The path towards prediction

An exploration of the evolutionary dynamics of the yeast polarity network

Kingma, E.

DOI
10.4233/uuid:07f4234d-4772-4c57-a59c-ac7068ed2f37

Publication date
2023

Document Version
Final published version

Citation (APA)
Kingma, E. (2023). *The path towards prediction: An exploration of the evolutionary dynamics of the yeast polarity network*. [Dissertation (TU Delft), Delft University of Technology].
https://doi.org/10.4233/uuid:07f4234d-4772-4c57-a59c-ac7068ed2f37

The path towards prediction

An exploration of the evolutionary dynamics of the yeast polarity network

The path towards prediction

An exploration of the evolutionary dynamics of the yeast polarity network

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus, Prof.dr.ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates
to be defended publicly on
Wednesday 8 November 2023 at 15:00 o'clock

by

Enzo Kingma

Master of Science in Applied Physics,
Delft University of Technology, the Netherlands,
born in Willemstad, Curaçao.

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus,	chairman
Dr. ir. L. Laan,	Technische Universiteit Delft, *promotor*
Dr. T. Idema,	Technische Universiteit Delft, *promotor*
Independent members:	
Prof. dr. A.M. Dogterom	Technische Universiteit Delft
Prof. dr. B. Kornmann	University of Oxford, England
Dr. S. Kryazhimskiy	University of California San Diego, United States
Dr. D. McCusker	Institut européen de chimie et biologie, France
Dr. M. Bauer	Technische Universiteit Delft
Prof. dr. G. Koenderink	Technische Universiteit Delft, *reserve member*

Keywords: Evolution, cell polarity, yeast, transposon mutagenesis, biological networks, pleiotropy, epistasis

Printed by: Gildeprint

Front & Back: "*Puffins on a boat adventure next to the cliffs of Moher*" in the style of Caravaggio. Conceived by Marieke Glazenburg.

ISBN 000-00-0000-000-0

An electronic version of this dissertation is available at http://repository.tudelft.nl/.

Contents

Summary

All living systems possess the ability to evolve. This ability has allowed life to adopt a large diversity of shapes, colors, sizes and lifestyles. However, despite being a fundamental property of life, our knowledge of what makes living systems evolvable is limited. The fact that evolution can be sometimes be hopelessly unpredictable while in other cases it follows only a small number of predictable paths has puzzled evolutionary biologists for decades. An important step towards resolving this problem has been the discovery that non-additive effects between mutations, a phenomenon coined epistasis, can act as a source of evolutionary constraint. Importantly, the limited number of accessible adaptive pathways in conditions where constraints are pervasive provides opportunities for prediction. In this thesis, we explore how these constraints can arise from the biochemical interactions within the cell. Aside from answering fundamental questions in biology, the ability to predict evolution will open up unprecedented possibilities, ranging from the engineering of living systems to the development of novel treatments for disease.

We use polarity establishment in the budding yeast *Saccharomyces cerevisiae* as a model system to study evolutionary dynamics. Budding yeast proliferates exclusively through asymmetric cell divisions and establishing an axis of polarity is therefore a crucial part of the cell cycle. Perturbing the machinery responsible for polarity establishment by deleting an important polarity gene (*BEM1*) strongly diminishes the chances of survival. Interestingly, it has been shown these defects can be largely compensated during evolution by inactivating other genes (*BEM3* and *NRP1*) in a conserved order. Furthermore, this evolutionary trajectory occurs repeatably in during experiments, which suggests that it could have been predicted. Here, we present our progress towards understanding how the accessibility of this pathway is shaped by the molecular interactions in the cell.

In **chapter 2**, we review examples of studies that experimentally show how epistatic interactions can arise from underlying molecular interactions, with a focus on regulatory networks. We discuss these examples at three levels of biological organization: (1) within regulatory interactions, (2) within regulatory networks and (3) between regulatory networks. Comparing these different studies reveals that the necessary information to make predictions varies depending on the scale of the system.

In addition to genetic architecture, the selective environment plays a crucial role during adaptation. Importantly, cells must adequately respond to unpredictable environmental shifts to increase their chances of survival. Responsiveness to environmental change requires integrating signaling pathways with other cellular functions. This can lead to pleiotropy, a phenomenon where a single gene influences multiple traits. In **Chapter 3**, we study how the adaptive response to a pleiotropic mutation is affected by environmental complexity. We find that regardless of whether the environment is static or dynamic, the responsiveness of cell polarity to environmental signals can recover during adaptation. This is surprising, since cells adapted in a static environment do not experience direct selective pressures to improve this trait. We propose that pleiotropic interactions may be an important structural component of cellular networks that alters the phenotypic response to selection pressures, leading to unexpected results.

The relation between genotype and fitness is often presented in the form of fitness landscapes. Their ability to capture the number of accessible evolutionary pathways leading to higher fitness makes them of central interest for predictive models of evolution. Unfortunately, experimentally constructing large and unbiased fitness landscapes remains challenging. In **chapter 4**, we introduce a method to quantify the fitness of gene disruption mutants using SATAY, a transposon mutagenesis screen that was recently developed for *S. cerevisiae*. We show that the averaging over the sequencing read counts of different transposon insertion sites provides a proxy for fitness that is robust across replicate SATAY experiments. However, our fitness estimates differ substantially from those obtained using other experimental procedures. We discuss several possible explanations for these inconsistencies.

In **chapter 5**, we go one step beyond prediction and explore the possibilities to exercise control over the accessibility of evolutionary pathways. To do so, we make use of the high level of reproducibility of the adaptive pathway that is taken after the deletion of the polarity protein Bem1. We investigate the possibilities to control the accessibility of this trajectory through the rational design of epistatic interactions. Our results indicate that the gene encoding for an actin-binding protein (*ABP1*) has the desired epistatic interaction pattern. We hypothesize how the ability of *ABP1* to act as a control knob for pathway accessibility might have been predicted from its functional relation to cell polarity.

During evolution, the deleterious effects of mutations at one locus can often be compensated by additional mutations elsewhere in the genome. It has previously been shown that defects in cell polarity caused by the loss of Bem1 can be compensated through the additional deletion of Bem3 and Nrp1. The fact that functionality can be restored through additional gene deletions suggests that the protein interaction network is rewired such that an alternative pathway for polarity establishment is formed. In **chapter 6**, we investigate how compensatory evolution of the polarity module affects the molecular interactions in the cell. Our results show that the three gene deletions have genome-wide consequences for the tolerance of genes to transposon disruptions. These changes extend beyond the originally perturbed pathway and impact multiple, seemingly unrelated, cellular processes. We discuss the cellular components that are likely involved in the alternative pathway for polarity establishment.

Taken together, our work demonstrates the complexity of evolution: the effects of mutations resonate throughout the interaction network, affecting multiple cellular functions at once. Moving towards evolutionary predictions will therefore likely involve taking a genome-wide approach to understand how molecular interactions in the cell promote or inhibit adaptation. The results presented here encourage further investigation into the mechanisms by which gene loss can lead to functional changes of other genes.

Samenvatting

Alle levende systemen beschikken over het vermogen om te evolueren. Deze eigenschap heeft het leven in staat gesteld om een grote diversiteit aan vormen, kleuren, groottes en levensstijlen aan te nemen. Hoewel evolutie een fundamentele eigenschap van het leven is, is onze kennis van wat levende systemen evolueerbaar maakt beperkt. Het feit dat evolutie soms hopeloos onvoorspelbaar kan zijn, terwijl het in andere gevallen slechts een klein aantal voorspelbare paden volgt, heeft evolutionaire biologen de afgelopen decenia voor een raadsel gesteld. Een belangrijke stap naar het oplossen van deze puzzel is de ontdekking dat niet-additieve interacties tussen mutaties, een fenomeen dat epistase wordt genoemd, de mogelijkheden voor evolutie kan beperken. Vooral condities waarin er door deze beperkingen maar een klein aantal toegankelijke evolutionaire paden zijn bieden mogelijkheden voor het doen van voorspellingen. In dit proefschrift onderzoeken we hoe deze beperkingen kunnen ontstaan uit de biochemische interacties binnen de cel. Naast het beantwoorden van fundamentele vragen in de biologie zal het vermogen om evolutie te voorspellen ongekende mogelijkheden openen, variërend van de ontwikkeling van levende systemen tot de creatie van nieuwe behandelingen voor ziekten.

We gebruiken cel polarisatie in de gist *Saccharomyces cerevisiae* als een modelsysteem om evolutionaire dynamiek te onderzoeken. Gistcellen vermenigvuldigen zich uitsluitend door middel van asymmetrische celdelingen en daardoor is het vormen van een polariteitsas een cruciaal onderdeel van de celcyclus. Het verstoren van het mechanisme voor polarisatie door een belangrijk gen (*BEM1*) te verwijderen vermindert sterk de overlevingskansen. Wat interessant is, is dat deze verstoring tijdens de evolutie grotendeels gecompenseerd kan worden door andere genen (*BEM3* en *NRP1*) in een vaste volgorde te inactiveren. Daarnaast is dit evolutionary traject zeer reproduceerbaar tijdens experimenten, wat suggereert dat het voorspeld had kunnen worden. Hier presenteren we onze voortgang in het begrijpen van hoe de toegankelijkheid van dit pad wordt beïnvloed door de moleculaire interacties in de cel.

In **hoofdstuk 2** bespreken we voorbeelden van studies die experimenteel aantonen hoe epistatische interacties kunnen ontstaan vanuit moleculaire interacties, met een focus op regulerende netwerken. We bespreken drie verschillende niveaus van biologische organisatie: (1) binnen regulerende interacties, (2) binnen regulerende netwerken en (3) tussen regulerende netwerken. Uit het vergelijken van deze verschillende studies blijkt dat de benodigde informatie om evolutionaire voorspellingen te doen afhangt van het niveau waarop gekeken wordt.

Naast de genetische opmaak speelt ook de omgeving een belangrijke rol tijdens de evolutie. Cellen moeten namelijk adequaat reageren op onvoorspelbare veranderingen in hun omgeving om hun overlevingskansen te vergroten. Reactivermogen op een veranderende omgeving vereist dat signaalroutes met andere celprocessen worden geïntegreerd. Dit kan leiden tot pleiotropie, een fenomeen waarbij een enkel gen invloed heeft op meerdere eigenschappen van de cel. In **Hoofdstuk 3** onderzoeken we hoe de adaptieve respons op een pleiotropische mutatie wordt beïnvloed door de complexiteit van de omgeving. We constateren dat, ongeacht of de omgeving statisch of dynamisch is, de responsiviteit van de celmorfologie op omgevingssignalen tijdens adaptatie kan herstellen. Dit is verrassend, aangezien cellen die zijn aangepast aan

een statische omgeving geen directe selectiedruk ervaren om deze eigenschap te verbeteren. We stellen voor dat pleiotropische interacties een belangrijk structureel onderdeel van cellulaire netwerken kunnen zijn dat de fenotypische respons op selectiedruk kan veranderen, wat leidt tot onverwachte resultaten.

De relatie tussen genotype en fitheid wordt vaak gepresenteerd in de vorm van fitnesslandschappen. Hun vermogen om het aantal toegankelijke evolutionaire paden naar een hogere fitness vast te leggen, maakt ze van centraal belang voor voorspellende modellen van evolutie. Helaas blijft het experimenteel construeren van grote en onvertekende fitnesslandschappen een uitdaging. In **hoofdstuk 4** introduceren we een methode om de fitheid van mutanten met een verstoord gen te kwantificeren met behulp van SATAY, een transposon-mutagenese screening die recentelijk is ontwikkeld voor *S. cerevisiae*. We laten zien dat het gemiddelde aantal verkregen DNA-sequenties per transposon insertie een fitheidswaarde geeft die consistent is tussen herhaalde SATAY experimenten. We bespreken een aantal mogelijke verklaringen voor deze verschillen.

In **hoofdstuk 5** gaan we een stap verder dan voorspelling en onderzoeken we de mogelijkheden om controle uit te oefenen over de toegankelijkheid van evolutionaire paden. Om dit te doen, maken we gebruik van de hoge mate van reproduceerbaarheid van het adaptieve pad dat wordt gevolgd na het verwijderen van het polariteitseiwit Bem1. We bekijken de mogelijkheden om de toegankelijkheid van dit traject te controleren door het rationeel ontwerpen van epistatische interacties. Onze resultaten geven aan dat het gen dat codeert voor een actine-bindend eiwit (*ABP1*) het gewenste epistatische interactiepatroon vertoont. We stellen een hypothese op over hoe de mogelijkheid van *ABP1* om als controleknop te fungeren voor de toegankelijkheid van het evolutionaire pad voorspeld had kunnen worden op basis van zijn functionele relatie met de celpolariteit.

Tijdens de evolutie kunnen de schadelijke effecten van mutaties op één locatie vaak gecompenseerd worden door aanvullende mutaties elders in het genoom. Eerder is aangetoond dat defecten in celpolarisatie die zijn veroorzaakt door het verlies van Bem1 gecompenseerd kunnen worden door Bem3 en Nrp1 uit te schakelen. Het feit dat deze functie hersteld kan worden door extra genen uit te schakelen suggereert dat de eiwitinteracties in de cel zich herstructureren zodat er een alternatief mechanisme voor celpolarisatie ontstaat. In **hoofdstuk 6** bekijken we hoe compensatie van defecten in celpolarisatie de moleculaire interacties in de cel beïnvloedt. Onze resultaten laten zien dat het uitschakelen van de drie genen consequenties heeft voor de tolerantie voor transposon-interrupties van een groot aantal genen in het genoom. Deze consequenties beperken zich niet alleen tot genen van de initieel verstoorde functie, maar raken ook genen uit andere, ogenschijnlijk niet-gerelateerde, cellulaire processen.

Tot slot toont ons werk de complexiteit van evolutie: de effecten van mutaties resoneren door het hele interactienetwerk, waarbij tegelijkertijd meerdere cellulaire functies worden beïnvloed. Het streven naar het doen van evolutionaire voorspellingen zal waarschijnlijk een genoomwijde benadering vereisen om te begrijpen hoe moleculaire interacties in de cel de aanpassing bevorderen of belemmeren. De hier gepresenteerde resultaten zijn een aanmoediging voor verder onderzoek aan naar de mechanismen waardoor genverlies kan leiden tot functionele veranderingen van andere genen.

1

I think nature's imagination is so much greater than man's,
she's never going to let us relax

— Richard P. Feynman

1

Life is perplexingly diverse. Looking only at the phylogenetic tree of fungi, we can already see a large variety of shapes, colors, sizes and lifestyles (Figure 1.1). Amazingly, this diversity has originated from a single ancestor through a process of random mutation and selection. This evolutionary process has allowed living systems to change their characteristics in such a way that they better fit their environment. But how do you change something as complex as a living cell, without breaking it? Somehow, living systems are able to walk the fine line between being hopelessly unstable at short timescales and too robust to evolve on long timescales. The ability to evolve is undoubtedly rooted in the set of physical interactions that define a living cell. However, we are still profoundly unaware as to how the design of these interaction networks allows them to reconcile robustness with evolvability. In this thesis, we link the evolutionary dynamics of a cellular function to the molecular mechanism that generates this function. Here, we give an introduction of the known mechanisms that affect evolution and how they relate to intracellular organization.

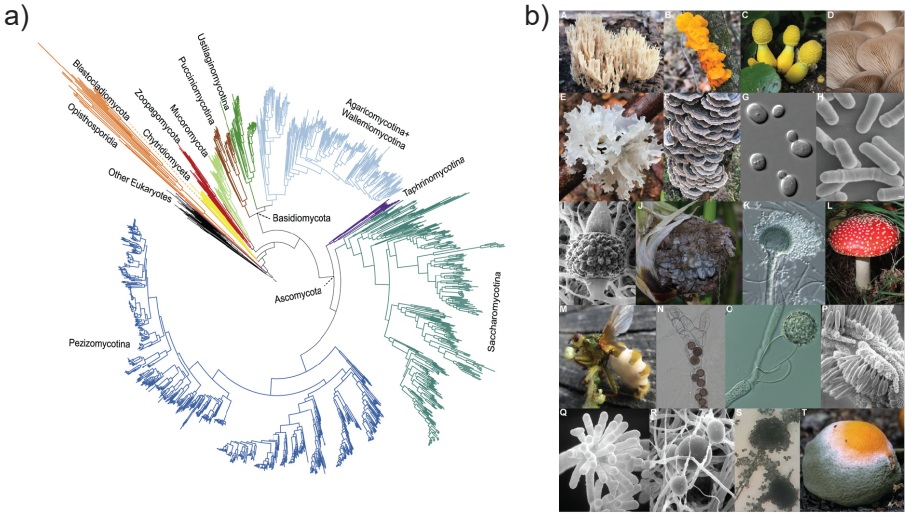


Figure 1.1. Diversity in the fungal tree of life.(**a**) Phylogenetic tree of the fungal kingdom. (**b**) Representative species from several major fungal lineages. (A) Crown coral. (B) Witch's butter. (C) Flowerpot parasol. (D) Pearl oyster mushroom. (E) Snow fungus. (F) Turkey tail. (G) Baker's yeast. (H) Fission yeast. (I) *Mucor mucedo*. (J) Corn smut. (K) *Aspergillus oerlinghausenensis*. (L) Fly agaric *Amanita muscaria*. (M) *Entomophthora muscae*. (N) *Rozella allomycis* parasitizing the chytrid *Allomyces*. (O) *Monoblepharis macrandra*. (P) *Coemansia braziliensis*. (Q) *Piptocephalis repens*. (R) *Mortierella elongata*. (S) *Rhizopus spp.*. (T) *Penicillium digitatum*. Images reprinted from Li et al. (2021) with permission conveyed through Copyright Clearance Cente

1.1. Evolution by natural selection

To understand the factors that facilitate and constrain evolution, we must know the mechanism that drives evolutionary change. Arguably the most important mechanism through which populations can evolve to better fit their environment (adaptation) is *natural selection* (Darwin 1809-1882, 1859). The theory of natural selection was developed in the 19th century during the famous expedition of Charles Darwin. While controversial at first, the theory eventually became so well accepted that natural selection and evolution started to be used interchangeably (Fisher, 1930). However, it is important to realize that not all changes in a trait during evolution are the result

of natural selection. Several other mechanisms of evolution have since then been identified, and importantly, these do not always result in adaptation. For a trait to evolve through natural selection, the following four conditions must be met (Godfrey-Smith, 2007):

1. There must be *variation*. Within a population, different versions of a trait must exist for the trait to evolve through natural selection. These variations typically arise from random genetic mutations.

2. Variations in a trait must lead to differences in the *fitness* of individuals. Consequently, certain individuals within the population are more successful in reproduction than others.

3. The trait must be *heritable*. There needs to be a correlation between the variant of the trait in a parent and that in the offspring it produces.

4. There must be *competition*. Individuals within the population must compete for limited resources.

The combination of these four conditions will cause the individuals with the highest reproductive success (fitness) to eventually take over the population. Thus, through natural selection a population will evolve to better fit its environment. This process requires no rational intervention or tinkering: genetic variants that have higher fitness are simply generated by random genetic mutations and pass on their characteristics to the next generation in the form of genes. However, despite the random origin of mutations, what variants are possible and how they can increase fitness is subject to constraints. The factors that act as a source for these evolutionary constraints are discussed in the next sections.

1.2. Genetic constraints

Evolving organisms are bound by constraints. If we would examine the evolutionary history of an organism, we would see that only a subset of the countless forms we might envision are realized. Thus, while often inventive, the possibilities for evolution are not limitless. Constraints play a pivotal role in our ability to predict evolution: the less options there are, the more certain we can be about the course that evolution will follow.

Some constraints exist simply because living systems must obey the laws of physics (Kempes et al., 2019). For instance, arthropods cannot exceed a certain maximum body size because they would not be able to support the excessive weight of their exoskeleton. A second form of constraint stems from the way in which the genotype of an organism is linked to its fitness. It is this type of evolutionary constraint, called *genetic constraint*, that is the primary focus of this thesis. In this section, we will discuss two forms of genetic constraint: pleiotropy and epistasis.

Pleiotropy

Organisms are often complex systems that need to adapt to their environment in multiple different ways. For example, the flowering success of a tree will depend on both its ability to withstand the weather conditions and its ability to utilize the soil as a source of nutrients. Pleiotropy occurs when a single mutation affects multiple aspects of an organism's phenotype. This is a consequence of a non-modular mapping from the genotype of an organism to its phenotype. For example, a mutation will have pleiotropic effects when it inactivates a protein that acts in pathways related to two distinct traits. If this causes the traits to negatively correlate, this generates

adaptive trade-offs as both traits cannot be simultaneously optimized (Guillaume & Otto, 2012; Kessi-Pérez et al., 2016; MacLean et al., 2004). Importantly, it has been mathematically shown that populations adapt more slowly and through mutations of smaller effect size as the extent of pleiotropy increases (Fisher, 1930; Orr, 2000). Because we tend to perceive non-modularity as complex, this effect of pleiotropy has been labelled as the *cost of complexity* (Orr, 2000; Welch & Waxman, 2003). Intuitively, the cost of complexity can be understood using the analogy of a microscope that needs to be focused on a plane (Welch & Waxman, 2003). If this microscope is complex (it has many knobs and settings) and its components are interconnected (each knob affects several settings), then large adjustments are likely to adversely affect at least one of the settings. Instead, only small changes have a reasonable chance of enhancing focus. The degree of pleiotropy in real organisms and the extent to which the cost of complexity affects their evolution is, however, still a matter of active debate (Hill & Zhang, 2012; McGuigan et al., 2014; Wagner & Zhang, 2011).

Epistasis

The same mutation can have different phenotypic effects depending on the genetic context in which it occurs. An intriguing example of this effect can be found in the inherited blood disorder thalassemia. Inheritance of either the allelic version of the gene encoding for α- or β- thalessemia leads to blood disorders (Nagel, 2005). However, co-inheritance of both alleles ameliorates the severity of these disorders (Penman et al., 2009). Such a dependency of the phenotypic effect of a mutation on the allelic status of other loci is referred to as epistasis (Bateson et al., 1909; Phillips, 1998). Whether epistasis suppresses or aggravates the effect of a mutation depends on the molecular cause of the gene-gene interaction (Lehner, 2011). For example, if two proteins are both necessary to form a molecular complex then the effect of their simultaneous deletion will not be much different from deleting only one of them: in both cases, the complex is destroyed. Alternatively, if two proteins separately catalyze the same reaction, the deletion of one of them will have only mild effects on fitness, but their simultaneous deletion can be lethal.

From the perspective of evolutionary constraints, *sign epistasis* is a particularly important form of epistasis (Kvitek & Sherlock, 2011; Weinreich et al., 2005). Rather than changing the effect size, sign epistasis alters the sign of the fitness effect that a mutation has (Kogenaru et al., 2009; Phillips, 2008). Identifying the molecular causes of sign epistasis has proven to be challenging (Nghe et al., 2018), but a common metaphor used to describe how it can emerge is that of a matching key and lock (Taute et al., 2014). Changing either the key or the lock individually is deleterious, as this leads to a loss of recognition. However, simultaneously mutating both the key and lock allows the system to explore novel functionalities that increase fitness without having loss of recognition as an intermediate step. Thus, under the condition that every mutation must improve fitness to contribute to adaptation, only trajectories in which both the lock and key mutate together are accessible. Similar to this example, sign epistasis generally constrains the order in which mutations can occur during adaptive evolution (Anderson et al., 2015; Kvitek & Sherlock, 2011; Weinreich, 2006). Specifically, pervasive sign epistasis can render the majority of evolutionary trajectories to higher fitness inaccessible (Kvitek & Sherlock, 2011; Weinreich et al., 2005), thereby increasing the predictability of the course that evolution will take.

1.3. From molecules to networks

Cells are composed of a variety of molecules ranging from proteins and DNA to lipids and metabolites. These molecules do not work in isolation, but act together in a dynamic network to perform different cellular functions (Barabási & Oltvai, 2004; Han, 2008). Importantly, what these networks look like has implications for evolution. First of all, the molecular interactions in cellular networks are the cause of epistasis and pleiotropy, which were discussed in the previous section. Second, the global topology of the network is one of the factors that controls the robustness of the cell to perturbations. Here, we discuss the current understanding of the link between the global network topology and robustness.

Advances in molecular biology have made it possible to map the interactions between biological molecules on a system wide level (Bader et al., 2008; Ghadie et al., 2018; Legrain et al., 2001). Studies that have created such maps for protein-protein interactions (PPI) have found that these networks typically contain a few proteins with many interactions, and many proteins with few interactions (Barabási & Oltvai, 2004; Spirin & Mirny, 2003). In fact, the number of proteins P with k interactions in a PPI network was found to approximately follow a power law:

$$P(k) \sim k^{-\gamma}. \tag{1.1}$$

Here, γ is the network's degree exponent that determines the amount of variation in the number of interaction partners (also know as the *degree*) of the proteins in the network: as γ becomes smaller, the variation increases asymptotically towards a uniform distribution. For biological networks, γ is typically found to be in the range of $2 < \gamma < 3$ (Albert, 2005; Barabási & Albert, 1999). Because there is no typical degree that can be used to describe all nodes in the network, networks that follow a power law are described as *scale free*. Why biological networks tend to have a scale free structure is currently unclear: some have proposed that it is the result of the mechanism through which networks expand during evolution (Barabási & Albert, 1999), while others have argued that it is due to differences in the intrinsic ability of proteins to gain new interactions (Caldarelli et al., 2002).

An important structural feature of scale free networks is that it contains several sub-networks consisting of nodes that are more densely connected to each other than to other nodes (Han et al., 2004; Spirin & Mirny, 2003). These densely connected sub-networks are, in turn, connected to each other through high-degree proteins (*hubs*). Interestingly, studies have shown that hub proteins are more likely to be essential for survival than low degree proteins (Batada et al., 2006; Hahn & Kern, 2005; Jeong et al., 2001; Peng et al., 2015). This correlation between essentiality and degree might appear to be an obvious consequence of the central role hub proteins play in the overall connectivity of the network, but several other explanations have been proposed (see, for example, He and Zhang (2006)).

The scale-free structure of networks, where a few highly connected proteins take a central position, has been reported to increase the robustness of the network to random perturbations (Albert et al., 2000; Kitano, 2002; Whitacre, 2012). If a protein is removed from the network at random, the chance that this protein is an essential hub protein is low. Thus, most random mutations will not greatly impact the topology of the network, suggesting that they will have a limited effect on phenotype. This demonstrates that the structure of biological networks can already impact evolutionary dynamics.

1.4. Transposon mutagenesis screens

Central to predicting evolution is understanding how genotype relates to fitness (de Visser et al., 2018). A relatively new technique to explore the fitness contribution of genes on a genome-wide scale is transposon insertion site sequencing (TIS) (Cain et al., 2020; Goodman et al., 2011; Langridge et al., 2009; van Opijnen & Camilli, 2013). TIS makes use of transposons (also called jumping genes), which are pieces of DNA that can migrate between different locations in the genome (McClintock, 1950). Transposons occur naturally in the genomes of many organisms, where they play an important role in generating genetic diversity (Bourque et al., 2018; Kidwell & Lisch, 1997; Makałowski et al., 2019; O'Donnell & Burns, 2010). TIS methods have harnessed these mutagenic properties of transposons to generate complete libraries of gene deletion mutants in a single step (Schrevens & Sanglard, 2021). The exact details of the technique differ across versions of TIS, but here we give an overview of the three steps that are common to (nearly) all TIS screens: library generation, library expansion and high-throughput sequencing.

Library generation

The collection of all mutants that are used in a TIS screen is called the *mutant library* (Jacobs et al., 2003; Michel et al., 2017; van Opijnen et al., 2009). Mutant libraries are created by introducing an exogenous transposon system into a strain of interest (Cain et al., 2020). Here, the fact that this system is exogenous is important to prevent the interference of possible endogenous transposons (Kawakami et al., 2017; Schrevens & Sanglard, 2021). The transposon system consists of two components: a protein (called the transposase) and the genetic sequence that embodies the transposon (Lorenzo et al., 1998; Schrevens & Sanglard, 2021). Mutations are generated by inducing the transposase-mediated translocation of the transposon from its donor sequence (typically a plasmid) to a random location in the host genome. By doing this in a controlled manner, a collection of mutants is created that each carry a single transposon in their genome at a unique location. The insertion of the transposon into the genome disrupts the endogenous sequence and is typically assumed to cause a loss-of-function mutation when it occurs in a coding region (Michel et al., 2017; van Opijnen et al., 2009). The number of independent mutants that are generated through this process is called the *complexity* of the library (Chao et al., 2016; Mahmutovic et al., 2020). Thus, a more complex library means that it consists of a more genetically diverse population of mutants. The degree to which the mutant library reflects all possible mutations that could have been realized by transposon mutagenesis is referred to as the *saturation* of the library. However, how the term saturation is used in practice tends to depend on the aim with which the TIS screen is performed. For example, while some use it to refer to the fraction of all possible transposon insertion sites that are occupied (Chao et al., 2016), others use it to describe the fraction of possible gene disruption mutants that are represented in the TIS screen (Michel et al., 2017).

Library expansion

The fitness of the mutants that are present in the library is assessed through a competitive fitness assay. This fitness assay is referred to as the *library expansion*. It is important to note that the library expansion step does not generate new mutants, but rather allows the existing mutant population to "expand" through reproduction (Chao et al., 2016; Michel et al., 2017; van Opijnen et al., 2009). The principle behind the library expansion step is that fitter mutants will increase in abundance relative to less fit mutants. Thus, after the library expansion, the relative fitness of

a mutant is expected to correlate with its abundance.

High-throughput sequencing

After the library expansion step, the mutant population is pooled and the genomic DNA is extracted. The relative abundance of each mutant in the library is subsequently estimated through sequencing of the pooled genomic DNA sample. To increase the efficiency of sequencing, TIS relies on methods to identify and enrich the transposon-genome junctions such that targeted sequencing of these junctions can be performed. Examples of methods that are used to enrich the transposon-genome junctions are polymerase chain reaction (PCR) amplification (Michel et al., 2017; van Opijnen et al., 2009) and bead based extraction (Goodman et al., 2009; Goodman et al., 2011). After sequencing the transposon genome junctions, mutant abundance can be estimated from the number of sequencing reads that map to a genomic location. Genomic locations with a high number of sequencing reads are considered to be dispensable regions, while those that have a low read count or are void of reads correspond to (nearly) essential regions.

1.5. Thesis aim and outline

In this thesis, we study how the functional organization of the cell at the molecular level can facilitate and constrain evolution. To do so, we take advantage of an evolutionary trajectory in *Saccharomyces cerevisiae* that has been demonstrated to be highly reproducible. Our work on this subject is organized into the following chapters:

Chapter 2: Predicting evolution using regulatory architecture.

Epistasis is an important source of evolutionary constraint that can profoundly increase the predictability of evolutionary trajectories. However, it is often unclear how observed epistatic interactions relate to the genetic makeup of an organism. In this chapter, we review recent advancements in the identification of the mechanistic causes of epistasis across different levels of biological organization. We do so with a particular focus on gene regulatory networks.

Chapter 3: Pleiotropy can drive adaptive changes in traits required for survival in novel environments.

Evolving the ability to respond to environmental change should depend on whether it provides a fitness advantage. Responsiveness is therefore expected to be a specific feature of evolution in a dynamic environment. However, pleiotropic effects of mutations can change the correlation between how traits evolve and selection pressures from the environment. In this chapter, we evolve a polarity mutant that is unable to appropriately respond to environmental change in both a static and dynamic environment. We find that regardless of whether the environment is static or dynamic, the responsiveness to environmental change recovers during evolution. We propose that pleiotropic interactions may be an important structural component of cellular networks that allows cells to improve their fitness in unseen environments.

Chapter 4: SATAY as a method to determine the fitness effect of gene disruptions on a genome-wide scale.

One requirement to predict evolution is knowledge of the mapping between the genotype of an organism and its fitness. By definition, this information is embedded in the concept of the fitness landscape. However, fitness landscapes have a highly dynamic structure that changes across

genetic backgrounds and environments. Most methods that are currently used to systematically map the structure of fitness landscapes are either not flexible enough to capture these dynamics or can only be used to map small regions of the fitness landscape. In this chapter, we develop a method to estimate the fitness of gene disruption mutants on a system-wide level using the TIS screen SATAY.

Chapter 5: Epistasis allows simple control over the accessibility of evolutionary trajectories.

Epistasis has an important role in generating genetic constraints that make evolutionary trajectories more reproducible. At the same time, epistasic interactions might be a potential engineering tool that would allow control over evolutionary pathways. In this chapter, we test the feasibility of exploiting epistasis for this purpose. Using the TIS screen SATAY, we show that the deletion of a single gene might be sufficient to block the preferred evolutionary trajectory.

Chapter 6: Compensatory evolution changes the global structure of the genotype-to-fitness map.

Bem1 is a core polarity protein of *Saccharomyces cerevisiae*. Loss of Bem1 is sub-lethal, but cells can recover to almost the wild-type fitness level through the step-wise deletion of two other proteins (Bem3 and Nrp1). The fact that compensation can be achieved through additional gene deletions suggests that the protein interaction network rewires in such a way that an alternative pathway is established. In this chapter, we show that compensatory evolution through gene deletions causes global changes in the fitness contribution of genes. Based on the functional annotations of the genes that have an altered fitness contribution, we propose the cellular pathways that are likely involved in the alternative pathway for polarity establishment.

Bibliography

Albert, R. (2005). Scale-free networks in cell biology. https://doi.org/10.1242/jcs.02714

Albert, R., Jeong, H., & Barabási, A.-L. (2000). Error and attack tolerance of complex networks. *Nature*, *406*(6794), 378–382. https://doi.org/10.1038/35019019

Anderson, D. W., McKeown, A. N., & Thornton, J. W. (2015). Intermolecular epistasis shaped the function and evolution of an ancient transcription factor and its DNA binding sites. *eLife*, *4*(JUNE2015). https://doi.org/10.7554/ELIFE.07864

Bader, S., Kühner, S., & Gavin, A. C. (2008). Interaction networks for systems biology. *FEBS Letters*, *582*(8), 1220–1224. https://doi.org/10.1016/J.FEBSLET.2008.02.015

Barabási, A.-L., & Albert, R. (1999). Emergence of Scaling in Random Networks. *Science*, *286*(5439), 509–512. https://doi.org/10.1126/science.286.5439.509

Barabási, A.-L., & Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, *5*(2), 101–113. https://doi.org/10.1038/nrg1272

Batada, N. N., Hurst, L. D., & Tyers, M. (2006). Evolutionary and Physiological Importance of Hub Proteins. *PLOS Computational Biology*, *2*(7), e88. https://doi.org/10.1371/JOURNAL.PCBI.0020088

Bateson, W., Mendel, G., & Leighton, A. G. (1909). *Mendel's principles of heredity, by W. Bateson*. University Press. https://www.biodiversitylibrary.org/item/15713

Bourque, G., Burns, K. H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izsvák, Z., Levin, H. L., Macfarlan, T. S., Mager, D. L., & Feschotte, C. (2018). Ten things you should know about transposable elements. *Genome Biology*, *19*(1), 199. https://doi.org/10.1186/s13059-018-1577-z

Cain, A. K., Barquist, L., Goodman, A. L., Paulsen, I. T., Parkhill, J., & van Opijnen, T. (2020). A decade of advances in transposon-insertion sequencing. *Nature Reviews Genetics*, *21*(9), 526–540. https://doi.org/10.1038/s41576-020-0244-x

Caldarelli, G., Capocci, A., De Los Rios, P., & Muñoz, M. A. (2002). Scale-Free Networks from Varying Vertex Intrinsic Fitness. *Physical Review Letters*, *89*(25), 258702. https://doi.org/10.1103/PhysRevLett.89.258702

Chao, M. C., Abel, S., Davis, B. M., & Waldor, M. K. (2016). The design and analysis of transposon insertion sequencing experiments. *Nature Reviews Microbiology*, *14*(2), 119–128. https://doi.org/10.1038/nrmicro.2015.7

Darwin 1809-1882, C. (1859). *On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life*. London : John Murray, 1859. https://search.library.wisc.edu/catalog/9934839413602122

de Visser, J. A. G. M., Elena, S. F., Fragata, I., & Matuszewski, S. (2018). The utility of fitness landscapes and big data for predicting evolution. *Heredity*, *121*(5), 401–405. https://doi.org/10.1038/s41437-018-0128-4

Fisher, R. A. (1930). *The genetical theory of natural selection*. Clarendon Press. https://doi.org/10.5962/bhl.title.27468

Ghadie, M. A., Coulombe-Huntington, J., & Xia, Y. (2018). Interactome evolution: insights from genome-wide analyses of protein–protein interactions. *Current Opinion in Structural Biology*, *50*, 42–48. https://doi.org/10.1016/J.SBI.2017.10.012

Godfrey-Smith, P. (2007). Conditions for Evolution by Natural Selection. *Journal of Philosophy*, *104*(10), 489–516. https://doi.org/10.5840/jphil2007104103

Goodman, A. L., McNulty, N. P., Zhao, Y., Leip, D., Mitra, R. D., Lozupone, C. A., Knight, R., & Gordon, J. I. (2009). Identifying Genetic Determinants Needed to Establish a Human Gut Symbiont in Its Habitat. *Cell Host & Microbe*, *6*(3), 279–289. https://doi.org/10.1016/j.chom.2009.08.003

Goodman, A. L., Wu, M., & Gordon, J. I. (2011). Identifying microbial fitness determinants by insertion sequencing using genome-wide transposon mutant libraries. *Nature Protocols*, *6*(12), 1969–1980. https://doi.org/10.1038/nprot.2011.417

Guillaume, F., & Otto, S. P. (2012). Gene Functional Trade-Offs and the Evolution of Pleiotropy. *Genetics*, *192*(4), 1389–1409. https://doi.org/10.1534/genetics.112.143214

Hahn, M. W., & Kern, A. D. (2005). Comparative Genomics of Centrality and Essentiality in Three Eukaryotic Protein-Interaction Networks. *Molecular Biology and Evolution*, *22*(4), 803–806. https://doi.org/10.1093/molbev/msi072

Han, J.-D. J., Bertin, N., Hao, T., Goldberg, D. S., Berriz, G. F., Zhang, L. V., Dupuy, D., Walhout, A. J. M., Cusick, M. E., Roth, F. P., & Vidal, M. (2004). Evidence for dynamically organized modularity in the yeast protein–protein interaction network. *Nature*, *430*(6995), 88–93. https://doi.org/10.1038/nature02555

Han, J.-D. J. (2008). Understanding biological functions through molecular networks. *Cell Research*, *18*(2), 224–237. https://doi.org/10.1038/cr.2008.16

He, X., & Zhang, J. (2006). Why Do Hubs Tend to Be Essential in Protein Networks? *PLOS Genetics*, *2*(6), e88. https://doi.org/10.1371/JOURNAL.PGEN.0020088

Hill, W. G., & Zhang, X.-S. (2012). Assessing pleiotropy and its evolutionary consequences: pleiotropy is not necessarily limited, nor need it hinder the evolution of complexity. *Nature Reviews Genetics*, *13*(4), 296–296. https://doi.org/10.1038/nrg2949-c1

Jacobs, M. A., Alwood, A., Thaipisuttikul, I., Spencer, D., Haugen, E., Ernst, S., Will, O., Kaul, R., Raymond, C., Levy, R., Chun-Rong, L., Guenthner, D., Bovee, D., Olson, M. V., & Manoil, C. (2003). Comprehensive transposon mutant library of Pseudomonas aeruginosa. *Proceedings of the National Academy of Sciences*, *100*(24), 14339–14344. https://doi.org/10.1073/pnas.2036282100

Jeong, H., Mason, S. P., Barabási, A.-L., & Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature*, *411*(6833), 41–42. https://doi.org/10.1038/35075138

Kawakami, K., Largaespada, D. A., & Ivics, Z. (2017). Transposons As Tools for Functional Genomics in Vertebrate Models. *Trends in Genetics*, *33*(11), 784–801. https://doi.org/10.1016/j.tig.2017.07.006

Kempes, C. P., Koehl, M. A., & West, G. B. (2019). The scales that limit: The physical boundaries of evolution. *Frontiers in Ecology and Evolution*, *7*(AUG), 242. https://doi.org/10.3389/FEVO.2019.00242/BIBTEX

Kessi-Pérez, E. I., Araos, S., García, V., Salinas, F., Abarca, V., Larrondo, L. F., Martínez, C., & Cubillos, F. A. (2016). RIM15 antagonistic pleiotropy is responsible for differences in fermentation and stress response kinetics in budding yeast (I. Pretorius, Ed.). *FEMS Yeast Research*, *16*(3), fow021. https://doi.org/10.1093/femsyr/fow021

Kidwell, M. G., & Lisch, D. (1997). Transposable elements as sources of variation in animals and plants. *Proceedings of the National Academy of Sciences*, *94*(15), 7704–7711. https://doi.org/10.1073/pnas.94.15.7704

Kitano, H. (2002). Computational systems biology. *Nature*, *420*(6912), 206–210. https://doi.org/10.1038/nature01254

Kogenaru, M., de Vos, M. G. J., & Tans, S. J. (2009). Revealing evolutionary pathways by fitness landscape reconstruction. *Critical Reviews in Biochemistry and Molecular Biology*, *44*(4), 169–174. https://doi.org/10.1080/10409230903039658

Kvitek, D. J., & Sherlock, G. (2011). Reciprocal Sign Epistasis between Frequently Experimentally Evolved Adaptive Mutations Causes a Rugged Fitness Landscape. *PLOS Genetics*, *7*(4), 1–11. https://doi.org/10.1371/journal.pgen.1002056

Langridge, G. C., Phan, M.-D. D., Turner, D. J., Perkins, T. T., Parts, L., Haase, J., Charles, I., Maskell, D. J., Peters, S. E., Dougan, G., Wain, J., Parkhill, J., & Turner, A. K. (2009). Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. *Genome Research*, *19*(12), 2308–2316. https://doi.org/10.1101/gr.097097.109

Legrain, P., Wojcik, J., & Gauthier, J. M. (2001). Protein-protein interaction maps: A lead towards cellular functions. *Trends in Genetics*, *17*(6), 346–352. https://doi.org/10.1016/S0168-9525(01)02323-X

Lehner, B. (2011). Molecular mechanisms of epistasis within and between genes. *Trends in Genetics*, *27*(8), 323–331. https://doi.org/10.1016/J.TIG.2011.05.007

Li, Y., Steenwyk, J. L., Chang, Y., Wang, Y., James, T. Y., Stajich, J. E., Spatafora, J. W., Groenewald, M., Dunn, C. W., Hittinger, C. T., Shen, X.-X., & Rokas, A. (2021). A genome-scale phylogeny of the kingdom Fungi. *Current Biology*, *31*(8), 1653–1665. https://doi.org/10.1016/j.cub.2021.01.074

Lorenzo, V., Herrero, M., SÃ¡nchez, J. M., & Timmis, K. N. (1998). Mini-transposons in microbial ecology and environmental biotechnology. *FEMS Microbiology Ecology*, *27*(3), 211–224. https://doi.org/10.1111/J.1574-6941.1998.TB00538.X

