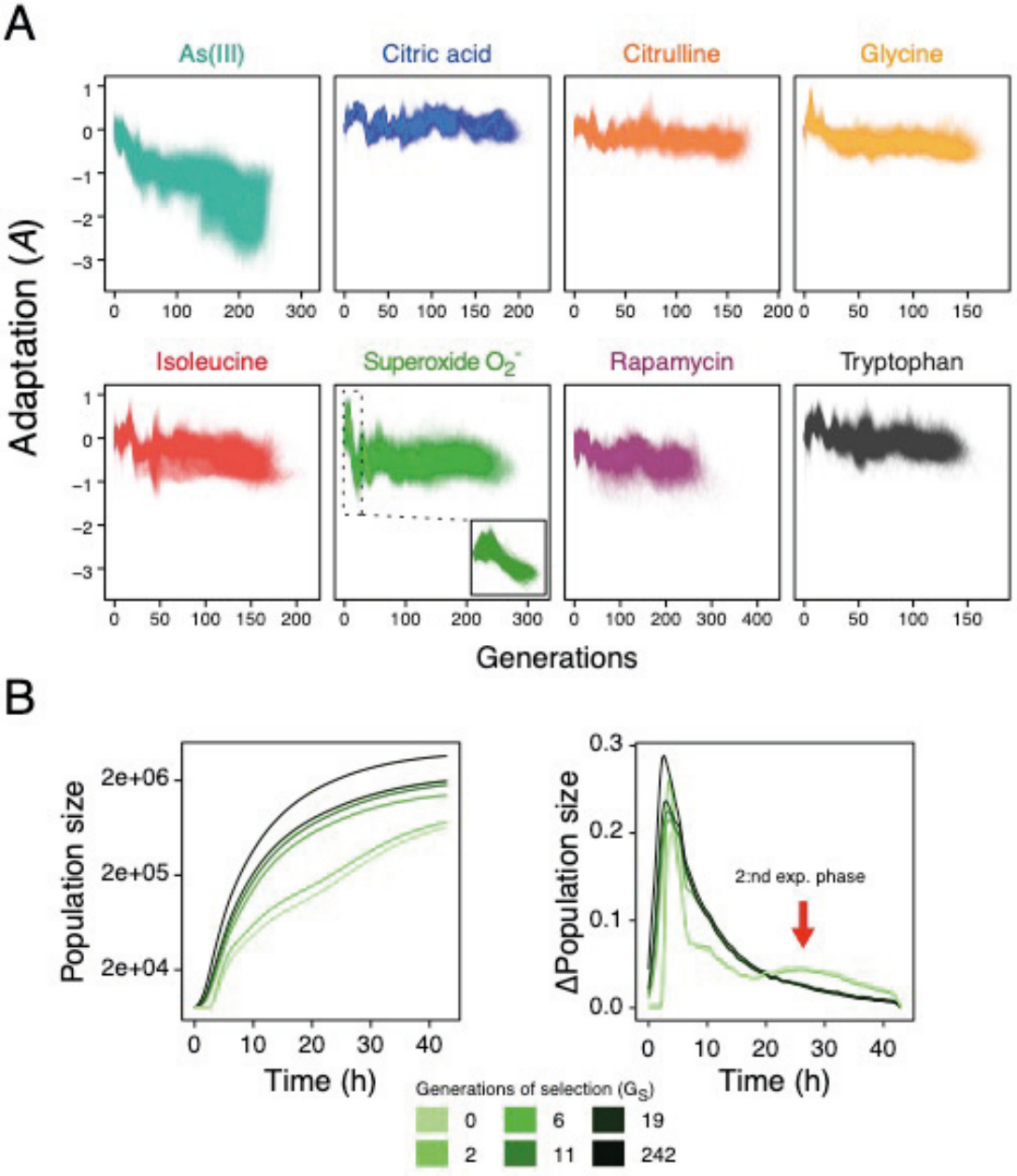

A
As(III)
Citric acid
Citrulline
Glycine
Isoleucine
Superoxide O_2^-
Rapamycin
Tryptophan
Adaptation (A)
Generations
B
Population size
2e+06
2e+05
2e+04
Time (h)
ΔPopulation size
2:nd exp. phase
Time (h)
Generations of selection (G_S)
0
2
6
11
19
242