MacLean, R. C., Bell, G., & Rainey, P. B. (2004). The evolution of a pleiotropic fitness tradeoff in Pseudomonas fluorescens. *Proceedings of the National Academy of Sciences*, *101*(21), 8072–8077. https://doi.org/10.1073/pnas.0307195101

Mahmutovic, A., Abel zur Wiesch, P., & Abel, S. (2020). Selection or drift: The population biology underlying transposon insertion sequencing experiments. *Computational and Structural Biotechnology Journal*, *18*, 791–804. https://doi.org/10.1016/J.CSBJ.2020.03.021

Makałowski, W., Gotea, V., Pande, A., & Makałowska, I. (2019). Transposable Elements: Classification, Identification, and Their Use As a Tool For Comparative Genomics. In *Methods in molecular biology (clifton, n.j.)* (pp. 177–207). Humana Press Inc. https://doi.org/10.1007/978-1-4939-9074-0{_}6

McClintock, B. (1950). The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Sciences of the United States of America*, *36*(6), 344–355. https://doi.org/10.1073/PNAS.36.6.344

McGuigan, K., Collet, J. M., McGraw, E. A., Ye, Y. H., Allen, S. L., Chenoweth, S. F., & Blows, M. W. (2014). The nature and extent of mutational pleiotropy in gene expression of male Drosophila serrata. *Genetics*, *196*(3), 911–921. https://doi.org/10.1534/genetics.114.161232

Michel, A. H., Hatakeyama, R., Kimmig, P., Arter, M., Peter, M., Matos, J., De Virgilio, C., & Kornmann, B. (2017). Functional mapping of yeast genomes by saturated transposition. *eLife*, *6*. https://doi.org/10.7554/eLife.23570

Nagel, R. L. (2005). Epistasis and the genetics of human diseases. *Comptes Rendus Biologies*, *328*(7), 606–615. https://doi.org/10.1016/j.crvi.2005.05.003

Nghe, P., Kogenaru, M., & Tans, S. J. (2018). Sign epistasis caused by hierarchy within signalling cascades (2018/04/15). *Nature Communications*, *9*(1), 1451. https://doi.org/10.1038/s41467-018-03644-8

O'Donnell, K. A., & Burns, K. H. (2010). Mobilizing diversity: transposable element insertions in genetic variation and disease. *Mobile DNA*, *1*(1), 21. https://doi.org/10.1186/1759-8753-1-21

Orr, H. A. (2000). ADAPTATION AND THE COST OF COMPLEXITY. *Evolution*, *54*(1), 13–20. https://doi.org/10.1111/j.0014-3820.2000.tb00002.x

Peng, X., Wang, J., Wang, J., Wu, F.-X., & Pan, Y. (2015). Rechecking the Centrality-Lethality Rule in the Scope of Protein Subcellular Localization Interaction Networks (P. Holme, Ed.). *PLOS ONE*, *10*(6), e0130743. https://doi.org/10.1371/journal.pone.0130743

Penman, B. S., Pybus, O. G., Weatherall, D. J., & Gupta, S. (2009). Epistatic interactions between genetic disorders of hemoglobin can explain why the sickle-cell gene is uncommon in the Mediterranean. *Proceedings of the National Academy of Sciences*, *106*(50), 21242–21246. https://doi.org/10.1073/pnas.0910840106

Phillips, P. C. (1998). The Language of Gene Interaction. *Genetics*, *149*(3), 1167–1171. https://doi.org/10.1093/genetics/149.3.1167

Phillips, P. C. (2008). Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Reviews Genetics*, *9*(11), 855–867. https://doi.org/10.1038/nrg2452

Schrevens, S., & Sanglard, D. (2021). Hijacking Transposable Elements for Saturation Mutagenesis in Fungi. *Frontiers in Fungal Biology*, *2*, 11. https://doi.org/10.3389/ffunb.2021.633876

Spirin, V., & Mirny, L. A. (2003). Protein complexes and functional modules in molecular networks. *Proceedings of the National Academy of Sciences*, *100*(21), 12123–12128. https://doi.org/10.1073/pnas.2032324100

Taute, K. M., Gude, S., Nghe, P., & Tans, S. J. (2014). Evolutionary constraints in variable environments, from proteins to networks. *Trends in Genetics*, *30*(5), 192–198. https://doi.org/10.1016/j.tig.2014.04.003

van Opijnen, T., Bodi, K. L., & Camilli, A. (2009). Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nature Methods*, *6*(10), 767–772. https://doi.org/10.1038/nmeth.1377

van Opijnen, T., & Camilli, A. (2013). Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nature Reviews Microbiology*, *11*(7), 435–442. https://doi.org/10.1038/nrmicro3033

Wagner, G. P., & Zhang, J. (2011). The pleiotropic structure of the genotype–phenotype map: the evolvability of complex organisms. *Nature Reviews Genetics*, *12*(3), 204–213. https://doi.org/10.1038/nrg2949

Weinreich, D. M. (2006). Darwinian Evolution Can Follow Only Very Few Mutational Paths to Fitter Proteins. *Science*, *312*(5770), 111–114. https://doi.org/10.1126/science.1123539

Weinreich, D. M., Watson, R. A., & Chao, L. (2005). PERSPECTIVE:SIGN EPISTASIS AND GENETIC CONSTRAINT ON EVOLUTIONARY TRAJECTORIES. *Evolution*, *59*(6), 1165–1174. https://doi.org/10.1554/04-272

Welch, J. J., & Waxman, D. (2003). MODULARITY AND THE COST OF COMPLEXITY. *Evolution*, *57*(8), 1723–1734. https://doi.org/10.1111/j.0014-3820.2003.tb00581.x

Whitacre, J. M. (2012). Biological Robustness: Paradigms, Mechanisms, and Systems Principles. *Frontiers in Genetics*, *3*. https://doi.org/10.3389/fgene.2012.00067

2

Nothing in biology makes sense except in the light of evolution

— Theodosius Dobzhansky

Predicting evolution using regulatory architecture

Philippe Nghe*, Marjon G. J. de Vos *, Enzo Kingma*, Manjunatha Kogenaru, Frank J. Poelwijk, Liedewij Laan, Sander J. Tans

* equal contributions

Abstract The limits of evolution have long fascinated biologists. However, the causes of evolutionary constraint have remained elusive due to a poor mechanistic understanding of studied phenotypes. Recently, a range of innovative approaches have leveraged mechanistic information on regulatory networks and cellular biology. These methods combine systems biology models with population and single-cell quantification and with new genetic tools, and they have been applied to a range of complex cellular functions and engineered networks. In this article, we review these developments which are revealing the mechanistic causes of epistasis at different levels of biological organization —in molecular recognition, within a single regulatory network, and between different networks— providing first indications of predictable features of evolutionary constraint.

2.1. Introduction

Elucidating the range of possibilities and limitations of evolutionary adaptation has been one of the most evocative and complex problems in biology (Bell et al., 2010; Wright, 1932). Evolution is sometimes strikingly rapid but can also display long-term stagnation for reasons that often remain obscure. Resolving this conundrum is central to understanding natural and laboratory evolution, and important to harnessing evolutionary optimization in protein and cellular engineering applications. At a conceptual level, a wide variety of causes have been invoked to explain why new functions may fail to evolve: Those functions may be impossible biochemically (Beardmore et al., 2011) or physically (Dill et al., 2011), a sufficiently strong selection may be absent (Gavrilets & Hastings, 1993), or, alternatively, an organism's genetic makeup may hamper evolution (De Visser & Krug, 2014; Lobkovsky & Koonin, 2012). Indeed, functional improvements that require multiple genetic changes are difficult to acquire by fixing each mutation one by one. The first empirical studies of such genetic interdependencies were enabled by systematic genetic reconstruction of evolutionary intermediates and laboratory evolution (Bridgham et al., 2006; Lindsey et al., 2013; Lunzer et al., 2005; Poelwijk et al., 2007; Poelwijk et al., 2019; Weinreich, 2006). Still, we have only begun to scratch the surface of this multifaceted issue.

It has proven useful to break down the problem of genetic interdependence into a few elementary types of pairwise genetic interactions (De Visser & Krug, 2014; Lobkovsky & Koonin, 2012). Briefly, *reciprocal sign epistasis* refers to cases in which two independent disadvantageous mutations are simultaneously required for an improved phenotype or fitness (Poelwijk, Tănase-Nicola, et al., 2011). It is reciprocal, because both mutations influence each other's effect, and it is described using the term "sign", because the fitness effects of the mutations switch between

A version of this chapter has been published as Nghe P, de Vos MGJ, Kingma E, et al. Predicting Evolution Using Regulatory Architecture. *Annu Rev Biophys*. 2020;49:181-197.

negative and positive. It is this type of interaction that constrains adaptive evolution the most, since it implies that both mutations must be fixed simultaneously in a selective sweep – a topic that is discussed further in Section 2.2. Evolution is less constrained when only one of the two mutations switches its effect between negative and positive, which is referred to as regular *sign epistasis* (Weinreich, 2006). In this case, some evolutionary pathways are inaccessible to adaptive evolution, but others remain possible. Finally, one can distinguish between two cases that do not restrict adaptive pathways: *magnitude epistasis*, where the occurrence of one mutation alters the magnitude of the fitness effect of another mutation, and *no epistasis*, where mutational effects on fitness are additive (Poelwijk et al., 2007).

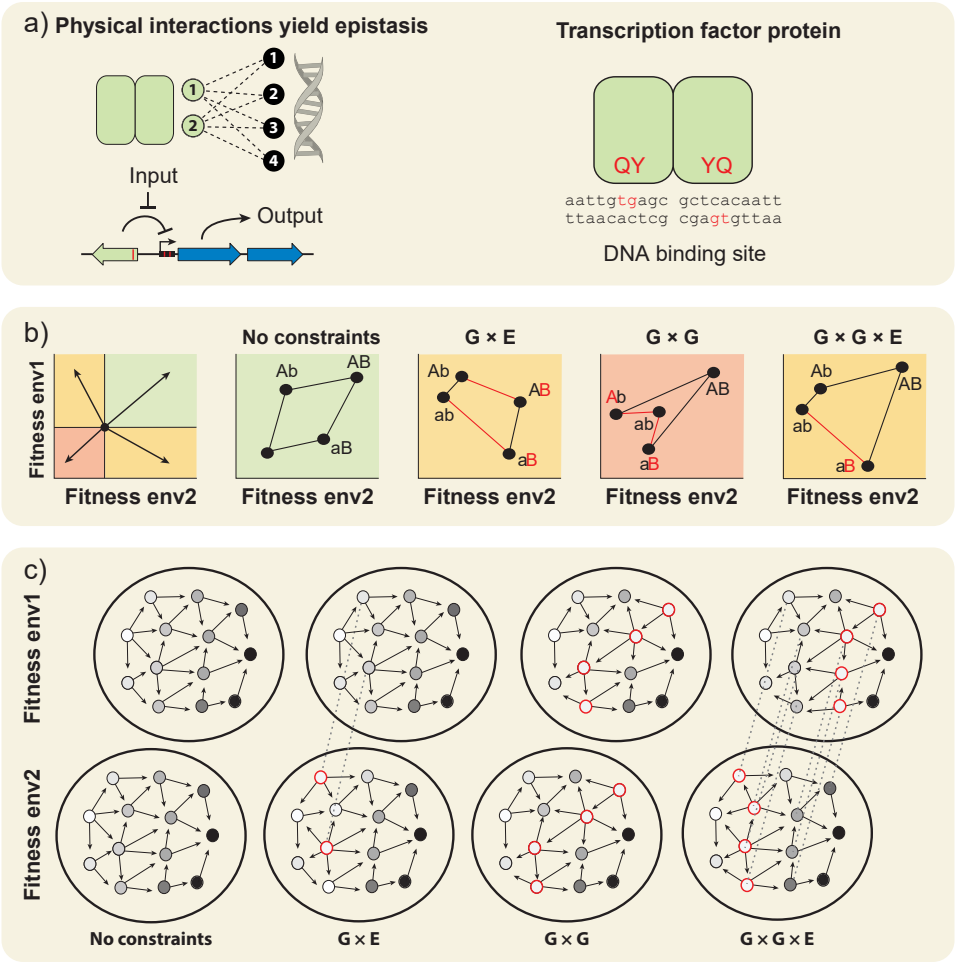
One of the major challenges in current evolutionary research is to go beyond description and toward prediction. In this context, the notion of epistasis is useful because it provides the capacity to classify and quantify evolutionary constraints. However, it only provides a part of the picture, since it does not address the underlying molecular mechanisms. Recently, a series of studies exploited knowledge about the architecture of regulatory networks to begin filling this void. The rationale for these studies is that this type of knowledge provides a mechanistic basis for notions such as constraint and epistasis, which by themselves are mechanism independent. One can determine how the quantifiable network properties, like topology, expression levels of constituent genes, or molecular affinities, affect phenotype and fitness. Moreover, mathematical modeling may be used to extensively explore the range of possible phenotypes, thus opening up the possibility to predict constraints and epistasis.

This early stage of exploration of constraints in network evolution is characterized by a combination of experimental and theoretical innovations, and has focused on elementary questions. For instance, do regulatory trade-offs limit evolution, or can they also accelerate it? Are downstream regulatory elements constrained differently than upstream elements? Does pleiotropy within networks frustrate or facilitate their evolution? In this review, these efforts are organized into three sections, each considering a different level of biological organization, ranging from single molecular interactions to highly interconnected networks.

2.2. Epistasis in regulatory interactions

Intermolecular binding is at the heart of all regulatory networks, whether it occurs between membrane-associated effector proteins and kinases to transduce signals (Podgornaia & Laub, 2015) or between transcription factors and their DNA binding sites to regulate gene expression (Balleza et al., 2009; Poelwijk, de Vos, et al., 2011). In this section, we will consider studies of how the physical binding of macromolecules impacts epistasis in a regulatory system, how environmental changes modulate epistasis, and the relationship of these factors to the predictability of epistasis and evolutionary constraint.

To function, transcription factors must bind their own cognate DNA binding site while avoiding others. It has been proposed that such specific molecular recognition represents an architectural feature that gives rise to epistasis (Poelwijk, Heyning, et al., 2011). The rationale is that changing such lock-key systems requires modifications in both key (transcription factor) and lock (binding site), as changing only one of them yields nonmatching combinations Figure 2.1a). Since scenarios where two (or more) genetic changes occur in the same (selection) period are rare, such an architecture could prohibit mutational trajectories to improved phenotypes under positive selection.



(a) Molecular recognition in cellular regulation. The specific binding of a dimeric transcription factor (light green) to a DNA binding site, which allows expression control of a downstream gene (blue), can be seen as a lock-key interaction. Mutating both binding partners produces new lock-key combinations, while mutating either yields non-functional ones, as shown by systematic mutagenesis (Nghe et al., 2018). (**b**) Schematic diagrams illustrating how environmental change can affect the epistatic interactions between two mutations a-to-A and b-to-B. Mutations are depicted as vectors, and denote the change in fitness in environment 1 (Env1) and environment 2 (Env2). No genetic constraints: Both mutations improve fitness, independent of the environment and the genetic background. GXE interaction: The environment changes the sign of a mutational effect (of b-to-B), independently of the genetic background. The mutation b-to-B lowers fitness in Env1 and hence is inaccessible by adaptive evolution, but becomes accessible when changing to Env2. GxG interaction: The mutational effects depend on the genetic background, but not on the environment. Here, both mutational effects change sign depending on the genetic background (but not the environment), and thus correspond to reciprocal sign epistasis. GxGxE interaction: Both the environment and the genetic background determine the mutational effect. Here, the effect of b-to-B changes sign depending on the genetic background, but only in Env1. (**c**) Schematic of an adaptive landscape in two environments. Nodes (circles) indicate genotypes, arrows indicate mutational steps of increasing fitness. Gray scale depicts fitness, with darker tones indicating higher fitness. Note that this schematic depiction does not include the full genetic multidimensionality of an adaptive landscape, e.g. it does not contain all pair wise genetic interactions which may form epistatic motifs. ***(Caption continues on next page).***

(Continued from previous page). Without genetic constraints, all mutations are additive and hence all trajectories from the lowest fitness genotype (white) to the optimum (black) are accessible. GxE interaction: Mutational effects deviate from additivity in Env2 (red circles, detrimental mutants). Yet the landscape can be crossed in Env1, by alternating between Env1 and Env2 (dotted lines). GxG interaction: In both environments the mutations are detrimental, with correlated effect in both environments, block direct access to the optimum by the creation of an adaptive valley. The landscape cannot be crossed by single step mutations under positive selection. GxGxE interaction: In both environments, mutations deviate from additivity as detrimental mutations block direct access to the optimum by the formation of an adaptive valley in each environment. Yet the location of these mutants is different in both environments and a part of the landscapes are anti-correlated. Therefore, these landscapes can be crossed by single step mutations under fluctuating selection in Env1 and Env2 (dotted lines).

These core ideas are testable experimentally, as has been done using the archetypal model system of transcriptional regulation: the *lac* operon in *Escherichia coli* (de Vos et al., 2015). Owing to decades of mutational and physiological study of this regulatory system, it is known that a few key operator base pairs and a few amino acid residues in the binding interface of the *lac* repressor determine binding specificity (Lehming et al., 1990; Sartorius et al., 1989). Consistent with the predictions, these key residues indeed display reciprocal sign epistasis: Mutations in both the DNA binding site and repressor allowed binding improvements, while mutations in either one alone only led to deterioration. Six key sites in the transcription factor and the binding site were mutated. Of the 720 possible mutational pathways going from one specifically binding transcription factor-operator pair to another, none of the trajectories contained only mutations that improved the phenotype. From an adaptive landscape perspective, these data thus indicated local optima separated by a valley (de Vos et al., 2015).

Regulation allows cells to respond to environmental cues. The *lac* repressor, for instance, allows repression of the *lac* operon in the absence of lactose, and expression in its presence, by inducing a conformational change in the transcription factor that lowers its affinity to the DNA binding site. The *lac* repressor should thus be able to not only bind the operator, but also to efficiently release it in the presence of lactose. Analysis of the 720 possible mutational trajectories of the *lac* repressor-operator combinations in this second environment also showed that none of the mutational trajectories allowed continuous improvements. However, *alternating* between the two environments did open up adaptive trajectories with constant improvements for each mutation. With a computational method that describes the mutational and environmental transitions as a Markov process, the crossing rate from the initial to the final genotype for all trajectories in the landscape, including detours, could be determined. Interestingly, this rate is found to be maximal when the rate of environmental switches compares with the mutation rate (de Vos et al., 2015).

Cross-environmental trade-offs appeared to be responsible for the adaptive accessibility of adaptive trajectories: Sequences that were suboptimal peaks in one environment were transformed into valleys in the other environment, thus allowing escape from a suboptimum (Figure 2.1b, c). In other cases, inaccessible downward slopes were turned into accessible ascending slopes upon environmental change, allowing adaptive trajectories to surf (Mustonen & Lässig, 2009) these slopes with positive selective coefficients. Evolutionary constraints can thus be overcome by environment-dependent ratcheting that allows the crossing of otherwise inaccessible regions in sequence space (de Vos et al., 2015; Steinberg & Ostermeier, 2016).

This highlights the major that the environment plays in the accessibility of biological func-

tions during selection, by modulating genotype-genotype interactions (Figure 2.1b). This is important for more than just regulatory systems, as constraints due to mutations in nonregulatory but coding sequences can be affected as well by environmental change. A study that focused on environment-dependent fitness effects, constructed the genotype space of five mutations in the genome of *E. coli* and measured the phenotype of these genotypes in 1,920 environments (Flynn et al., 2013). The fitness effects of the mutations significantly changed in 203 environments. Moreover, by focusing on the adaptive landscapes involving all interactions among these five mutations in the three environments with the most distinct effects, they observed significant changes in the topography of the adaptive landscape; thus epistatic interactions also differed in the different environments.

Environment-dependent epistasis can also affect the ability of a population to adapt to an environment that is gradually becoming more challenging. In laboratory evolution experiments, the rate of environmental change (which modulates the selective pressure), as well as the chemical nature of the environment (which determines the genotypes that may confer a benefit), determined whether evolving populations could keep up with the imposed environmental change (Gorter et al., 2018; Lindsey et al., 2013). Environmental circumstances can thus alter the sign of a mutational effect, its epistatic interactions with other genetic changes (Figure 2.1b), and thus the course of evolution (de Vos et al., 2018).

The level of ruggedness of genotype-phenotype landscapes does not appear to be specific to the *lac* repressor system. A computational analysis based on the *in vitro* affinity between transcription factors and their binding sites in eukaryotes found that most of these landscapes were relatively rugged (Aguilar-Rodríguez et al., 2017), that is, they were neither as rugged as those obtained from randomly shuffled genotypes, nor purely additive. Nonetheless, many of these landscapes were highly navigable, with mutational trajectories in which binding affinity increased at each mutational step. Does this imply a form of evolutionary optimality of the transcription factor – binding site combinations found in nature? Starr *et al*. found hundreds of alternative transcription factor protein sequences that use diverse binding mechanisms, but perform their function at least as well as the transcription factor that has historically evolved. As they noted, this indicates that "the outcome of evolution depends on a serial chain of compounding chance events" (Starr et al., 2017, p. 409). Thus, they argue that, if evolution had begun from a different ancestral starting point in sequence space, then different genetic and biochemical forms would probably have evolved.

The evolution of regulatory functions can also be constrained by multiple physical interactions. In a regulatory system containing DNA-binding sites for both the RNA polymerase and a transcription factor, both of these proteins compete for binding to an overlapping binding site on the DNA (Lagator, Paixão, et al., 2017). Lagator, Paixão, et al. (2017) measured the phenotypes of both single and double mutants; based on a thermodynamic model in which the sign and magnitude of the individual mutational effects served as input, they could predict the effects of double mutations. Such models, which take into account key functional parameters, may set the stage for the further prediction of epistasis and, thus, the course of evolution based on the data available for single mutants.

2.3. Epistasis in regulatory pathways

In this section, we discuss recent studies that have sought to quantify epistasis between genes that act within a regulatory pathway. Unlike the proteins discussed in the previous section, these

proteins do not necessarily interact physically, but rather by performing one regulatory function together. We examine the generality and consequences of such functional yet nonphysical interactions and propose interpretations of recent experiments based on phenotype-to-fitness models, also called geometric models. A first interpretation is that network structure can indeed explain how epistasis between two genes arises. However, it is not a direct proxy: One cascade structure can display different types of sign epistasis, depending on other details such as the nature of the variable environment. Second, geometric models can provide a unified framework to interpret both the evolutionary and the more classical phenotypic interpretations of epistasis, such as mutations that can mask the phenotypic effects of other mutations. Finally, one can define general conditions for sign epistasis to arise in any system: It does so when the optimal value of one phenotypic parameter, like the binding constant of a transcription factor, depends on another phenotypic parameter within the network. These findings highlight how functional dependencies within regulatory networks can induce strong constraints in fitness landscapes.

Epistasis was originally used for scenarios in which certain genetic backgrounds masked mutation-induced phenotypic variation (Phillips, 2008). Such epistatic interactions between loci across the genome are expected from a purely functional basis, without necessarily implying direct physical interactions between mutated residues, as is, for instance, typically the case for genes within developmental pathways (Peter & Davidson, 2011). Mechanistic biochemical models indicate that genes in parallel metabolic pathways tend to interact negatively, as the flux catalyzed by a gene can be compensated by flux in a parallel branch, redundantly allowing production of a same final metabolite. However, genes in series within a chain should interact positively, as the removal of any of the catalytic species would strongly reduce the overall metabolic flux (Szathmáry, 1993). This suggests a direct relationship between the network wiring and the observed epistasis (Lehner, 2011; Segrè et al., 2005). However, the sign changes that are key to adaptive fixation were not considered in any of these studies. In parallel, the idea of Fisher geometric models has gained momentum to explain epistasis from biological mechanisms and their function. In this case, generic Gaussian functions describe how fitness depends on a few phenotypic parameters. Although it is heuristic, this assumption can reproduce statistical distributions of epistasis with few parameters (G. Martin et al., 2007) and generate a large variety of epistasis distributions (Gros et al., 2009; Weinreich et al., 2005). This phenotypic view of epistasis and the classical genetic view are starting to be reconciled using geometric models that describe phenotype–fitness relationships in mechanistic terms, using information on the network in question. This approach has so far mainly been applied to small, well-characterized networks (Corson & Siggia, 2012; Cotterell & Sharpe, 2013; Nghe et al., 2018; Szathmáry, 1993). While lacking the full cellular context, these studies have revealed general causes of epistasis, and could serve as a basis for more phenomenological long-term evolutionary models (G. Martin, 2014).

As genes typically represent distinct DNA regions, mutations in one gene logically do not affect the biochemical parameters of another gene, such as its binding constant or enzymatic activity. This genetic modularity has been exploited recently to predict epistasis between genes (Nghe et al., 2018). For instance, in the lac operon, a mutation in the transcription factor binding region impacts expression level, and a mutation in the *LacZ* gene affects catalytic rates independently, although they do both participate in the same physiological function. The independence of mutation effects on different genes applies to regulatory cascades, a ubiquitous regulatory motif in cells, where an upstream gene *y* regulates the expression of a downstream gene *x*, which itself regulates an output gene (Figure 2.2a). Crucially for epistasis, mutational steps that affect the

phenotypic parameters *X* are orthogonal to mutational steps that affect the phenotypic parameters *Y* within the phenotype space. Thus, either *X* or *Y* changes at each mutational step, but not simultaneously.

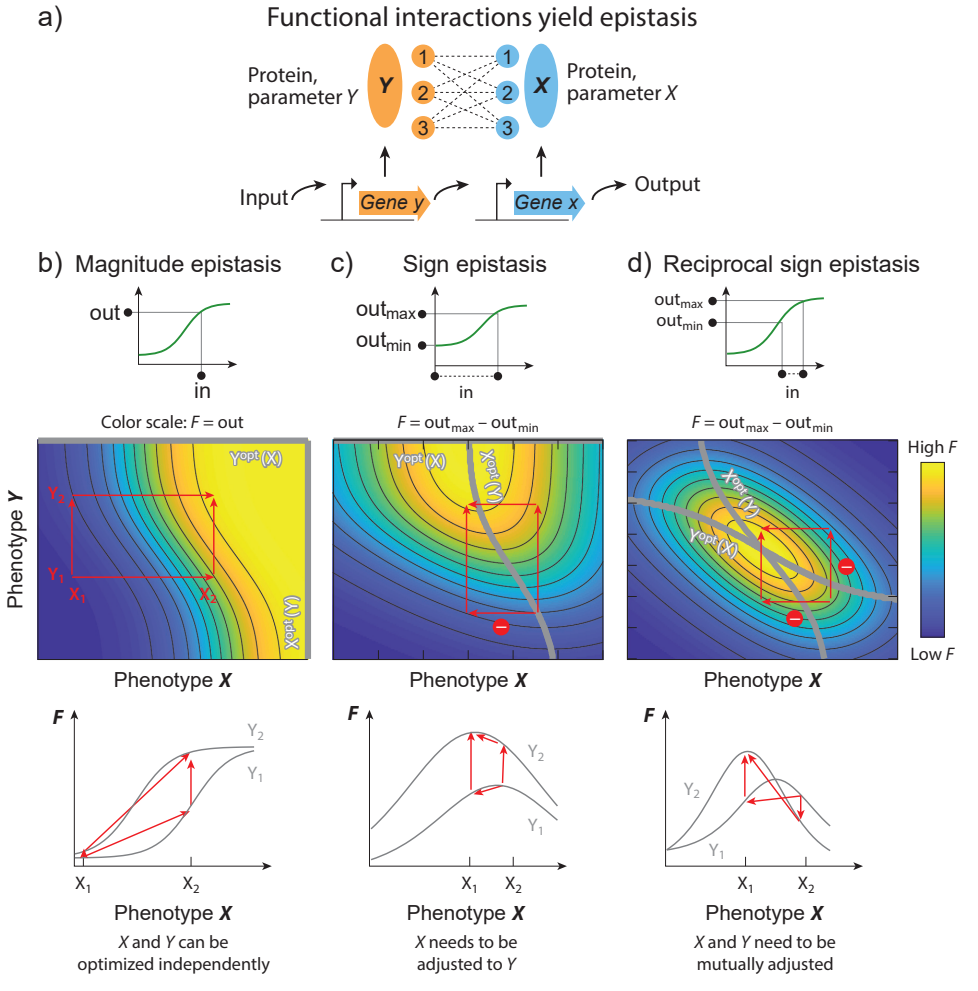
The resulting epistasis predictions could be verified experimentally by systematically combining mutations within the different transcription factors that together form a regulatory cascade and quantifying their input–output relationships, as in Reference (Nghe et al., 2018). Based on that study, we discuss how functional relationships between genes produce epistasis and how they can be explained by the shape of the phenotype-fitness functions. We distinguish three major classes of fitness functions.

First, consider a fitness that varies monotonically with phenotypic parameters *X* and *Y*, such as the binding affinity of the transcription factor for its operator (Figure 2.2a). This case arises, for example, when the fitness (or performance) would correspond directly to the output for any fixed input that does not vary in time. We recover the classical notion of phenotypic epistasis, as illustrated in Figure 2.2b: A mutation with a strong effect on *X* (in the most extreme case a knock-out) cancels any observable variation of the output that could be caused by changes in *Y*. Additionally, it is not possible to generate sign epistasis in this scenario. As discussed further below, this is generally the case when the optimal value for *X* does not depend on *Y* and vice versa.

A second, qualitatively distinct, scenario is when the optimum of *Y* does depend on *X*, while the optimal value of *X* does not depend on *Y* (Figure 2.2c). This situation is found when the fitness corresponds to the dynamic range of the cascade, as quantified by the difference between the minimal and maximal output expression levels, in response to widely varying input signals, as has been shown in Reference (Nghe et al., 2018) . The landscape of this fitness function can generate sign epitasis, with the specific property that only mutations in gene x can lead to decreasing fitness. This is most easily seen when *Y* is at its optimum but *X* is not: Reaching a better *X*-*Y* combination cannot be achieved by mutating *Y* first, as it is already at its maximum in the current *X* background. However, fitness can increase in a stepwise manner, as *X* is not yet at its maximum and thus can be improved. In this example, the evolutionary hierarchy reflects the functional hierarchy: The upstream gene x must be mutated first because the optimum of the downstream gene y must be well tuned to accommodate X expression but not the converse.

The last case is when the optimum of *Y* depends on *X* and the optimum of *X* depends on *Y* (Figure 2.2d). This scenario is observed, for example, when the fitness again corresponds to the dynamic range of the output but in response to input signals with variations over a smaller range. By the same geometric reasoning as above, there are starting phenotypes leading to a decrease in fitness when mutating *x* first or when mutating *y* first. This “or” relationship can become an “and”, as exemplified in the mutational trajectory of Figure 2.2d, where reaching a better combination X-Y in a stepwise manner requires first decreasing the performance of the cascade, independently of the mutated phenotype. This scenario of reciprocal sign epistasis can be understood as a purely functional version of lock-key constraints. It should be noted that geometric models exist where both x-related and y-related sign epistasis exist, but they do not combine into reciprocal sign epistasis. For example, as shown in Reference (Nghe et al., 2018), slightly tilted Gaussian geometric models do not combine mere sign epistasis into reciprocal sign epistasis, but strongly tilted Gaussian models do.

Hierarchy within a cascade can be applied to interpretation of epistasis in a negative feedback loop in the galactose regulatory system in yeast (Peng et al., 2015): Mutants of the downstream



Predicted sign epistasis in a regulatory cascade. (**a**) A transcriptional cascade, where an input signal, such as an inducer, modulates the expression level of the upstream gene *y*, which itself expresses a transcription factor that regulates a downstream gene *x*, which in turn regulates the level of an output gene. Systematic combinations of mutations in genes *x* and *y* allow testing for the effect of combined changes in the phenotypic parameters respectively called X and Y, as has been explored in Reference (Nghe et al., 2018). For example, X may be the binding constant of regulatory protein x to the promoter of the output gene. (**b**) Scenario yielding magnitude epistasis. (Top) The performance or fitness of the system is measured as the output level in response to a single input that is fixed in time, as it would be in a constant environment. (Middle) The corresponding phenotype-to-fitness relationship is computed as a function of parameters X and Y, for example, the binding constants of the transcription factors x and y to their target promoters, using a mathematical model (Nghe et al., 2018). Red arrows represent the two ways by which mutations leading from X_1 to X_2 and from Y_1 to Y_2 can be combined to optimize fitness. Importantly, given the independent effects of mutations on X and Y, trajectories are parallel to the axis. Given that, in this case, fitness increases with both X and Y, the maximum values of X as a function of Y and of Y as a function of X, respectively denoted X^{opt} and Y^{opt}, are the straight thick gray lines on the right-hand side and top of the landscape, respectively. (Bottom) Given that X^{opt} and Y^{opt} are independent of, respectively, Y and X, the geometric model can only generate magnitude epistasis. (**c**) Scenario yielding regular sign epistasis. (Top) Fitness is the total output range $F=out_{max} - out_{min}$ in response to a variable environment providing wide input variations, with the input going down to zero (signal is absent). ***(Caption continues on next page).***

(Continued from previous page). (Middle) Fitness is optimal for intermediate values X^{opt} of X for fixed Y and X^{opt} is the thick gray curved line that varies as a function of Y. In the example mutational trajectories (red arrows), one path leads to a decreasing step when mutating X first (circled minus sign). (Bottom) The starting phenotype X_1 is optimal given Y_1; thus, mutating it can only lead to decreased fitness, causing the sign epistasis. (**d**) Scenario yielding reciprocal sign epistasis. (Top) The output is evaluated in response to a more restricted range of input signals (the minimum input does not reach zero). (Middle) Both X^{opt} and Y^{opt} are curved, and can lead to decreasing fitness when mutating X or Y first. (Bottom) The mutual dependence of the optima of X and Y on each other's values can generate reciprocal sign epistasis patterns.

GAL80 gene mask mutations in the upstream *GAL3* gene. This corresponds to the geometric model in Figure 2.2b, where the output corresponds to the intensity of the feedback, and mutations cause strong knock-down effects. Epistatic effects from regulatory structure also appear when integrating signals at the same promoter, as studied by mutating a *lambda* phage promoter repressed by the protein CI (Lagator, Sarikas, et al., 2017). In this case, loss-of-functions in polymerase recruitment mask loss-of-functions in CI repression, but CI binding allows tuning expression in the presence of mutated but functional polymerase recruitment, corresponding again to the scheme in Figure 2.2b. The main point to note is that epistasis is predominantly explained by the regulatory logic, as opposed to the pleiotropy caused by physical interactions (Lagator, Sarikas, et al., 2017), indicating a modularity in mutational effects, even within a single regulatory sequence.

Epistasis caused by the functional dependence between the components of a network is also observed in incoherent feedforward network motifs (Schaerli et al., 2018), which integrate an environmental signal via two regulatory loops into a single output promoter that defines the phenotype. In this case, the mutational effects were measured in regulatory loops comprising either a double activator or a double repressor cascade. When mutations were introduced into the cis-regulatory regions of the networks at each node separately, not all of the possible phenotypic states were accessible. This suggests that the optimal phenotypic parameters of one gene depend on the phenotypic parameters of other genes. However, when mutations were introduced in the cis-regulatory regions of the individual nodes simultaneously, this barrier could be overcome through epistasis between the cis-regulatory region mutations. Overall, these results support the idea that epistasis originates from the tuning of genetically independent phenotypic parameters with respect to each other.

Overall, we have seen how to use functional dependences to identify causes of sign epistasis: Sign epistasis arises when the optimum of a module needs to be adjusted to the state of another module. Generalization to a large number of phenotypic dimensions is possible (Nghe et al., 2018) and suggests that phenotypes whose phenotypic optima are mutually dependent (independent) generate ruggedness (smoothness) in genotype-to-fitness landscapes. The interplay between phenotypes and their optimality is also crucial during network rewiring, where the accessibility of evolutionary paths requires finely tuned steps to preserve function (Sorrells et al., 2015). So far, most mutational scanning studies can be interpreted in large part through loss of function within networks. Revealing sign epistasis systematically will instead require the combination of mutations with mild phenotypic effects in the relative proximity of their optima (Gros et al., 2009).

2.4. Epistasis between networks

In the previous sections, we discuss how epistasis constrains the evolution of regulatory interactions and networks. A focus on the physical (Diss & Lehner, 2018; Starr & Thornton, 2016) or functional (Boucher & Jenna, 2013; Yang et al., 2017) features of these regulatory systems allowed prediction of epistasis within a network of interest. Another situation arises when selection drives interactions between networks. An interesting example is a recent experimental evolution study in *Saccharomyces cerevisiae* (Laan et al., 2015). In *S. cerevisiae*, formation of a polarized spot of the GTPase Cdc42, a protein that cycles between an active GTP-bound state and an inactive GDP-bound state, is an essential part of the cell cycle. As polarization serves as a paradigm for symmetry breaking, a large amount of experimental and theoretical work has been dedicated to identifying the major components of this network and their interactions (Halatek et al., 2018; S. G. Martin, 2015). The knowledge that this work has generated makes this network an attractive system to study adaptive pathways. Laan et al. (2015) applied a strong perturbation to this module by deleting a scaffold protein with a central role in symmetry breaking. The subsequent adaptation to the loss of this component was restricted by sign epistasis, which constrained the order in which other network components were mutated, leading to reproducible mutational pathways across multiple parallel lineages. Surprisingly, this mutational pathway consisted of the inactivation of proteins rather than changes to their biochemical properties, which was confirmed by reconstructing the mutational pathway observed from natural evolution by synthetically deleting the genes. Another puzzling finding was that the epistatic interactions that determined the adaptive pathway were not exclusively among known network components, but also included a protein that was not considered to be part of the polarization network. Thus, despite its well-studied nature, information about the interaction network was insufficient to explain the relevant epistasis. This example highlights a generic challenge when the system increases in size. Many selective pressures, particularly in response to strong perturbations, elicit mutations throughout the genome, which limits predictive approaches.

The above discussion indicates that a trait can be restored by co-opting networks with some functional overlap, rather than by reconstructing lost components. How the particular gene deletions that facilitate this process in the polarity network do so remains elusive, but it is known that networks involved in other cellular processes can affect polarity establishment (Slaughter et al., 2013; Woods et al., 2016) This mode of adaptation appears to not be specific to the example described above: A large-scale study that tracked the adaptive response of *S. cerevisiae* to 187 different single-gene knock-outs showed that compensatory mutations following gene deletions often partially restore affected traits without restoring the original genomic expression pattern (Szamecz et al., 2014). Thus, evolution can exploit other networks to rescue perturbed cellular functions (Hottes et al., 2013; Rojas Echenique et al., 2019) by changing the connections between redundant and connected networks, rather than by restoring the original network (Figure 2.3a-d).

Accessible pathways thus do not always depend only on the topology of the perturbed network, but also on that of compensating networks and on their interconnections. When insight into the relevant network features is incomplete, it is not straightforward to functionally explain the observed epistasis, let alone offer predictions. Another consequence of networks being interconnected is that they typically give rise to pleiotropy: Not only is one trait affected by multiple networks or mutations, but one network or mutation also affects multiple traits. Such pleiotropic properties have long been thought to have important consequences for evolving populations: For example, they are considered to maintain genetic variation (Zhang & Hill, 2005), produce

trade-offs (Cooper & Lenski, 2000; MacLean et al., 2004) and be an important cause of the persistence of genetic diseases (Amell et al., 2018; Carter & Nguyen, 2011).

The questions that naturally arise are: How prevalent is pleiotropy in cellular networks? Which adaptive processes cannot be considered without the context of the entire cellular interaction network? The view of universal pleiotropy, where a mutation can potentially affect all selectable traits, is implicit in the original geometric model from Fisher (Fisher, 1930) and has been the dominant view for many years (Stearns, 2010). Within this view, one may consider whether the number of evolutionary constraints increases as the number of traits increases, seemingly contradicting the emergence and adaptability of complex organisms (Orr, 2000). Due to practical challenges in empirically determining the degree of pleiotropy in organisms, the discussion remained mainly based on theoretical models. However, new molecular biology techniques and interaction network databases have sparked attempts to empirically quantify the number of pleiotropic genes in actual biological systems. For example, Wang et al. (2010) used large data sets containing effects of gene deletions on different traits to determine the total pleiotropy on a genome-wide scale for *S. cerevisiae*, *C. elegans* and *M. musculus* (Wang et al., 2010). Comparison to random gene-trait relationships suggested that cellular networks are mostly modular, with only a small percentage of traits (1-9% of the traits included for analysis) affected by pleiotropic genes. This result would advocate the view of modular pleiotropy, where sets of traits covary, and only a subset of genes within a network exhibit pleiotropic effects.

However, a common objection is that pleiotropic interactions can remain undetected due to experimental noise and detection limits. The absence of a standardized methodology for extracting statistically significant interactions from experimental data has led to significant variations in the estimates of genes with pleiotropic interactions, with a possible bias for modular pleiotropy (Hill & Zhang, 2012). The development of new, more sensitive methods for the extraction of significant interactions from databases is required to settle this debate (Koch et al., 2017; Tyler et al., 2016).

Although the discussion on the extent of pleiotropy is far from resolved, the notion that pleiotropic interactions exist and impact evolutionary trajectories is widely accepted (Wagner & Zhang, 2011). As discussed above, pleiotropy can lead to epistatic interactions between components of different networks, and both epistasis and pleiotropy make evolution dependent on the genetic background. To understand how pleiotropy affects evolution, it is useful to focus on this interplay between pleiotropic and epistatic interactions. Epistasis is broadly believed to guide evolution by imposing constraints, and from a network perspective, antagonistic pleiotropy (AP) is considered to play a central role in constraining the evolution of networks. The negative correlation between different traits in AP can make it difficult to simultaneously optimize multiple traits and can thus restrict the number of accessible mutational pathways (Figure 2.3d). The extent of AP in yeast was examined by Qian et al. (2012). They tested 4,642 nonessential genes for antagonistic pleiotropy by performing competition experiments of null mutants together with the wild type in different environmental conditions (Qian et al., 2012). At least 13.6% of the analyzed genes displayed AP in the considered environments, indicating its importance in restricting mutational pathways. Interestingly, they found signs that antagonistic interactions between networks could be mitigated by changes in trans-regulatory molecules that regulate gene expression, rather than in DNA regulatory or coding sequences. The ability of trans-regulatory molecules to serve as a source for alleviating constraints is surprising, considering that their evolution is typically regarded to be heavily constrained itself due to their extensive interaction networks (Prud'homme et al., 2007;

Voordeckers et al., 2015).

How *trans*-acting elements can resolve AP and alleviate constraints is illustrated by a study of the pheromone response pathways between different yeast species (Sorrells et al., 2015). Haploid yeast cells can exist in two different mating types, **a** and **α**. In response to sensing pheromones of the opposite mating type, each of these upregulates the expression of genes required for the pheromone response pathway. Although the upregulation of some of these genes is mating type specific (expression in **a** or **α**), a large portion overlaps in both mating types (expression in **a** and **α**). In both mating types, the upregulation of these genes is induced by the conserved transcription factor Ste12. Despite the conserved function of Ste12, Sorrells et al. (2015) found that different species maintained different network structures for the upregulation of mating type specific genes: In *Saccharomyces*, mating type **a** specific genes (**a**sgs) contain motifs for the direct binding of Ste12, while *Kluyveromyces* and *Candida* required transcriptional coregulators for induction of the same genes. However, evolution of a regulatory network for **a**sgs where Ste12 is recruited by coregulators to one where Ste12 directly binds **a**sg promoters appeared to be inaccessible: Introduction of the Ste12 binding sites of the *Saccharomyces* clade into the *Kluyveromyces* clade resulted in a loss of regulation (Figure 2.3e-f). Instead, the addition of a repressor for **a**sg expression to the regulatory network of mating type **α** cells was required prior to introducing Ste12 binding sites to prevent their (mis)expression in **α** cells (Figure 2.3g-i).

Note that understanding why the pheromone response pathway in **a** cells cannot directly evolve additional binding sites for **a**sgs requires knowledge of the structure of the pheromone response pathway in **α** cells. Otherwise, the requirement for evolving a repressor for **a**sgs would appear as a hidden parameter in the epistatic landscape, similar to those that appear as hidden parameters for the polarization network of *S. cerevisiae* described above. The way in which interactions between these pheromone response networks constrain evolution shows surprising similarities to sign epistasis between components within a single module: The adaptation in one network becomes beneficial only when the structure of the other network changes (Figure 2.3e). It is tempting to consider the analogy between AP and sign epistasis, but how concepts from epistasis relate to pleiotropy in networks remains to be investigated. Apart from AP, which constrains evolution, interactions that result in a positive correlation in the fitness level of different traits have been found to drive coevolution of traits that are not under selection. For example, Jerison et al. (2017) found that populations of budding yeast adapting to growth at high temperature also improved growth at standard growth temperatures, although the molecular basis of this was not elaborated. This shows how pleiotropic networks and the environment can interact to give surprising evolutionary results. Unraveling these interactions at the molecular level, similar to what has been done for epistatic interactions of the *lac* operon (de Vos et al., 2015), can reveal new concepts that would explain adaptive pathways at both the inter- and the intranetwork levels.

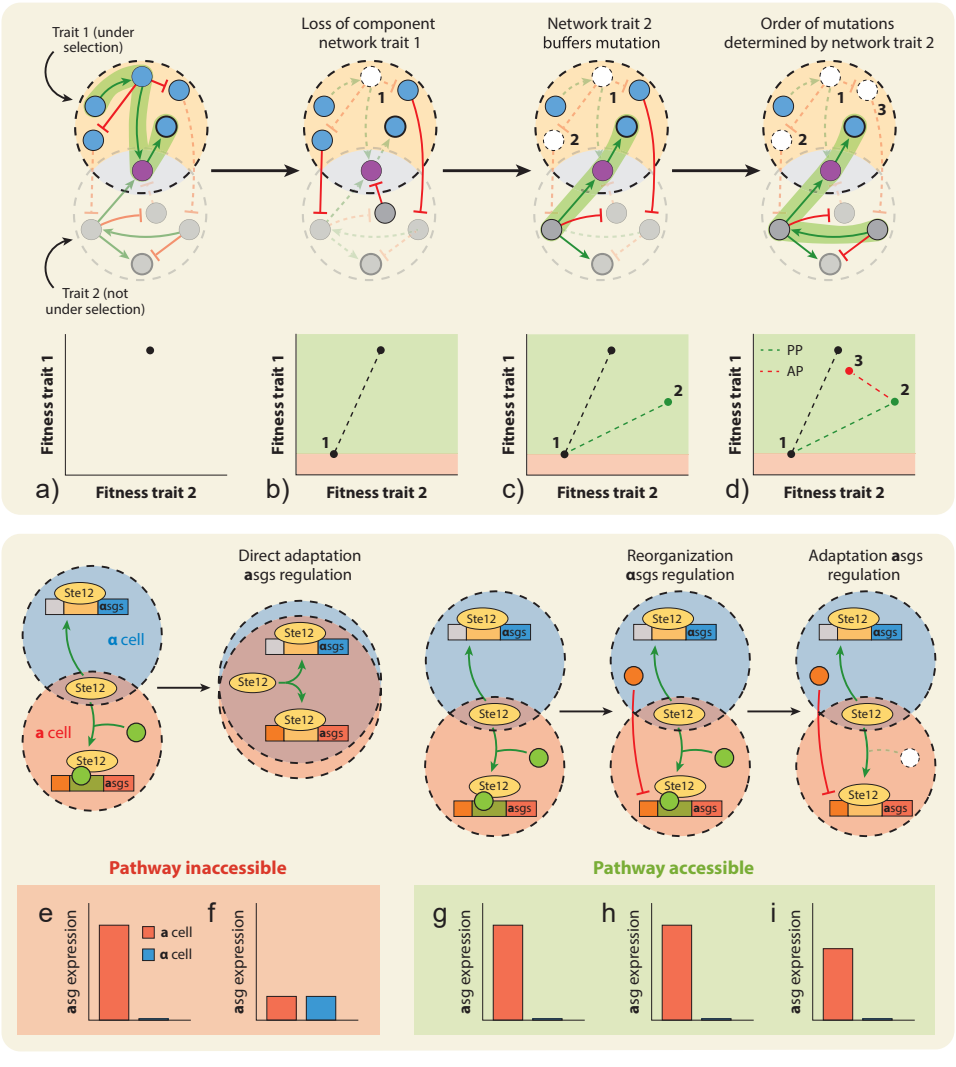


Figure 2.3. Pleiotropic interactions can facilitate and constrain evolution. (**a-d**) When the networks of two traits share interactions, but only one network is under selection, the network not under selection can compensate for mutations in the other network. (**a**) The initial pathway (green highlight) leads to activation of a trait-defining node (encircled blue node) through a shared component (purple node). (**b**) Deletion of a central component in the network of trait 1 decreases fitness of the trait. (**c**) The network of trait 2 can buffer this deletion by taking over some of the interactions of network 1 through the deletion of other components, (**d** with the order in which these deletions can take place being constrained by the interactions between the network of trait 1 and the network of trait 2, which can provide a reconstruction the activation pathway. Only pathways that improve trait 1 are accessible (green shaded area); they can have both positive pleiotropic (PP, from mutation 1 to 2) or antagonistic pleiotropic (AP, from mutation 2 to 3) effects on trait 2. (**e-i**) When both traits are under selection, connected networks can lead to additional constraints. (**e**) The two transcription networks regulating mating pathways in yeast both use the transcription factor Ste12, causing them to partially overlap. (**f**) This overlap prevents the evolution of Ste12 binding motifs in a cells, as this causes misexpression of asgs in αcells. (**g-i**) To maintain correct regulation of asgs during evolution, the transcription network of αcells must first reorganize to obtain an inhibitor for asg expression, which releases the constraint on the evolution of Ste12 binding motifs in a cells.

2.5. Conclusions

It is evident that evolutionary constraints are inherently interconnected with phenotypes. Indeed, the genetic interactions that underlie constraint are quantified by their impact on phenotypes and fitness (De Visser & Krug, 2014; Lobkovsky & Koonin, 2012). As highlighted in this chapter, epistasis offers a route to predict evolutionary constraint and potential – one of the major goals of evolution research. At the same time, owing to the overwhelming complexity of this phenotypic puzzle, and its many missing pieces, such prediction insights have been difficult to achieve. This humbling reality remains to a large extent, as is also clear from the studies reviewed above. At all levels of biological organization that we address, one encounters unknowns that pose limits to general predictive frameworks. For instance, it is not known if the presence of adaptive valleys is general for molecular interactions other than the well-studied *lac* system, how the geometric landscape prediction method can be applied to naturally occurring pathways, and which redundant networks will be able to compensate for the loss of core cellular functions.

Nonetheless, this new wave of quantitative studies provided the first tools to predict key epistasis features. Within the wide range of studied systems, notable parallels and differences were observed. Specific recognition was found to produce reciprocal sign epistasis, but its limiting effects on adaptation could be mitigated by environmental interactions. The last section showed that interactions between networks can overcome constraint by providing components that are coopted into altered functions. In both cases, additional interactions could alleviate existing constraints, even as the type of interaction differed. This makes intuitive sense, since those extra interactions (with the environment and with other networks) can be seen as dimensions that are orthogonal to the initial genotypic space, and thus can allow escape from suboptima. Antagonistic pleiotropy played a role in both processes, with mutations having contrasting effects on different phenotypes and networks.

The level of prediction was found to depend on the scale of the system, as well as on the selective pressures acting on it. At the smallest scale – that of molecular binding – epistasis was merely implied by the specific nature of the recognition and ultimately depended on the details of the molecular binding interface; the resulting genotype-phenotype map required experimental reconstruction (de Vos et al., 2015). Notably, such molecular details no longer appeared relevant at the intermediate scale of a single pathway, where epistasis instead emerged in a more predictable fashion from the network topology (Lagator, Paixão, et al., 2017; Lagator, Sarikas, et al., 2017; Nghe et al., 2018). The key changes in binding affinity could be achieved by a wide spectrum of mutations, suggesting that a coarse-grained description suffices when studying evolution at the network scale. This is a promising realization: As larger systems are considered (Rojas Echenique et al., 2019), the basis of constraint does not necessarily become less predictable. However, for the largest scale considered in this chapter – that of networks interacting in core cellular functions – it was shown that predictions are more challenging because such networks are more interconnected with unknown factors (Laan et al., 2015; Szamecz et al., 2014).

Approaches similar to the ones discussed here, which exploit any functional information for the prediction of constraint, can be applied more broadly. For instance, it is intriguing to consider epistatic constraints in RNA molecules (Li & Zhang, 2018) and metabolic networks (Bachmann et al., 2017), as well as in the evolution of complex traits (Carlborg & Haley, 2004), ecosystems (Kassen & Rainey, 2004) and disease (Moore, 2003).

Bibliography

Aguilar-Rodríguez, J., Payne, J. L., & Wagner, A. (2017). A thousand empirical adaptive landscapes and their navigability. *Nature Ecology & Evolution*, *1*(2), 0045. https://doi.org/10.1038/s41559-016-0045

Amell, A., Roso-Llorach, A., Palomero, L., Cuadras, D., Galvan-Femenia, I., Serra-Musach, J., Comellas, F., de Cid, R., Pujana, M. A., & Violan, C. (2018). Disease networks identify specific conditions and pleiotropy influencing multimorbidity in the general population. *Sci Rep*, *8*(1), 15970. https://doi.org/10.1038/s41598-018-34361-3

Bachmann, H., Molenaar, D., Branco dos Santos, F., & Teusink, B. (2017). Experimental evolution and the adjustment of metabolic strategies in lactic acid bacteria. *FEMS Microbiology Reviews*, *41*(Supp_1), S201–S219. https://doi.org/10.1093/femsre/fux024

Balleza, E., López-Bojorquez, L. N., Martínez-Antonio, A., Resendis-Antonio, O., Lozada-Chávez, I., Balderas-Martínez, Y. I., Encarnación, S., & Collado-Vides, J. (2009). Regulation by transcription factors in bacteria: beyond description. *FEMS Microbiology Reviews*, *33*(1), 133–151. https://doi.org/10.1111/j.1574-6976.2008.00145.x

Beardmore, R. E., Gudelj, I., Lipson, D. A., & Hurst, L. D. (2011). Metabolic trade-offs and the maintenance of the fittest and the flattest. *Nature*, *472*(7343), 342–346. https://doi.org/10.1038/nature09905

Bell, M. A., Futuyma, D. J., Eanes, W. F., & Levinton, J. S. (2010). *Evolution since Darwin : the first 150 years*. Sinauer Associates.

Boucher, B., & Jenna, S. (2013). Genetic interaction networks: better understand to better predict. *Frontiers in Genetics*, *4*(290), 290. https://doi.org/10.3389/fgene.2013.00290

Bridgham, J. T., Carroll, S. M., & Thornton, J. W. (2006). Evolution of hormone-receptor complexity by molecular exploitation. *Science*, *312*(5770), 97–101. https://doi.org/10.1126/science.1123348

Carlborg, Ö., & Haley, C. S. (2004). Epistasis: too often neglected in complex trait studies? *Nature Reviews Genetics*, *5*(8), 618–625. https://doi.org/10.1038/nrg1407

Carter, A. J., & Nguyen, A. Q. (2011). Antagonistic pleiotropy as a widespread mechanism for the maintenance of polymorphic disease alleles. *BMC Medical Genetics*, *12*(1), 160. https://doi.org/10.1186/1471-2350-12-160

Cooper, V. S., & Lenski, R. E. (2000). The population genetics of ecological specialization in evolving Escherichia coli populations. *Nature*, *407*(6805), 736–739. https://doi.org/10.1038/35037572

Corson, F., & Siggia, E. D. (2012). Geometry, epistasis, and developmental patterning. *Proceedings of the National Academy of Sciences*, *109*(15), 5568–5575. https://doi.org/10.1073/pnas.1201505109

Cotterell, J., & Sharpe, J. (2013). Mechanistic Explanations for Restricted Evolutionary Paths That Emerge from Gene Regulatory Networks (A. Palsson, Ed.). *PLoS ONE*, *8*(4), e61178. https://doi.org/10.1371/journal.pone.0061178

De Visser, J. A. G. M., & Krug, J. (2014). Empirical fitness landscapes and the predictability of evolution. https://doi.org/10.1038/nrg3744

de Vos, M. G. J., Dawid, A., Sunderlikova, V., & Tans, S. J. (2015). Breaking evolutionary constraint with a tradeoff ratchet. *Proceedings of the National Academy of Sciences*, *112*(48), 14906–14911. https://doi.org/10.1073/pnas.1510282112

de Vos, M. G. J., Schoustra, S. E., & de Visser, J. A. G. M. (2018). Ecology dictates evolution? About the importance of genetic and ecological constraints in adaptation. *EPL (Europhysics Letters)*, *122*(5), 58002. https://doi.org/10.1209/0295-5075/122/58002

Dill, K. A., Ghosh, K., & Schmit, J. D. (2011). Physical limits of cells and proteomes. *Proceedings of the National Academy of Sciences*, *108*(44), 17876–17882. https://doi.org/10.1073/pnas.1114477108

Diss, G., & Lehner, B. (2018). The genetic landscape of a physical interaction (N. Barkai, Ed.). *eLife*, *7*, e32472. https://doi.org/10.7554/eLife.32472

Fisher, R. A. (1930). *The genetical theory of natural selection*. Clarendon Press. https://doi.org/10.5962/bhl.title.27468

Flynn, K. M., Cooper, T. F., Moore, F. B.-G., & Cooper, V. S. (2013). The Environment Affects Epistatic Interactions to Alter the Topology of an Empirical Fitness Landscape (J. C. Fay, Ed.). *PLoS Genetics*, *9*(4), e1003426. https://doi.org/10.1371/journal.pgen.1003426

Gavrilets, S., & Hastings, A. (1993). Maintenance of genetic variability under strong stabilizing selection: a two-locus model. *Genetics*, *134*(1), 377–386. https://doi.org/10.1093/genetics/134.1.377

Gorter, F. A., Aarts, M. G. M., Zwaan, B. J., & De Visser, J. A. G. M. (2018). Local fitness landscapes predict yeast evolutionary dynamics in directionally changing environments. *Genetics*, *208*(1), 307–322. https://doi.org/10.1534/GENETICS.117.300519/-/DC1

Gros, P.-A., Le Nagard, H., & Tenaillon, O. (2009). The Evolution of Epistasis and Its Links With Genetic Robustness, Complexity and Drift in a Phenotypic Model of Adaptation. *Genetics*, *182*(1), 277–293. https://doi.org/10.1534/genetics.108.099127

Halatek, J., Brauns, F., & Frey, E. (2018). Self-organization principles of intracellular pattern formation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1747), 20170107. https://doi.org/10.1098/rstb.2017.0107

Hill, W. G., & Zhang, X.-S. (2012). Assessing pleiotropy and its evolutionary consequences: pleiotropy is not necessarily limited, nor need it hinder the evolution of complexity. *Nature Reviews Genetics*, *13*(4), 296–296. https://doi.org/10.1038/nrg2949-c1

Hottes, A. K., Freddolino, P. L., Khare, A., Donnell, Z. N., Liu, J. C., & Tavazoie, S. (2013). Bacterial Adaptation through Loss of Function (I. Matic, Ed.). *PLoS Genetics*, *9*(7), e1003617. https://doi.org/10.1371/journal.pgen.1003617

Jerison, E. R., Kryazhimskiy, S., Mitchell, J. K., Bloom, J. S., Kruglyak, L., & Desai, M. M. (2017). Genetic variation in adaptability and pleiotropy in budding yeast. *eLife*, *6*. https://doi.org/10.7554/eLife.27167

Kassen, R., & Rainey, P. B. (2004). The Ecology and Genetics of Microbial Diversity. *Annual Review of Microbiology*, *58*(1), 207–231. https://doi.org/10.1146/annurev.micro.58.030603.123654

Koch, E. N., Costanzo, M., Deshpande, R., Andrews, B., Boone, C., & Myers, C. L. (2017). Systematic identification of pleiotropic genes from genetic interactions. *bioRxiv*, 112326. https://doi.org/10.1101/112326

Laan, L., Koschwanez, J. H., & Murray, A. W. (2015). Evolutionary adaptation after crippling cell polarization follows reproducible trajectories. *eLife*, *4*. https://doi.org/10.7554/elife.09638

Lagator, M., Paixão, T., Barton, N. H., Bollback, J. P., & Guet, C. C. (2017). On the mechanistic nature of epistasis in a canonical cis-regulatory element (A. Sanchez, Ed.; 2017/05/19). *eLife*, *6*, e25192. https://doi.org/10.7554/eLife.25192

Lagator, M., Sarikas, S., Acar, H., Bollback, J. P., & Guet, C. C. (2017). Regulatory network structure determines patterns of intermolecular epistasis (P. J. Wittkopp, Ed.; 2017/11/14). *eLife*, *6*, e28921. https://doi.org/10.7554/eLife.28921

Lehming, N., Sartorius, J., Kisters-Woike, B., von Wilcken-Bergmann, B., & Müller-Hill, B. (1990). Mutant lac repressors with new specificities hint at rules for protein–DNA recognition. *The EMBO Journal*, *9*(3), 615–621. https://doi.org/10.1002/j.1460-2075.1990.tb08153.x

Lehner, B. (2011). Molecular mechanisms of epistasis within and between genes. *Trends in Genetics*, *27*(8), 323–331. https://doi.org/10.1016/J.TIG.2011.05.007

Li, C., & Zhang, J. (2018). Multi-environment fitness landscapes of a tRNA gene. *Nature Ecology & Evolution*, *2*(6), 1025–1032. https://doi.org/10.1038/s41559-018-0549-8

Lindsey, H. A., Gallie, J., Taylor, S., & Kerr, B. (2013). Evolutionary rescue from extinction is contingent on a lower rate of environmental change. *Nature*, *494*(7438), 463–467. https://doi.org/10.1038/nature11879

Lobkovsky, A. E., & Koonin, E. V. (2012). Replaying the Tape of Life: Quantification of the Predictability of Evolution. *Frontiers in Genetics*, *3*, 246. https://doi.org/10.3389/fgene.2012.00246

Lunzer, M., Miller, S. P., Felsheim, R., & Dean, A. M. (2005). The Biochemical Architecture of an Ancient Adaptive Landscape. *Science*, *310*(5747), 499–501. https://doi.org/10.1126/science.1115649

MacLean, R. C., Bell, G., & Rainey, P. B. (2004). The evolution of a pleiotropic fitness tradeoff in Pseudomonas fluorescens. *Proceedings of the National Academy of Sciences*, *101*(21), 8072–8077. https://doi.org/10.1073/pnas.0307195101

Martin, G. (2014). Fisher's geometrical model emerges as a property of complex integrated phenotypic networks. *Genetics*, *197*(1), 237–255. https://doi.org/10.1534/GENETICS.113.160325/-/DC1

Martin, G., Elena, S. F., & Lenormand, T. (2007). Distributions of epistasis in microbes fit predictions from a fitness landscape model. *Nature Genetics*, *39*(4), 555–560. https://doi.org/10.1038/ng1998

Martin, S. G. (2015). Spontaneous cell polarization: Feedback control of Cdc42 GTPase breaks cellular symmetry. *BioEssays*, *37*(11), 1193–1201. https://doi.org/10.1002/bies.201500077

Moore, J. H. (2003). The Ubiquitous Nature of Epistasis in Determining Susceptibility to Common Human Diseases. *Human Heredity*, *56*(1-3), 73–82. https://doi.org/10.1159/000073735

Mustonen, V., & Lässig, M. (2009). From fitness landscapes to seascapes: non-equilibrium dynamics of selection and adaptation. *Trends in Genetics*, *25*(3), 111–119. https://doi.org/10.1016/j.tig.2009.01.002

Nghe, P., Kogenaru, M., & Tans, S. J. (2018). Sign epistasis caused by hierarchy within signalling cascades (2018/04/15). *Nature Communications*, *9*(1), 1451. https://doi.org/10.1038/s41467-018-03644-8

Orr, H. A. (2000). ADAPTATION AND THE COST OF COMPLEXITY. *Evolution*, *54*(1), 13–20. https://doi.org/10.1111/j.0014-3820.2000.tb00002.x

Peng, W., Liu, P., Xue, Y., & Acar, M. (2015). Evolution of gene network activity by tuning the strength of negative-feedback regulation. *Nature Communications*, *6*(1), 6226. https://doi.org/10.1038/ncomms7226

Peter, I. S., & Davidson, E. H. (2011). Evolution of Gene Regulatory Networks Controlling Body Plan Development (2011/03/19). *Cell*, *144*(6), 970–985. https://doi.org/10.1016/j.cell.2011.02.017

Phillips, P. C. (2008). Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Reviews Genetics*, *9*(11), 855–867. https://doi.org/10.1038/nrg2452

Podgornaia, A. I., & Laub, M. T. (2015). Pervasive degeneracy and epistasis in a protein-protein interface. *Science*, *347*(6222), 673–677. https://doi.org/10.1126/science.1257360

Poelwijk, F. J., de Vos, M. G., & Tans, S. J. (2011). Tradeoffs and Optimality in the Evolution of Gene Regulation. *Cell*, *146*(3), 462–470. https://doi.org/10.1016/j.cell.2011.06.035

Poelwijk, F. J., Heyning, P. D., de Vos, M. G., Kiviet, D. J., & Tans, S. J. (2011). Optimality and evolution of transcriptionally regulated gene expression. *BMC Systems Biology*, *5*(1), 128. https://doi.org/10.1186/1752-0509-5-128

Poelwijk, F. J., Kiviet, D. J., Weinreich, D. M., & Tans, S. J. (2007). Empirical fitness landscapes reveal accessible evolutionary paths. *Nature*, *445*(7126), 383–386. https://doi.org/10.1038/nature05451

Poelwijk, F. J., Socolich, M., & Ranganathan, R. (2019). Learning the pattern of epistasis linking genotype and phenotype in a protein. *Nature Communications*, *10*(1), 4213. https://doi.org/10.1038/s41467-019-12130-8

Poelwijk, F. J., Tănase-Nicola, S., Kiviet, D. J., & Tans, S. J. (2011). Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. *Journal of Theoretical Biology*, *272*(1), 141–144. https://doi.org/10.1016/j.jtbi.2010.12.015

Prud'homme, B., Gompel, N., & Carroll, S. B. (2007). Emerging principles of regulatory evolution. *Proceedings of the National Academy of Sciences*, *104*(suppl_1), 8605–8612. https://doi.org/10.1073/pnas.0700488104

Qian, W., Ma, D., Xiao, C., Wang, Z., & Zhang, J. (2012). The Genomic Landscape and Evolutionary Resolution of Antagonistic Pleiotropy in Yeast. *Cell Reports*, *2*(5), 1399–1410. https://doi.org/10.1016/j.celrep.2012.09.017

Rojas Echenique, J. I., Kryazhimskiy, S., Nguyen Ba, A. N., & Desai, M. M. (2019). Modular epistasis and the compensatory evolution of gene deletion mutants. *PLOS Genetics*, *15*(2), e1007958. https://doi.org/10.1371/JOURNAL.PGEN.1007958

Sartorius, J., Lehming, N., Kisters, B., von Wilcken-Bergmann, B., & Müller-Hill, B. (1989). lac repressor mutants with double or triple exchanges in the recognition helix bind specifically to lac operator variants with multiple exchanges. *The EMBO Journal*, *8*(4), 1265–1270. https://doi.org/10.1002/j.1460-2075.1989.tb03500.x

Schaerli, Y., Jiménez, A., Duarte, J. M., Mihajlovic, L., Renggli, J., Isalan, M., Sharpe, J., & Wagner, A. (2018). Synthetic circuits reveal how mechanisms of gene regulatory networks constrain evolution. *Molecular Systems Biology*, *14*(9), e8102. https://doi.org/10.15252/msb.20178102

Segrè, D., DeLuna, A., Church, G. M., & Kishony, R. (2005). Modular epistasis in yeast metabolism. *Nature Genetics*, *37*(1), 77–83. https://doi.org/10.1038/ng1489

Slaughter, B. D., Unruh, J. R., Das, A., Smith, S. E., Rubinstein, B., & Li, R. (2013). Non-uniform membrane diffusion enables steady-state cell polarization via vesicular trafficking. *Nature Communications*, *4*(1), 1380. https://doi.org/10.1038/ncomms2370

Sorrells, T. R., Booth, L. N., Tuch, B. B., & Johnson, A. D. (2015). Intersecting transcription networks constrain gene regulatory evolution. *Nature*, *523*, 361–365. https://doi.org/10.1038/nature14613

Starr, T. N., Picton, L. K., & Thornton, J. W. (2017). Alternative evolutionary histories in the sequence space of an ancient protein. *Nature*, *549*(7672), 409–413. https://doi.org/10.1038/nature23902

Starr, T. N., & Thornton, J. W. (2016). Epistasis in protein evolution. *Protein Science*, *25*(7), 1204–1218. https://doi.org/10.1002/pro.2897

Stearns, F. W. (2010). One Hundred Years of Pleiotropy: A Retrospective. *Genetics*, *186*(3), 767–773. https://doi.org/10.1534/genetics.110.122549

Steinberg, B., & Ostermeier, M. (2016). Environmental changes bridge evolutionary valleys. *Science Advances*, *2*(1), e1500921. https://doi.org/10.1126/sciadv.1500921

Szamecz, B., Boross, G., Kalapis, D., Kovács, K., Fekete, G., Farkas, Z., Lázár, V., Hrtyan, M., Kemmeren, P., Groot Koerkamp, M. J. A., Rutkai, E., Holstege, F. C. P., Papp, B., & Pál, C. (2014). The Genomic Landscape of Compensatory Evolution (N. H. Barton, Ed.). *PLoS Biology*, *12*(8), e1001935. https://doi.org/10.1371/journal.pbio.1001935

Szathmáry, E. (1993). Do deleterious mutations act synergistically? Metabolic control theory provides a partial answer. *Genetics*, *133*(1), 127–132. https://doi.org/10.1093/genetics/133.1.127

Tyler, A. L., Crawford, D. C., & Pendergrass, S. A. (2016). The detection and characterization of pleiotropy: discovery, progress, and promise. *Briefings in Bioinformatics*, *17*(1), 13–22. https://doi.org/10.1093/bib/bbv050

Voordeckers, K., Pougach, K., & Verstrepen, K. J. (2015). How do regulatory networks evolve and expand throughout evolution? *Current Opinion in Biotechnology*, *34*, 180–188. https://doi.org/10.1016/j.copbio.2015.02.001

Wagner, G. P., & Zhang, J. (2011). The pleiotropic structure of the genotype–phenotype map: the evolvability of complex organisms. *Nature Reviews Genetics*, *12*(3), 204–213. https://doi.org/10.1038/nrg2949

Wang, Z., Liao, B.-Y., & Zhang, J. (2010). Genomic patterns of pleiotropy and the evolution of complexity. *Proceedings of the National Academy of Sciences*, *107*(42), 18034–18039. https://doi.org/10.1073/pnas.1004666107

Weinreich, D. M. (2006). Darwinian Evolution Can Follow Only Very Few Mutational Paths to Fitter Proteins. *Science*, *312*(5770), 111–114. https://doi.org/10.1126/science.1123539

Weinreich, D. M., Watson, R. A., & Chao, L. (2005). PERSPECTIVE:SIGN EPISTASIS AND GENETIC CONSTRAINT ON EVOLUTIONARY TRAJECTORIES. *Evolution*, *59*(6), 1165–1174. https://doi.org/10.1554/04-272

Woods, B., Lai, H., Wu, C. F., Zyla, T. R., Savage, N. S., & Lew, D. J. (2016). Parallel Actin-Independent Recycling Pathways Polarize Cdc42 in Budding Yeast. *Current Biology*, *26*(16), 2114–2126. https://doi.org/10.1016/j.cub.2016.06.047

Wright, S. (1932). The Roles of Mutation, Inbreeding, crossbreeding and Selection in Evolution. *Proceedings of the XI International Congress of Genetics*, 209–222.

Yang, Y.-F., Cao, W., Wu, S., & Qian, W. (2017). Genetic Interaction Network as an Important Determinant of Gene Order in Genome Evolution. *Molecular Biology and Evolution*, *34*(12), 3254–3266. https://doi.org/10.1093/molbev/msx264

Zhang, X.-S., & Hill, W. G. (2005). Genetic variability under mutation selection balance. *Trends in Ecology & Evolution*, *20*(9), 468–470. https://doi.org/10.1016/j.tree.2005.06.010

3

The most exciting phrase to hear in science, the one that heralds new discoveries, is not 'Eureka!' but '*That's funny...*'

— Isaac Asimov

Pleiotropy drives evolutionary repair of the responsiveness of polarized cell growth in to environmental cues

Enzo Kingma, Eveline T. Diepeveen, Leila Iñigo de la Cruz, Liedewij Laan

Abstract The ability of cells to translate different extracellular cues into different intracellular responses is vital for their survival in unpredictable environments. In *Saccharomyces cerevisiae*, cell polarity is modulated in response to environmental signals which allows cells to adopt varying morphologies in different external conditions. The responsiveness of cell polarity to extracellular cues depends on the integration of the molecular network that regulates polarity establishment with networks that signal environmental changes. The coupling of molecular networks often leads to pleiotropic interactions that can make it difficult to determine whether the ability to respond to external signals emerges as an evolutionary response to environmental challenges or as a result of pleiotropic interactions between traits. Here, we study how the propensity of the polarity network of *S. cerevisiae* to evolve toward a state that is responsive to extracellular cues depends on the complexity of the environment. We show that the deletion of two genes, *BEM3* and *NRP1*, disrupts the ability of the polarity network to respond to cues that signal the onset of the diauxic shift. By combining experimental evolution with whole-genome sequencing, we find that the restoration of the responsiveness to these cues correlates with mutations in genes involved in the sphingolipid synthesis pathway and that these mutations frequently settle in evolving populations irrespective of the complexity of the selective environment. We conclude that pleiotropic interactions make a significant contribution to the evolution of networks that are responsive to extracellular cues.

3.1. Introduction

Polarity establishment, the ability to generate an asymmetric distribution of cellular constituents, plays an important role in many of the biological functions that are observed throughout the tree of life (Jabbarzadeh, 2019). The dynamics of polarity establishment is regulated by an intricate network of molecular interactions, many of which are evolutionary conserved (Chiou et al., 2017; Etienne-Manneville, 2004; Thompson, 2013). What allows these networks to be versatile while maintaining a relatively high degree of conservation is their ability to generate different responses to various extracellular signals (Dickinson, 2008; Saito, 2010; Waltermann & Klipp, 2010). This feature makes it possible for the polarized appearance of cells to vary between environmental contexts (Granek et al., 2011).

Responsiveness to extracellular signals requires the integration of the polarity network with other molecular networks that either directly or indirectly translate these signals into an intracellular response (Broach, 2012; Granek et al., 2011; Mutavchiev et al., 2016; Saito, 2010; Salat-Canela et al., 2021; Waltermann & Klipp, 2010). An issue of integrated networks is that the

A version of this chapter has been published as Kingma E, Diepeveen ET, Iñigo de la Cruz L, Laan L. Pleiotropy drives evolutionary repair of the responsiveness of polarized cell growth to environmental cues. *Front Microbiol*. 2023;14:1076570.

decrease in modularity that arises when networks become coupled can frustrate evolvability (Fisher, 1930; Hartwell et al., 1999; Kirschner & Gerhart, 1998; Wagner & Altenberg, 1996; Wagner & Zhang, 2011). Because coupled networks become interdependent, the likelihood that a single mutation affects multiple phenotypic traits, an effect known as pleiotropy (Fisher, 1930; Wagner & Zhang, 2011), increases. Such pleiotropic effects are indeed frequently reported for genes involved in the establishment of cell polarity (Bauer et al., 1993; Prunskaite-Hyyryläinen et al., 2014; Zou et al., 2008). As antagonistic effects, where a mutation that is beneficial to one trait negatively affects a second trait (Austad & Hoffman, 2018; Mauro & Ghalambor, 2020; Paaby & Rockman, 2013), are considered to be more common than synergistic effects, pleiotropy is generally expected to constrain the number of accessible mutations during evolution in complex environments that select on multiple traits (Fisher, 1930; Orr, 2000; Waxman & Peck, 1998; J. J. Welch & Waxman, 2003). In turn, evolution in simple environments may not be constrained by pleiotropic interactions, but can instead lead to the deterioration of networks regulating unused traits (Fraebel et al., 2017; MacLean et al., 2004; Qian et al., 2012; Rose & Charlesworth, 1980). Thus, the molecular details of adaptive evolution of the polarity network are expected to depend on the environment: complex environments only allow mutations that preserve the integrity of coupled networks, while the released constraint in simple environments allows the system to explore alternative evolutionary pathways, but at the cost of the disintegration of unused networks and a loss of the ability to respond to environmental cues. However, whether these theoretical expectations form a general rule for the evolution of pleiotropically connected traits and if exceptions can be identified based on the molecular basis of their pleiotropic interactions is still a point of discussion (Agrawal & Stinchcombe, 2009; Jerison et al., 2020).

An attractive system to study the effect of pleiotropic interactions on the evolution of cell polarity is the yeast *Saccharomyces cerevisiae*. *S. cerevisiae* has adopted asymmetric cell division as its main mode of proliferation and must therefore establish an axis of polarity once per cell cycle (Chiou et al., 2017; Martin & Arkowitz, 2014). In addition, its polarity network is integrated with several different signaling networks to allow different growth modes in response to environmental cues, such as those that signal cell cycle progression (Yoshida & Pellman, 2008), filamentous growth (Cullen & Sprague Jr., 2012) and the activation of stress response pathways (Saito, 2010; Waltermann & Klipp, 2010). Here, we study whether the polarity network can restore its coupling to signaling networks after this coupling has been lost due to a genetic perturbation and how this restoration depends on selective pressures from the environment. In addition, we discuss whether known connections between the polarity network and other signaling pathways are able to explain our observed patterns of adaptation. To do this, we use a *bem3*Δ*nrp1*Δ strain of *S. cerevisiae* that has previously been demonstrated to be defective in polarity establishment during vegetative growth (Laan et al., 2015). We show that this genetic perturbation also disrupts the responsiveness of the polarity network to an environmental shift that induces cells to change their metabolic program. Using a combination of experimental evolution and whole-genome sequencing, we find that adaptive mutations that restore the responsiveness of the polarity network to this environmental insult emerge frequently and reproducibly in evolving populations and that their occurrence is surprisingly insensitive to the complexity of the environment.

3.2. Results

3.2.1. Deletion of *BEM3* and *NRP1* distorts cellular adaptation during the diauxic shift

The combined deletion of *BEM3* and *NRP1* has been shown previously to cause defects in polarity establishment that exceed the summed effects of their individual deletion (Laan et al., 2015), meaning they exhibit epistasis (Phillips, 2008). The existence of epistatic interactions between these mutations suggests a functional relation between Bem3 and Nrp1. This is surprising, because while Bem3 is known as a GTP Activating Protein (GAP) for Cdc42, the master regulator of cell polarity (Etienne-Manneville, 2004), Nrp1 has never been implicated to be involved in polarity establishment before. Instead, based on the current knowledge about its function, Nrp1 is best described as a prion forming protein that localizes to stress granules formed under conditions of glucose stress (Buchan et al., 2008; Kroschwald et al., 2015). This led us to hypothesize that the deletion of *BEM3* and *NRP1* may have consequences for the ability of the polarity network to respond to environmental cues that signal different growth modes.

We tested this hypothesis in the context of the ability of *S. cerevisiae* to perform diauxic growth between glucose and ethanol. In the presence of extracellular glucose, *S. cerevisiae* maintains a rapid mode of growth by alcoholic fermentation of glucose. The ethanol produced during alcoholic fermentation can be used as an alternative energy source when extracellular glucose drops below a critical level, but only in the presence of extracellular oxygen. The transition from the fermentation of glucose to the respiration of ethanol, a growth phase known as the diauxic shift, is characterized by several physiological changes (Galdieri et al., 2010), which includes changes in the polarized distribution of the actin cytoskeleton (De Virgilio & Loewith, 2006; Galdieri et al., 2010).

We qualitatively determined the effects of deleting *BEM3* and *NRP1* on the coupling of (diauxic growth) glucose sensing to cell polarity by imaging *bem3*Δ*nrp1*Δ cells during the diauxic shift (Figure 3.1a). The diauxic shift was induced by switching from growth media containing glucose as the sole carbon source to one where ethanol was the sole carbon source using a microfluidic device. A wild-type strain subjected to these conditions displayed the expected behavior, which consisted of rapid growth on glucose followed by a short growth pause at the time of the media switch, after which growth was resumed on ethanol media, but at a slower rate compared to growth on glucose media (Brauer et al., 2005). Overall, *bem3*Δ*nrp1*Δ cells followed the same pattern that we observed for the wild-type cells, but critically failed to produce buds during growth on ethanol media. Instead, isotropic growth was sustained in these cells up to the point where it induced cell death by lysis. Based on the link between polarity defects and an increase in cell size, we deduced that our observations for the *bem3*Δ*nrp1*Δ phenotype are the result of the inability of the polarity network to respond appropriately to the physiological changes that occur during the diauxic shift. Specifically, while *bem3*Δ*nrp1*Δ mutants are generally less fit than the wild-type strain, the cellular defect that leads to a lower fitness differs between conditions of standard vegetative growth and conditions where the cells must respond to the diauxic shift. During vegetative growth (2% glucose in Figure 3.1a) *bem3*Δ*nrp1*Δ cells proliferate, but do so at a slower rate than wild-type cells. After the media switch (transition from 2% glucose to 3% ethanol in Figure 3.1a) *bem3*Δ*nrp1*Δ cells enlarge, but are unable to divide.

Next, we quantified the effect of deleting *BEM3* and *NRP1* on the diauxic shift using Optical Density (OD) measurements of population growth in order to obtain growth curves for each

strain (Figure 3.1b, d). The diauxic shift was clearly visible in the growth curves as a transition period between two exponential growth phases with different growth rates. For technical reasons (see Supplementary Figure S3.1), we used media containing a high glucose concentration (2%) to quantify growth before the diauxic shift and a lower glucose concentration (0.1%) to quantify growth after the diauxic shift. We extracted the growth rate during the exponential phase before and after the occurrence of the diauxic shift by calculating the slope of the linear portion of the growth curve when plotted on a semi-log scale. These values were subsequently converted into their corresponding doubling times ($T_{PRE-shift}$ and $T_{POST-shift}$). This analysis revealed that *bem3*Δ*nrp1*Δ populations have a significantly longer doubling time than the wild-type both before and after onset of the diauxic shift (Figure 3.1c, e). While it is expected that the overall fitness defect of *bem3*Δ*nrp1*Δ mutants will lead to longer doubling times both before and after the diauxic shift, we argue based on our microfluidic experiment (Figure 3.1A) that the physiological cause that leads to a lower doubling time is different between the two conditions. Before the diauxic shift, *bem3*Δ*nrp1*Δ cells divide at a slower rate than the wild-type strain due to a defect in polarity establishment which causes a longer $T_{PRE-shift}$. In contrast, $T_{POST-shift}$ is affected by both the slower division rate and the higher death rate of *bem3*Δ*nrp1*Δ cells as the polarity defect becomes much more severe at the onset of the diauxic shift. In support of this idea, we found that *bem3*Δ*nrp1*Δ populations stop growing at a significantly lower OD than wild-type populations after the diauxic shift (ratio wild-type: *bem3*Δ*nrp1*Δ = 2.25, Figure 3.1f), while both strains enter the diauxic shift at approximately the same density (ratio wild-type: *bem3*Δ*nrp1*Δ = 1.15, Figure 3.1f). We therefore interpret these results as indications that the defects in polarity establishment caused by the deletion of *BEM3* and *NRP1* makes the polarity network insensitive to environmental cues that signal the onset of the diauxic shift. The loss of responsiveness to these cues causes an inability to establish a polarity site when the physiological changes related to the diauxic shift have taken place, leading to prolonged isotropic growth and an increase in cell size.