Figure S2

Figure S3

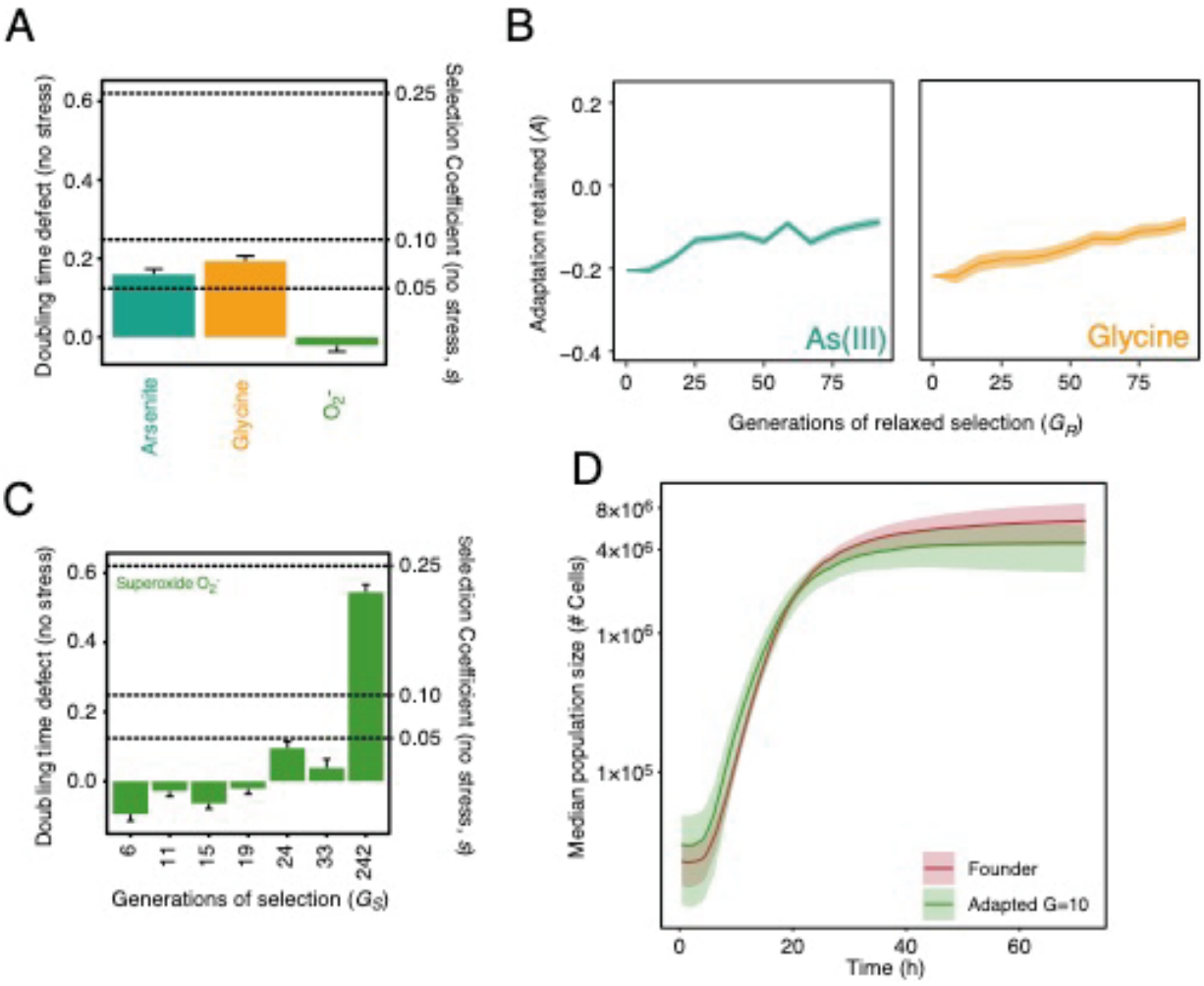

A
Doubling time defect (no stress)
0.6
0.4
0.2
0.0
0.25
0.10
0.05
Selection Coefficient (no stress, s)
Arsenite
Glycine
O_2^-
B
Adaptation retained (A)
0.2
0.0
−0.2
−0.4
As(III)
Glycine
0
25
50
75
Generations of relaxed selection (G_R)
C
Doubling time defect (no stress)
Superoxide O_2^-
Selection Coefficient (no stress, s)
6
11
15
19
24
33
242
Generations of selection (G_S)
D
Median population size (# Cells)
8x10^6
4x10^6
1x10^6
1x10^5
Founder
Adapted G=10
0
20
40
60
Time (h)