3.2.2. Recoupling of polarity establishment to sensing networks does not require a complex environment

We sought to determine whether the environment is the decisive factor that controls the adaptive value of restoring the cellular response to the diauxic shift during evolution. To do this, we took an experimental approach and evolved several parallel wild-type and *bem3*Δ*nrp1*Δ populations in two frequently used set-ups for experimental evolution (Figure 3.2). In the first setup, the batch culture, nutrient levels vary over time and cells experience periods of glucose depletion several times throughout the experiment (Brauer et al., 2005; Gresham & Dunham, 2014). Mutations that allow cells to correctly coordinate the physiological changes necessary to pass through the diauxic shift with those that regulate polarity establishment are therefore expected to be beneficial during evolution in a batch culture set-up, as this extends the overall number of progeny that a cell can produce before each passage. In the second set-up, the glucose limited continuous culture, nutrient concentrations remain constant after a steady state is reached and growth is maintained at a constant rate (Brauer et al., 2005; Gresham & Dunham, 2014). These constant environmental conditions have the consequence that cells do not induce the majority of the cellular responses that are associated with the diauxic shift (Brauer et al., 2005). Thus, the ability to perform diauxic growth appears as a dispensable trait during evolution in a continuous culture. Based on the theoretical assumptions that traits that do not experience selective pressure (1) tend to deteriorate and (2) are unlikely to fix mutations that improve their function during

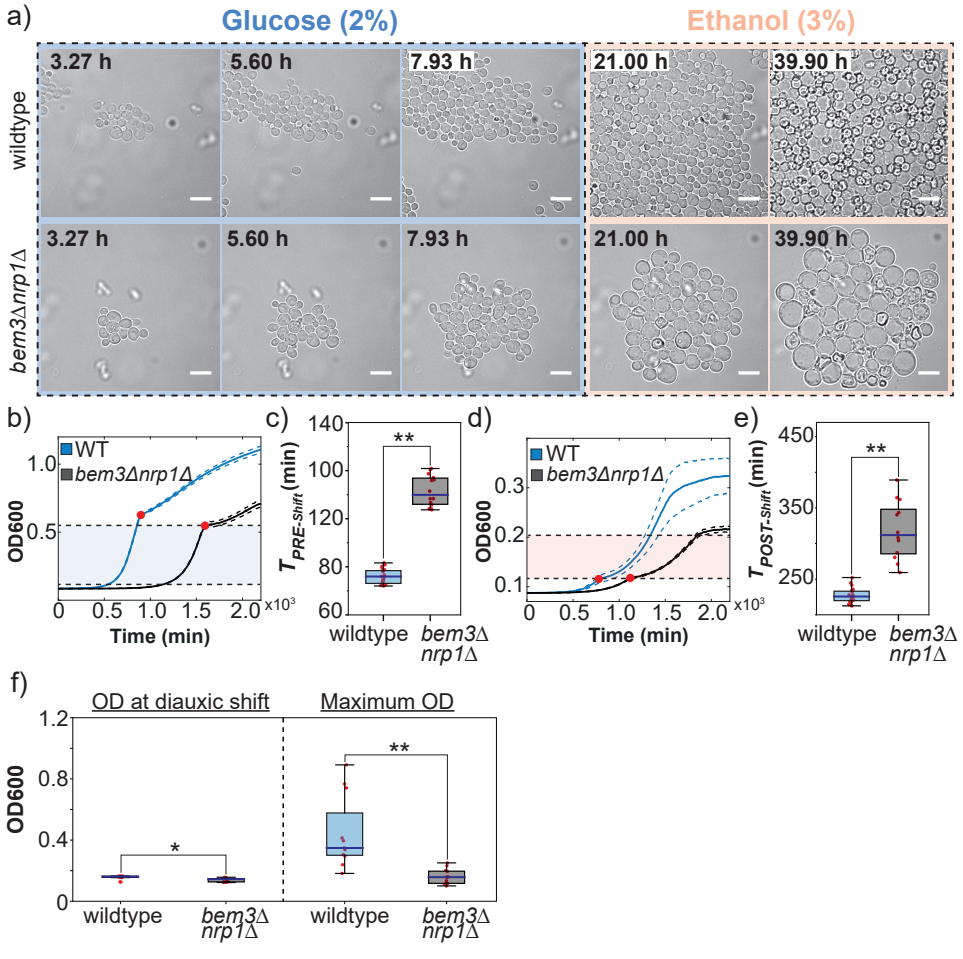


Figure 3.1. The *bem3*Δ*nrp1*Δ mutation causes defects in pre- and post-diauxic growth. (**a**) Time-lapse series of the diauxic shift. The wild-type and *bem3*Δ*nrp1*Δ strain were subjected to a switch from 2% glucose media to 3% ethanol media after 8 hours in 2% glucose media. The images show that while the wild-type strain is able to resume growth, the *bem3*Δ*nrp1*Δ cells increase in size without producing daughter cells. Scale bars represent 10 µm. (**b**) Growth curves of a wild-type (blue) and the *bem3*Δ*nrp1*Δ mutant (black) when grown in 2% glucose media. This data was used to obtain a measure for $T_{PRE-shift}$. Red dots indicate the point of diauxic shift, dashed lines represent the Standard Error of the Mean (SEM). (**c**) Doubling time of the wild-type strain and the *bem3*Δ*nrp1*Δ mutant during growth before the diauxic shift is entered ($T_{PRE-shift}$). (**d**) Growth curves of a wild-type (blue) and the *bem3*Δ*nrp1*Δ mutant (black) when grown in 0.1% glucose media. This data was used to obtain a measure for $T_{POST-shift}$. Red dots indicate the point of diauxic shift, dashed lines represent the SEM. (**e**) Doubling time of the wild-type strain and the *bem3*Δ*nrp1*Δ mutant during growth after passing through diauxic shift is entered ($T_{POST-shift}$). (**f**) Comparison of the OD at which the *bem3*Δ*nrp1*Δ mutant and the wild-type strain enter the diauxic shift and their OD at stationary phase when grown in YP + 0.1% glucose. The plot shows that while both strains enter diauxic shift at around the same density, the final density of the populations differ. *p-value < 0.05, **p-value < 0.005, Welch's t-test.

evolution, we expect restoration of diauxic growth by *bem3*Δ*nrp1*Δ populations to emerge only during batch culture evolution.

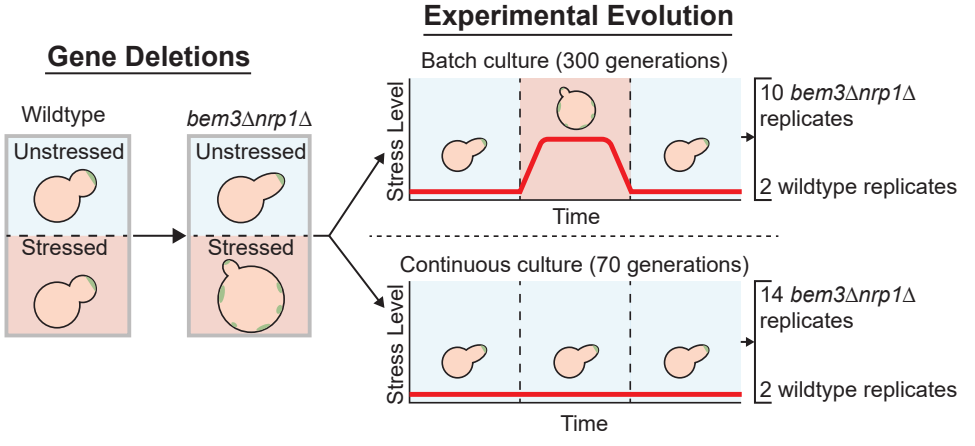


Figure 3.2. Overview of the scheme for experimental evolution. A polarity mutant that displays sensitivity to environmental stress is obtained after the deletion of the genes *BEM3* and *NRP1*. To assess the role the environment plays during the evolution of a network that is responsive to environmental signals, this mutant is evolved in an environment in which the stress level fluctuates (batch culture) and an environment where the stress level is constant (continuous culture).

We evolved a total of 14 *bem3*Δ*nrp1*Δ populations and 2 wild-type populations in the glucose limited continuous culture for 70 generations. The parameter $T_{PRE-shift}$ was used as a proxy for adaptations that restore the polarity defect caused by the deletion of *BEM3* and *NRP1*, but that do not necessarily improve the ability of the polarity network to respond to cues that signal the onset of the diauxic shift. Alternatively, $T_{POST-shift}$ was used as a proxy for adaptations that improve the response of cells to the diauxic shift. The values of $T_{PRE-shift}$ and $T_{POST-shift}$ of the evolved cell lines were determined by reviving the evolved population from a frozen stock and measuring the change in OD over time in media containing 2% and 0.1% glucose, respectively. This procedure is the same as what was done to determine $T_{PRE-shift}$ and $T_{POST-shift}$ for the ancestral wild-type and *bem3*Δ*nrp1*Δ populations (see Figures 3.1 1b, d and the section 3.4). Comparison of $T_{PRE-shift}$ between the evolved populations and their ancestor (Figure 3.3a) revealed that all evolved populations had either a similar or lower value for $T_{PRE-shift}$ relative to their ancestor. In contrast, we find that for $T_{POST-shift}$ half of the evolved *bem3*Δ*nrp1*Δ populations (7/14) had a lower doubling time, while the other half (7/14) had a longer doubling time relative to their ancestor, indicating that changes in diauxic growth do not affect fitness in a continuous culture. A similar trend for $T_{PRE-shift}$ and $T_{POST-shift}$ was visible for our two evolved wild-type populations. Taken together, these observations support our initial view that a continuous culture selects for faster vegetative growth, but not diauxic growth.

The finding that some of *bem3*Δ*nrp1*Δ populations evolved in the continuous culture show improvements in $T_{POST-shift}$ could be explained by a possible interdependence of $T_{PRE-shift}$ and $T_{POST-shift}$: improvements in $T_{PRE-shift}$ may be caused by mutations that increase the overall rate of cell division and these mutations will therefore also lead to improvements in $T_{POST-shift}$. However, these mutations do not necessarily also resolve the high death rate of *bem3*Δ*nrp1*Δ mutants at the start of the diauxic shift (Figure 3.1a), which may be a major factor that determines

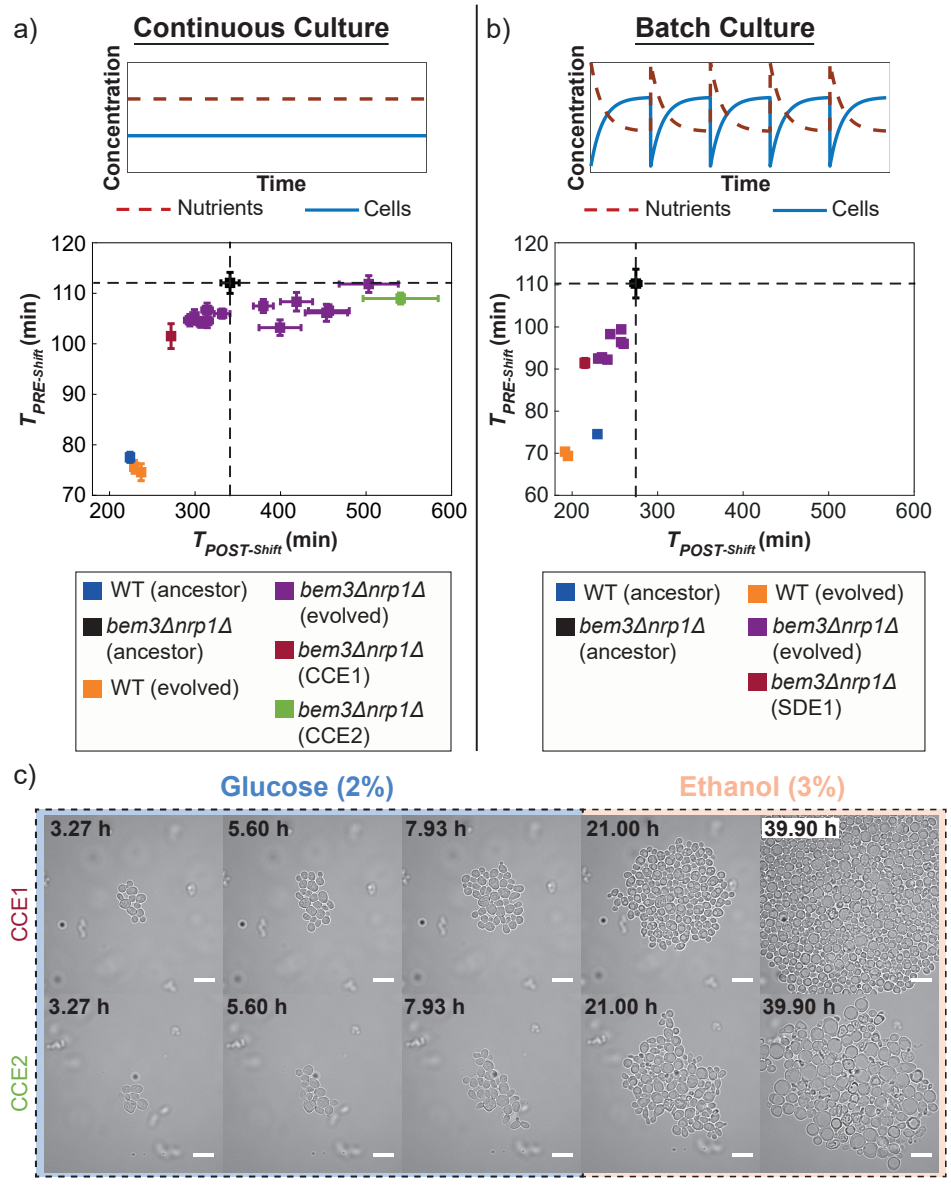


Figure 3.3. Experimental evolution of *bem3*Δ*nrp1*Δ mutants in a constant and variable environment. (**a**) (Top) In a continuous culture, both nutrient concentration and cell density remain constant over time. (Bottom) Scatter plot of $T_{PRE-Shift}$ against $T_{POST-Shift}$ 14 evolved *bem3*Δ*nrp1*Δ lines and 2 wild-type populations after 70 generations of evolution in a continuous culture. Dashed lines indicate the values of $T_{PRE-Shift}$ and $T_{POST-Shift}$ of ancestral *bem3*Δ*nrp1*Δ strain. Error bars show the SEM. (**b**) (Top) In a batch culture there are periodic fluctuations over time in nutrient concentration and cell density. (Bottom) Scatter plot of $T_{PRE-Shift}$ against $T_{POST-Shift}$ for 8 evolved *bem3*Δ*nrp1*Δ lines and 2 wild-type lines after 300 generations of evolution in a batch culture. Dashed lines indicate the values of $T_{PRE-Shift}$ and $T_{POST-Shift}$ of ancestral *bem3*Δ*nrp1*Δ strain. Error bars show the SEM. (**c**) Time-lapse of evolved lines CCE1 and CCE2 (continuous culture) during a sudden switch from 2% glucose media to 3% ethanol media (dashed red line). The images show that evolved line CCE1 contains cells that have a response to this environmental change that is phenotypically similar to the response of the wild-type strain. Evolved line CCE2 has a response that resembles the response of the ancestral *bem3*Δ*nrp1*Δ, but with a smaller increase in cell size (see Figure 3.1a). Scale bars represent 10 μm.

$T_{POST-shift}$. To verify that a decrease in $T_{POST-shift}$ relates to adaptations that resolve the high death rate, we imaged cells from the evolved population with the lowest (fastest growing, CCE1) and highest (slowest growing, CCE2) value for $T_{POST-shift}$ during the diauxic shift (Figure 3.3c) using the same microfluidic set-up we used in Figure 3.1. In agreement with our expectations, the results showed that the phenotype of CCE1 after switching to ethanol media was qualitatively more similar to that of our ancestral wild-type strain and CCE1 cells were able to resume proliferation after the diauxic shift. Alternatively, the phenotype of CCE2 was more similar to the ancestral *bem3*Δ*nrp1*Δ strain, as CCE2 cells enlarged and were frequently unable to divide after the onset of the diauxic shift.

We evolved 8 *bem3*Δ*nrp1*Δ and 2 wild-type populations in a batch culture with a daily passaging procedure. We initially maintained the same number of generations for evolution as we had done for the continuous culture (70 generations), but after assessing our proxies for fitness we were unable to identify any significant changes in the values of $T_{PRE-shift}$ and $T_{POST-shift}$ between the evolved populations and their ancestors (Supplementary Figure S3.2). We assumed that this is due to the frequent population bottlenecks that occur during the passaging of the populations, which can slow down the rate of adaptation by purging beneficial mutations from the population (Wein & Dagan, 2019). We provide an estimate of the effect of population bottlenecks on the fixation dynamics of beneficial mutations in Supplementary Section 3.5.1, which shows that bottlenecks vastly increase the expected number of generations that are required before a beneficial mutation that fixates in the population will emerge. To compensate for this effect of population bottlenecks, we allowed our batch culture experiment to run for an additional 230 generations such that the total number of generations was 300.

We found that all evolved populations grew faster than their ancestors, both before and after the diauxic shift (Figure 3.3b). The fact that we do not observe populations that evolve toward a state where the doubling time after the diauxic shift becomes longer suggests that these pathways are inaccessible during evolution in a batch culture. Taken together, these results imply that the environmental variability that exists in the batch culture imposes constraints on the diauxic growth pattern that can be attained during evolution, allowing only those where growth on both nutrients is improved, while the stable environment of the continuous culture releases some of these constrains. As a result, phenotypes that have evolved to perform well during the diauxic shift, presumably through evolutionary repair of the polarity defects caused by deleting *BEM3* and *NRP1*, only reproducibly emerge in a batch culture. However, although the degree of reproducibility is lower, similar phenotypes do frequently evolve in a continuous culture. This indicates that evolutionary constraints imposed by the environment are not sufficient to explain the restoration of the responsiveness of the polarity network to cues of the diauxic shift during evolution.

3.2.3. Populations with a restored responsiveness to extracellular cues accumulate mutations in genes related to the sphingolipid synthesis pathway

To understand the molecular basis of the different adaptations of $T_{PRE-shift}$ and $T_{POST-shift}$ we observed in our continuous and batch cultures, we performed Whole Genome Sequencing (WGS) on the 22 evolved *bem3*Δ*nrp1*Δ lines and the 4 evolved wild-type controls and compared them to the genome of their wild-type ancestor (see section 3.4). We looked for patterns of parallel evolution by restricting our analysis to genes that were mutated in at least 2 different populations evolved in the same environment. This resulted in a total of 88 genes that acquired non-synonymous

mutations or indels in 2 or more evolved populations (including the wild-type lines).

The most notable environment-specific mutations were the early stop codons in *WHI2* that frequently occurred in the populations evolved in the continuous culture: 12 out of 14 evolved *bem3*Δ*nrp1*Δ and both evolved wild-type controls had mutated *WHI2*. Disruptive mutations in *WHI2* have also been reported in other experimental evolution studies that used nutrient limited continuous cultures (Hong & Gresham, 2014; Kvitek & Sherlock, 2013) and these mutations therefore likely provide a general advantage during adaptation to nutrient-limiting conditions.

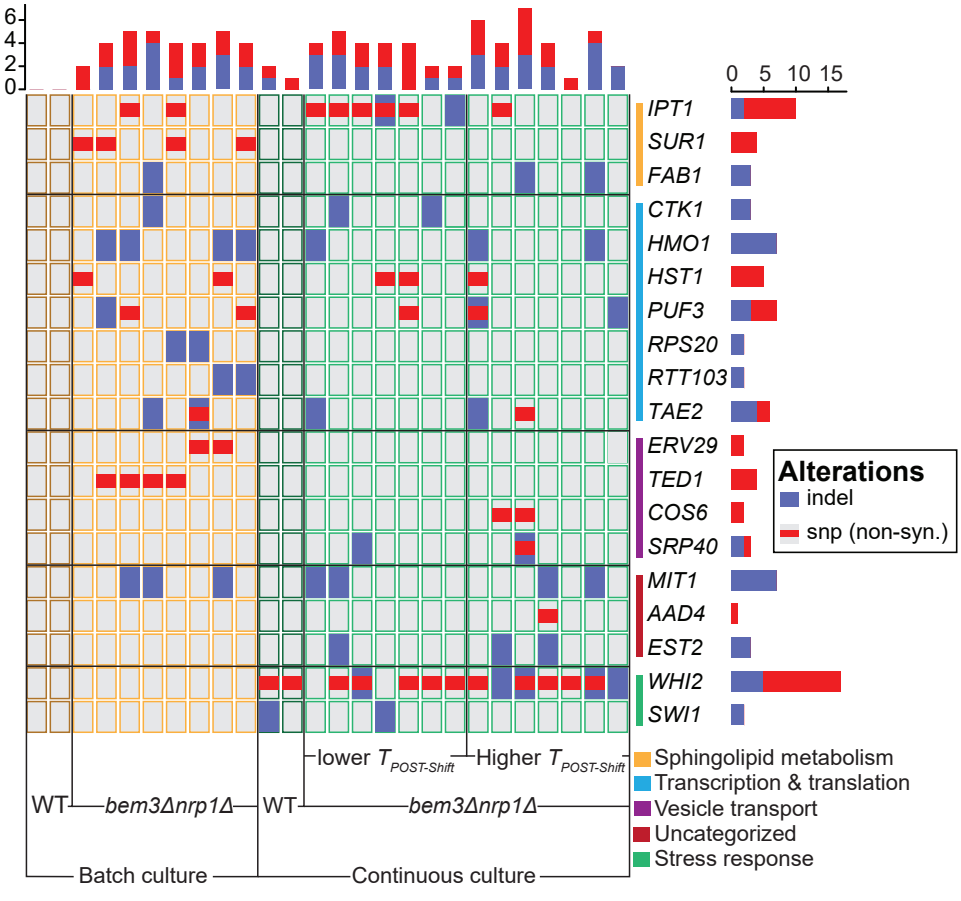


Figure 3.4. The mutational spectrum of different phenotypic subgroups that emerged after experimental evolution of *bem3*Δ*nrp1*Δ populations. The mutant specific mutations found in each gene for evolved continuous culture lines that decreased their respiration rate, evolved continuous culture lines that increased their respiration rate and evolved batch culture lines. Genes are grouped according to their cellular process GO-term. All genes shown were only mutated in the *bem3*Δ*nrp1*Δ populations and not in the wild-type populations, with the exception of *WHI2* and *SWI1*, which were also found to be mutated in the wild-type populations evolved in the continuous culture.

Because we saw the same phenotype emerge in the batch culture and continuous culture (populations that decreased $T_{POST-shift}$), we questioned whether the molecular basis of these adaptations were similar. In total, 22 genes were mutated in at least 2 of the *bem3*Δ*nrp1*Δ lines evolved in each environment. We grouped these genes according to their Biological Process

Gene Ontology (GO) annotation on the Saccharomyces Genome Database. This revealed that populations evolved in a continuous culture had more mutations in genes involved in the stress response, while populations from the batch culture had slightly more mutations in genes related to transcription and translation. We then split the evolved populations into two groups: those that evolved to decrease $T_{POST-shift}$ (7/14 populations of the continuous culture and 8/8 populations of the batch culture) and those that evolved to increase $T_{POST-shift}$ (7/14 populations of the continuous culture and 0/8 populations of the batch culture). Interestingly, populations with a decreased $T_{POST-shift}$ had more mutations in lipid metabolic genes than those that did not decrease $T_{POST-shift}$. Of the 14 *bem3*Δ*nrp1*Δ populations that were evolved in the continuous culture, 6/7 populations with a decreased $T_{POST-shift}$ had mutations in the *IPT1*, while we only found mutations in this gene in 1/7 populations with an increased $T_{POST-shift}$ (Figure 3.4). In the batch culture populations, 2/8 had acquired mutations in *IPT1*, while 3/8 had mutations in *SUR1*. Notably, Ipt1 acts directly downstream of Sur1 in the pathway for the synthesis of complex sphingolipids (Dickson et al., 2006; Morimoto & Tani, 2015; Thevissen et al., 2000).

Based on this correlation we hypothesize that, after the deletion of Bem3 and Nrp1, the robustness of the polarity module during the diauxic shift can be (partially) restored by changes in the lipid composition of the plasma membrane. Interestingly, this strategy appears to be dominant for repairing the defect caused by the deletion of *BEM3* and *NRP1* regardless of whether diauxic growth is part of the selective environment.

3.3. Discussion

The ability to respond to environmental cues is a crucial factor for the survival of organisms in complex environments. For example, studies have indicated that pathogens increase the likelihood of successfully infecting a host by adjusting their physiology to match the host's circadian rhythm (Kahl Lisa et al., 2022). Here, we used a genetically perturbed strain of *S. cerevisiae* to investigate the contribution of the environment in shaping a polarity network that can translate the extracellular signals for diauxic growth into an intracellular response. We show that the deletion of *BEM3* and *NRP1* has previously unknown consequences for polarity establishment that diminishes its capacity to respond to these extracellular signals and impedes the ability of cells to successfully navigate through the diauxic shift. Which molecular mechanisms are affected by the deletion of *BEM3* and *NRP1* in such a manner that it leads to the observed phenotype are not addressed in this study. However, the results from several other studies that have looked at the relationship between environmental stress and cell morphology allow us to formulate a hypothesis on how the deletion of *BEM3* and *NRP1* causes the decoupling of cell polarity from diauxic growth. The link between cell morphology and environmental stress is frequently proposed to be a consequence of the loss of polarity of the actin cytoskeleton induced by stress factors (Balguerie et al., 2002; Homoto & Izawa, 2018; Sivadon et al., 1995; Uesono et al., 2004). Failure to repolarize the actin cytoskeleton following environmental stress, either due to the severity of the stress conditions (Homoto & Izawa, 2018) or due to the loss of a genetic component required for repolarization (Balguerie et al., 2002; Sivadon et al., 1995), results in enlarged cells. The similarity of these hypertrophied cells under conditions of environmental stress to the phenotype of *bem3*Δ*nrp1*Δ mutants we observe during a transition from glucose-containing media to ethanol-containing media suggests they are caused by a defect in a similar pathway. Indeed, the depletion of glucose, one of the hallmark cues for entry into the diauxic shift (Brauer et al., 2005), has also been shown to cause the rapid and transient depolarization of actin in wild-

type cells (Uesono et al., 2004; Vasicova et al., 2016). The repolarization of actin in the context of glucose depletion depends on the activation of the respiratory metabolism (Uesono et al., 2004), as cells with dysfunctional mitochondria do not repolarize actin (Uesono et al., 2004; Vasicova et al., 2016). Thus, one possibility is that the deletion of *BEM3* and *NRP1* causes defects in respiration. However, we consider this unlikely based on our observation that *bem3*Δ*nrp1*Δ cells are still able to grow, although only isotropically, in media containing ethanol as the only carbon source. In addition, if mitochondrial dysfunction were the cause of the observed phenotype, mutations related to mitochondrial function would be expected to arise during our evolution experiments, but this was not the case.

Instead, our results suggest that the defects in diauxic growth of *bem3*Δ*nrp1*Δ mutants are suppressed by mutations in the sphingolipid synthesis pathway. Interestingly, the genes *IPT1* and *SUR1*) that were frequently mutated in evolved *bem3*Δ*nrp1*Δ populations with a (partially) restored ability to pass through the diauxic shift are also known to suppress the sensitivities to stress and starvation that arise after the deletion of genes that encode for the amphiphysin-like proteins Rvs161 and Rvs167 (Balguerie et al., 2002; Desfarges et al., 1993). Rvs161 and Rvs167 have a direct role in regulating the polarity of the actin cytoskeleton (Amberg et al., 1995; Breton & Aigle, 1998; Munn et al., 1995; Sivadon et al., 1995; Sivadon et al., 1997) and their loss causes defects in the depolarization and repolarization dynamics of actin during stress in an equivalent manner as has been described for glucose stress in the section above (Bauer et al., 1993; Crouzet et al., 1991; Sivadon et al., 1995). Suppression of these defects through the deletion of *IPT1* or *SUR1* has been reported to act by preventing the full depolymerization of actin under stressful conditions (Balguerie et al., 2002), thereby relieving some of the consequences of an inability to repolarize actin. Extrapolating these findings to *bem3*Δ*nrp1*Δ mutants, this implies that the evolutionary repair of diauxic growth in *bem3*Δ*nrp1*Δ populations acts by directly modulating actin dynamics using sphingolipid synthesis as a control knob. Similarly, the pleiotropic effects resulting from the deletion *BEM3* and *NRP1* are therefore likely a consequence of the dual role of actin in polarized growth and stress response pathways (Ho & Bretscher, 2001; Leadsham et al., 2010; Smethurst et al., 2014) that couples polarity establishment to the diauxic shift.

It remains unclear why the consequences of deleting *BEM3* and *NRP1* are different for the actin dynamics required during vegetative growth and the actin dynamics under stressed conditions. Much alike to what happens under conditions of stress, the actin cytoskeleton must depolarize and repolarize during the cell cycle to switch between modes of isotropic and polarized growth (Ahn et al., 2001; Bi & Park, 2012; Lew & Reed, 1993; Pruyne & Bretscher, 2000; M. D. Welch et al., 1994). However, our results show that polarized growth during the vegetative cell cycle is not strongly affected by the deletion of *BEM3* and *NRP1*, while polarized growth after the stress response of the diauxic shift is strongly diminished. This suggests that cell cycle-related polarization of actin may be regulated by a different pathway than the polarization of actin during the stress response. We find that, despite that they may be regulated by different pathways, the ability to perform polarized growth in both contexts can be restored by mutations in genes related to sphingolipid synthesis. Surprisingly, the fixation of these mutations that restore both vegetative and diauxic growth does not strongly depend on the complexity of the environment. Instead, we frequently see them emerge in populations evolved under constant conditions where improved diauxic growth appears to have no selective benefit, as is supported by our result that nearly all populations evolved in the continuous culture inactivate *WHI2*, which encodes for a protein that initiates the stress response during nutrient depletion (Kaida et al., 2002; Radcliffe et al., 1997;

Saul et al., 1985; Sudbery et al., 1980). Interestingly, a recent study investigating the adaptive response of *Escherichia coli* to different temperature fluctuation regimes also found that the same mutations frequently evolve in parallel in a manner that does not depend on the dynamics of the selective environment (Lambros et al., 2021). A large-scale phenotypic assay revealed that the evolved strains generally became closer to the phenotype of their ancestor under a large number of conditions, leading to the hypothesis that an innate evolutionary response of an organism in a stressful environment is to evolve in such a way that their physiology resembles that of their (fitter) ancestor in unstressed conditions. Overall our results agree with this hypothesis, as we find that genetically perturbed cells frequently evolve to better match the cellular response of their ancestor, even in unseen environmental conditions. Possible explanations for why the fixation of mutations that restore the cellular response to conditions beyond those experienced during adaptation would be preferred are that (1) these mutations might occur more frequently in the population because they constitute mutational hotspots or (2) their fixation is purely driven by the fitness benefit that they confer to vegetative growth and the restored diauxic growth is merely a side effect of a pleiotropic interaction network. In conclusion, our results demonstrate that the evolution of interaction networks that can sense and respond to different environmental signals should not always be interpreted as adaptive, but may instead be a consequence of a strong integration between different interaction networks regulating different cellular functions. Such an integration of different interaction networks may also be able to explain observations of the seemingly purposeless emergence of phenotypic plasticity, the ability of an organism to adjust its phenotype to its environment, during evolution in constant environmental conditions (Fraebel et al., 2020).

3.4. Methods

3.4.1. Yeast strains and media preparation

All strains used in this study are derived from the W303 background and are *MATa* haploid cells. We used yLL132a as our wild-type strain and yLL143a as our *bem3*Δ*nrp1*Δ strain (Laan et al., 2015), which has the same genetic background as yLL132a, but with *BEM3* and *NRP1* replaced with, respectively, the natMX4 (clonNAT-Nourseothricin resistance) and hphMX4 (Hygromycin B resistance) cassettes. For batch culture evolution experiments, standard rich media (10 g/L Yeast Extract, 20 g/L Peptone and 20 g/L Dextrose) was used and was prepared by dissolving 50 g/L from a premixed batch of ingredients (Sigma-Aldrich) in H_2O. For chemostat evolution experiments the same premix was used, but supplemented with 19 g/L extra Yeast Extract and 9.5 g/L extra Peptone to obtain a final dextrose concentration of 1 g/L. A total of 0.1 mg/mL Ampicillin was added to the chemostat media as a safeguard against bacterial contamination. Microscopy experiments were performed in Synthetic Complete (SC) media prepared from Complete Supplement Mixture without amino acids, riboflavin and folic acid (750 mg/L), Yeast Nitrogen Base (6.9 g/L) and either Dextrose (2% w/v) or Ethanol (3% v/v) as a carbon source. All media was filter sterilized to avoid degradation of components during autoclaving.

3.4.2. Experimental Evolution of Continuous Cultures

Multiplexed Chemostat Array Set-Up

We performed our evolution experiments in a dextrose limited chemostat environment by setting up a multiplexed chemostat array of 16 cultures according to the protocol from Miller et al. (2013). YP 0.1%D media was filter sterilized directly into a 10 L glass carboy. During the run,

fresh media was provided to the cultures from this carboy by using a peristaltic pump fitted with Marprene tubing. The correlation between rotation speed and media flow rate was empirically determined by measuring the effluent volume at different rotation speeds. Aquarium pumps were used to maintain the positive pressure inside the culture chambers required for the removal of excess culture volume, to keep the cultures aerated and mixed. To minimize evaporation and maintain sterility, air from the pumps was first routed through gas washing bottles and 0.45 µm PFTE filters before entering the culture chambers. The temperature was regulated at 30°C using heat bocks.

Initialization of Multiplexed Arrays

We initialized our multiplexed chemostat arrays by allowing the culture chambers to fill with media until the volume exceeded 20 mL. We dissolved cells from a glycerol stock in YP 0.1 %D media and used to inoculate the cultures by aseptically injecting 4 mL into each culture chamber. In total, 14 *bem3*Δ*nrp1*Δ cultures and 2 wild-type cultures were inoculated using this procedure. With the peristaltic pump turned off and the aquarium pumps turned on, the *bem3*Δ*nrp1*Δ cultures were left to grow for 4 days and the wild-type cultures were left to grow for 2 days until they reached saturation (batch phase growth). After the cultures reached saturation, the culture volume was set at 20±1 mL while performing the zero time point sampling.

Sampling Regimen

All cultures were sampled twice a week. Samples were taken by replacing the effluent bottles with sterile sampling bottles and collecting the effluent on ice over a period of approximately 2 hours. Directly after sampling, 1 mL of each collected sample was mixed with 500 µL glycerol and stored at -80°C. Optical Density (OD) measurements at 600 nm were taken of each sample in 10 mm plastic cuvettes using a photospectrometer (Nanodrop 2000C). When necessary, samples were diluted with YP to obtain a final OD of between 0.1-1.5. All samples were diluted in the same media used for blanking the photospectrometer. Effluent volumes were measured daily with a graduated cylinder from which the volume could be read with 0.5 ml precision. On days that sampling took place, the effluent volume of samples was determined after the standard procedure for sampling (glycerol stocks and OD measurements).

Calculation of Dilution Rates and Generation Times

We calculated the dilution rate D of each sample in our multiplexed chemostat array from the effluent volume using the following formula:

$$D = \frac{V_{Eff}}{t \cdot V_{Cult}}. \tag{3.1}$$

Here, V_{Eff} is the measured effluent volume, t is the time that has passed since the last sampling and V_{Cult} is the culture volume. At steady state, the growth rate equals the dilution rate (Gresham & Dunham, 2014), allowing the number of generations G that have passed to be calculated by:

$$G = t \cdot \frac{D}{\ln 2}. \tag{3.2}$$

Experimental Evolution of Batch Cultures

Batch culture evolution experiments were started with 10 *bem3*Δ*nrp1*Δ and 2 wild-type cultures. The cultures were derived from a single *bem3*Δ*nrp1*Δ and a single wild-type liquid culture initiated from a glycerol stock and grown to saturation for 2 days in YP 2 %D in a roller drum (set at 40 RPM) at 30°C. After the cultures reached saturation, 10 µL of each starter culture was diluted into 10 mL of fresh YP 2 %D and were placed back into the roller drum. The cultures were diluted by 10 µL into 10 mL of fresh YP 2 %D every 24±2 h. After each dilution, the OD at 600 nm of the remaining culture was measured using the same procedure as described above for the chemostat evolution experiment. Because batch cultures involve frequent population bottlenecks that can reduce genetic variation and possibly purge beneficial mutations (Gresham & Dunham, 2014; Gresham & Hong, 2014), it might take longer for an adaptive mutation to settle in the population. To compensate for this effect, the number of generations that the populations were evolved in a batch culture setting was increased to 300 generations (an additional 230 generations compared to the populations evolved in a continuous culture).

Growth Curve Measurements

Growth curves were obtained by measurements using a plate reader (Tecan Infinite 200 Pro). Cells were inoculated from a glycerol stock in YP 0.1 %D liquid media and grown to saturation for 2 days in a roller drum at 30°C. On the day of the measurement, the saturated cultures were diluted 1000X into fresh YP 2 %D to measure the pre-diauxic doubling time ($T_{PRE-shift}$) or fresh YP 0.1 %D to measure the post-diauxic doubling time ($T_{POST-shift}$). 100 µL of this culture was pipetted into each well of a sterile 96-well plate (Nunc™Edge 2.0, Thermo Scientific™) with the edge moats filled with 1.7 mL of sterile H_2O. Each plate contained multiple technical replicates of each sample. As a control for contamination and to allow for background subtraction for downstream processing, 8 wells were filled with blank medium. Measurements were taken during incubation at 30°C in the plate reader using the following protocol: First, the cells were shaken for 1000 s (linear shaking, 1 mm amplitude) without measurement. After this, the absorbance of each well was measured every 7 min with intermittent shaking (260 s, linear, 1 mm amplitude) for 48 h.

Growth Parameter Calculations

Doubling times for pre-diauxic ($T_{PRE-shift}$) and an post-diauxic ($T_{POST-shift}$) growth were extracted from the growth curve measurement in YP 2 %D and YP 0.1 %D, respectively. First, the measured OD values were blanked using the time average value of one of the wells containing blank media. Then, the data was converted to semi-log data by taking the natural logarithm of the blanked OD values. A home written MATLAB script was used to fit a line to the linear portion of the semi-log data to obtain the growth rate during pre-diauxic or post-diauxic growth (figure S3.1). These growth rates were converted into doubling times using the following relation:

$$T_d = \frac{\ln 2}{\mu},$$

where T_d is the doubling time corresponding to the growth rate μ obtained from the slope of the linear fit

3.4.3. Microscopy & Microfluidics

Cells were grown to log phase in SC media containing 2% dextrose. Clumps of cells were dissociated prior to imaging by sonicating (Q500 Sonicator, QSonica) in a sealed Eppendorf tube

using a cup horn at 70% amplitude for 2 minutes (cycle of 30 seconds pulse on, 15 seconds pulse off). After sonication, each sample was diluted to the same optical density in fresh Synthetic Complete media containing 2% dextrose. Cells were trapped in a microfluidic culture chamber (CellASIC ONIX Y04C-02, Merck – Millipore) after flushing the culture chambers with fresh media for 20 minutes using a pressure of 8 psi. Brightfield images were taken with a Nikon Eclipse Ti-E inverted Microscope using a 60x objective (Plan Apo λ 60X oil, NA: 1.40) with 1 minute intervals. During imaging, cells were maintained in a constant flow of media using a pressure of 1 psi. Cells were subjected to a media switch by changing from an inlet with SC media containing 2% dextrose to an inlet with SC media with 3% ethanol after 8 hours of imaging.

3.4.4. DNA extraction, Illumina library preparation & Whole Genome Sequencing

We extracted genomic DNA from liquid cultures grown for two overnights for each of the 16 chemostat samples, 10 serial dilution samples and a non-evolved yLL132a ancestor with the MasterPure Yeast DNA Purification Kit (Epicentre, Madison, WI, USA) following the manufacturer's protocol. We included a RNase A (Qiagen, Hilden, Germany) treatment step in the protocol and collected DNA in a final volume up to 30 µL H_2O. We pooled up to three extractions per sample using the Genomic DNA clean & Concentrator kit (Zymo Research, Irvine, CA, USA), following the supplied protocol. We eluted DNA in a final volume of 30 µL. We assessed DNA quality by 0.8 % agarose gel electrophoresis and quantity by fluorometry using a Qubit 4.0 Fluorometer (Invitrogen, Carlsbad, CA, USA). Samples were individually barcoded and pooled into a single library with the NEB Next Ultra DNA Library Prep Kit (New England Biolabs, Ipswich, MA, USA) and sequenced on a HiSeq machine (Illumina, San Diego, CA, USA) by Novogene (Bejing, China).

3.4.5. WGS Data analysis

We first checked raw paired-end reads (150 bp) for quality with the FASTQC toolkit (version 0.11.7).[1] We removed low quality ends (Quality scores <20; and first 9 bases of all reads), and removed duplicates with the FastX toolkit (version 0.0.14).[2] We downloaded the R64-1-1 S. cerevisiae genome from the Saccharomyces Genome Database (SGD)[3] and used it as our reference. We indexed the reference genome with the Burrows-Wheeler Aligner [BWA; version 0.7.17; (Li & Durbin, 2010)], and SAMtools [version 1.8; (Li & Durbin, 2009; Li, 2011)], and generated a dictionary with Picard (version 2.18.5).[4] We mapped sequences from all samples individually to the reference with BWA-MEM sorted and indexed mapped reads into a BAM file with SAMtools. We performed multisample SNP calling and additional indexing with SAMtools and BCFtools (version 1.8). We plotted and checked statistics, e.g., TS/TV and quality of sites and read depth, with BCFtools. These statistics were used to filter out SNPs and Indels with low quality sites (QUAL > 30), low read depth (DP > 20), and variants in close proximity to gaps (SnpGAP 10). We annotated the VCF file with snpEff [version 4.3T; (Cingolani et al., 2012)] with R64- 1-1.86. We then retrieved variants (SNPs and indels) of interest through comparison of variants between the reference strain, the ancestor strains, and the evolved strains. We excluded variants that were different between R64 and all our W303 samples, as these merely display differences between the two genetic backgrounds [see e.g., Ralser et al. (2012)]. Synonymous variants, variants in non-coding regions, and stop retained variants were excluded. Mutations in telomeric regions and in Long

[1] https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
[2] http://hannonlab.cshl.edu/fastx_toolkit/
[3] http://www.yeastgenome.org/ (accessed September 2018)
[4] https://broadinstitute.github.io/picard/

Terminal Repeats (LTRs) were excluded from analysis due to the natural variation that occurs in the genomic sequence of these regions. To find causative mutations, we looked for genes that mutated in at least two evolved lines, excluding those that appeared only in the mutant line(s) from one environment and the wild-type line(s) of the other environment. From the resulting list of genes, genes corresponding to dubious or uncharacterized Open Reading Frames (ORFs) were removed according to their description on SGD. Two genes (*RPS29B* and *ECM33*) had acquired the same mutation across all 22 parallel evolved *bem3*Δ*nrp1*Δ lines that sweeped the population, suggesting that these mutations were acquired in the ancestor before the different cell lines were split. Although these mutations might have some fitness benefit in the *bem3*Δ*nrp1*Δ background, they do not explain the adaptation we observe during our evolution experiments and we therefore excluded them from further analysis. We used the OncoPrint function from the ComplexHeatmap package (Gu et al., 2016) available in R (version 4.2.3) to visualize the relevant mutations in our evolved lines as a heatmap.