Figure S4

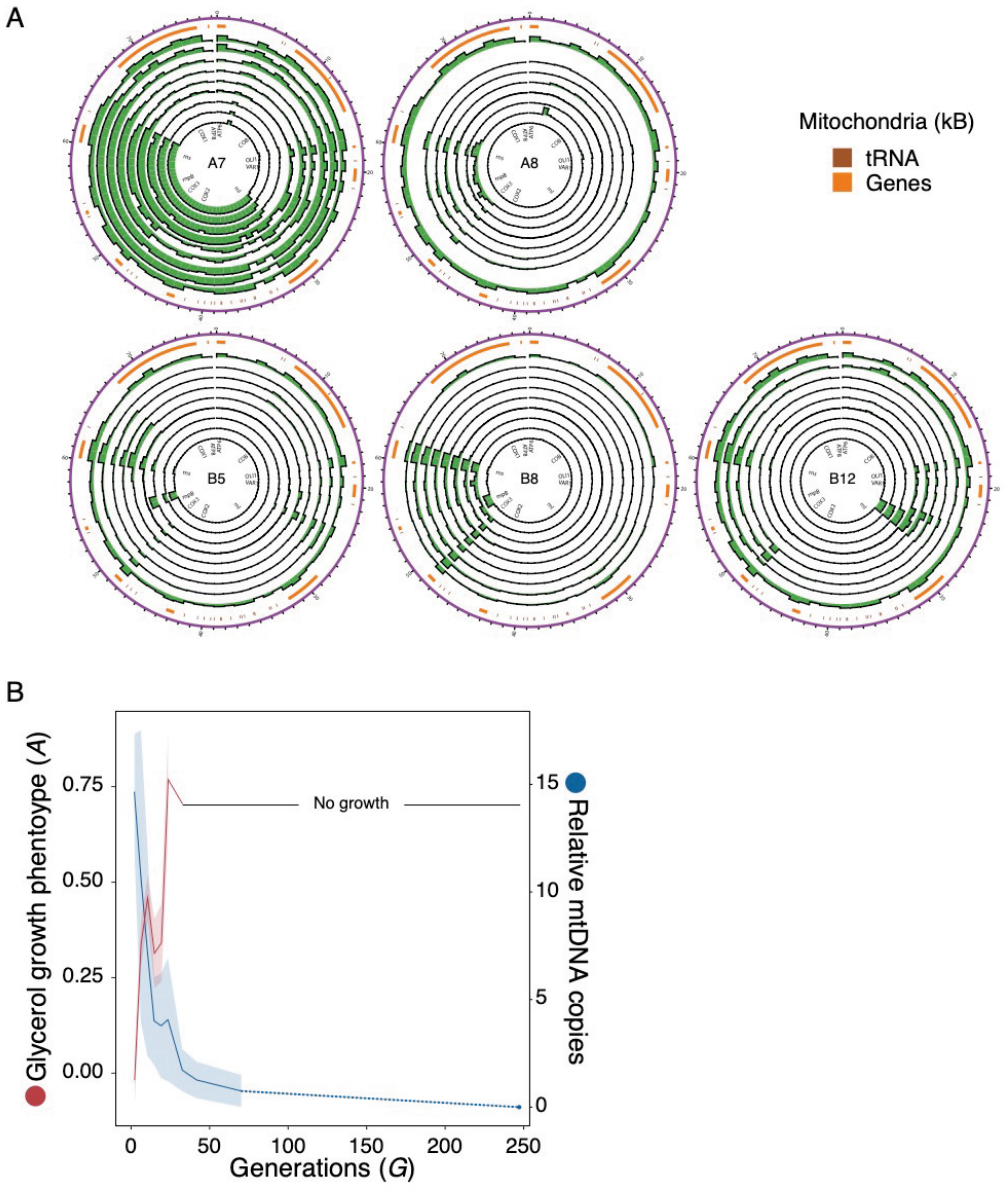

A
A7
A8
B5
B8
B12
Mitochondria (kB)
tRNA
Genes
B
Glycerol growth phentoype (A)
0.75
0.50
0.25
0.00
No growth
15
10
5
0
Relative mtDNA copies
0
50
100
150
200
250
Generations (G)