Bibliography

Agrawal, A. F., & Stinchcombe, J. R. (2009). How much do genetic covariances alter the rate of adaptation? *Proc Biol Sci*, *276*(1659), 1183–1191. https://doi.org/10.1098/rspb.2008.1671

Ahn, S. H., Tobe, B. T., Fitz Gerald, J. N., Anderson, S. L., Acurio, A., & Kron, S. J. (2001). Enhanced cell polarity in mutants of the budding yeast cyclin-dependent kinase Cdc28p. *Mol Biol Cell*, *12*(11), 3589–3600. https://doi.org/10.1091/mbc.12.11.3589

Amberg, D. C., Basart, E., & Botstein, D. (1995). Defining protein interactions with yeast actin in vivo. *Nature Structural Biology*, *2*(1), 28–35. https://doi.org/10.1038/nsb0195-28

Antal, T., & Scheuring, I. (2006). Fixation of Strategies for an Evolutionary Game in Finite Populations. *Bulletin of Mathematical Biology*, *68*(8), 1923–1944. https://doi.org/10.1007/s11538-006-9061-4

Austad, S. N., & Hoffman, J. M. (2018). Is antagonistic pleiotropy ubiquitous in aging biology? *Evolution, Medicine, and Public Health*, *2018*(1), 287–294. https://doi.org/10.1093/emph/eoy033

Balguerie, A., Bagnat, M., Bonneu, M., Aigle, M., & Breton, A. M. (2002). Rvs161p and Sphingolipids Are Required for Actin Repolarization following Salt Stress. *Eukaryotic Cell*, *1*(6), 1021. https://doi.org/10.1128/EC.1.6.1021-1031.2002

Barlukova, A., & Rouzine, I. M. (2021). The evolutionary origin of the universal distribution of mutation fitness effect. *PLOS Computational Biology*, *17*(3), e1008822. https://doi.org/10.1371/journal.pcbi.1008822

Bauer, F., Urdaci, M., Aigle, M., & Crouzet, M. (1993). Alteration of a yeast SH3 protein leads to conditional viability with defects in cytoskeletal and budding patterns. *Mol Cell Biol*, *13*(8), 5070–5084. https://doi.org/10.1128/mcb.13.8.5070-5084.1993

Bi, E., & Park, H. O. (2012). Cell polarization and cytokinesis in budding yeast. *Genetics*, *191*(2), 347–387. https://doi.org/10.1534/genetics.111.132886

Brauer, M. J., Saldanha, A. J., Dolinski, K., & Botstein, D. (2005). Homeostatic adjustment and metabolic remodeling in glucose-limited yeast cultures. *Molecular Biology of the Cell*, *16*(5), 2503–2517. https://doi.org/10.1091/mbc.e04-11-0968

Breton, A. M., & Aigle, M. (1998). Genetic and functional relationship between Rvsp, myosin and actin in Saccharomyces cerevisiae. *Current Genetics*, *34*(4), 280–286. https://doi.org/10.1007/s002940050397

Broach, J. R. (2012). Nutritional control of growth and development in yeast. *Genetics*, *192*(1), 73–105. https://doi.org/10.1534/genetics.111.135731

Buchan, J. R., Muhlrad, D., & Parker, R. (2008). P bodies promote stress granule assembly in Saccharomyces cerevisiae. *The Journal of Cell Biology*, *183*(3), 441–455. https://doi.org/10.1083/jcb.200807043

Campos, P. R. A., & Wahl, L. M. (2009). THE EFFECTS OF POPULATION BOTTLENECKS ON CLONAL INTERFERENCE, AND THE ADAPTATION EFFECTIVE POPULATION SIZE. *Evolution*, *63*(4), 950–958. https://doi.org/10.1111/j.1558-5646.2008.00595.x

Chiou, J. G., Balasubramanian, M. K., & Lew, D. J. (2017). Cell Polarity in Yeast. *Annual review of cell and developmental biology*, *33*, 77. https://doi.org/10.1146/ANNUREV-CELLBIO-100616-060856

Crouzet, M., Urdaci, M., Dulau, L., & Aigle, M. (1991). Yeast mutant affected for viability upon nutrient starvation: Characterization and cloning of theRVS161 gene. *Yeast*, *7*(7), 727–743. https://doi.org/10.1002/yea.320070708

Cullen, P. J., & Sprague Jr., G. F. (2012). The regulation of filamentous growth in yeast. *Genetics*, *190*(1), 23–49. https://doi.org/10.1534/genetics.111.127456

Day, A., Schneider, C., & Schneider, B. L. (2004). Yeast Cell Synchronization. In H. B. Lieberman (Ed.), *Cell cycle checkpoint control protocols* (pp. 55–76). Humana Press. https://doi.org/10.1385/1-59259-646-0:55

De Virgilio, C., & Loewith, R. (2006). Cell growth control: little eukaryotes make big contributions. *Oncogene*, *25*(48), 6392–6415. https://doi.org/10.1038/sj.onc.1209884

Desai, M. M., & Fisher, D. S. (2007). Beneficial Mutation–Selection Balance and the Effect of Linkage on Positive Selection. *Genetics*, *176*(3), 1759–1798. https://doi.org/10.1534/genetics.106.067678

Desai, M. M., Fisher, D. S., & Murray, A. W. (2007). The Speed of Evolution and Maintenance of Variation in Asexual Populations. *Current Biology*, *17*(5), 385–394. https://doi.org/https://doi.org/10.1016/j.cub.2007.01.072

Desfarges, L., Durrens, P., Juguelin, H., Cassagne, C., Bonneu, M., & Aigle, M. (1993). Yeast mutants affected in viability upon starvation have a modified phospholipid composition. *Yeast*, *9*(3), 267–277. https://doi.org/10.1002/yea.320090306

Dickinson, J. R. (2008). Filament formation in Saccharomyces cerevisiae — a review. *Folia Microbiologica*, *53*(1), 3–14. https://doi.org/10.1007/s12223-008-0001-6

Dickson, R. C., Sumanasekera, C., & Lester, R. L. (2006). Functions and metabolism of sphingolipids in Saccharomyces cerevisiae. *Progress in Lipid Research*, *45*(6), 447–465. https://doi.org/https://doi.org/10.1016/j.plipres.2006.03.004

Etienne-Manneville, S. (2004). Cdc42 - the centre of polarity. *Journal of Cell Science*, *117*(8), 1291–1300. https://doi.org/10.1242/JCS.01115

Fisher, R. A. (1930). *The genetical theory of natural selection*. Clarendon Press. https://doi.org/10.5962/bhl.title.27468

Fraebel, D. T., Mickalide, H., Schnitkey, D., Merritt, J., Kuhlman, T. E., & Kuehn, S. (2017). Environment determines evolutionary trajectory in a constrained phenotypic space. *Elife*, *6*. https://doi.org/10.7554/eLife.24669

Fraebel, D. T., Gowda, K., Mani, M., & Kuehn, S. (2020). Evolution of Generalists by Phenotypic Plasticity. *iScience*, *23*(11), 101678. https://doi.org/https://doi.org/10.1016/j.isci.2020.101678

Galdieri, L., Mehrotra, S., Yu, S., & Vancura, A. (2010). Transcriptional Regulation in Yeast during Diauxic Shift and Stationary Phase. *OMICS: A Journal of Integrative Biology*, *14*(6), 629–638. https://doi.org/10.1089/omi.2010.0069

Gerrish, P. J., & Lenski, R. E. (1998). The fate of competing beneficial mutations in an asexual population. *Genetica*, *102*(0), 127–144. https://doi.org/10.1023/A:1017067816551

Good, B. H., Rouzine, I. M., Balick, D. J., Hallatschek, O., & Desai, M. M. (2012). Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations. *Proceedings of*

the National Academy of Sciences, *109*(13), 4950–4955. https://doi.org/10.1073/pnas.1119910109

Granek, J. A., Kayıkçı, Ö., & Magwene, P. M. (2011). Pleiotropic signaling pathways orchestrate yeast development. *Current Opinion in Microbiology*, *14*(6), 676–681. https://doi.org/https://doi.org/10.1016/j.mib.2011.09.004

Gresham, D., & Dunham, M. J. (2014). The enduring utility of continuous culturing in experimental evolution. *Genomics*, *104*(6), 399–405. https://doi.org/10.1016/j.ygeno.2014.09.015

Gresham, D., & Hong, J. (2014). The functional basis of adaptive evolution in chemostats. *FEMS Microbiology Reviews*, *39*(1), 2–16. https://doi.org/10.1111/1574-6976.12082

Gu, Z., Eils, R., & Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, *32*(18), 2847–2849. https://doi.org/10.1093/bioinformatics/btw313

Haldane, J. B. S. (1927). A Mathematical Theory of Natural and Artificial Selection, Part V: Selection and Mutation. *Mathematical Proceedings of the Cambridge Philosophical Society*, *23*(7), 838–844. https://doi.org/10.1017/S0305004100015644

Hartwell, L. H., Hopfield, J. J., Leibler, S., & Murray, A. W. (1999). From molecular to modular cell biology. *Nature*, *402*(S6761), C47–C52. https://doi.org/10.1038/35011540

Ho, J., & Bretscher, A. (2001). Ras regulates the polarity of the yeast actin cytoskeleton through the stress response pathway. *Mol Biol Cell*, *12*(6), 1541–1555. https://doi.org/10.1091/mbc.12.6.1541

Homoto, S., & Izawa, S. (2018). Persistent actin depolarization caused by ethanol induces the formation of multiple small cortical septin rings in yeast. *Journal of Cell Science*, *131*(15), jcs217091. https://doi.org/10.1242/jcs.217091

Hong, J., & Gresham, D. (2014). Molecular Specificity, Convergence and Constraint Shape Adaptive Evolution in Nutrient-Poor Environments. *PLOS Genetics*, *10*(1), e1004041. https://doi.org/10.1371/journal.pgen.1004041

Jabbarzadeh, E. (2019). Polarity as a physiological modulator of cell function. *Frontiers in Bioscience*, *24*(3), 4728. https://doi.org/10.2741/4728

Jerison, E. R., Nguyen Ba, A. N., Desai, M. M., & Kryazhimskiy, S. (2020). Chance and necessity in the pleiotropic consequences of adaptation for budding yeast. *Nature Ecology & Evolution*, *4*(4), 601–611. https://doi.org/10.1038/s41559-020-1128-3

Kahl Lisa, J., Eckartt Kelly, N., Morales Diana, K., Price-Whelan, A., & Dietrich Lars, E. P. (2022). Light/Dark and Temperature Cycling Modulate Metabolic Electron Flow in Pseudomonas aeruginosa Biofilms. *mBio*, *13*(4), 01407–22. https://doi.org/10.1128/mbio.01407-22

Kaida, D., Yashirod, H., Toh-E, A., & Kikuchi, Y. (2002). Yeast Whi2 and Psr1-phosphatase form a complex and regulate STRE-mediated gene expression. *Genes to Cells*. https://doi.org/10.1046/j.1365-2443.2002.00538.x

Kassen, R., & Bataillon, T. (2006). Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria. *Nature Genetics*, *38*(4), 484–488. https://doi.org/10.1038/ng1751

Kirschner, M., & Gerhart, J. (1998). Evolvability. *Proceedings of the National Academy of Sciences*, *95*(15), 8420–8427. https://doi.org/10.1073/pnas.95.15.8420

Kroschwald, S., Maharana, S., Mateju, D., Malinovska, L., Nüske, E., Poser, I., Richter, D., & Alberti, S. (2015). Promiscuous interactions and protein disaggregases determine the material

state of stress-inducible RNP granules. *eLife*, *4*, e06807–e06807. https://doi.org/10.7554/eLife.06807

Kvitek, D. J., & Sherlock, G. (2013). Whole Genome, Whole Population Sequencing Reveals That Loss of Signaling Networks Is the Major Adaptive Strategy in a Constant Environment. *PLOS Genetics*, *9*(11), e1003972. https://doi.org/10.1371/journal.pgen.1003972

Laan, L., Koschwanez, J. H., & Murray, A. W. (2015). Evolutionary adaptation after crippling cell polarization follows reproducible trajectories. *eLife*, *4*. https://doi.org/10.7554/elife.09638

Lambros, M., Pechuan-Jorge, X., Biro, D., Ye, K., & Bergman, A. (2021). Emerging Adaptive Strategies Under Temperature Fluctuations in a Laboratory Evolution Experiment of Escherichia Coli. *Frontiers in Microbiology*, *12*. https://www.frontiersin.org/articles/10.3389/fmicb.2021.724982

Leadsham, J. E., Kotiadis, V. N., Tarrant, D. J., & Gourlay, C. W. (2010). Apoptosis and the yeast actin cytoskeleton. *Cell Death & Differentiation*, *17*(5), 754–762. https://doi.org/10.1038/cdd.2009.196

Lew, D. J., & Reed, S. I. (1993). Morphogenesis in the yeast cell cycle: regulation by Cdc28 and cyclins. *J Cell Biol*, *120*(6), 1305–1320. https://doi.org/10.1083/jcb.120.6.1305

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btp698

M Altrock, P., & Traulsen, A. (2009). Fixation times in evolutionary games under weak selection. *New Journal of Physics*, *11*(1), 013012. https://doi.org/10.1088/1367-2630/11/1/013012

MacLean, R. C., Bell, G., & Rainey, P. B. (2004). The evolution of a pleiotropic fitness tradeoff in Pseudomonas fluorescens. *Proceedings of the National Academy of Sciences*, *101*(21), 8072–8077. https://doi.org/10.1073/pnas.0307195101

Martin, S. G., & Arkowitz, R. A. (2014). Cell polarization in budding and fission yeasts. *FEMS Microbiology Reviews*, *38*(2), 228–253. https://doi.org/10.1111/1574-6976.12055

Mauro, A. A., & Ghalambor, C. K. (2020). Trade-offs, Pleiotropy, and Shared Molecular Pathways: A Unified View of Constraints on Adaptation. *Integrative and Comparative Biology*, *60*(2), 332–347. https://doi.org/10.1093/icb/icaa056

Miller, A. W., Befort, C., Kerr, E. O., & Dunham, M. J. (2013). Design and Use of Multiplexed Chemostat Arrays. *Journal of Visualized Experiments*, (72). https://doi.org/10.3791/50262

Morimoto, Y., & Tani, M. (2015). Synthesis of mannosylinositol phosphorylceramides is involved in maintenance of cell integrity of yeast Saccharomyces cerevisiae. *Molecular Microbiology*, *95*(4), 706–722. https://doi.org/10.1111/mmi.12896

Munn, A. L., Stevenson, B. J., Geli, M. I., & Riezman, H. (1995). end5, end6, and end7: mutations that cause actin delocalization and block the internalization step of endocytosis in Saccharomyces cerevisiae. *Molecular Biology of the Cell*, *6*(12), 1721–1742. https://doi.org/10.1091/mbc.6.12.1721

Mutavchiev, D. R., Leda, M., & Sawin, K. E. (2016). Remodeling of the Fission Yeast Cdc42 Cell-Polarity Module via the Sty1 p38 Stress-Activated Protein Kinase Pathway. *Current Biology*, *26*(21), 2921–2928. https://doi.org/https://doi.org/10.1016/j.cub.2016.08.048

Orr, H. A. (2000). ADAPTATION AND THE COST OF COMPLEXITY. *Evolution*, *54*(1), 13–20. https://doi.org/10.1111/j.0014-3820.2000.tb00002.x

Orr, H. A. (2003). The Distribution of Fitness Effects Among Beneficial Mutations. *Genetics*, *163*(4), 1519–1526. https://doi.org/10.1093/genetics/163.4.1519

Paaby, A. B., & Rockman, M. V. (2013). The many faces of pleiotropy. *Trends in genetics : TIG*, *29*(2), 66. https://doi.org/10.1016/J.TIG.2012.10.010

Phillips, P. C. (2008). Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Reviews Genetics*, *9*(11), 855–867. https://doi.org/10.1038/nrg2452

Prunskaite-Hyyryläinen, R., Shan, J., Railo, A., Heinonen, K. M., Miinalainen, I., Yan, W., Shen, B., Perreault, C., & Vainio, S. J. (2014). Wnt4, a pleiotropic signal for controlling cell polarity, basement membrane integrity, and antimüllerian hormone expression during oocyte maturation in the female follicle. *Faseb j*, *28*(4), 1568–1581. https://doi.org/10.1096/fj.13-233247

Pruyne, D., & Bretscher, A. (2000). Polarization of cell growth in yeast. *Journal of Cell Science*, *113*(4), 571–585. https://doi.org/10.1242/jcs.113.4.571

Qian, W., Ma, D., Xiao, C., Wang, Z., & Zhang, J. (2012). The Genomic Landscape and Evolutionary Resolution of Antagonistic Pleiotropy in Yeast. *Cell Reports*, *2*(5), 1399–1410. https://doi.org/10.1016/j.celrep.2012.09.017

Radcliffe, P., Trevethick, J., Tyers, M., & Sudbery, P. (1997). Deregulation of CLN1 and CLN2 in the Saccharomyces cerevisiae whi2 mutant. *Yeast*, *13*(8), 707–15. https://doi.org/10.1002/(SICI)1097-0061(19970630)13:8<707::AID-YEA130>3.0.CO;2-9

Ralser, M., Kuhl, H., Ralser, M., Werber, M., Lehrach, H., Breitenbach, M., & Timmermann, B. (2012). The Saccharomyces cerevisiae W303-K6001 cross-platform genome sequence: insights into ancestry and physiology of a laboratory mutt. *Open Biology*, *2*(8), 120093. https://doi.org/10.1098/rsob.120093

Rose, M., & Charlesworth, B. (1980). A test of evolutionary theories of senescence. *Nature*, *287*(5778), 141–142. https://doi.org/10.1038/287141a0

Saito, H. (2010). Regulation of cross-talk in yeast MAPK signaling pathways. *Current Opinion in Microbiology*, *13*(6), 677–683. https://doi.org/10.1016/j.mib.2010.09.001

Salat-Canela, C., Carmona, M., Martín-García, R., Pérez, P., Ayté, J., & Hidalgo, E. (2021). Stress-dependent inhibition of polarized cell growth through unbalancing the GEF/GAP regulation of Cdc42. *Cell Reports*, *37*(5), 109951. https://doi.org/https://doi.org/10.1016/j.celrep.2021.109951

Saul, D. J., Walton, E. F., Sudbery, P. E., & Carter, B. L. A. (1985). Saccharomyces cerevisiae whi2 Mutants in Stationary Phase Retain the Properties of Exponentially Growing Cells. *Microbiology*, *131*(9), 2245–2251. https://doi.org/https://doi.org/10.1099/00221287-131-9-2245

Sivadon, P., Bauer, F., Aigle, M., & Crouzet, M. (1995). Actin cytoskeleton and budding pattern are altered in the yeast rvs161 mutant: the Rvs161 protein shares common domains with the brain protein amphiphysin. *Molecular and General Genetics MGG*, *246*(4), 485–495. https://doi.org/10.1007/BF00290452

Sivadon, P., Crouzet, M., & Aigle, M. (1997). Functional assessment of the yeast Rvs161 and Rvs167 protein domains. *FEBS Letters*, *417*(1), 21–27. https://doi.org/https://doi.org/10.1016/S0014-5793(97)01248-9

Smethurst, D. G. J., Dawes, I. W., & Gourlay, C. W. (2014). Actin – a biosensor that determines cell fate in yeasts. *FEMS Yeast Research*, *14*(1), 89–95. https://doi.org/10.1111/1567-1364.12119

Sudbery, P. E., Goodey, A. R., & Carter, B. L. A. (1980). Genes which control cell proliferation in the yeast Saccharomyces cerevisiae. *Nature*, *288*(5789), 401–404. https://doi.org/10.1038/288401a0

Thevissen, K., Cammue, B. P. A., Lemaire, K., Winderickx, J., Dickson, R. C., Lester, R. L., Ferket, K. K. A., Van Even, F., Parret, A. H. A., & Broekaert, W. F. (2000). A gene encoding a sphingolipid biosynthesis enzyme determines the sensitivity of Saccharomyces cerevisiae to an antifungal plant defensin from dahlia (Dahlia merckii). *Proceedings of the National Academy of Sciences*, *97*(17), 9531. https://doi.org/10.1073/pnas.160077797

Thompson, B. J. (2013). Cell polarity: models and mechanisms from yeast, worms and flies. *Development*, *140*(1), 13–21. https://doi.org/10.1242/DEV.083634

Tkadlec, J., Pavlogiannis, A., Chatterjee, K., & Nowak, M. A. (2019). Population structure determines the tradeoff between fixation probability and fixation time. *Communications Biology*, *2*(1), 138. https://doi.org/10.1038/s42003-019-0373-y

Uesono, Y., Ashe, M. P., & Toh-e, A. (2004). Simultaneous yet Independent Regulation of Actin Cytoskeletal Organization and Translation Initiation by Glucose in Saccharomyces cerevisiae. *Molecular Biology of the Cell*, *15*(4), 1544–1556. https://doi.org/10.1091/mbc.e03-12-0877

Vasicova, P., Rinnerthaler, M., Haskova, D., Novakova, L., Malcova, I., Breitenbach, M., & Hasek, J. (2016). Formaldehyde fixation is detrimental to actin cables in glucose-depleted S. cerevisiae cells. *Microbial Cell*, *3*(5), 206–214. https://doi.org/10.15698/mic2016.05.499

Wagner, G. P., & Altenberg, L. (1996). Complex adaptations and the evolution of evolvability. *Evolution*, *50*(3), 967–976. https://doi.org/10.1111/J.1558-5646.1996.TB02339.X

Wagner, G. P., & Zhang, J. (2011). The pleiotropic structure of the genotype–phenotype map: the evolvability of complex organisms. *Nature Reviews Genetics*, *12*(3), 204–213. https://doi.org/10.1038/nrg2949

Wahl, L. M., Gerrish, P. J., & Saika-Voivod, I. (2002). Evaluating the impact of population bottlenecks in experimental evolution. *Genetics*, *162*(2), 961–971. https://pubmed.ncbi.nlm.nih.gov/12399403%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1462272/

Waltermann, C., & Klipp, E. (2010). Signal integration in budding yeast. *Biochemical Society Transactions*, *38*(5), 1257–1264. https://doi.org/10.1042/BST0381257

Waxman, D., & Peck, J. R. (1998). Pleiotropy and the Preservation of Perfection. *Science*, *279*(5354), 1210–1213. https://doi.org/10.1126/science.279.5354.1210

Wein, T., & Dagan, T. (2019). The Effect of Population Bottleneck Size and Selective Regime on Genetic Diversity and Evolvability in Bacteria. *Genome Biology and Evolution*, *11*(11), 3283–3290. https://doi.org/10.1093/gbe/evz243

Welch, J. J., & Waxman, D. (2003). MODULARITY AND THE COST OF COMPLEXITY. *Evolution*, *57*(8), 1723–1734. https://doi.org/10.1111/j.0014-3820.2003.tb00581.x

Welch, M. D., Holtzman, D. A., & Drubin, D. G. (1994). The yeast actin cytoskeleton. *Current Opinion in Cell Biology*, *6*(1), 110–119. https://doi.org/10.1016/0955-0674(94)90124-4

Yoshida, S., & Pellman, D. (2008). Plugging the GAP between cell polarity and cell cycle. *EMBO reports*, *9*(1), 39–41. https://doi.org/10.1038/SJ.EMBOR.7401142

Zou, L., Sriswasdi, S., Ross, B., Missiuro, P. V., Liu, J., & Ge, H. (2008). Systematic Analysis of Pleiotropy in C. elegans Early Embryogenesis. *PLOS Computational Biology*, *4*(2), e1000003. https://doi.org/10.1371/journal.pcbi.1000003

3.5. Supplement

3.5.1. Estimation of relative fixation times

Serial dilution approaches for experimental evolution often include severe population bottlenecks which affect the fixation time of beneficial mutations. Here, we provide an estimate of how the fixation time of a beneficial mutation in a batch culture population relates to the fixation time of the same mutation in a continuous culture. The total fixation time T_{fix} can be subdivided in the following two components:

$$T_{fix} = T_m + T_s, \tag{3.3}$$

where T_m is the time it takes for a mutation to emerge that will eventually fix and T_s is the time it takes for this mutation to sweep the population. For the classical case of a large population of fixed size, it has been derived that the fixation probability of a beneficial mutation is proportional to the selective coefficient s (Haldane, 1927). Because population size is large ($N_c \sim 10^9$) and remains roughly constant in a continuous culture, we use this result to approximate the fixation probability of mutations for samples evolved in the chemostat:

$$P(chemostat) \approx 2s. \tag{3.4}$$

Which is valid for small values of s ($s \ll 1$). In the batch culture evolved samples, population size does not remain constant, but rather fluctuates during each serial passaging between the values N_0 ($\sim 10^6$) and N_f ($\sim 10^9$). Here, N_f is the number of individuals just before reseeding the population in fresh media and N_0 is the number of individuals just after reseeding, such that $N_0 = D \cdot N_f$ with D the dilution factor. Wahl et al. (2002) have shown that periodic bottlenecks caused by serial passaging reduce the fixation probability by a factor $D[\ln D]^2$, such that the fixation probability in a batch culture can be approximated as:

$$P(batch) \approx 2s \cdot D[\ln D]^2 \tag{3.5}$$

If we now calculate the relative fixation probabilities between the two environment using the $D = 0.001$ dilution factor we have used in our experiments, we find:

$$\frac{P(chemostat)}{P(batch)} \approx \frac{2s}{2s \cdot D[\ln D]^2} = \frac{1}{D[\ln D]^2} \approx 21. \tag{3.6}$$

Thus, for an equal rate of beneficial mutations in both conditions we can expect that it will take approximately 21 fold more generations for a mutation to emerge that eventually fixes in the batch culture relative to the continuous culture. We can therefore write that:

$$T_{m,batch} \approx 21 \cdot T_{m,chemostat}. \tag{3.7}$$

For the sweeping times (in generations) of mutations that fix in the batch culture, we use the approximation for large populations that are evolved through serial passaging given by Campos and Wahl (2009):

$$T_{s,batch} \approx 2/s \ln N_0. \tag{3.8}$$

In the continuous culture the change in mutant frequency can be modelled as a Moran process of a well-mixed population, for which the fixation time is of the order (Antal & Scheuring, 2006; M Altrock & Traulsen, 2009; Tkadlec et al., 2019):

$$T_{s,chemostat} \approx (1 + \frac{2}{s}) \ln N_c \approx \frac{2}{s} \ln N_c, \tag{3.9}$$

where we have used that $s \ll 1$ for the last approximation on the right hand side. Using this equation, we can calculate the ratio of seeping times in the two environments:

$$\frac{T_{s,batch}}{T_{s,chemostat}} \approx \frac{2/s \ln N_0}{2/s \ln N_c} = \frac{\ln N_0}{\ln N_c} = \frac{2}{3}, \tag{3.10}$$

where we have used that $N_0 = 10^6$ and $N_c = 10^9$. Taken together, we see that for the total fixation time in the batch culture we obtain:

$$T_{fix,batch} = T_{m,batch} + T_{s,batch} \approx 21 \cdot T_{m,chemostat} + \frac{2}{3} T_{s,chemostat}. \tag{3.11}$$

Thus, the difference in fixation times of beneficial mutations in the batch culture and chemostat is dominated by the larger number of mutations that need to be sampled in a batch culture before a mutation emerges that fixates in the population.

To roughly estimate the expected number of generations it would take for the for a beneficial mutation to reach fixation in the chemostat, we consider a mutation that has a selection coefficient $s = 0.1$. For comparison, this mutation would decrease the doubling time of a *bem3*Δ*nrp1*Δ mutant from $T_d = 110$ to $T_d = 100$ minutes. The time for a *de novo* generated mutation with this selection coefficient to arise in the population can be estimated by:

$$T_{m,chemostat} = \frac{1}{N_c \cdot U_B \cdot \rho(s) \cdot P(chemostat)} \approx \frac{1}{N_c \cdot U_B \cdot \rho(s) \cdot 2s}. \tag{3.12}$$

Here, U_B is the rate at which beneficial mutations occur per cell division and $\rho(s)$ is the probability that a beneficial mutation has a selection coefficient of s or higher. To determine $\rho(s)$, we assume that the distribution of fitness effects follows an exponential distribution (Barlukova & Rouzine, 2021; Good et al., 2012; Kassen & Bataillon, 2006; Orr, 2003) with a mean of $\sigma = 0.01$:

$$\rho(s) = \int_s^\infty \frac{1}{\sigma} e^{-s/\sigma} \tag{3.13}$$

Taking the beneficial mutation rate to be of the order $U_B \sim 10^{-5}$ per cell per generation, we obtain the result that the expected number of generations it will take for a beneficial mutation to occur that escapes drift is

$$T_{m,chemostat} \sim 11 \text{ generations}. \tag{3.14}$$

The time it takes for this mutation to sweep the population can be calculated using equation 3.9, which for $s = 0.1$ gives:

$$T_{s,chemostat} \sim 207 \text{ generations}. \tag{3.15}$$

These results show that while we expect that beneficial mutations arise after only few generations in the chemostat, these mutations sweep through the population at a relatively slow rate. Converting the values of T_m and T_s of the chemostat to the expected values for the batch culture using equation 3.11 results in:

$$T_{m,batch} \sim 231 \text{ generations} \tag{3.16}$$

and

$$T_{s,batch} \sim 138 \text{ generations.} \tag{3.17}$$

Thus, the situation is reversed in the batch culture: beneficial mutations emerge only sparingly due to the loss of genetic variation caused by the population bottlenecks, but once they emerge these mutations sweep through the population faster than in a continuous culture. This result is in agreement with our finding that *bem3*Δ*nrp1*Δ populations evolved through serial dilutions show no significant difference in their phenotype compared to their ancestor after 70 generations. Comparing the total fixation times to the true number of generations that we evolved the *bem3*Δ*nrp1*Δ populations in the two conditions indicates that beneficial mutations in the batch culture (evolved for 300 generations, expected $T_{fix} = 369$ generations) are expected to be closer to fixation than beneficial mutations in the chemostat (evolved for 70 generations, expected $T_{fix} = 218$ generations). An important remark is that in our calculation of the fixation times, we have assumed that beneficial mutations are rare enough that they only emerge in succession and never simultaneously, thereby ignoring clonal interference (Desai & Fisher, 2007; Desai et al., 2007; Gerrish & Lenski, 1998). Clonal interference is likely to play a larger role in continuous culture than in batch cultures due to the larger amount of genetic variation and may further reduce the speed of evolution.

Estimation of population sizes

The approximate population sizes were determined from the optical density (OD600) measurements taken of the growing populations in the two environments. These measurements show that the OD600 is ~ 3 for populations growing in the chemostat and ~ 10 for populations growing in the batch culture just before passaging. Using the approximation that a population with an OD600 of 1 contains of the order 10^7 cells/ml (Day et al., 2004) and the total culture volumes of 20 ml for the chemostat and 10 ml for the batch culture, we find that both the chemostat and batch culture populations contain of the order of 10^9 individuals.

3.5.2. Supplementary figures

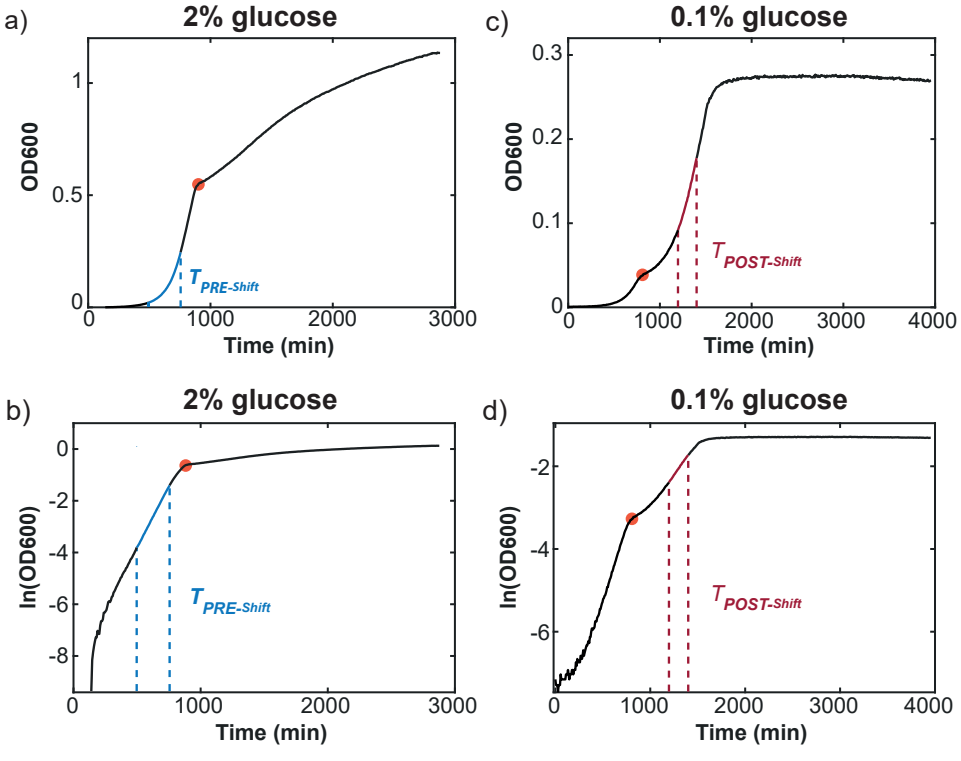


Figure S3.1. Low glucose media induces earlier diauxic shift, allowing better quantification of growth after the diauxic shift. (**a**,**b**) When grown on high glucose (2%) media, most of the biomass is produced by fermentation. (**c**,**d**) Lowering the glucose concentration to 0.1% induces an earlier diauxic shift and improves visualization and quantification of growth after the shift. Dashed lines indicate the period of exponential growth for growth in YP+2% glucose (blue) and for growth in YP+0.1% glucose (red). Although the diauxic shift phase is clearly visible in both cases as a temporary cessation of growth (indicated by the red dots), the high OD at which the diauxic shift occurs in 2% glucose media makes it unsuitable to quantify growth beyond this point due to the possible non-linear relationship between cell density and absorbance at high OD values. Conversely, in 0.1% glucose media the diauxic shift occurs at a relatively low density, decreasing the amount of signal relative to the noise during pre-diauxic growth. Therefore, population doubling times before passage through the diauxic shift ($T_{PRE-Shift}$) were determined from growth data in 2% glucose media, while population doubling times after passage through the diauxic shift ($T_{POST-Shift}$) were obtained from growth data 0.1% glucose media.

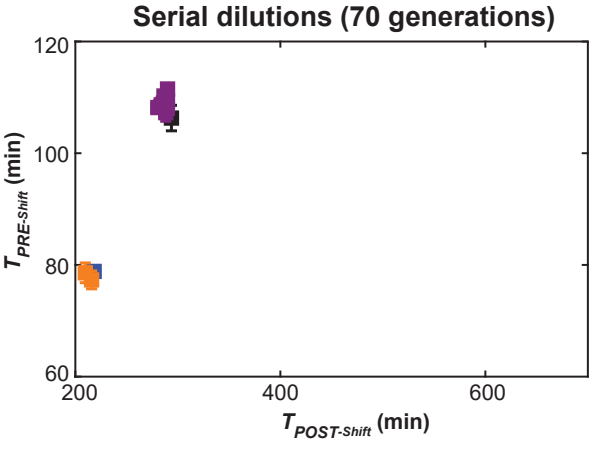


Figure S3.2. Scatter plot of $T_{PRE-Shift}$ vs $T_{POST-Shift}$ after 70 generations in a batch culture. After 70 generations of evolution in a batch culture the evolved cell lines were still phenotypically highly similar to their ancestral strains. This was the case for both the *bem3*Δ*nrp1*Δ mutants and the wildtype strains.

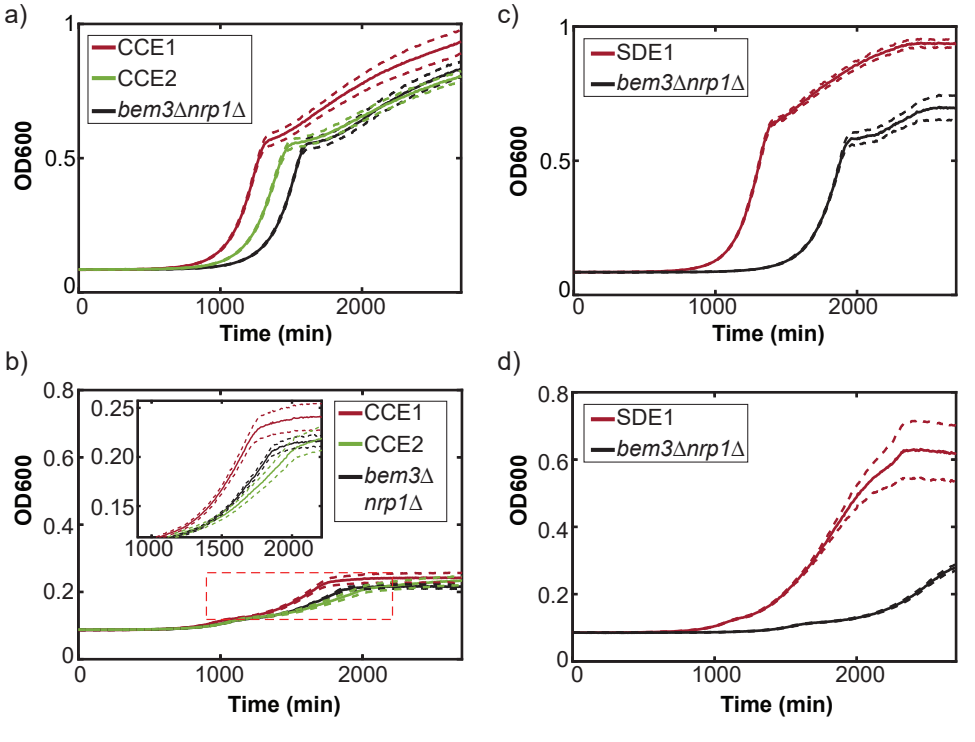


Figure S3.3. Growth curves of selected evolved lines CCE1, CCE2 and SDE1. (**a**,**b**) Bulk measurements in (a) 2% dextrose and (b) 0.1% dextrose media of selected evolved mutant lines CCE1 and CCE2 from the continuous culture experiment together with the mutant ancestor strain. (**c**,**d**) Bulk measurements in 2% dextrose (**c**) and 0.1% dextrose (**d**) media selected evolved mutant line SDE1 from the batch culture experiment. Dashed lines display the SEM.

4

A police officer finds a drunk man late at night crawling on his hands and knees on a sidewalk under a streetlight. Questioned, the drunk man tells her he's looking for his wallet. When the officer asks if he's sure that he dropped the wallet here, the man replies that he thinks he more likely dropped it across the street. Then why are you looking over here? asks the befuddled officer. Because the light's better here, explains the drunk man.

— David H. Freedman

Science is a bit like the joke about the drunk who is looking under a lamppost for a key that he has lost on the other side of the street, because that's where the light is. It has no other choice.

— Noam Chomsky

Saturated transposon analysis in yeast as a one-step method to quantify the fitness effects of gene disruptions on a genome-wide Scale

Abstract Transposon insertion site sequencing (TIS) is an extremely powerful tool that has greatly advanced our knowledge of functional genomics. However, studies using TIS often focus on gene essentiality and neglect possibly interesting but subtle differences in the importance of genes for fitness. As shown by other studies, expanding the analysis of TIS data to allow a quantitative estimate of fitness has important applications in genetics and evolutionary biology. Here, we present a method to estimate the fitness of gene disruption mutants on a quantitative level using data obtained from a TIS screen developed for the yeast *Saccharomyces cerevisiae* called SATAY. We show that using the average read count per transposon insertion site provides a metric for fitness that is robust across biological and technical replicate experiments. Importantly, the ability to resolve differences between gene disruption mutants with low fitness crucially depends on the inclusion of insertion sites that are not captured by the sequencing data to determine the mean read count. In contrast, increasing the sequencing depth is an ineffective strategy to improve the resolution of the fitness estimates, indicating that SATAY datasets typically reside far from the condition where every insertion site is represented by at least one read. While our method generates reproducible results across replicate SATAY datasets, the obtained fitness distribution differs substantially from those obtained using other techniques. Currently, it is unclear whether these inconsistencies caused by technical or biological differences. Our findings underscore the importance of exploring the variability in fitness distributions across different genetic backgrounds and environments.

4.1. Introduction

Measuring the phenotype of gene deletion mutants has been instrumental to our understanding of cell and evolutionary biology. In particular, the relation between genotype and fitness is a key element to move from descriptive to predictive evolutionary models. This mapping from genotype to fitness is typically conceptualized in the form of a fitness landscape. Importantly, the degree to which evolution can be predicted depends on the structure of this fitness landscape. Theoretical work has shown that the ruggedness of this landscape (that is, the number of fitness peaks) is an important feature that controls the predictability of evolutionary pathways (Bank et al., 2016; Franke et al., 2011; S. Kauffman & Levin, 1987; S. A. Kauffman & Weinberger, 1989). However, the construction of empirical fitness landscapes remains challenging. Traditional methods to determine fitness are based on growth measurements of reconstructed mutants harbouring the mutation of interest. These approaches are generally low-throughput, allowing only a handful of mutations to be assessed each time (Chou et al., 2011; Hietpas et al., 2011; Lozovsky et al., 2009; Mira et al., 2015; Weinreich, 2006). As the field of evolutionary biology is moving towards a more holistic view, there has been an increased demand in the past decades for techniques that allow evaluation of a large number of mutants in a single assay (De Visser & Krug, 2014; de Visser

et al., 2018).

While complete fitness maps at the resolution of single point mutations are still experimentally infeasible (Johnson et al., 2023), considerable progress has been made in methods to analyze fitness in large libraries of gene disruption or deletion mutants. One of the earliest examples of these high-throughput methods is the Synthetic Genetic Array (SGA) which was developed for yeast (Baryshnikova, Costanzo, Dixon, et al., 2010; Baryshnikova, Costanzo, Kim, et al., 2010; Yan Tong & Boone, 2006) and for bacteria (Butland et al., 2008). A striking achievement of the SGA was the construction of a global map of genetic interactions between non-essential genes of the *Saccharomyces cerevisiae* genome (Costanzo et al., 2010; Costanzo et al., 2016; Usaj et al., 2017). However, the automated SGA workflow relies on robotics for library construction and fitness assays (Baryshnikova, Costanzo, Dixon, et al., 2010; Giaever & Nislow, 2014; Kuzmin et al., 2021), which profoundly reduces its general accessibility due to associated costs. As a consequence, SGA has only been used sparingly and the majority of the reported fitness values are derived from a single mutant library. This is a crucial limitation of SGA, as it is known that fitness effects can strongly depend on genetic background. For example, a large scale study by van Leeuwen et al., 2020 showed that an astonishing 17% of the annotated essential genes of budding yeast were found to be dispensable in a different genetic context. Thus, methods to estimate the fitness effect of gene disruptions on a genome wide scale should ideally be easily applicable across different genetic backgrounds and environmental conditions.

Newer and more flexible techniques for high-throughput fitness measurements are based on pooled fitness assays, followed by next-generation sequencing to detect changes in mutant frequencies (Smith et al., 2009). In Barcode sequencing (Bar-seq), mutant genomes are tagged with a unique nucleotide sequence that allows their identification after sequencing (D. G. Robinson et al., 2014). Read counts are then typically converted into a metric for fitness by calculating the log-frequency slope of each barcode measured between two timepoints (Johnson et al., 2019; Venkataram et al., 2016; Wetmore et al., 2015). Although the pooled assay of Bar-seq greatly facilitates the fitness assessment, it has the same issues as SGA with regard to the laborious steps required for mutant library construction (Wetmore et al., 2015). In addition, the number of mutants that can be assessed in a single Bar-seq experiment is relatively low (typically between 1,000-5,000) due to the limited availability of unique barcodes (Johnson et al., 2019; Venkataram et al., 2016; Wetmore et al., 2015). An alternative to Bar-seq that allows simple *de-novo* generation of mutant libraries and is not limited by barcode availability is transposon insertion sequencing (TIS). TIS methods utilize the ability of transposons to randomly translocate between different molecules of DNA to generate a library of gene disruption mutants (Cain et al., 2020; van Opijnen et al., 2009; van Opijnen & Camilli, 2013). Typically, transposon mutagenesis is efficient enough to produce libraries consisting of more than 100,000 mutants. A single library will often contain multiple mutants carrying disruptions at different locations in the same gene (Michel et al., 2017; Rifat et al., 2021; van Opijnen et al., 2009). As mutants are identified based on the the transposon insertion site by sequencing the transposon-genome junction, the genomic library preparation and bioinformatics analysis of TIS data are relatively complex. However, protocols have been developed that combine TIS with Bar-seq to simplify these steps (Wetmore et al., 2015).