Figure S5

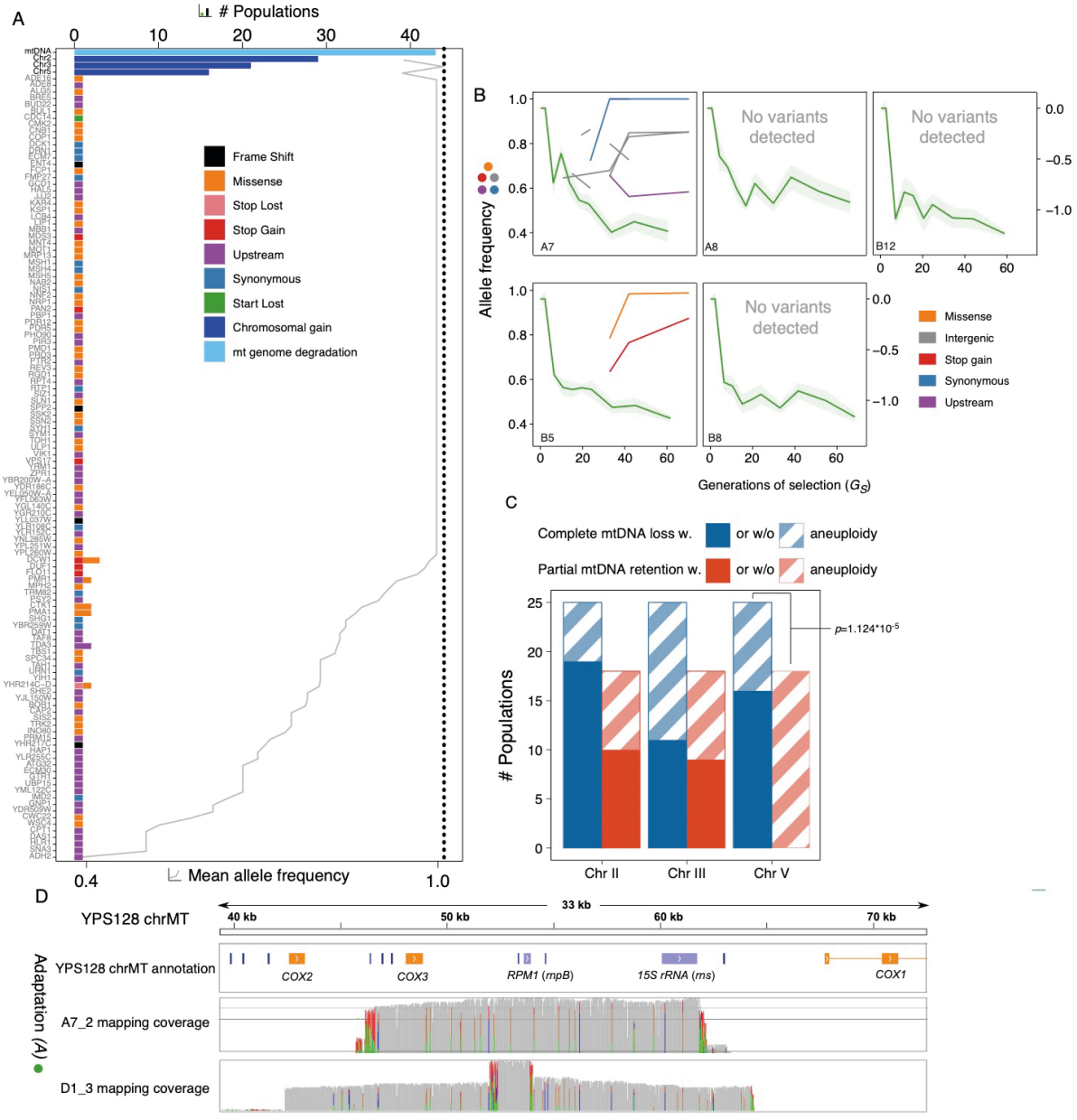
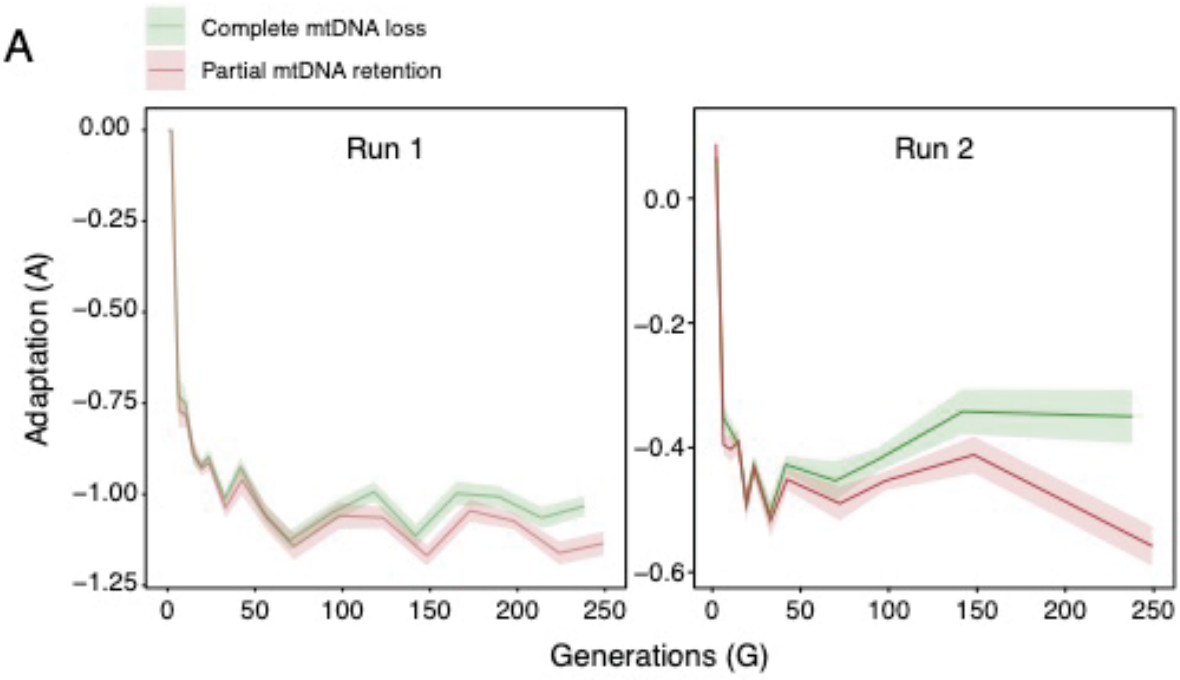