Despite its experimental flexibility, TIS has only scarcely been used to generate quantitative fitness maps of gene disruptions on a global scale (van Opijnen et al., 2009; van Opijnen & Camilli, 2012; van Opijnen et al., 2014). In essence, the approach of estimating gene disruption fitness from read count frequencies is similar for TIS and Bar-seq. (van Opijnen et al., 2009) have indeed

used the log-fold change in read count frequency to quantify fitness from TIS data of *Streptococcus pneumoniae* with results that were in good agreement with measured growth rate of gene deletion strains. However, methods developed to estimate fitness for one version of TIS can in general not directly be applied to other TIS variants for the following two reasons. First, different variants of TIS may suffer from their own type of insertional bias that depends on the specific type of transposon system that is used (Chao et al., 2016; Green et al., 2012). The ubiquity of insertion biases in TIS is becoming increasingly clear, as deep analyses are uncovering biases for transposons with a previously assumed uniform insertion profile (DeJesus et al., 2017; Green et al., 2012). Second, some transposons used in TIS can only insert at specific nucleotide sequences (Goodman et al., 2009; van Opijnen et al., 2009), while others can insert anywhere in the genome (Biery, 2000; Michel et al., 2017; van Opijnen & Levin, 2020). This difference in accessible insertion sites can lead to marked differences in the expected complexity (the number of different mutants present) of the TIS library. As a consequence, the typical number of reads acquired per mutant can vary substantially between different TIS protocols.

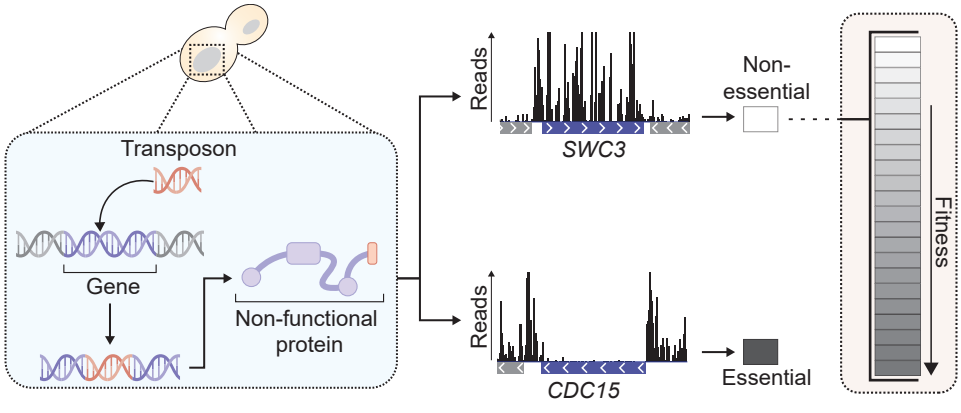


Figure 4.1. Illustration of the aim of this chapter. While existing methods for analyzing TIS data primarily categorize genes as either essential or non-essential based on transposon insertion density, our goal is to establish a method that quantifies the fitness effect of losing a non-essential gene on a more gradual scale.

Recently, a TIS method named SAturated Transposon Analysis in Yeast (SATAY) has been developed for the yeast *Saccharomyces cerevisiae*. In their original paper, Michel et al., 2017 used SATAY to identify changes in the fitness contribution of genes in different genetic and environmental contexts through a one-on-one comparison of their transposon insertion density. While effective, this approach can only identify fitness changes for the same gene across datasets, and not for different genes within the same dataset. In addition, it neglects possible subtle differences that are reflected in read count but not in the insertion density. Here, we describe an approach to generate quantitative fitness maps from read count data obtained from a single SATAY dataset (Figure 4.1) which is based on models frequently used to analyze count data from RNA-seq experiments (D. G. Robinson et al., 2014). We demonstrate that including transposon insertions that are not visible through sequencing due to low abundance is crucial for resolving fitness differences in the lower end of the spectrum. Furthermore, we show that our method generates reproducible results across technical and biological replicate experiments, but has substantial differences with fitness and genetic interaction maps produced with Bar-seq and SGA in other studies.

4.2. Results

4.2.1. Bias correction methods

Insertions near gene edges are less likely to result in loss of gene function

The read counts in TIS datasets are generally expected to correlate with mutant fitness, as fitter mutants will increase in frequency during the growth assay. However, in practice this correlation between fitness and read counts becomes obscured due to noise derived from stochastic growth trajectories and population sampling. An effective way to reduce the influence of these biological and technical noise sources on the fitness estimates is to average over several replicate measurements of the same mutant. Studies using TIS often assume transposon insertions to result in complete loss of gene function (van Opijnen et al., 2009). Under this assumption, read counts obtained from different insertion sites within the same gene can be considered replicate measurements of the same gene deletion mutant and their averaging is justified. However, in other TIS systems it has been found that insertions near the 5' and 3' end of a gene are less likely complete loss of gene function (Jacobs et al., 2003). Similarly, studies using SATAY have reported that genes considered to be essential can sometimes tolerate insertions close to the gene ends while central regions remain empty (Michel et al., 2017). If this higher tolerance to insertions near gene ends is a general phenomenon affecting all genes, averaging the read counts of the entire coding region of a gene would create a bias in the fitness estimate.

We examined whether there exists a general trend of higher read counts for insertions close to gene edges that can be observed on a genome-wide level. To do so, we segmented the open reading frames of all annotated genes in the genome into 20 equally sized bins, such that each bin amounts to 5% of the coding sequence (Figure 4.2a). For each segment, we then calculated the average number of reads per transposon insertion that mapped to the respective segment. For genes annotated as non-essential, we observed that insertions within the first and last 10% of a gene tend to acquire more reads than insertions in the central 80% of the gene, although this effect is weak (Figure 4.2b). However, this non-uniformity was clearly visible for genes annotated as essential (Figure 4.2c). This difference between essential and non-essential genes is expected, as the fitness difference between a complete and partial knock-out should typically be larger for essential genes. Interestingly, the read count profile of essential genes shows that the bias towards higher read counts is strongest for insertions close to the stop codon. This likely reflects the mechanism for gene inactivation of the MiniDS transposon, which is based on creating gene truncations by introducing several early stop codons in the open reading frame (Michel et al., 2017) . Thus, insertions that lead to truncations close to the C-terminal part of the protein will often still allow the protein to (partially) retain its functionality.

Since insertions that do not lead to full gene knockouts invalidate the averaging over the read counts of different insertion sites, we excluded all insertions that map within the first or last 10% of a coding region. However, even after removal of these insertion sites we find that the average read count distributions for essential and non-essential genes still overlap to some extent (Figure 4.2d). This indicates that using only the average read count as a metric for fitness does not allow us to distinguish essential from non-essential genes. A possible cause of this effect is that insertions in the essential regions of a gene remain unobserved because the corresponding mutants are lost before they can be sequenced. Neglecting these unobserved insertions means that our fitness estimates would be based completely on insertion sites that, for biological or technical reasons, have a higher read count than the typical insertion site. As a result, the fitness

estimates become biased towards higher values and the ability to resolve fitness differences for low-fitness mutants is lost. In support of the idea that unobserved insertion sites play a role in this bias, we find that the consecutive span of insertion sites that appear unoccupied tend to be longer for essential genes (Figure 4.2e). Thus, essential genes are therefore ideally identified using other analysis methods, for example those based on the insertion-free span (DeJesus et al., 2013; Griffin et al., 2011).

Correcting for the preferential insertion of MiniDS at pericentromeric sites

Mutants that are rare after the fitness assay, either because they grow poorly or have a high death rate, are likely to be lost during the sampling steps leading up to sequencing. Hence, the insertion sites associated with these mutants remain undiscovered in the final SATAY dataset (Figure 4.3a). While this loss of rare transposon mutants forms the basis for the identification of essential genes, it can distort the relation between mutant fitness and the read counts per insertion site. Specifically, including only insertion events that are observable from the sequencing data increases the susceptibility of the fitness estimate to gene disruption mutants that behave differently than average. In addition, it puts the lower boundary for the average read count at 1, which severely reduces the number of distinct fitness levels that can be resolved for lower fitness values.

One approach to reduce these distortions is to include the undiscovered insertion sites in the fitness estimate as insertion sites with a read count of zero. This makes the fitness value of a gene dependent on the combination of the average read count per insertion site and the fraction of insertion sites that yield a read count larger than zero. To determine the number of insertion sites with zero reads for each gene, we use the global insertion density post-library expansion to infer an expected insertion rate. However, an issue with this approach is that the probability of insertion of the MiniDS transposon in SATAY depends on the distance of the insertion site to the centromeric regions of the chromosome. Because the MiniDS transposon preferentially inserts close to its excision site, the experimental design of SATAY has been reported to yield higher transposon densities in pericentromeric regions when compared to more distal regions. To verify the existence this centromere bias in our insertion data, we plotted the cumulative insertion count as a function of varying distances to the centromere (Figure 4.3b). This analysis indeed revealed an enrichment of transposons in genomic regions that are closer than approximately 200 kb from a centromere. We confirmed that this effect is not due to a lower density of essential genes compared to non-essential genes in pericentromeric regions (Figure S4.1). Thus, if left unaddressed, this bias would lead to underestimation of the expected insertion rate for genes close to centromeres.

Ideally, a bias-corrected curve for the transposon insertion rate should follow the trend of the empirical data, while smoothing out any high frequency fluctuations. This is because such fluctuations are likely caused by the specific genetic content of the different genomic regions. We assessed two methods to model and correct for the centromeric bias: (1) a power-law and (2) a 4^{th} order polynomial (Figure 4.3c). While the plot of cumulative insertion count against distance seems to approach a power-law relationship, inspection of the residuals revealed that the fit systematically over- and underestimates different sections of the curve (Figure 4.3c and d). The increased flexibility of a 4^{th} order polynomial provided a better approximation for distances smaller than 200 kb, but began to fluctuate wildly for larger distances (Figure 4.3c and d).

Based on these findings, we decided to use the polynomial function to model the changes

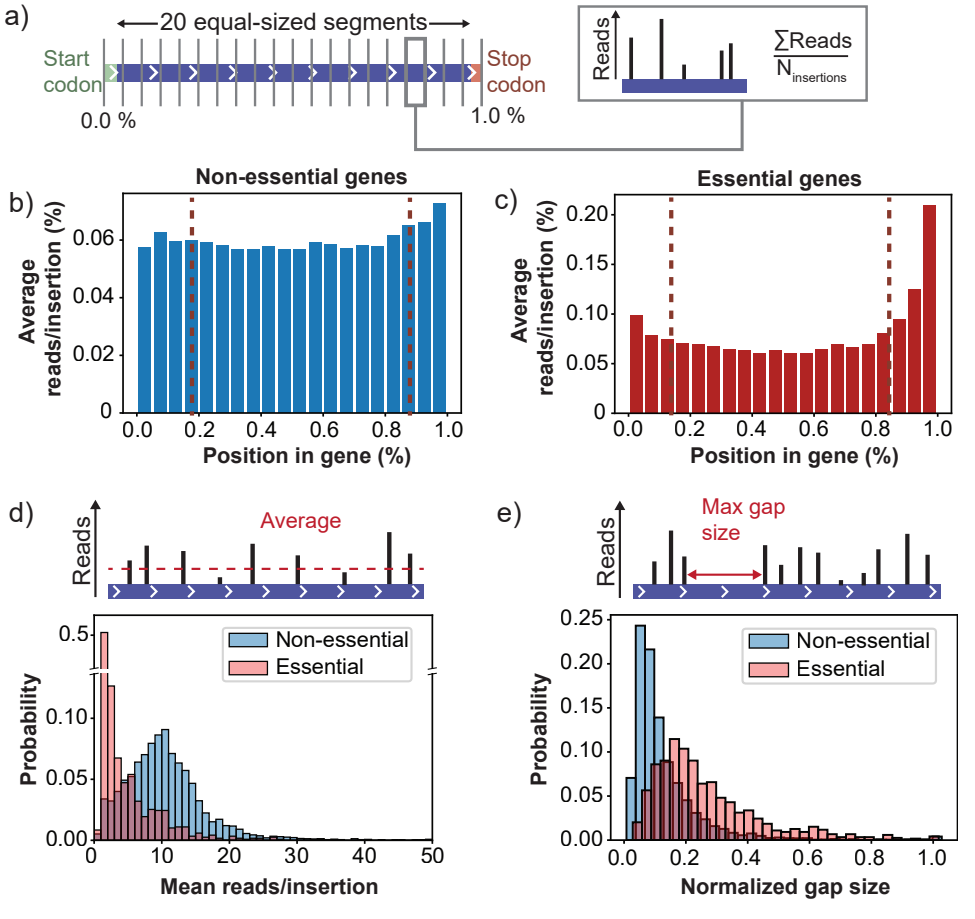


Figure 4.2. Transposon insertions near the start and end of a gene are less likely to generate a gene knockout. (**a**) To identify a possible dependence of the read counts on the position of a transposon insertion within a gene, we split the open reading frame of each gene in 20 equally sized segments. Each segment therefore covers 5% of the coding region of a gene. For every segment we calculated the average read count per insertion site to obtain a profile along the coding region. (**b**) Profile of the read counts against the position within the coding sequence, averaged over all non-essential genes. The profile shows that insertions near the start and stop codon of a gene obtain a slightly higher read count than insertions in the central 80% of the coding region. (**c**) This effect is visibly stronger for genes that are annotated as essential, indicating that insertions near the gene edges do not cause a full gene knockout. (**d**) The read count per insertion averaged over the central 80% of the coding region for essential and non-essential genes. While essential genes typically have a lower average read count per insertion site than non-essential genes, their distributions still overlap. (**e**) Histogram of the largest span that appears free of insertions, expressed as a fraction of the gene's length. The distribution for essential genes has a tail towards longer insertion free spans (up to the full length of the coding region), while the spans are typically shorter for non-essential genes.

in the insertion rate up to distances 200 kb from the centromere while assuming a constant insertion rate beyond 200 kb. This assumption is based on the idea that the centromeric bias is a 'memoryless' feature, such that the insertion rate no longer depends on the distance to the centromere beyond a certain point. Furthermore, we note that the insertion rate at 200 kb inferred from the polynomial fit closely matched the insertion rate obtained when fitting a linear curve to the portion of the curve for distances of 200 kb and larger (Figure 4.3b and d). By taking the derivative of the polynomial fit with respect to the distance from the centromere, we obtain the global trend of changes in the insertion rate. Importantly, this estimate does not capture the local fluctuations in insertion rate, which are likely caused by the specific properties of genes at different positions along the chromosome. Overall, the insertion rate varied between approximately 0.12 bp^{-1} for regions close to the centromere to 0.067 bp^{-1} for more distal regions. Because centromere bias should not play a significant role for regions very far from the centromere, we artificially flattened the fitted curve by setting a constant rate for positions located more than 200 kb away from the centromere. In summary, we use the following equation to determine a bias corrected insertion rate λ:

$$\lambda(r_c) = \begin{cases} a_0 + a_1 r_c + a_2 r_c^2 + a_3 r_c^3, & \text{for } r_c < 2 \cdot 10^5 \\ \lambda(r_c = 2 \cdot 10^5), & \text{for } r_c \geq 2 \cdot 10^5 \end{cases} \tag{4.1}$$

Where r_c is the distance from the centromere in base pair and a_{0-3} are the coefficients obtained from least squares polynomial fit. Because we assume that the insertion rate remains approximately constant over the span of a gene's coding sequence, the expected number of insertions for a gene g is calculated by multiplying the insertion rate with gene size:

$$E(X_g) = \lambda(r_{c-g}) \cdot L_g, \tag{4.2}$$

with r_{c-g} the distance from gene g to the centromere measured from the gene's center and L_g the length of the gene in basepairs. The number of zero read count sites is then estimated to be equal to the difference between $E(X_g)$ and the number of observed insertions $O(X_g)$ when $E(X_g) > O(X_g)$ (Figure 4.3a). When the number of observed insertions exceeds their expected value, we conclude that no unobserved insertions sites exist for that gene. We find that this procedure is able to correct for the skew in the expected number of insertion sites that produce zero reads for genes close to the centromere (Figure S4.2).

To determine whether our model of the expected insertion rate was effective at identifying the larger fraction of undiscovered insertion sites in annotated essential genes, we plotted the difference in the expected and observed insertion counts for all genes (Figure 4.3e). Here, we consider this difference to reflect the number of transposon insertions that remain undiscovered after sequencing. It can be seen that for nearly all essential genes, the expected insertion density is higher than what is found from the sequencing data. Alternatively, for non-essential genes, the difference between expected and observed insertion counts follows a Gaussian-like distribution around zero that partially overlaps with the distribution for essential genes. Thus, including undiscovered insertion sites provides additional information on essentialty, although not sufficient to distinguish essential from non-essential genes. Nevertheless, our results show that including these unobserved insertion sites to estimate fitness is crucial to resolve differences between low-fitness mutants.

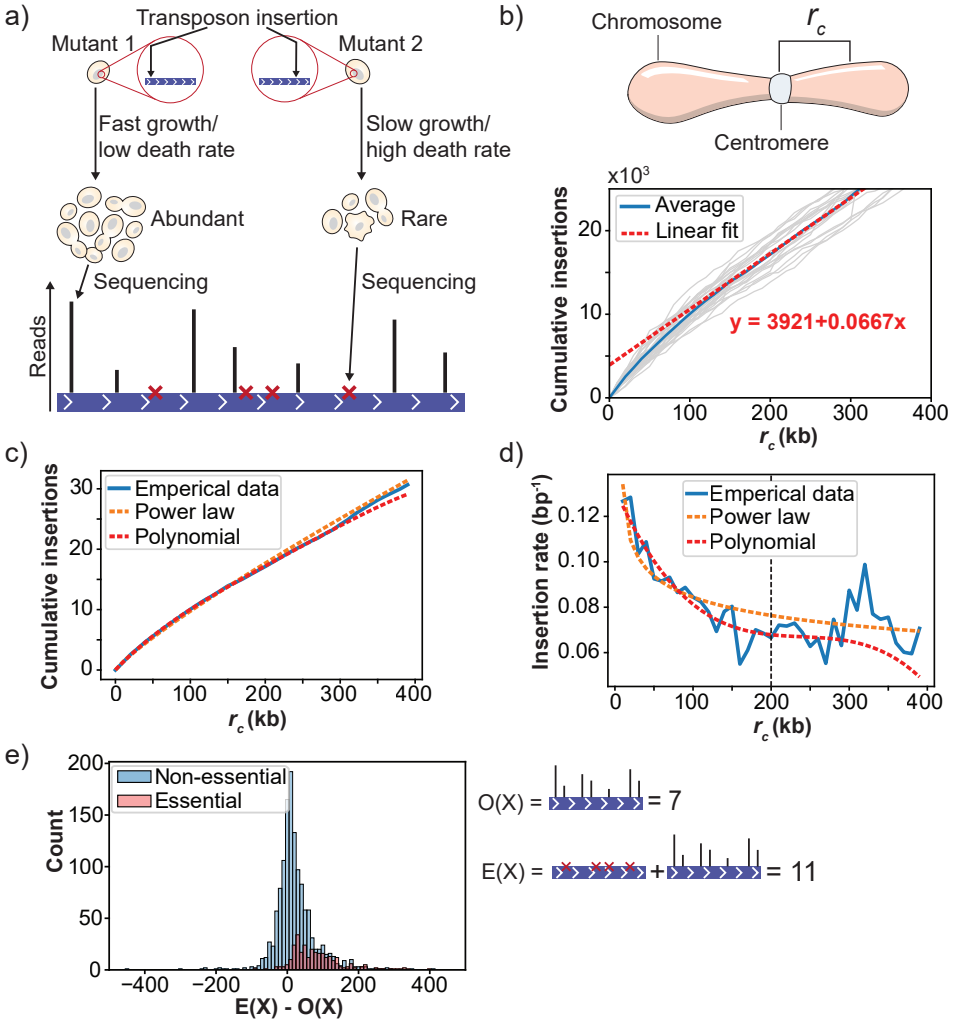


Figure 4.3. Correcting for the insertion bias near centromeres to determine the expected transposon insertion rate of genes. (**a**) The goal of calculating the expected insertion rate is to estimate the number of insertion sites that produce no reads because the mutants were lost during the sampling of the population due to their low abundance. (**b**) The empirical insertion rate depends on the distance of a gene to the centromere. To visualize this bias we determined the number of transposons that mapped within a distance r_c from the chromosome centromere for different values of r_c, as done previously by Michel et al. (2017). The plots of the cumulative insertions for the two halves of each individual chromosome are shown as gray lines, the averaged curve is shown as a blue line. The non-zero intercept of a linear fit of the portion of the average curve for distances r_c >300 kb demonstrates the existence of the centromere bias in our dataset. (**c**) The averaged cumulative plot for distances R_c <400 kb was fitted with an exponential function and a 4th order polynomial. While the power law function does approximate the shape of the curve, it systematically under- and overfits different portions of the curve. Overall, the 4th order polynomial better approximates the curve. (**d**) The approximated empirical insertion rate by the fitted power law and polynomial functions. The plot shows that while the approximation by the polynomial function is better for r_c <200 kb, the polynomial starts to oscillate for larger distances. (**e**) The difference between the expected (E(X)) and observed (O(X)) transposon insertion rates for essential and non-essential genes. While the average difference is close to zero for non-essential genes, it becomes positive for essential genes.

4.2.2. Estimating fitness from read counts in SATAY

Calculating mean mutant fitness

Estimating the fitness effect of gene disruptions based on SATAY data relies on the relation between the observed read counts and mutant abundance. Specifically, fitter mutants proliferate more rapidly than less fit mutants, increase their relative abundance in the population and finally produce more read counts. The challenge of determining fitness from SATAY data can therefore be formulated as estimating the fitness parameter μ_g, which we will refer to as the growth rate, using the observed read counts $y_{g,i}$ at the insertion location i within gene g. While we will refer to and treat μ_g as a growth rate, we note that the interpretation of μ_g is significantly different from the conventional meaning of a growth rate in biology. Typically, growth rate is measured as the rate of population expansion during the exponential growth phase. Hence, by proposing a relation between growth rate and read counts we implicitly assume that the population remains in the exponential phase for the complete duration of the fitness assay. In reality, the length of the different growth phases (lag, exponential, stationary) can vary substantially between genetic backgrounds. All these variations will impact mutant abundance and are implicitly incorporated in the parameter μ_g rather than being explicitly modeled. However, we treat μ_g as the rate at which the mutant would have grown *if* it had remained in the exponential phase. Importantly, we assume that the abundance of a mutant with fitness μ_g increases over time according to the Malthusian growth model:

$$N_{g,i}(t) = N_{g,i}(t_0) \cdot 2^{\mu_g t}. \tag{4.3}$$

Here, $N_{g,i}(t)$ denotes the number of mutant cells that carry a transposon insertion at position i in gene g at time t. We use the notation μ_g rather than $\mu_{g,i}$ to emphasize that each gene deletion mutant can be characterized by a single growth rate. Mutant abundance after a growth period t is inferred from the number of reads that map to the location $y_{g,i}$. The partial distortion caused by the sampling and PCR amplification steps before sequencing yields an additional noise term ϵ in the relationship between mutant abundance and read count:

$$y_{g,i} = N_{g,i} \pm \epsilon = N_{g,i}(t_0) \cdot 2^{\mu_g t} \pm \epsilon. \tag{4.4}$$

Since read count values are limited to non-negative integers, we expect the distribution of $y_{g,i}$ of a gene g to be Poisson-like. The Maximum Likelihood Estimator (MLE) of a Poisson distribution is equal to the sample mean, giving:

$$\hat{y}_g = \frac{1}{n_g} \sum_{i \in g} y_{g,i}. \tag{4.5}$$

Where n_g is the total number of transposons that are mapped to gene g and $\hat{y}_g$ is the MLE estimator of the read counts y_g that would be obtained for the gene deletion mutant in the absence of noise. We emphasize that $\hat{y}_g$ will only be the MLE estimator of y_g if it is ensured that all $y_{g,i}$ in equation 4.5 are different realisations of sampling the same mutant with deleted gene g. Using this estimator for y_g, we can rewrite equation 4.4 to get the following expression for μ_g:

$$\mu_g = \frac{\log_2(\hat{y}_g)}{t}. \tag{4.6}$$

Here, we have omitted the term $\log_2(N_{g,i}(t_0))$ that would appear in equation 4.6 based on the assumption that the likelihood of multiple transposons inserting into the same genomic location is low, such that $N_{g,i}(t_0) = 1$ and $\log_2(N_{g,i}(t_0)) = 0$.

To be able to compare the fitness values obtained from different experimental repetitions, possibly with different growth times, the growth rates must be scaled to that of the non-mutated ancestral strain. As other studies have shown that the majority of gene deletions have only a minor effect on fitness, we take the median of the distribution of growth rates to represent the fitness of the ancestral strain:

$$w_g = \frac{\mu_g}{\mu_{ref}} = \frac{\log_2(\hat{y}_g)}{\mu_{ref}} \text{ ,with } \mu_{ref} = \text{median}[\log_2(\hat{y}_g)], \tag{4.7}$$

with w_g the scaled fitness value of a mutant with gene g deleted. The additional benefit of the expression shown in equation 4.7 is that it does not depend on time. Thus, our fitness estimate depends only on the average read count per insertion site.

Notice that for some genes the fitness value can be negative. These genes contain an excess of zero read count sites, resulting in an average read count value $\hat{y}_g$ smaller than one. Because many insertion mutants are lost during the sampling steps of SATAY, a negative fitness value is not necessarily caused by a higher death rate of the gene deletion mutant. Instead, it should be interpreted to mean that the low effective growth rate of the mutant is so low that it is frequently lost from the population after the sampling steps. As such, the abundance of genes that are ascribed a negative fitness value will depend on the size of the population bottleneck and the difference in growth rate between the fastest and slowest growing mutants.

Variance of the fitness estimates

The identification of fitness differences that are statistically significant requires a method to estimate the uncertainty of the found fitness values. In the previous section, we proposed that the distribution of read counts obtained from insertions in the same gene ($y_{g,i}$) is Poisson-like. However, other studies have shown that the variance in read counts from independent replicates are generally overdispersed compared to the Poisson distribution (Anders & Huber, 2010; Limdi & Baym, 2023). This overdispersion is caused by biological noise that is superimposed the sequencing noise. Thus, if insertions mapping to different sites within the same gene represent different biological replicates, we would expect to observe this overdispersion in our read counts. We tested for overdispersion by plotting the average read count per insertion site against the variance for all genes (Figure 4.4c). This mean-variance relationship revealed that our data indeed displayed overdispersion compared to a Poisson model. To account for this overdispersion, we decided to use a Negative Binomial to model count noise, which is frequently used in the analysis of RNA-seq data (Love et al., 2014; M. D. Robinson et al., 2010; M. D. Robinson & Smyth, 2007a, 2007b; Zhang et al., 2013). We used the following parameterization of the Negative Binomial which relates its mean and variance through the overdispersion parameter α:

$$\begin{cases} E(Y_g) = \mu_g. \\ Var(Y_g) = \sigma_g^2 = \mu_g + \alpha\mu_g^2. \end{cases} \tag{4.8}$$

In this equation, α regulates the degree of overdispersion relative to the Poisson model: as α shrinks to 0, our model becomes equivalent to a Poisson. Hence, the term $\alpha\mu_g^2$ can conveniently be interpreted as the variance due to biological noise that is added to the sampling noise.

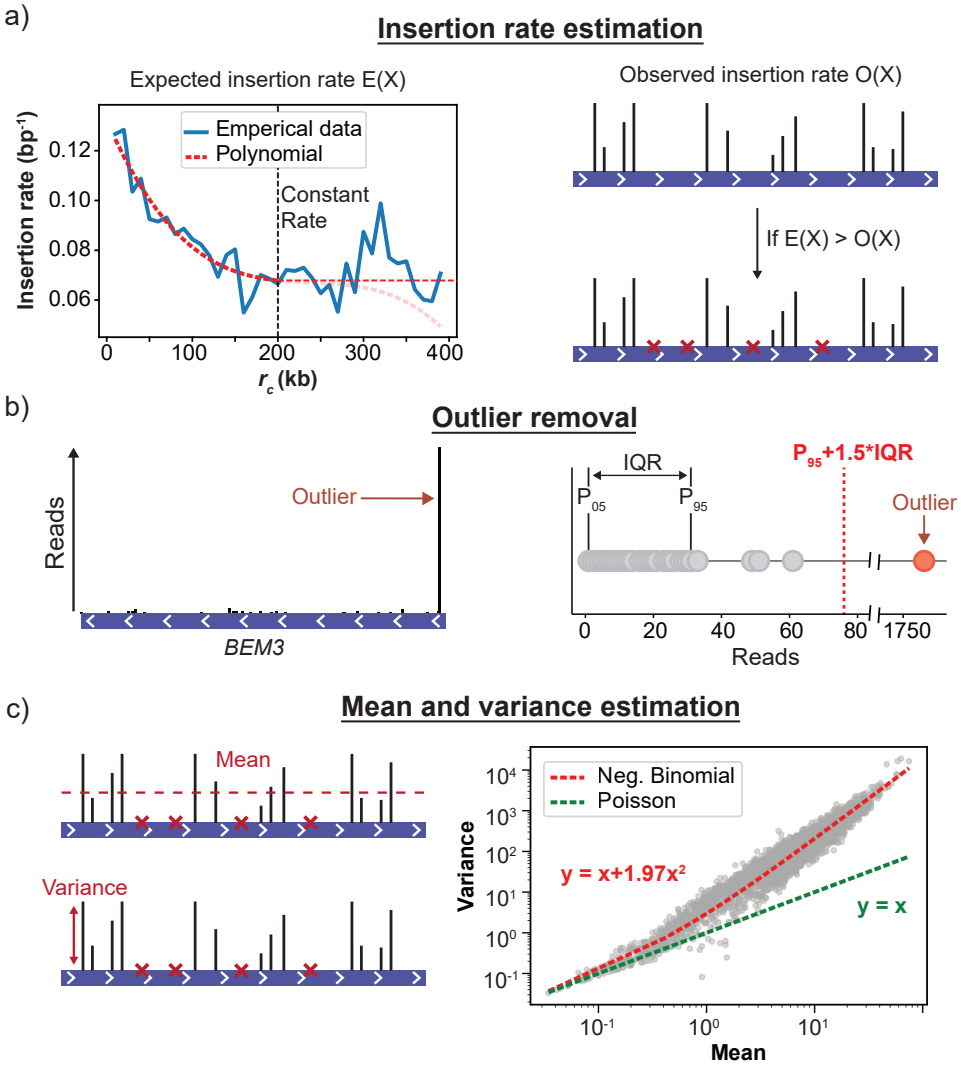


Figure 4.4. The steps of the procedure to calculate fitness from transposon insertion data. (**a**) First, the expected number of insertion events is calculated for each gene, using the global insertion profile to correct for the centromere bias. The expected number of insertions is then compared to the observed number of insertions to determine the number of sites that have an insertion but produce no reads. These sites are included as sites that have zero read counts. (**b**) After adding the zero read count sites, outliers are removed using the range between the 5th and 95th percentiles of the data. (**c**) Finally, the mean and variance of the read counts at different insertion sites are used to determine the average and uncertainty, respectively, of the fitness of a gene deletion mutant. To provide a robust robust estimate for the variance, information is shared between genes by fitting the global mean-variance relationship with an overdispersed Poisson model. The resulting fit allows us to determine the variance based on the mean read count of a gene based on the assumption that this mean-variance relationship is a property of the dataset.

To estimate α we assumed that (1) it can be estimated independently from the distribution mean and (2) that all genes in a single dataset share the same value for α. Our assumption that α and μ can be determined independently is justified by the fact that our fitness estimates are based on a single library and all read counts are therefore conditional on the same sequencing depth. Under these conditions, the MLE estimator for μ will always be the sample mean, regardless of the value of α. We use the following regression equation to estimate α using Ordinary Least Squares (OLS) regression of the empirical mean-variance relationship:

$$\frac{(y_{g,i} - \hat{y}_g)^2 - \hat{y}_g}{\hat{y}_g} = \alpha \hat{y}_g. \tag{4.9}$$

Where the value of the estimator $\hat{y}_g$ is found using equation 4.5. An example of the resulting fit for the mean-variance relationship obtained with this regression model is shown in Figure 4.4c. To account for the possibility that, for biological reasons, some genes have a higher variance than what is found when using the trended dispersion fit, we take the maximum of each genes individual sample variance and the variance obtained from equation 4.8 to prevent underestimation of the true variance:

$$\mathrm{Var}(Y_g) = \max\left(\sigma^2_{g,trend}, \sigma^2_{g,sample}\right). \tag{4.10}$$

4.2.3. Reproducibility of the fitness estimates

Fitness values are reproducible across replicate SATAY experiments

An important requirement for the applicability of SATAY for global fitness maps of gene disruption mutants is that the fitness estimates are reproducible between replicates of the same genetic background. To test this robustness, we created 11 replicate SATAY datasets using a wild type strain of *Saccharomyces cerevisiae* from the W303 background. All 11 datasets are derived from the same wild-type transformed with a plasmid carrying the machinery to induce transposon mutagenesis (Michel et al., 2017; Michel et al., 2019), but were subsequently split at different steps of the experimental procedure (Figure 4.6). Specifically, from the resulting transformation plate four colonies were picked (indicated as replicates B1-B4 in Figure 4.5) and used to generate four independent mutant libraries. For replicates B1 and B2, we sequenced the mutant library obtained after library expansion respectively six times (replicates B1_T1-T6) and three times (replicates B2_T1-T3) (Figure 4.5). We refer to datasets obtained from individual colonies on the transformation plate as biological replicates, as the difference between these libraries are (although not exclusively) caused by differences in population composition and random birth/death events during library expansion. Alternatively, datasets obtained by re-sequencing the same library several times are referred to as technical replicates, as these exclusively contain the noise caused by random sampling during the sequencing process. Thus, this should allow us to distinguish the relative contribution of biological noise from the technical noise that results from limited sequencing of the samples.

To determine the reproducibility of our fitness estimates obtained from SATAY datasets using the procedure described in the previous sections, we compared the fitness values for each non-essential gene obtained from different technical (B1_T1 vs. B1_T2) and biological replicates (B1 vs. B2). We excluded genes for which the expected number of transposon insertions was less than five ($E(X) < 5$) on the basis that we have too little information from these genes to reliably estimate fitness. The results in Figure 4.6 show that fitness values are well correlated between

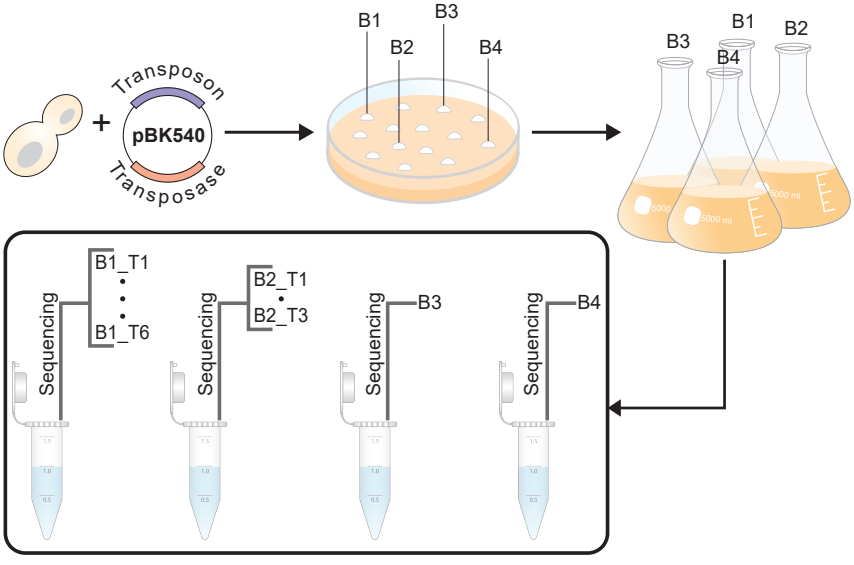


Figure 4.5. Overview of the experimental procedure used to determine the reproducibility of the fitness values across different replicate experiments. Biologial replicates B1-B4 are different clones of a single wild-type strain transformed with plasmid pBK549. For two of the biological replicates (B1 and B2), the extracted genomic DNA was sampled and sequenced multiple times, yielding the technical replicates B1_T1-6 and B2_T1-3.

technical replicates ($\rho = 0.97$, Figure 4.6a) and biological replicates ($\rho = 0.91$, Figure 4.6c), although the variance appears slightly higher between biological replicates.

In addition to a strong correlation of the fitness values between replicates, we furthermore find that for all samples the Distribution of Fitness Effects (DFE) contains a single peak with a skewed tail with a skew towards lower fitness values (4.6b and e). This is in agreement with the expected shape of the DFE when advantageous mutations are rare and most mutations have neutral or deleterious effects. We therefore conclude that fitness estimates based on read counts from SATAY datasets are reproducible across replicates of the same genetic background.

Fitness estimates correlate poorly with estimates based on the Yeast Gene Deletion collection

Although we find that the fitness estimates strongly correlate between replicate experiments (Figure 4.6), we wanted to investigate whether we would be able to reproduce the fitness values and epistatic interactions reported by other studies. Of particular interest for this purpose is the study by Qian et al., 2012, as they used Bar-seq to determine the fitness effects of gene disruptions. While fitness estimates are determined from read counts in both Bar-seq and SATAY, we need to account for the slight difference in the definition of fitness to compare our results to the results of Qian et al., 2012. Specifically, Qian et al., 2012 use the following definition:

$$w_g = \left(\frac{P' P_{wt}}{P P'_{wt}} \right)^{1/t}. \tag{4.11}$$

Where P' and P'_{wt} are the mutant and wild-type frequencies, respectively, at the start of the competition experiment, P and P_{wt} are their frequencies at the end and t is the number of gen-

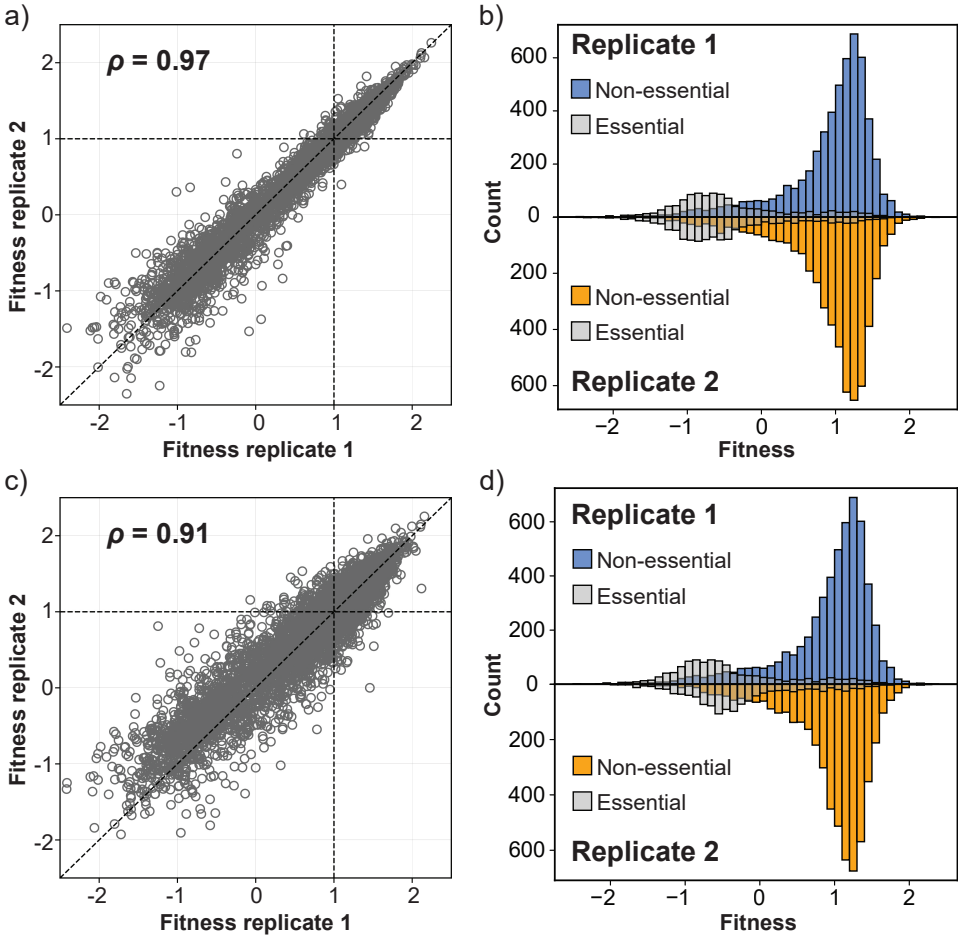


Figure 4.6. Fitness estimates from SATAY are reproducible across replicate experiments. (**a**) The estimated fitness effect of gene disruptions for all genes of technical replicate 1 plotted against its estimated value in technical replicate 2. The identity lined (red dashed line) is shown as a reference for perfect correlation between the two replicates. (**b**) The fitness distributions of technical replicate 1 (top) and technical replicate 2 (bottom). Essential genes are represented by gray transparent bars. (**c**) Same plot as in (a), but for two biological replicates. (**d**) The fitness distributions of biological replicate 1 (top) and biological replicate 2 (bottom). Annotations are the same as in (b).

erations over which the competition is evaluated. We therefore converted our fitness estimates to the same scale by setting P equal to $\hat{y}_g$ (equation 4.2), P_{wt} equal to $\text{median}[\hat{y}_g]$ and both primed variables equal to 1. The number of generations spent in library expansion is estimated to be approximately 10. Our adjusted fitness equation therefore becomes:

$$w_g = \left(\frac{\hat{y}_g}{\text{median}[\hat{y}_g]} \right)^{1/10}. \tag{4.12}$$

Figure 4.7a shows that there is a very poor correlation between the fitness values obtained by the two methods. In particular, the DFE obtained by Qian et al., 2012 is much more centered around neutral fitness effects (that is, a fitness value of 1) and contains less spread compared to the DFE obtained from SATAY. Thus, many genes that would be considered to have a near-neutral fitness cost when deleted based on the data by Qian et al., 2012 are classified as deleterious in our fitness estimates from SATAY. These differences indicate that the fitness values obtained from SATAY datasets may be condition specific and cannot be generalized without further considerations of the experimental design.

Despite the low correlation with the data from Qian et al., 2012, we wanted to examine if the fitness values obtained from SATAY could nevertheless be used to identify known epistatic interactions between genes. To this end, we compared the fitness values obtained from a wild-type strain with those from a *bem3*Δ mutant (Figure 4.7b). Interestingly, we do observe more variation between these two samples than we see for biological replicates of the same genetic backgrounds (Figure 4.6). In theory, genes that interact with *BEM3* will lie further away from the identity line than those that do not, since their fitness effect will either be suppressed (positive epistasis) or aggravated (negative epistasis) in the *bem3*Δ genetic background relative to their effect in the wild-type background. We annotated the positive and negative genetic interactors of *BEM3* that have been identified by an SGA screen (Costanzo et al., 2016) in Figure 4.7b to see if we could identify any correlation between the distance of the datapoint from the identity line and its annotation as a genetic interactor. Note that although a different metric is used to calculate fitness (colony size), this SGA screen makes use of the same collection of gene deletion mutants (Giaever & Nislow, 2014) as Qian et al., 2012 to screen for genetic interactions. We find that most annotated genetic interactors do not deviate more from the identity line than other, non-interacting, genes and are therefore not clearly identifiable. Hence, we would not be able to recover all of the genetic interactions of *BEM3* reported by Costanzo et al., 2016 if we would base our analysis solely on the fitness values we obtain through our analysis of SATAY datasets. Therefore, similar to what we found for the fitness values obtained by different methods, the ability to predict epistatic interactions using the method for fitness estimation from SATAY data used here is either poor or inconsistent with other datasets.