A
Populations
0
10
20
30
40
Frame Shift
Missense
Stop Lost
Stop Gain
Upstream
Synonymous
Start Lost
Chromosomal gain
mt genome degradation
0.4
Mean allele frequency
1.0
B
Allele frequency
No variants detected
A7
A8
B12
B5
B8
Missense
Intergenic
Stop gain
Synonymous
Upstream
Generations of selection (G_S)
C
Complete mtDNA loss w. or w/o aneuploidy
Partial mtDNA retention w. or w/o aneuploidy
Populations
p=1.124*10^{-5}
Chr II
Chr III
Chr V
D
YPS128 chrMT
33 kb
40 kb
50 kb
60 kb
70 kb
YPS128 chrMT annotation
COX2
COX3
RPM1 (mpB)
15S rRNA (rns)
COX1
A7_2 mapping coverage
D1_3 mapping coverage
Adaptation (A)

Figure S6


A
Complete mtDNA loss
Partial mtDNA retention
Run 1
Run 2
Adaptation (A)
Generations (G)
0.00
−0.25
−0.50
−0.75
−1.00
−1.25
0.0
−0.2
−0.4
−0.6
0
50
100
150
200
250

ISBN: 978-82-575-1586-7
ISSN: 1894-6402

Norwegian University
of Life Sciences

Postboks 5003
NO-1432 Ås, Norway
+47 67 23 00 00
www.nmbu.no