4.2.4. Fitness resolution does not improve with increasing sequencing depth

While our fitness estimates are substantially different from those presented by Qian et al., 2012 and Costanzo et al., 2016, our results in Figure 4.6 demonstrate that they exhibit minimal variation across replicate SATAY datasets. However, some applications may have specific requirements for the level of uncertainty in the fitness estimates. For example, if one wishes to resolve subtle differences in fitness it is desirable to design the experiment such that the highest possible accuracy is obtained. In these cases, it is useful to know how the experimental design can be modified to improve the accuracy of the fitness estimates. Intuitively, a simple way to improve

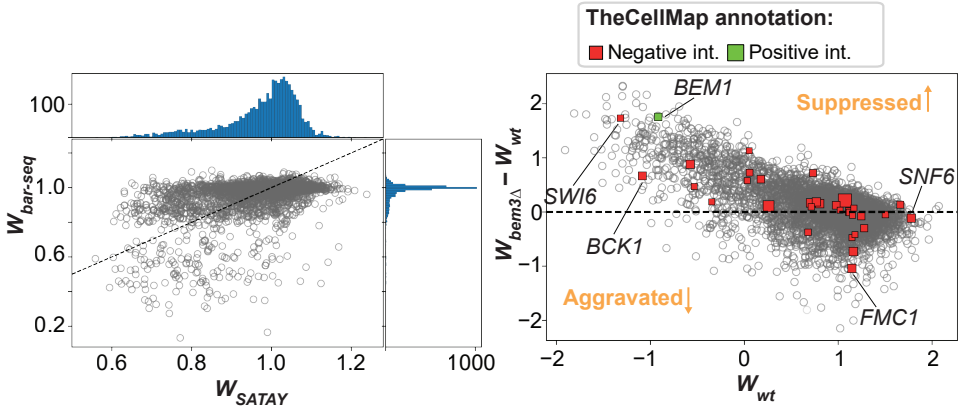


Figure 4.7. The fitness values obtained from SATAY datasets only weakly correlate with the values reported by other studies. (**a**) Plot of the fitness values of gene deletion mutants reported by Qian et al., 2012 against the fitness values obtained from a SATAY dataset in this study. The top and side panels show the DFE densities. The dashed line represents the identity line. (**b**) Plot of the fitness values of gene deletion mutants generated using SATAY in a *bem3*Δ genetic background against the fitness values of the same gene deletion mutants generated in a wild-type (WT) genetic background (both from this study). Positive and negative genetic interactors of *BEM3*, as annotated by Costanzo et al., 2016, are shown as green and red datapoints, respectively.

the accuracy of the fitness estimates would be by increasing the sequencing depth. Specifically, deeper sequencing should allow better resolution of deleterious fitness effects (Limdi & Baym, 2023), as low-abundance mutants are often lost when sampling the population. As sequencing depth increases, an important point is where all insertions are represented by at least one read. We refer to this as the saturation point of sequencing. When sequencing saturation is reached, the information on mutant fitness is completely contained in the read counts and should no longer depend on the insertion density, as all unique insertion sites have been identified.

We attempted to determine the point of sequencing saturation for SATAY libraries by merging the datasets of technical replicates B1_T1-6 (Figure 4.8a). Because these technical replicates are resamplings of the same DNA sample, their merging results in a dataset that is equivalent to a deeper sequencing of the sample in a single sequencing run. To estimate which portion of this merged dataset needs to be sampled in order to obtain at least one read count for each unique transposon insertion event, we computationally sampled from this merged dataset using a hypergeometric model. Figure 4.8b shows how the number of mapped transposon insertions changes as the number of sampled reads increases. The rate at which new transposon insertions are found slightly decreases when sampling more reads. However, even with the exceptionally high sequencing depth of the entire dataset (approximately 100 million reads), we do not observe evident signs of saturation. This is also reflected by the fact that the median read count per transposon barely increases as the sequencing depth increases (4.8b,c), which indicates that the majority of insertion sites are represented by a relatively small number of reads. The inability to reach the saturation point limits the benefit of acquiring additional reads for the accuracy of fitness estimation. As is shown in Figure 4.8d, there is no clear improvement in the relative squared error of the mean of our fitness estimates when using datasets with a higher total read count. Although saturation may be reached for even further increased sequencing depth, we

consider going beyond 100 million reads impractical for any general application. We therefore conclude that deeper sequencing of samples is an ineffective strategy to improve the resolution of the fitness estimates.

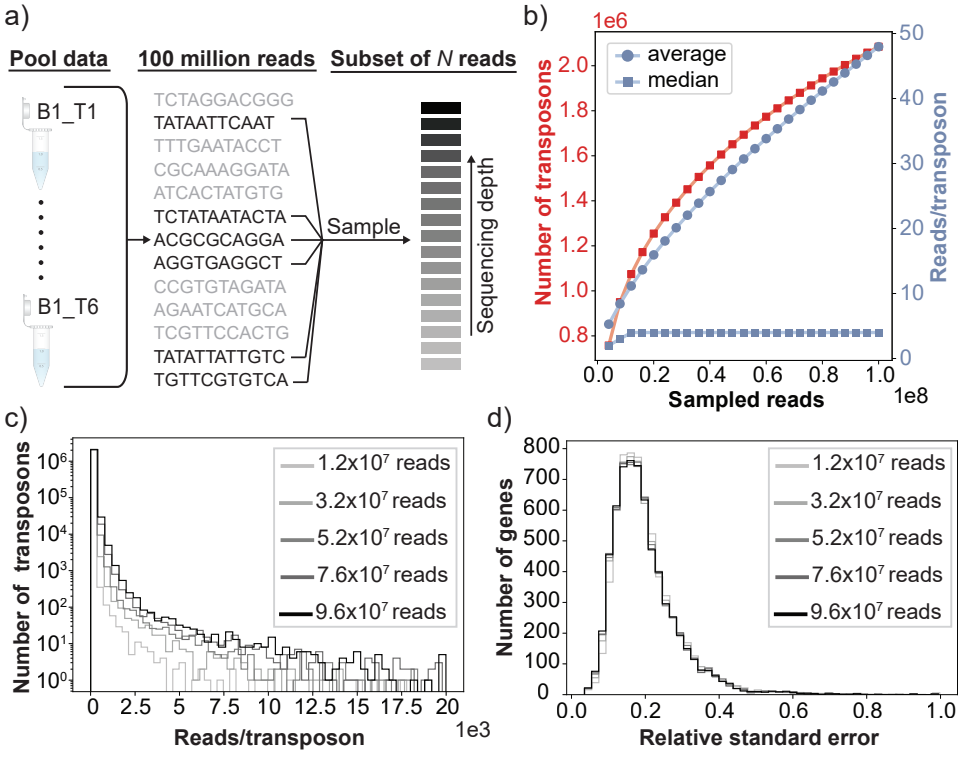


Figure 4.8. Increasing sequencing depth does not improve the accuracy of fitness estimate. (**a**) The effect of sequencing depth on the accuracy of the fitness estimates was determined by pooling the transposon insertion data of the six technical replicates B1_T1-B1_T6 and randomly sampling a subset of n reads without replacement from this pooled dataset. (**b**) The number of observed independent transposon insertions (red) and the average and median read count per insertion site (blue) as a function of the number of sampled reads. (**c**) The read count distribution for varying levels of the sequencing depth. The distribution has been cut off at a maximum value of $20{\times}10^3$ reads/transposon. (**d**) The distribution of the relative standard error of the fitness estimates across different genes for different levels of sequencing depth.

4.3. Discussion

We developed a method to quantify the fitness of gene deletion mutants based on data from the transposon mutagenesis screen SATAY. Our approach differs from other methods that determine fitness from pooled assays in that our estimates are based on measurements of a single timepoint rather than the log-frequency change between two successive timepoints. Estimating fitness from a single timepoint required the assumption that all mutants are equally abundant at the start of library expansion (Equation 4.6). It is true that this assumption neglects the effect of altered growth conditions as the population is transferred from one media type to another during the SATAY protocol (Michel & Kornmann, 2022). However, we anticipate that the impact of this media transfer on the growth trajectories is limited, as most transposition events occur dur-

ing the saturated phase in induction media (Michel et al., 2019). As a result, each mutant should only pass though a limited number of cell cycles before the transfer to the library expansion media. Moreover, it is important to recognize that sampling the population to measure mutant frequencies inevitably influences their growth trajectories. The distortion in growth caused by sampling can be particularly significant during early timepoints when mutant frequencies tend to be low.

To compare the fitness values obtained from different datasets, we rely on normalization to the median fitness of the population. This approach implicitly assumes that the majority of gene disruptions have a neutral effect on fitness. While we find that normalization to the median works well when the fitness distribution is unimodal, it is unlikely to produce reliable results for multimodal distributions. Conceptually, multimodal fitness distributions may arise when there are a small number of mutants that have a significantly higher fitness than the starting strain. Thus, the ability to compare fitness values from different datasets will depend on the fitness of the different genotypes present in the mutant library. This limitation has been addressed by Li and Zhang (2018), who showed that while providing accurate fitness ranks, fold-enrichment measures can yield biased proxies for fitness. Although computationally more intensive, approaches that explicitly model the growth trajectories of mutants during library expansion may be a robust alternative for normalization of multimodal distributions (Schlecht et al., 2017).

Gene deletion mutants from pre-constructed gene deletion libraries have been found to contain secondary mutations that compensate for the fitness defects caused by deleting a gene (Giaever & Nislow, 2014; Teng et al., 2013; van Leeuwen et al., 2016). In SATAY, mutant libraries can be easily generated *de novo* for each experiment, making the fitness estimates less susceptible to the accumulation of secondary mutations. However, the complex relation between read count and fitness does increase the risk of introducing bias in the fitness values for the following reasons.

First, gene disruptions resulting from transposition may not always be equivalent to a full gene deletion, which can lead to underestimation of the fitness effect of a gene deletion. We indeed observed that insertions close to the 3' and 5' ends of a gene are often associated with a higher read count (Figure 4.2a,b). Excluding insertion events that map close to the gene edges is a relatively simple method to correct for this bias. However, it has been shown that the fitness effect of a transposon insertion also depends on the specific protein domain that it affects (Michel et al., 2017). The presence of multiple domains in an open reading frame therefore causes a more complex relation between read count and the position of an insertion within a gene. Correcting for such domain effects requires the grouping of insertions according to which domain they affect. While this would allow a more refined mapping of fitness based on domain-wise rather than a gene-wise fitness values, the spatial clustering of insertions based on their read count value is a complex problem that is beyond the scope of this study. Hence, the method presented here does not account for these domain effects and the reported fitness values should be interpreted as an average of the domain-wise fitness.

Second, using the average read count over all insertion sites within a gene as a metric to determine fitness makes the fitness estimate dependent on the mapping accuracy of the reads. Misalignment of reads derived from the same transposon insertion event causes them to become spread out over a larger genomic region. This leads to lower fitness values, as the reads mapping to a gene will now be divided by a larger number of transposons to obtain the average. Because the miniDS transposon used in the SATAY screen can insert at any basepair, it has a rel-

atively higher susceptibility to this artefact compared to other TIS methods that use transposons that can only insert at specific nucleotide sequences. A possible sign that mapping accuracy plays a role in our datasets is the fact that we are unable reach the point of sequencing saturation of the library, where each unique transposon insertion is represented by at least one read (4.8b,c). In addition, the number of identified unique transposon insertions is for a given total read count is systematically higher than what is reported by other studies using SATAY (Michel et al., 2017; Michel et al., 2019). One way to address the misalignment of reads is to discard reads with an alignment score below a certain threshold. However, this may disproportionately affect genes with a coding sequence that does not allow accurate alignment (such as those with repeated regions) which can artificially lower their read count.

We have shown that the method to estimate fitness presented in this chapter generates reproducible results across replicate SATAY experiments performed in the same genetic background. However, there was only a weak correlation with the fitness values reported by other studies and we were unable to retrieve epistatic interactions between genes that have been annotated based on SGA screens. We provide two explanations for this lack of consistency. First, we compared our results to the fitness values that were obtained from a pre-constructed gene deletion collection (Costanzo et al., 2016; Giaever et al., 2002; Qian et al., 2012). In pre-constructed gene deletion collections, mutants usually pass through several rounds of replication after their construction before they are subjected to a fitness assay. This gives the opportunity for secondary mutations that can potentially mask the defects caused by the primary gene deletion to set in the population. It has indeed been shown that several mutants in the yeast gene collection harbour such compensatory mutations (Teng et al., 2013; van Leeuwen et al., 2016). Libraries created with SATAY are less likely to suffer from secondary mutations because the fitness assay usually follows directly after library creation. As a consequence, fitness values based on measurements of the yeast gene deletion collection may overestimate the number of neutral genes compared to results obtained with SATAY, which is in agreement with our observation that the DFE obtained by Qian et al., 2012 is much narrower than ours (Figure 4.7a). The second explanation is that the dissimilarities are caused by the different genetic background used by other studies (S288C (Costanzo et al., 2016; Giaever & Nislow, 2014; Qian et al., 2012), whereas we have used W303). Because the fitness effect of a gene deletion can depend on the allelic status of genes elsewhere in the genome, the observed inconsistencies may simply reflect a fundamental difference of the fitness landscape of the two genetic backgrounds. It is difficult to determine the extent to which the fitness landscape can be expected to diverge between genetic backgrounds, as relatively few studies have assessed fitness on the genome wide scale and most of these are based on the same collection of gene knock-outs. However, comparative studies of the genetic interaction maps of different species have shown that their structure can strongly vary (Tischler et al., 2008), although some level of conservation does appear to exist (Deshpande et al., 2013; Dixon et al., 2008; Roguev et al., 2008).

In conclusion, the method we present here allows reproducible quantification of fitness from the read count data obtained from SATAY experiments. However, we hope the considerable differences between our results and fitness and genetic interaction maps produced with other methods will encourage initiatives to determine the amount of variation in these maps across genetic backgrounds and environments.

4.4. Methods

4.4.1. Strains

All strains used in this study are of the W303 genetic background and are derived from a single parental strain (see table 4.1). The parental strain was made heterotrophic for adenine (*ade*⁻) by replacing the *ADE2* gene with the *URA3* marker using homologous recombination, followed by counterselection against the *URA3* marker with 5-fluoro-orotic acid to obtain a clean gene deletion. *bem3*Δ and *nrp1*Δ strains were created using homologous recombination to replace the endogenous genes with the natMX4 and hphMX4 cassettes, respectively. Strains were stored at -80°C as frozen stocks in 40% (v/v) glycerol.

4.4.2. Media

Standard culturing and growth assays were performed in YPD (10g/L Yeast extract, 20 g/L Peptone, 20 g/L dextrose), SC (6.9 g/L Yeast nitrogen base, 0.75 g/L Complete supplement mixture, 20 g/L dextrose). For *ade*⁻ strains, standard growth media was supplemented with 20 mg/L adenine just before incubation. Liquid media for the preculture and induction steps of SATAY were prepared according to the recipe in table 4.2. After preparation, the media was filter sterilized using Rapid-Flow Sterile Disposable Filter Units (Nalgene) and stored at 4°C until use. Liquid media for the reseed step of SATAY was prepared by autoclaving 2.6 L of MiliQ water in a 5 L flask. 400 ml of a 7.5X concentrated solution of the nutrients was prepared separately and filter sterilized. To prevent the degradation of media components, this concentrate was stored in the dark at 4°C until used. On the day of reseed, the concentrate was aseptically added to the 5 L flask containing 2.6 L of MiliQ water and mixed. Solid media was prepared by adding 20 g/L agar and 30 mM Tris-HCl (pH 7.0) to the liquid media recipe and autoclaving the mixture for 20 minutes at 121°C. 20 mg/L adenine was aseptically added after autoclaving, unless plates were intended to be selective for adenine auxotrophy.

4.4.3. SAturated Transposon Analysis in Yeast (SATAY)

Library generation

SATAY libraries were generated based on the procedure described by Michel et al., 2019, which is a modification of the original protocol (Michel et al., 2017) to allow transposition to occur in liquid media. *ade*⁻ cells were transformed with plasmid pBK549 (Michel et al., 2019), which was a kind gift from Benoît Kornmann, according to a lithium acetate transformation protocol (Gietz & Schiestl, 2007). To screen for clones transformed with the intact version of plasmid pBK549 (see Michel and Kornmann, 2022 for details on the different species of pBK549), 12-24 colonies were picked from the transformation plate, re-streaked on fresh SD-ADE and SD-URA plates and incubated for 3 days at 30°C. For clones that showed full growth on SD-URA plates while producing a small number of colonies on SD-ADE plates, cells were scraped from the SD-URA plate and used to inoculate 25 ml of preculture media (table 4.2) at an OD600 of 0.20-0.28. Precultures were grown on an orbital platform shaker at 160 rpm, 30 °C until the OD600 was between 5-7 (~20h). The saturated precultures were used to inoculate 200 ml of induction media at an OD600 of 0.10-0.27 and grown for 52 hours to allow transposition to occur. The efficiency of transposition was monitored by plating samples of the liquid induction cultures on SD-ADE at T=0 and T=52 hours and scoring the number of colonies on these plates after 3 days of incubation at 30 °C. After 52 hours of induction, the resulting transposon mutagenesis libraries were reseeded in 3 liters of reseed media at an OD600 of 0.21-0.26. Typically, this meant that around 7 million transposon

mutants were reseeded per library. Reseeded libraries were grown for 92 hours at 140 rpm, 30 °C. At the end of reseed, cells were harvested by centrifugation of the reseed cultures at 5000 xg for 30 minutes. Cell pellets were stored at -20 °C.

Genomic DNA extraction

A 500 mg frozen pellet was resuspended in 500 µl cell breaking buffer[1] and distributed into 280 µl aliquots. 300 µl of 0.4-0.6 mm glass beads (Sigma-Aldrich, G8772) and 200 µl of Phenol:Chloroform:isoamyl alcohol 25:24:1 (Sigma-Aldrich, P2069) were added to each aliquot and cells were lysed by vortexing the samples with a Vortex Genie 2 at maximum speed at 4°C for 10 minutes. 200 µl of TE buffer was added to each lysate, after which the samples were centrifuged at 16100x g, 4°C for 5 minutes. After centrifugation, the upper layer (~400 µl) was transferred to a clean eppendorf tube. 2.5 volumes of 100% absolute ethanol was added to each sample and mixed by inversion to precipitate the genomic DNA. After precipitation, the DNA was pelleted by centrifugation at 16100x g, 20°C for 5 minutes. The supernatant was removed and the DNA pellet was resuspended in 200 µl of 250 µg/ml RNAse A solution (qiagen, Cat. No. 19101). the resuspended DNA pellets were incubated at 55°C for 15 minutes to allow digestion of the RNA. After digestion, 20 µl 3M, pH 5.2 sodium acetate (Merck) and 550 µl 100% absolute ethanol was added to each sample and mix by inversion. DNA was pelleted by centrifugation at 16100x g, 20°C for 5 minutes. Pellets were washed with 70% absolute ethanol and dried at 37°C for 10 minutes or until all ethanol had evaporated. The dried pellets were resuspended in a total volume of 100 µl MiliQ water and the concentration of the genomic DNA samples was quantified on a 0.6% agarose gel using the Eurogentec Smartladder 200bp-10kb as a reference. Prepared DNA samples were stored at -20°C or 4°C until used.

Library sequencing

To prepare genomic DNA samples for sequencing, 2x2 µg of DNA from each sample were transferred to non-stick microcentrifuge tubes and digested with 50 units of DpnII and NlaIII in a total volume of 50 µl for 17 hours at 37°C. After digestion, the restriction enzymes were heat-inactivated by incubating the samples at 65°C for 20 minutes. Digestion results were qualitatively assessed by visualization on a 1% agarose gel stained with Sybr-Safe. Successfully digested DNA samples were circularized in the same tube using 25 Weiss units of T4 DNA ligase (Thermo Scientific, Catalog #EL0011) at 22°C for 6 hours in a total volume of 400 µl. After ligation, the circularized DNA was precipitated using 1ml 100% absolute ethanol, 20 µl 3M, pH 5.2 sodium acetate (Merck) and 5 µg linear acrylamide (invitrogen, AM9520) as a carrier. DNA was precipitated for at least 2 days at -20°C. Precipitated DNA was pelleted by centrifugation for 20 minutes at 16100x g at 4°C and washed with 1 ml of 70% ethanol. After washing, the DNA was re-pelleted by centrifugation for 20 minutes at 16100x g at 20°C, the supernatant was removed and pellets were dried for 10 minutes a 37°C. Each dried pellet was resuspended in water and used as a template for 20 PCR reactions of 50 µl.

For samples sequenced on the Illumina HiSeq platform (yLIC136), PCR amplification of the transposon genome-junctions, sequencing and sequence alignment were performed by Agnès Michel and Benoît Kornmann (Oxford). For samples sequenced on the Illumina NovaSeq 6000 platform (yWT01a, yLIC137), the transposon-genome junctions were amplified using the barcoded primers 1 and 2 (table 4.3) for DpnII digested DNA or primers 3 and 4 (table 4.3) for

[1] 2% Triton X-100, 1% SDS, 100 mM NaCl, 100 mM Tris-HCl pH8.0, 1 mM EDTA

NlaIII digested DNA on a thermal cycler (Bio-Rad C1000 Touch) with the block settings shown in table 4.4. PCR amplified samples were purified using the NucleoSpin Gel and PCR cleanup kit (Macherey-Nagel) and quantified on the NanoDrop 2000 spectrophotometer (Thermo Scientific). For each sample, equal ratios (w/w) of DpnII and NlaIII digested DNA were pooled. Library preparation and sample sequencing were performed by Novogene (UK) Company Limited. Sequencing libraries were prepared with the NEBNext Ultra II DNA Library Prep Kit, omitting the size selection and PCR enrichment steps. Libraries were sequenced using Paired-End (PE) sequencing with a read length of 150 bp.

Sequence alignment

FASTQ files obtained from the HiSeq platform were analyzed by Agnès Michel and Benoît Kornmann (Oxford) using their in-house pipeline for sequence processing and alignment (Michel & Kornmann, 2022). FASTQ files obtained from the NovaSeq 6000 platform were demultiplexed into DpnII and NlaIII digested DNA samples based on the barcodes introduced during PCR amplification. Read pairs with non-matching barcodes were discarded. After demultiplexing, the forward read of each read pair was selected and the sequences upstream of primer 688_minidsSEQ1210 (Michel et al., 2017) and downstream of the DpnII (GATC) or NlaIII (CATG) restriction site were trimmed. All demultiplexing and trimming steps were executed with BBduk integrated into a home-written pipeline written in Bash. After trimming, the forward reads were aligned to the S288C reference genome (version R64-2-1_20150113) with the Transposonmapper pipeline (Iñigo de la Cruz et al., 2022 version v1.1.4) using the following settings:

- Data type: 'Single-end'
- Trimming software: 'donottrim'
- Alignment settings: '-t 1 -v 2'

4.4.4. Analysis of essential genes

Genes were marked as essential based on their annotation on the Saccharomyces Genome Database (Cherry et al., 2012, accessed on 03/17/2006)

4.4.5. Centromere bias correction

Centromere bias was estimated based on the global transposon insertion density profile (see Figure 4.3). The insertion density profile was fitted with a third-degree polynomial using the polyfit function from the Numpy package (version 1.21.5). The derivative of the fitted polynomial was determined with the polyder function from the Numpy package (version 1.21.5). To calculate the distance (in bp) of a gene to the corresponding chromosome centromere (r_{c-g} in equation 4.1), the coordinates of each centromere were obtained from the Saccharomyces Genome Database (accessed on 02/06/2022). Either the start or stop position of the centromere was used to calculate distance, depending on which was closest to the gene of interest. Because the transposon insertion count can only take on discrete values, the value of the expected number of transposon insertions in a gene ($E(X_g)$ in equation 4.2) was rounded down to the nearest integer.

4.4.6. Fitness and variance calculation

Fitness values were based on the mean read count of transposon insertions that mapped to the central 80% of the coding region of a gene. Reads that exceeded 1.5 times the 5-95 percentile

range of the read distribution of a gene were classified as outliers and removed. The 5-95 percentile range was determined using the stats module available from the Scipy package (Virtanen et al., 2020, version 1.7.3) If the number of insertion sites used to calculate fitness was less than 5 after outlier removal or if all remaining insertion sites had a read count of 0, the fitness value was set to undetermined and not used for comparisons with other datasets.

The variance of the fitness estimates was determined from the observed mean-variance relationship of all genes (Figure 4.4c). The mean-variance relationship was fitted with equation 4.9 using the OLS function available from the Statsmodels module. The obtained variance estimates were used to calculate the standard error of the fitness values with the following equation:

$$SE_g = \frac{\sqrt{V_g}}{\sqrt{n}} \tag{4.13}$$

Where SE_g is the standard error of gene g, V_g is the variance of gene g estimated according to equation 4.10 and n is the number of transposon insertions that have been mapped to the coding sequence of gene g.

4.4.7. Genetic interactions of *BEM3*

Genetic interactions of *BEM3* were downloaded from TheCellMap.org ((Usaj et al., 2017), accessed on 13/07/2022). A stringent cut-off was used for the significance level of negative (GI-score< -0.12, p-value< 0.05) and positive (GI-score> 0.16, p-value< 0.05) genetic interactions. Interactions annotated as dubious or those that were reported to have been affected by a suppressor mutation were excluded. In addition, only those interactions that were derived from fitness measurements of gene deletion strains (and not conditional knockouts) were used.

4.5. Contributions & Acknowledgements

The strains used in this chapter were constructed by Leila M. iñigo de la Cruz and Thomas Wisse. SATAY experiments were carried out by Floor Dolsma, Thomas Wisse, and Enzo Kingma. Data analysis was performed by Enzo Kingma and Floor Dolsma. Enzo Kingma was responsible for writing the chapter and creating the figures. We express our gratitude toBenoît Kornmann and Agnès H. Michel for their invaluable assistance in setting up and troubleshooting the SATAY experiments, as well as for engaging discussions. Additionally, we appreciate the insightful discussions with Wessel Teunisse, Gregory van Beek, Leila M. Iñigo de la Cruz, and Werner Daalman.

Bibliography

Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology*, *11*(10), R106. https://doi.org/10.1186/gb-2010-11-10-r106

Bank, C., Matuszewski, S., Hietpas, R. T., & Jensen, J. D. (2016). On the (un)predictability of a large intragenic fitness landscape. *Proceedings of the National Academy of Sciences*, *113*(49), 14085–14090. https://doi.org/10.1073/pnas.1612676113

Baryshnikova, A., Costanzo, M., Dixon, S., Vizeacoumar, F. J., Myers, C. L., Andrews, B., & Boone, C. (2010). Chapter 7 - Synthetic Genetic Array (SGA) Analysis in Saccharomyces cerevisiae and Schizosaccharomyces pombe. In *Guide to yeast genetics: Functional genomics, proteomics, and other systems analysis* (pp. 145–179). Academic Press. https://doi.org/https://doi.org/10.1016/S0076-6879(10)70007-0

Baryshnikova, A., Costanzo, M., Kim, Y., Ding, H., Koh, J., Toufighi, K., Youn, J.-Y., Ou, J., San Luis, B.-J., Bandyopadhyay, S., Hibbs, M., Hess, D., Gingras, A.-C., Bader, G. D., Troyanskaya, O. G., Brown, G. W., Andrews, B., Boone, C., & Myers, C. L. (2010). Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nature Methods*, *7*(12), 1017–1024. https://doi.org/10.1038/nmeth.1534

Biery, M. C. (2000). A simple in vitro Tn7-based transposition system with low target site selectivity for genome and gene analysis. *Nucleic Acids Research*, *28*(5), 1067–1077. https://doi.org/10.1093/nar/28.5.1067

Butland, G., Babu, M., Díaz-Mejía, J. J., Bohdana, F., Phanse, S., Gold, B., Yang, W., Li, J., Gagarinova, A. G., Pogoutse, O., Mori, H., Wanner, B. L., Lo, H., Wasniewski, J., Christopoulos, C., Ali, M., Venn, P., Safavi-Naini, A., Sourour, N., … Emili, A. (2008). eSGA: E. coli synthetic genetic array analysis. *Nature Methods*, *5*(9), 789–795. https://doi.org/10.1038/nmeth.1239

Cain, A. K., Barquist, L., Goodman, A. L., Paulsen, I. T., Parkhill, J., & van Opijnen, T. (2020). A decade of advances in transposon-insertion sequencing. *Nature Reviews Genetics*, *21*(9), 526–540. https://doi.org/10.1038/s41576-020-0244-x

Chao, M. C., Abel, S., Davis, B. M., & Waldor, M. K. (2016). The design and analysis of transposon insertion sequencing experiments. *Nature Reviews Microbiology*, *14*(2), 119–128. https://doi.org/10.1038/nrmicro.2015.7

Cherry, J. M., Hong, E. L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E. T., Christie, K. R., Costanzo, M. C., Dwight, S. S., Engel, S. R., Fisk, D. G., Hirschman, J. E., Hitz, B. C., Karra, K., Krieger, C. J., Miyasato, S. R., Nash, R. S., Park, J., Skrzypek, M. S., … Wong, E. D. (2012). Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Research*, *40*(D1), D700–D705. https://doi.org/10.1093/nar/gkr1029

Chou, H.-H., Chiu, H.-C., Delaney, N. F., Segrè, D., & Marx, C. J. (2011). Diminishing Returns Epistasis Among Beneficial Mutations Decelerates Adaptation. *Science*, *332*(6034), 1190–1192. https://doi.org/10.1126/science.1203799

Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E. D., Sevier, C. S., Ding, H., Koh, J. L., Toufighi, K., Mostafavi, S., Prinz, J., St. Onge, R. P., VanderSluis, B., Makhnevych, T., Vizeacoumar, F. J., Alizadeh, S., Bahr, S., Brost, R. L., Chen, Y., … Boone, C. (2010). The Genetic

Landscape of a Cell. *Science*, *327*(5964), 425–431. https://doi.org/10.1126/science.1180823

Costanzo, M., VanderSluis, B., Koch, E. N., Baryshnikova, A., Pons, C., Tan, G., Wang, W., Usaj, M., Hanchard, J., Lee, S. D., Pelechano, V., Styles, E. B., Billmann, M., van Leeuwen, J., van Dyk, N., Lin, Z.-Y., Kuzmin, E., Nelson, J., Piotrowski, J. S., … Boone, C. (2016). A global genetic interaction network maps a wiring diagram of cellular function. *Science*, *353*(6306), aaf1420–aaf1420. https://doi.org/10.1126/science.aaf1420

De Visser, J. A. G. M., & Krug, J. (2014). Empirical fitness landscapes and the predictability of evolution. https://doi.org/10.1038/nrg3744

DeJesus, M. A., Gerrick, E. R., Xu, W., Park, S. W., Long, J. E., Boutte, C. C., Rubin, E. J., Schnappinger, D., Ehrt, S., Fortune, S. M., Sassetti, C. M., & Ioerger, T. R. (2017). Comprehensive Essentiality Analysis of the Mycobacterium tuberculosis Genome via Saturating Transposon Mutagenesis (C. L. Stallings, Ed.). *mBio*, *8*(1). https://doi.org/10.1128/mBio.02133-16

DeJesus, M. A., Zhang, Y. J., Sassetti, C. M., Rubin, E. J., Sacchettini, J. C., & Ioerger, T. R. (2013). Bayesian analysis of gene essentiality based on sequencing of transposon insertion libraries. *Bioinformatics*, *29*(6), 695–703. https://doi.org/10.1093/bioinformatics/btt043

Deshpande, R., Asiedu, M. K., Klebig, M., Sutor, S., Kuzmin, E., Nelson, J., Piotrowski, J., Ho Shin, S., Yoshida, M., Costanzo, M., Boone, C., Wigle, D. A., & Myers, C. L. (2013). A Comparative Genomic Approach for Identifying Synthetic Lethal Interactions in Human Cancer. *Cancer Research*, *73*(20), 6128–6136. https://doi.org/10.1158/0008-5472.CAN-12-3956

de Visser, J. A. G. M., Elena, S. F., Fragata, I., & Matuszewski, S. (2018). The utility of fitness landscapes and big data for predicting evolution. *Heredity*, *121*(5), 401–405. https://doi.org/10.1038/s41437-018-0128-4

Dixon, S. J., Fedyshyn, Y., Koh, J. L. Y., Prasad, T. S. K., Chahwan, C., Chua, G., Toufighi, K., Baryshnikova, A., Hayles, J., Hoe, K.-L., Kim, D.-U., Park, H.-O., Myers, C. L., Pandey, A., Durocher, D., Andrews, B. J., & Boone, C. (2008). Significant conservation of synthetic lethal genetic interaction networks between distantly related eukaryotes. *Proceedings of the National Academy of Sciences*, *105*(43), 16653–16658. https://doi.org/10.1073/pnas.0806261105

Franke, J., Klözer, A., de Visser, J. A. G. M., & Krug, J. (2011). Evolutionary Accessibility of Mutational Pathways (C. O. Wilke, Ed.). *PLoS Computational Biology*, *7*(8), e1002134. https://doi.org/10.1371/journal.pcbi.1002134

Giaever, G., Chu, A. M., Ni, L., Connelly, C., Riles, L., Véronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., André, B., Arkin, A. P., Astromoff, A., El Bakkoury, M., Bangham, R., Benito, R., Brachat, S., Campanaro, S., Curtiss, M., Davis, K., … Johnston, M. (2002). Functional profiling of the Saccharomyces cerevisiae genome. *Nature*, *418*(6896), 387–391. https://doi.org/10.1038/nature00935

Giaever, G., & Nislow, C. (2014). The Yeast Deletion Collection: A Decade of Functional Genomics. *Genetics*, *197*(2), 451–465. https://doi.org/10.1534/GENETICS.114.161620

Gietz, R. D., & Schiestl, R. H. (2007). High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nature Protocols*, *2*(1), 31–34. https://doi.org/10.1038/nprot.2007.13

Goodman, A. L., McNulty, N. P., Zhao, Y., Leip, D., Mitra, R. D., Lozupone, C. A., Knight, R., & Gordon, J. I. (2009). Identifying Genetic Determinants Needed to Establish a Human Gut Symbiont in Its Habitat. *Cell Host & Microbe*, *6*(3), 279–289. https://doi.org/10.1016/j.chom.2009.08.003

Green, B., Bouchier, C., Fairhead, C., Craig, N. L., & Cormack, B. P. (2012). Insertion site preference of Mu, Tn5, and Tn7 transposons. *Mobile DNA*, *3*(1), 1–6. https://doi.org/10.1186/1759-8753-3-3/FIGURES/3

Griffin, J. E., Gawronski, J. D., DeJesus, M. A., Ioerger, T. R., Akerley, B. J., & Sassetti, C. M. (2011). High-Resolution Phenotypic Profiling Defines Genes Essential for Mycobacterial Growth and Cholesterol Catabolism (L. Ramakrishnan, Ed.). *PLoS Pathogens*, *7*(9), e1002251. https://doi.org/10.1371/journal.ppat.1002251

Hietpas, R. T., Jensen, J. D., & Bolon, D. N. A. (2011). Experimental illumination of a fitness landscape. *Proceedings of the National Academy of Sciences*, *108*(19), 7896–7901. https://doi.org/10.1073/pnas.1016024108

Iñigo de la Cruz, L., van Beek, G., & Kok, M. (2022). transposonmapper. https://doi.org/10.5281/ZENODO.5903081

Jacobs, M. A., Alwood, A., Thaipisuttikul, I., Spencer, D., Haugen, E., Ernst, S., Will, O., Kaul, R., Raymond, C., Levy, R., Chun-Rong, L., Guenthner, D., Bovee, D., Olson, M. V., & Manoil, C. (2003). Comprehensive transposon mutant library of Pseudomonas aeruginosa. *Proceedings of the National Academy of Sciences*, *100*(24), 14339–14344. https://doi.org/10.1073/pnas.2036282100

Johnson, M. S., Martsul, A., Kryazhimskiy, S., & Desai, M. M. (2019). Higher-fitness yeast genotypes are less robust to deleterious mutations. *Science*, *366*(6464), 490–493. https://doi.org/10.1126/science.aay4199

Johnson, M. S., Reddy, G., & Desai, M. M. (2023). Epistasis and evolution: recent advances and an outlook for prediction. *BMC Biology*, *21*(1), 120. https://doi.org/10.1186/s12915-023-01585-3

Kauffman, S., & Levin, S. (1987). Towards a general theory of adaptive walks on rugged landscapes. *Journal of Theoretical Biology*, *128*(1), 11–45. https://doi.org/10.1016/S0022-5193(87)80029-2

Kauffman, S. A., & Weinberger, E. D. (1989). The NK model of rugged fitness landscapes and its application to maturation of the immune response. *Journal of Theoretical Biology*, *141*(2), 211–245. https://doi.org/10.1016/S0022-5193(89)80019-0

Kuzmin, E., Rahman, M., VanderSluis, B., Costanzo, M., Myers, C. L., Andrews, B. J., & Boone, C. (2021). τ-SGA: synthetic genetic array analysis for systematically screening and quantifying trigenic interactions in yeast. *Nature Protocols*, *16*(2), 1219–1250. https://doi.org/10.1038/s41596-020-00456-3

Laan, L., Koschwanez, J. H., & Murray, A. W. (2015). Evolutionary adaptation after crippling cell polarization follows reproducible trajectories. *eLife*, *4*. https://doi.org/10.7554/elife.09638

Li, C., & Zhang, J. (2018). Multi-environment fitness landscapes of a tRNA gene. *Nature Ecology & Evolution*, *2*(6), 1025–1032. https://doi.org/10.1038/s41559-018-0549-8

Limdi, A., & Baym, M. (2023). Resolving Deleterious and Near-Neutral Effects Requires Different Pooled Fitness Assay Designs. *Journal of Molecular Evolution*, *91*(3), 325–333. https://doi.org/10.1007/s00239-023-10110-7

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Lozovsky, E. R., Chookajorn, T., Brown, K. M., Imwong, M., Shaw, P. J., Kamchonwongpaisan, S., Neafsey, D. E., Weinreich, D. M., & Hartl, D. L. (2009). Stepwise acquisition of pyrimethamine resistance in the malaria parasite. *Proceedings of the National Academy of Sciences*, *106*(29), 12025–12030. https://doi.org/10.1073/pnas.0905922106

Michel, A. H., Hatakeyama, R., Kimmig, P., Arter, M., Peter, M., Matos, J., De Virgilio, C., & Kornmann, B. (2017). Functional mapping of yeast genomes by saturated transposition. *eLife*, *6*. https://doi.org/10.7554/eLife.23570

Michel, A. H., & Kornmann, B. (2022). SAturated Transposon Analysis in Yeast (SATAY) for Deep Functional Mapping of Yeast Genomes. In *Methods in molecular biology (clifton, n.j.)* (pp. 349–379). Humana Press Inc. https://doi.org/10.1007/978-1-0716-2257-5{_}20

Michel, A. H., van Schie, S., Mosbach, A., Scalliet, G., & Kornmann, B. (2019). Exploiting homologous recombination increases SATAY efficiency for loss- and gain-of-function screening. *bioRxiv*, 866483. https://doi.org/https://doi.org/10.1101/866483

Mira, P. M., Crona, K., Greene, D., Meza, J. C., Sturmfels, B., & Barlow, M. (2015). Rational Design of Antibiotic Treatment Plans: A Treatment Strategy for Managing Evolution and Reversing Resistance (P. J. Planet, Ed.). *PLOS ONE*, *10*(5), e0122283. https://doi.org/10.1371/journal.pone.0122283

Qian, W., Ma, D., Xiao, C., Wang, Z., & Zhang, J. (2012). The Genomic Landscape and Evolutionary Resolution of Antagonistic Pleiotropy in Yeast. *Cell Reports*, *2*(5), 1399–1410. https://doi.org/10.1016/j.celrep.2012.09.017

Rifat, D., Chen, L., Kreiswirth, B. N., & Nuermberger, E. L. (2021). Genome-Wide Essentiality Analysis of Mycobacterium abscessus by Saturated Transposon Mutagenesis and Deep Sequencing (M. Sloan Siegrist, Ed.). *mBio*, *12*(3). https://doi.org/10.1128/mBio.01049-21

Robinson, D. G., Chen, W., Storey, J. D., & Gresham, D. (2014). Design and Analysis of Bar-seq Experiments. *G3 Genes|Genomes|Genetics*, *4*(1), 11–18. https://doi.org/10.1534/g3.113.008565

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139. https://doi.org/10.1093/BIOINFORMATICS/BTP616

Robinson, M. D., & Smyth, G. K. (2007a). Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics*, *9*(2), 321–332. https://doi.org/10.1093/biostatistics/kxm030

Robinson, M. D., & Smyth, G. K. (2007b). Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics*, *23*(21), 2881–2887. https://doi.org/10.1093/bioinformatics/btm453

Roguev, A., Bandyopadhyay, S., Zofall, M., Zhang, K., Fischer, T., Collins, S. R., Qu, H., Shales, M., Park, H.-O., Hayles, J., Hoe, K.-L., Kim, D.-U., Ideker, T., Grewal, S. I., Weissman, J. S., & Krogan, N. J. (2008). Conservation and Rewiring of Functional Modules Revealed by an Epistasis Map in Fission Yeast. *Science*, *322*(5900), 405–410. https://doi.org/10.1126/science.1162609

Schlecht, U., Liu, Z., Blundell, J. R., St.Onge, R. P., & Levy, S. F. (2017). A scalable double-barcode sequencing platform for characterization of dynamic protein-protein interactions. *Nature Communications*, *8*(1), 15586. https://doi.org/10.1038/ncomms15586

Smith, A. M., Heisler, L. E., Mellor, J., Kaper, F., Thompson, M. J., Chee, M., Roth, F. P., Giaever, G., & Nislow, C. (2009). Quantitative phenotyping via deep barcode sequencing. *Genome Research*, *19*(10), 1836–1842. https://doi.org/10.1101/gr.093955.109

Teng, X., Dayhoff-Brannigan, M., Cheng, W.-C., Gilbert, C. E., Sing, C. N., Diny, N. L., Wheelan, S. J., Dunham, M. J., Boeke, J. D., Pineda, F. J., & Hardwick, J. M. (2013). Genome-wide Consequences of Deleting Any Single Gene. *Molecular Cell*, *52*(4), 485–494. https://doi.org/10.1016/j.molcel.2013.09.026

Tischler, J., Lehner, B., & Fraser, A. G. (2008). Evolutionary plasticity of genetic interaction networks. *Nature Genetics*, *40*(4), 390–391. https://doi.org/10.1038/ng.114

Usaj, M., Tan, Y., Wang, W., VanderSluis, B., Zou, A., Myers, C. L., Costanzo, M., Andrews, B., & Boone, C. (2017). TheCellMap.org: A Web-Accessible Database for Visualizing and Mining the Global Yeast Genetic Interaction Network. *G3 Genes|Genomes|Genetics*, *7*(5), 1539–1549. https://doi.org/10.1534/g3.117.040220

van Leeuwen, J., Pons, C., Mellor, J. C., Yamaguchi, T. N., Friesen, H., Koschwanez, J., Ušaj, M. M., Pechlaner, M., Takar, M., Ušaj, M., VanderSluis, B., Andrusiak, K., Bansal, P., Baryshnikova, A., Boone, C. E., Cao, J., Cote, A., Gebbia, M., Horecka, G., … Boone, C. (2016). Exploring genetic suppression interactions on a global scale. *Science*, *354*(6312). https://doi.org/10.1126/science.aag0839

van Leeuwen, J., Pons, C., Tan, G., Wang, J. Z., Hou, J., Weile, J., Gebbia, M., Liang, W., Shuteriqi, E., Li, Z., Lopes, M., Ušaj, M., Dos Santos Lopes, A., van Lieshout, N., Myers, C. L., Roth, F. P., Aloy, P., Andrews, B. J., & Boone, C. (2020). Systematic analysis of bypass suppression of essential genes. *Molecular Systems Biology*, *16*(9), e9828. https://doi.org/10.15252/msb.20209828

van Opijnen, T., Bodi, K. L., & Camilli, A. (2009). Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nature Methods*, *6*(10), 767–772. https://doi.org/10.1038/nmeth.1377

van Opijnen, T., & Camilli, A. (2012). A fine scale phenotype–genotype virulence map of a bacterial pathogen. *Genome Research*, *22*(12), 2541–2551. https://doi.org/10.1101/gr.137430.112

van Opijnen, T., & Camilli, A. (2013). Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nature Reviews Microbiology*, *11*(7), 435–442. https://doi.org/10.1038/nrmicro3033

van Opijnen, T., Lazinski, D. W., & Camilli, A. (2014). Genome-Wide Fitness and Genetic Interactions Determined by Tn-seq, a High-Throughput Massively Parallel Sequencing Method for Microorganisms. *Current Protocols in Molecular Biology*, *106*(1). https://doi.org/10.1002/0471142727.mb0716s106

van Opijnen, T., & Levin, H. L. (2020). Transposon Insertion Sequencing, a Global Measure of Gene Function. *Annual Review of Genetics*, *54*(1), 337–365. https://doi.org/10.1146/annurev-genet-112618-043838

Venkataram, S., Dunn, B., Li, Y., Agarwala, A., Chang, J., Ebel, E. R., Geiler-Samerotte, K., Hérissant, L., Blundell, J. R., Levy, S. F., Fisher, D. S., Sherlock, G., & Petrov, D. A. (2016). Development of a Comprehensive Genotype-to-Fitness Map of Adaptation-Driving Mutations in Yeast. *Cell*, *166*(6), 1585–1596.e22. https://doi.org/10.1016/J.CELL.2016.08.002

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … Vázquez-Baeza, Y. (2020).

SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Weinreich, D. M. (2006). Darwinian Evolution Can Follow Only Very Few Mutational Paths to Fitter Proteins. *Science*, *312*(5770), 111–114. https://doi.org/10.1126/science.1123539

Wetmore, K. M., Price, M. N., Waters, R. J., Lamson, J. S., He, J., Hoover, C. A., Blow, M. J., Bristow, J., Butland, G., Arkin, A. P., & Deutschbauer, A. (2015). Rapid Quantification of Mutant Fitness in Diverse Bacteria by Sequencing Randomly Bar-Coded Transposons (M. A. Moran, Ed.). *mBio*, *6*(3), e00306–15. https://doi.org/10.1128/mBio.00306-15

Yan Tong, A. H., & Boone, C. (2006). Synthetic Genetic Array Analysis in Saccharomyces cerevisiae. In W. Xiao (Ed.), *Yeast protocol* (pp. 171–191). Humana Press. https://doi.org/10.1385/1-59259-958-3:171

Zhang, H., Pounds, S. B., & Tang, L. (2013). Statistical Methods for Overdispersion in mRNA-Seq Count Data. *The Open Bioinformatics Journal*, (7), 3–34.

4.6. Supplement

4.6.1. Supplemental figures

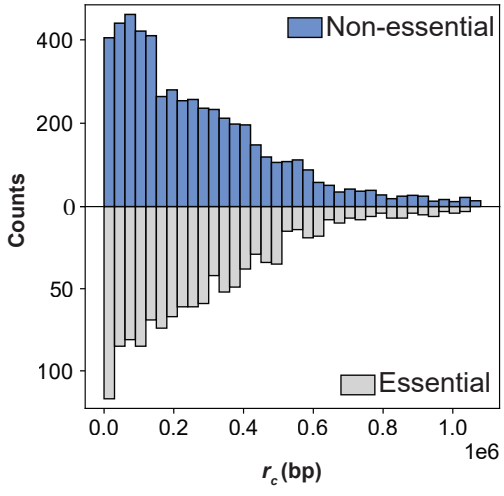


Figure S4.1. Essential and non-essential genes are distributed in a similar manner across the chromosome. Plot of the number of essential and non-essential genes for a specified distance r_c from the centromere. The distributions do not show a clear enrichment of essential genes over non-essential genes in pericentromeric regions.

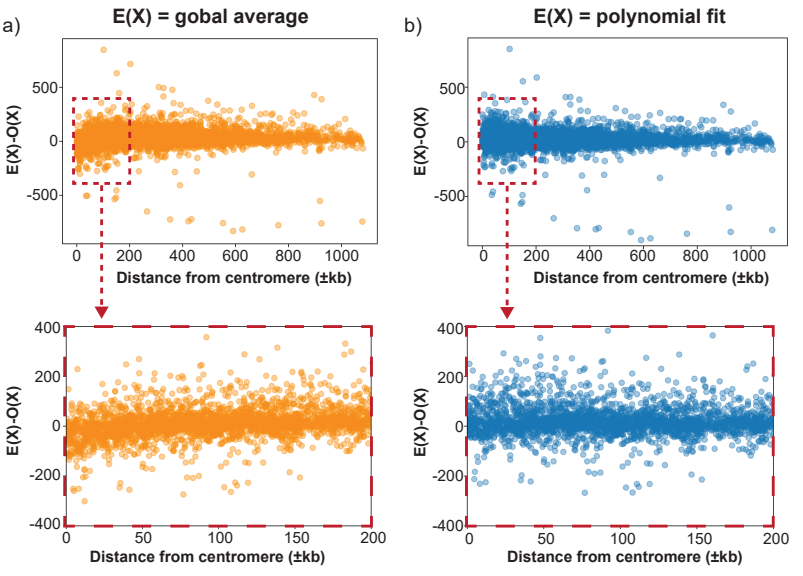


Figure S4.2. Polynomial fit of the observed transposon density across the genome corrects for the pericentromeric insertion bias of the MiniDS transposon. The difference between the expected (E(X)) and observed (O(X)) insertion density for all genes is plotted against the distance of the gene to the centromere. (**a**) Using the global average to estimate the expected insertion rate results in a systematic overestimation of the insertion density of genes close to the centromere. This overestimation is visible as a skew towards negative values for genes close to the centromere. (**b**) Using a polynomial fit of the observed insertion rate corrects for this skew.

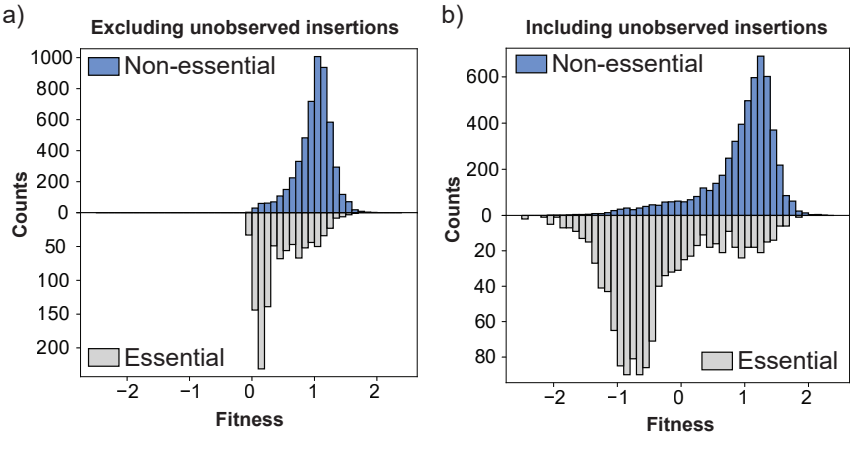


Figure S4.3. Inclusion of unobserved insertion sites improves the distinction between low-fitness genes. (**a**) The fitness distribution of gene disruptions when unobserved insertions are excluded during the fitness estimation. In this case, the fitness estimates depend only on the average read count per insertion site. (**b**) The fitness distribution of gene disruptions when unobserved insertions are excluded during the fitness estimation. The plot shows that the lower tail of the distribution has a larger spread when compared to (a), providing a better distinction between low-fitness genes. In this case, the fitness estimates of genes that have low fitness when disrupted depends on a combination of the average read count and the insertion density.

4.6.2. Supplementary tables

Table 4.1. List of strains used in this chapter

Name	Purpose	Ancestor	Genotype	Reference
yLIC135	Parental strain	yLL3a	*Matα can1-100 leu2-3,112 his3-11,15 ura3*^{0} *BUD4 from S288C ade2Δ::*	Laan et al., 2015
yLIC137	Identification of genetic interactions of *BEM3* (data Fig. 4.7b)	yLIC135	*Matα can1-100 leu2-3,112 his3-11,15 ura3*^{0} *BUD4 from S288C ade2Δ:: bem3::natMX4*	This study
yWT01a	Identification of the genetic interactions of *BEM3* (data Fig. 4.7b) Comparison with fitness values obtained from Bar-seq (data Fig. 4.7a)	yLIC135	*Matα can1-100 leu2-3,112 his3-11,15 ura3*^{0} *ade2Δ::*	This study

Table 4.2. Liquid media used for SATAY library generation. YNB: Yeast Nitrogen Base. CSM: Complete Supplement Mixture. Ura: Uracil. Ade: Adenine

Step	Media	Components
Preculture	SD-Ura+0.2% Glucose +2% Raffinose	• YNB w/o Amino Acids (6.8 g/L) • CSM -Ura (0.77 g/L) • Glucose (2 g/L) • Raffinose (20 g/L) • Adenine (20 mg/L)
Induction	SD-Ura+2% Galactose	• YNB w/o Amino Acids (6.8 g/L) • CSM -Ura (0.77 g/L) • Galactose (20 g/L) • Adenine (20 mg/L)
Reseed	SD-Ade+2% Glucose	• YNB w/o Amino Acids (6.8 g/L) • CSM -Ade (0.78 g/L) • Glucose (20 g/L)

Table 4.3. List of primers used in this chapter

#	Name	Barcode	Sequence (5'-3')
1	HT50_688_minidsSEQ1210	HT50	GCC ACA TAT TTA CCG ACC GTT ACC GAC CGT TTT CAT CCC TA
2	E2_HT48_MiniDS_RV	HT48	AGG TCA GTC ACA TGG TTA GGA CGC AGA GCT GAA ACG AAA ACG AAC GGG ATA AA
3	HT60_688_minidsSEQ1210	HT60	TAG GAT GAT TTA CCG ACC GTT ACC GAC CGT TTT CAT CCC TA
4	E2_HT49_MiniDS_RV	HT49	AGG TCA GTC ACA TGG TTA GGA CGC AGA TAG ACA ACG AAA ACG AAC GGG ATA AA

Table 4.4. PCR amplification protocol

Step	Temperature	Time
1	95 ºC	3 min
2	95 ºC	30 sec
3	55 ºC	30 sec
4	72 ºC	3 min
Go to step 2 (35x)		
6	72 ºC	10 min
7	4 ºC	Inf

Table 4.5. Characteristics of the SATAY libraries used in this chapter. Reads that map to the coding sequence of *ADE2* were excluded when determining the reported values.

Strain	**Biological Replicates**	**# clones**	**Technical Replicates**	**# of reads mapped**	**# of transposons mapped**
yWT01a	B1	$7.1x10^6$	B1_T1	15341620	964082
			B1_T2	14133484	942484
			B1_T3	15803098	948445
			B1_T4	16963682	955841
			B1_T5	21601243	1012544
			B1_T6	18749208	980674
	B2	$8.6x10^6$	B2_T1	21301610	1107947
			B2_T2	17979884	1062241
			B2_T3	22407804	1152751
	B3	$7.9x10^6$	—	21870779	1010354
	B4	$9.9x10^6$	—	18792166	1168784
yLIC137	—		—	11341842	597351

5

The best way to predict the future is to design it.

— Richard Buckminster Fuller

Controlling the accessibility of evolutionary trajectories with epistasis

Abstract The ability to control evolution has important applications in healthcare, industry and synthetic biology. An important theoretical and experimental finding has been that epistatic interactions between mutations can act as a major determinant for the accessibility of evolutionary pathways. Thus, epistatic interactions could, in principle, be used as a design element for the evolutionary control of biological systems. However, to our knowledge, exerting control over evolutionary pathways using epistatic interactions as a design element has not yet been attempted. Therefore, it is currently unclear whether the adaptive response of a cell can be controlled through the rational design of its genotype. In this chapter, we experimentally test the possibility of using epistasis to control evolution through a single genetic change. Specifically, we aim at finding genes that allow a weak form of control by rendering the most likely trajectory inaccessible. To this end, we exploit a previously mapped evolutionary trajectory of the polarity network of *Saccharomyces cerevisiae* for which it was shown that the inactivation of the protein Bem3 rapidly and reproducibly follows after the deletion of the protein Bem1. Using the genome-wide screen SATAY, we find two genes that may make this pathway inaccessible when deleted without strongly affecting the fitness of the wild-type genetic background. In addition, our data suggest that at least one of these two genes that can act as a switch for pathway accessibility could have been predicted based on a shared functional domain with Bem1. Our data therefore suggest that controlling the accessibility of evolutionary trajectories through epistasis can be achieved by single gene deletions.

5.1. Introduction

The effect of a genetic mutation depends on the genetic background in which it occurs (Carlborg & Haley, 2004; Fenster et al., 1997; Phillips, 2008). This context dependency, called epistasis (Bateson et al., 1909; Phillips, 1998), can cause the same oncogenic mutation in two healthy individuals to lead to the formation of cancer in one, but not the other individual (Nadeau, 2001). In a similar manner, epistatic interactions between mutations are an important determinant for which set of mutations improve fitness during evolution (De Visser & Krug, 2014; Domingo et al., 2019; Phillips, 2008). Specifically, theoretical and experimental studies have shown that epistasis can constrain the order in which mutations occur by making certain combinations inaccessible during natural selection (de Vos et al., 2015; Kauffman & Weinberger, 1989; Neidhart et al., 2014; Poelwijk et al., 2019; Sailer & Harms, 2017; Weinreich et al., 2005). These epistatic constraints often underlie patterns of reproducible evolutionary trajectories that have been observed both at intra-genic (de Vos et al., 2015; Weinreich, 2006) and inter-genic (Cooper et al., 2008; Laan et al., 2015; R. Woods et al., 2006) levels. As reproducibility implies the possibility of prediction, this raises the question of whether epistatic constraints can be used as a design element to control the mutational trajectory that will be taken in response to a given perturbation. This question is becoming increasingly relevant with the development of synthetic biological systems and the

increasing application of genetically modified organisms in industrial processes.

In this chapter, we explore the feasibility of controlling evolutionary trajectories through epistatic interactions between mutations. We focus on a weak form of control where a chosen mutational pathway is rendered inaccessible. Central to achieve this type of control are mutations that exhibit strong sign epistasis. Sign epistasis makes the sign of the fitness effect of a mutation dependent on genetic background (Kogenaru et al., 2009; Nghe et al., 2018; Poelwijk et al., 2011), which can lead to conditional essential genes in severe cases. The deletion of a gene that is conditionally essential between the subsequent mutational steps of an evolutionary trajectory should effectively block the mutational pathway. However, to be applicable to the design of biological systems the effect of deleting this conditionally essential gene should have minimal effects on the desired phenotype of the starting strain. It is unclear whether both these features can be reconciled through the deletion of a single gene or if this would require a more complex set of mutations.

To study the possibility of directing evolutionary pathways through a single genetic change, we use a previously mapped evolutionary trajectory of the polarity network of *Saccharomyces cerevisiae*. In *S. cerevisiae*, polarity establishment is a vital part of the cell cycle and failure to polarize leads to cell death (Irazoqui et al., 2003; Laan et al., 2015; Martin & Arkowitz, 2014; B. Woods et al., 2015). As in all eukaryotes, polarization is centered around the evolutionary conserved Rho GTPase Cdc42 which cycles between an inactive GDP-bound and an active GTP-bound form (Etienne-Manneville, 2004). In late G1, active Cdc42 accumulates at a single site on the cortical membrane where it induces the further asymmetric organization of internal cellular components (Chiou et al., 2017; Martin & Arkowitz, 2014). According to the current understanding of the mechanism for cell polarization, the scaffold protein Bem1 plays a key role in mediating the positive feedback loop for the accumulation of active Cdc42 on the membrane (Freisinger et al., 2013; Irazoqui et al., 2003; Klünder et al., 2013; Martin, 2015). Specifically, Bem1 plays a crucial role in connecting the localization of active Cdc42 to the guanine exchange factor (GEF) Cdc24. It achieves this by binding both proteins simultaneously at distinct domains, thereby enhancing the rate of Cdc42 activation (Chiou et al., 2017; Martin & Arkowitz, 2014). Deletion of Bem1 is not lethal, but severely diminishes the ability of cells to efficiently polarize (B. Woods et al., 2015). Interestingly, Laan et al. (2015) demonstrated that following the deletion of the Bem1, loss-of-function mutations in a gene coding for GTPase-activating protein Bem3 occur almost inevitably as compensatory mutations during experimental evolution. This high degree of reproducibility may allow us to control the adaptive response of cells to the loss of Bem1. Here, we asses the possibility to block this pathway through the deletion of genes that are nearly neutral in a *bem1*Δ genetic background, but become essential or strongly deleterious in a *bem1*Δ*bem3*Δ background (5.1). To do this, we use the genome-wide transposon mutagenesis screen SATAY (Michel et al., 2017) to look for genes that display the desired pattern of essentiality and dispensability. Furthermore, using the results of this screen we assess (1) how numerous the possibilities are for blocking such a mutational pathway and (2) whether genes that hold the desired properties have any functional relationship to the starting (*bem1*Δ) or the suppressor (*bem1*Δ*bem3*Δ) mutations.

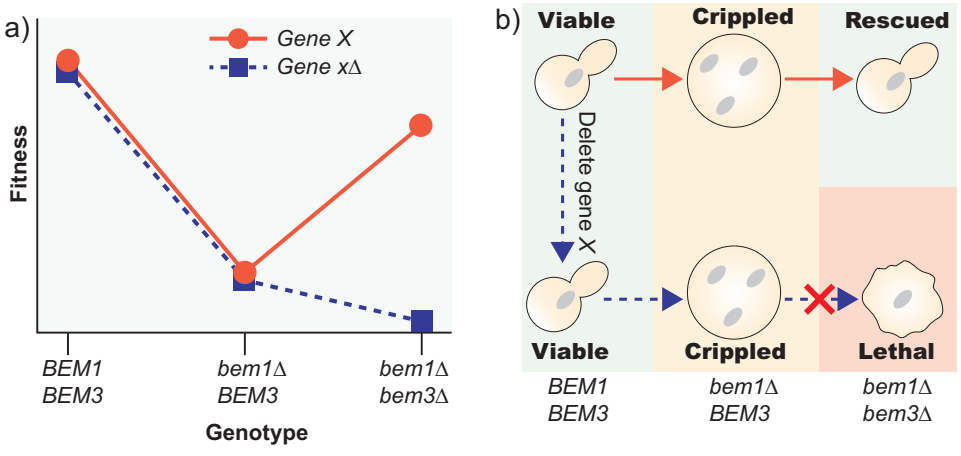


Figure 5.1. Illustration of the devised strategy to prevent the occurrence of compensatory mutations in *BEM3* after the loss of *BEM1*. (**a**) Fitness diagram of the evolutionary recovery after the loss of *BEM1*. In the known evolutionary trajectory (solid red line, circular markers), the fitness defects incurred by the loss of *BEM1* are largely recovered by the subsequent inactivation of *BEM3*. This pathway could be made inaccessible by the deletion of a hypothetical gene *X* if the loss of gene causes the *bem1*Δ *bem3*Δ mutant to have a lower fitness than the *bem1*Δ mutant (dashed blue line, square markers). Ideally, the loss of gene *X* has a negligible effect on fitness in the wild-type genetic background, which would allow this genotype to persist in natural populations. (**b**) Graphical depiction of the fitness landscape shown in panel (a). The loss of gene *X* should switch the *bem1*Δ *bem3*Δ from rescued to inviable, while leaving the fitness of the other genotypes unchanged.

5.2. Results

5.2.1. Construction of a conditional *bem1*Δ strain using the auxin-inducible degron

The deletion of *BEM1* has been reported to result in an exceptionally low proliferation rate (Chenevert et al., 1992; B. Woods et al., 2015). While SATAY is a versatile tool that can be used to determine the fitness effect of deleting a gene across genetic backgrounds, its application to genotypes that have a low fitness contains several pitfalls. First, the low proliferation rate complicates the construction of a mutant library. Second, low fitness strains are more likely to accumulate secondary mutations that compensate the fitness defect. Indeed, it has been shown that selective sweeps of compensatory mutations can occur within several generations in a *bem1*Δ population (Laan et al., 2015). As a consequence, the obtained results no longer represent the presumed genetic background.

Because the *bem1*Δ strain represents a low fitness genotype that will likely introduce the above-mentioned artifacts into the SATAY screen, we decided to construct a conditional knockdown of Bem1 using the auxin-inducible degron (AID) system (Nishimura et al., 2009; Papagiannakis et al., 2017). The AID-system uses the plant degradation signalling pathway to conditionally deplete a protein of interest. Briefly, the fusion of a protein to the AID tag results in the ubiquitination and subsequent degradation of the protein when plant hormone auxin is added to the growth media (Papagiannakis et al., 2017; Shetty et al., 2019). The concentration of auxin in the external media controls the protein degradation rate, with higher auxin concentrations resulting in a faster protein degradation rate. When the degradation rate exceeds the synthesis rate,

the protein is fully depleted, effectively generating a knockdown strain. Thus, the conditional depletion of Bem1 would allow us to construct the mutant library for SATAY in a strain that has a fitness similar to the wild-type, while performing the competitive fitness assay in a strain that is phenotypically a *bem1*Δ mutant.

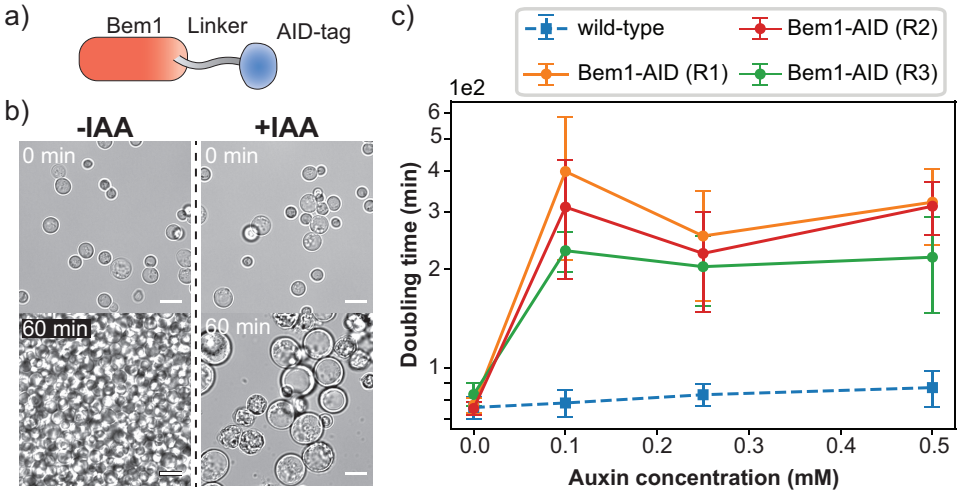


Figure 5.2. The strain carrying the Bem1-AID construct is phenotypically similar to a *bem1*Δ strain in the presence of auxin. (**a**) Schematic depiction of the construct used to fuse Bem1 to the AID tag. Due to the presence of the AID tag, Bem1 gets degraded in the presence of auxin (IAA). (**b**) Live-cell microscopy of the Bem1-AID strain growing in the absence (-) and presence (+) of IAA. In the absence of auxin, the strain proliferates rapidly and has morphology similar to that of a wild-type strain. In the presence of 0.1 mM indole-3-acetic acid (IAA), cells divide substantially slower and become abnormally large. (**b**) The doubling times of a wild-type strain and three replicate Bem1-AID strains (R1-R3) during growth in media with varying concentrations of IAA. Due to the large difference in doubling times between the strains, the y-axis is shown on a logarithmic scale. The graph shows that the doubling time plateaus in the Bem1-AID strains at the lowest IAA concentration tested (0.1mM), indicating that full depletion of Bem1-AID is achieved at this concentration. The doubling time of the wild-type strain is moderately affected by high concentrations of IAA, as has been reported by Nicastro et al. (2021).

A conditional Bem1 knockdown strain was created by fusing the C-terminus of endogenously expressed Bem1 to the AID tag (Figure 5.2a, top). We determined the sensitivity of this strain to auxin using live-cell microscopy (Figure 5.2a, bottom). Bem1-AID cells were grown either in the absence or in the presence of 0.1 mM natural auxin (Indole-3-acetic acid, IAA). The presence of IAA profoundly altered the morphology of the cells after 60 minutes of growth, causing them to grow into large unbudded cells. This phenotype is similar to what has been reported for germinating *bem1*Δ spores (Laan et al., 2015). We therefore conclude that our Bem1-AID construct is responsive to IAA degradation signalling.

To determine the auxin concentration at which Bem1 is fully depleted and therefore constitutes a Bem1 knockdown, we quantified the doubling time of Bem1-AID populations at various concentrations of IAA during growth over a period of 90 hours (Figure 5.2b). In the absence of IAA (0 mM), the growth curve of Bem1-AID strains was highly reproducible across replicate experiments and their doubling was similar to that of the wild-type strain. This confirms our observation by live-cell microscopy that the tagging of the Bem1 C-terminus has negligible effects on phenotype in the absence of IAA. In contrast, the addition of auxin caused highly variable

growth characteristics and strongly decreased the growth rate for all IAA concentrations tested. The doubling time plateaus at the lowest IAA concentration (0.1 mM), indicating that Bem1-AID is fully depleted at this concentration. However, to be certain that the effect of IAA lasts for the full duration of the competitive fitness assay of SATAY, we decided to use a concentration of 0.25 mM IAA in our SATAY screens.

5.2.2. Identifying candidate genes for evolutionary control through epistatic interactions

As stated in section 5.1, we are looking for genes that, when deleted, prevent the occurrence of loss-of-function mutations in *BEM3* after the loss of *BEM1*. We used SATAY in three genetic backgrounds to identify such genes: the wild-type, *BEM1-AID* and *BEM1-AID+bem3*Δ genetic backgrounds. Bem1 was depleted in the *BEM1-AID* and *BEM1-AID+bem3*Δ genetic backgrounds by adding 0.25 mM IAA the start of the competitive fitness assay. To maintain equivalent environmental conditions, the same amount of IAA was added during SATAY screens of the wild-type strain.

In principle, the read counts obtained from each transposon insertion reflects the abundance of a particular mutant and could be used to determine changes in gene deletion fitness across genetic backgrounds. However, we found that the average read count per transposon insertion site correlates poorly between replicate experiments of *BEM1-AID* and *BEM1-AID+bem3*Δ strains (Figure S5.1c,d). This poor correlation appears to be due to large differences in fitness between different mutants, which causes a large fraction of the reads to map to the coding region of a small number of genes. We therefore decided to not use read counts for our analysis and instead look for epistatic interactions based on the transposon insertion density, which shows a strong correlation between replicate experiments (Figure S5.1). To validate the use of insertion density as a metric for changes in gene dispensability, we compared our wild-type and *BEM1-AID* SATAY datasets to see if we could identify known synthetic rescue interactions of *BEM1*. We found that all annotated synthetic rescue genes were enriched for transposon insertions in the *BEM1-AID* dataset relative to the wild-type dataset (Figure S5.2).

The first requirement to render the trajectory from a *bem1*Δ mutant to a *bem1*Δ*bem3*Δ mutant inaccessible by the deletion of a third gene is that this third gene has a strongly altered degree of dispensability between the two genetic backgrounds. To identify genes with this characteristic, we created a volcano plot that compares the transposon insertion density of all annotated coding sequences between a *BEM1-AID* and a *BEM1-AID+bem3*Δ strain (Figure 5.3a). We considered the fitness effect of gene disruption to be strongly different between the two strains if the average insertion density of a gene was at least twice as high in the *BEM1-AID* strain relative to the *BEM1-AID+bem3*Δ strain and the p-value (student's t-test) was at least 0.1. In total, 330 genes were classified to be strongly enriched for transposon insertions in the *BEM1-AID* strain compared to the *BEM1-AID+bem3*Δ strain.

The second requirement is that the deleted gene should not strongly affect the fitness of the *bem1*Δ mutant. Due to the low fitness of the *bem1*Δ mutant, a gene deletion will be effectively lethal if it is deleterious in this background and can therefore not exist in nature. Alternatively, if the gene deletion is beneficial in the *bem1*Δ background, this mutation would not only block the trajectory to a *bem1*Δ*bem3*Δ genotype, but also redirect the *bem1*Δ mutant to a new trajectory towards higher fitness. Because we expect the fitness effects of individual gene disruptions to average out over larger genomic regions, we consider genes to be neutral when their insertion

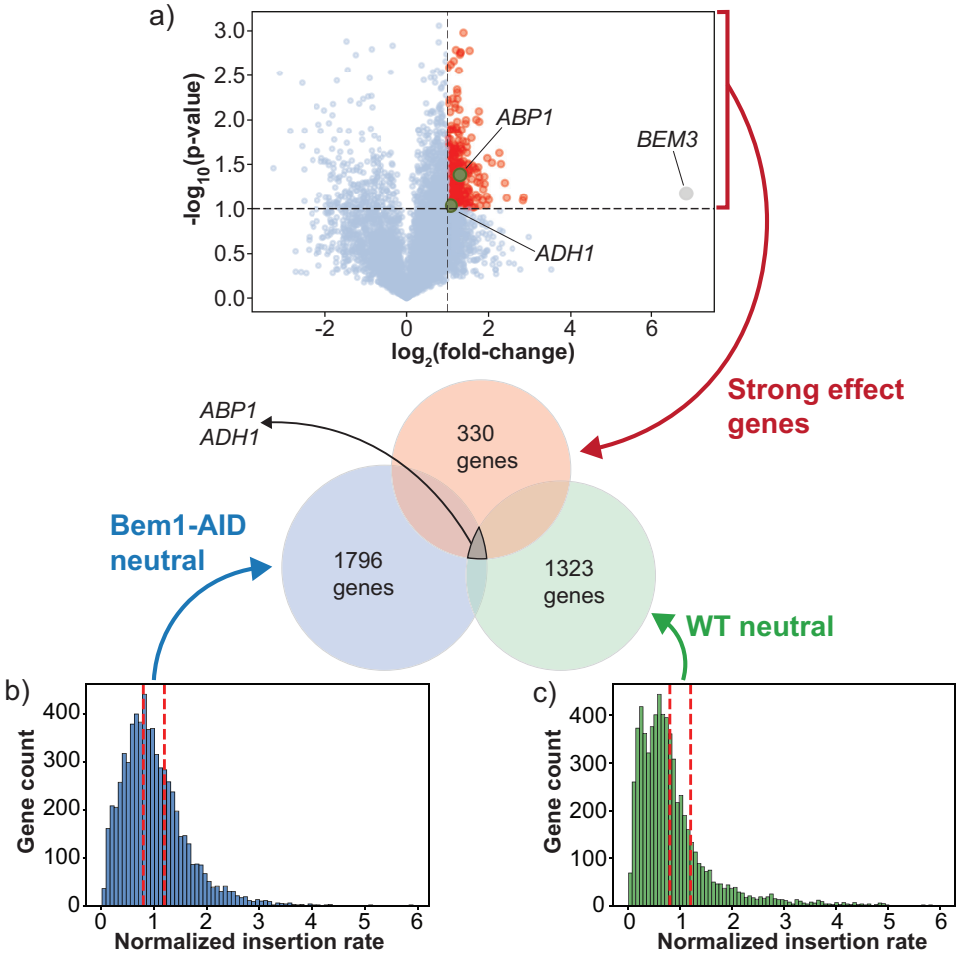


Figure 5.3. The genes *ABP1* and *ADH1* have the desired epistatic interaction profile to control evolution. (**a**) Volcano plot in which the transposon insertion density of genes in the *BEM1-AID* strain are compared to the insertion density of genes in the *bem3*Δ *BEM1-AID* strain. The fold-change in insertion density of a gene between the two genetic backgrounds is shown on the x-axis. The significance level (p-value, student's t-test) of this fold-change is shown y-axis. Each dot in the plot represents a single gene. Genes that have a positive fold-change are enriched for transposons in the *BEM1-AID* strain relative to the *bem3*Δ *BEM1-AID* strain. Genes were considered to have a strong difference in fitness effect between the two genetic backgrounds if their $\log_2$(fold-change) is larger than 1 and the p-value is larger than 0.1. A total of 330 genes satisfy these criteria. (**b**) Histogram of the transposon insertion density of genes in the *BEM1-AID* strain normalized to the chromosomal insertion density of the chromosome on which the gene is located. The average value of two replicate experiments is shown. Genes were considered to have a neutral effect on fitness if their normalized insertion density was between 0.8 and 1.2. A total of 1796 genes satisfy this criterion. (**c**) Histogram of the transposon insertion density of genes in the wild-type strain normalized to the chromosomal insertion density of the chromosome on which the gene is located. The average value of two replicate experiments is shown. Genes were considered to have a neutral effect on fitness if their normalized insertion density was between 0.8 and 1.2. A total of 1323 genes satisfy this criterion. The Venn diagram in the center shows that two genes, *ABP1* and *ADH1*, fall in all of the aforementioned categories.

density lies between 0.8 and 1.2 times the average density of the chromosome on which they are located (Figure 5.3b). Based on this criterion, we found 1796 genes that have a near neutral effect when deleted in the *BEM1-AID* strain.

The last requirement is that the gene deletion should be nearly neutral in the wild-type strain. A neutral fitness effect would allow this genotype to persist in natural environments where there may be competition with a wild-type population. In addition, neutrality suggests that the phenotype resulting from the gene deletion is equivalent to that of a wild-type strain, which may be a desirable property in the context of, for example, industrial applications. Using the same criterion for neutrality as used for the *BEM1-AID* strain, we found 1323 neutral genes in the wild-type (Figure5.3c).

There are two genes (out of 6605 genes) that fulfill all three described requirements for control of the evolutionary trajectory: *ADH1* and *ABP1*. Adh1 is an alcohol dehydrogenase that is required for the production of ethanol during fermentation and deletion of *ADH1* severely compromises anaerobic growth (Ida et al., 2012; Smidt et al., 2012). As its known molecular function is metabolic, the emergence of *ADH1* may be a consequence of a difference in growth conditions between the strains. Abp1 is an actin binding protein involved in the regulation of actin dynamics and endocytosis (Guo et al., 2018; Quintero-Monzon et al., 2005). Interestingly, actin mediated transport and endocyotosis are considered to be processes related to polarity establishment (Freisinger et al., 2013) and may therefore have a functional link with the defect incurred by the loss of *BEM1*. All together, our results suggest that controlling pathway accessibility based on epistatic interactions can be done through single gene deletions.

5.2.3. Domain analysis reveals functional relation between Abp1 and Bem1

SATAY uses the MiniDS transposon to randomly disrupt genes throughout the genome. Because the MiniDS transposon contains several stop codons in all frames of the transposon that truncate the expressed protein when inserted into a coding region, SATAY allows analysis at the level of protein domains (Michel et al., 2017). We used this feature to obtain a more detailed picture of the functional defect caused by the deletion of *BEM1* and how this relates to the function of the two candidate genes (*ADH1* and *ABP1*).

As expected, we find that the transposon insertion profile of *BEM1* shows a strong skew towards insertions close to the 3' end of the gene (Figure 5.4a, b). Mutants with gene truncations near the 3' end are able to escape the IAA induced depletion of Bem1 through expression of Bem1 without the AID-tag. Interestingly, insertions generating shorter truncations of Bem1 also appear to partially alleviate the fitness defect caused by the *BEM1* knockdown. Starting from the 3' gene end, the shortest fragment of *BEM1-AID* that still seems to tolerate transposon insertions end approximately halfway of the second SH3 domain (Figure 5.4a). This suggests that compensation of the fitness effects of *BEM1* knockdown relies on the two SH3 domains, while the PX and PB1 domains are (to some degree) dispensable. The transposon insertion density in the approximate region between the second SH3 domain and the PX domain decreases after the deletion of *bem3* in the *BEM1-AID* genetic background (compare the regions delineated by the read dashed box in Figure 5.4a and b). The deletion of *BEM3* in the *bem1*Δ background may therefore provide a fitness benefit by compensating for the loss of the SH3 domains of Bem1. Intriguingly, *ABP1*, which emerged as one of the candidate genes for evolutionary control, encodes for a protein with an SH3 domain at its C-terminus (Figure 5.4c,d). The ability of *ABP1* to control the accessibility of the adaptive pathway from a *bem1*Δ strain to a *bem1*Δ*bem3*Δ strain may therefore be a result

of its functional relation to the defect introduced by deleting *BEM1*. In conclusion, we propose that functional relations may provide a means to predict genes that control the accessibility of a given evolutionary trajectory through their epistatic interaction network.

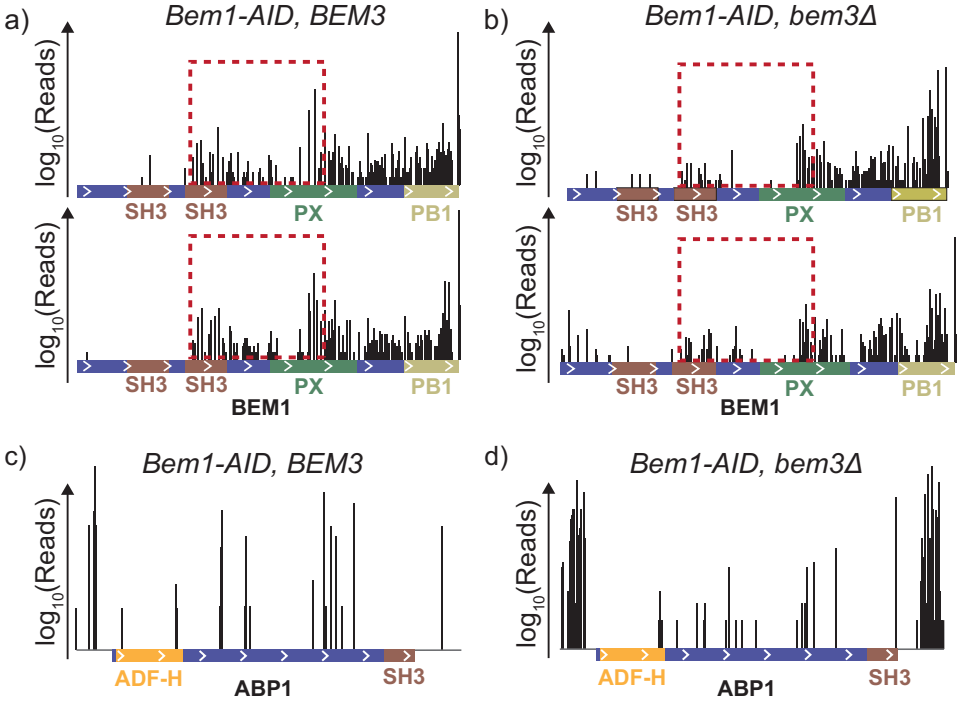


Figure 5.4. The deletion of *BEM3* alters the transposon insertion pattern across the domains of Bem1 in the *BEM1-AID* strain. (**a-b**) The transposon insertion profile along the *BEM1* gene in (**a**) two replicate *BEM1-AID* strains and (**b**) two replicate *BEM1-AID+bem3*Δ strains (b). The protein domains of Bem1 are shown at the bottom of each panel. The plots show that a region containing part of the PX domain and part of the second SH3 domain of Bem1 are enriched for transposon insertions in the *BEM1-AID* strains relative to the *BEM1-AID+bem3*Δ strains (region delineated by the dashed box). This suggests that compensation of the fitness defect incurred by the loss of *BEM1* through the subsequent inactivation of *BEM3* is related to the SH3 domain of Bem1. (**c-d**) The insertion pattern in the *ABP1* gene in (**c**) a *BEM1-AID* and (**d**) a *BEM1-AID+bem3*Δ background. The protein domain annotations show that Abp1 also contains an SH3 domain, indicating a possible functional link to Bem1.

5.3. Conclusions

Epistasis has been recognized as an important determinant for evolution since the introduction of the adaptive landscape metaphor by Sewall Wright (Wright, 1932). As the tools for mapping epistasis on a genome-wide scale become increasingly available, we can now study whether it is possible to reshape the structure of the adaptive landscape to exert control over adaptive pathways. In this chapter, we have used a reproducible evolution trajectory of the polarity network of *S. cerevisiae* to determine if evolutionary trajectories can be controlled through epistatic interactions. Specifically, we screened for genes with an epistatic interaction pattern that would prevent the occurrence of inactivating mutations in *BEM3* after the deletion of *BEM1*, but leave the wild-type phenotype unmodified. Using the genome-wide screen SATAY, we found two genes

that appear to fit these criteria: *ADH1* and *ABP1*.

The fact that we only find two genes (out of a total of 6600) that have the desired characteristics suggests that the options to control evolution through the deletion of a single gene may not be numerous. However, it is likely that our experimental design limits the sensitivity of our screen and we can therefore not formulate a conclusive statement on the prevalence of such genes from our results. In particular, we chose to perform the SATAY screen using a conditional knockdown of the *BEM1* gene (implemented with the AID degradation system) to prevent the accumulation of secondary mutations. Although this approach increases the interpretability of our results by strongly reducing the occurrence of secondary mutations, it has the drawback that the AID-degron system is an easy mutational target through which the conditional mutants can be reverted to the wild-type phenotype. As a result, large fitness differences exist between the mutants present in our SATAY screens, ranging from those that have a near lethal phenotype (*bem1*Δ) to those with a neutral phenotype (wild-type). Therefore, a small group with high fitness will quickly dominate the population (S5.3) which limits the amount of information we obtain on genes with neutral and deleterious fitness effects, as these mutants are rapidly diluted out of the population. This effect plays a role in both our screens of the *BEM1-AID* and the *BEM1-AID-bem3*Δ mutants, although to a lesser extent. In addition, the number of genes might have been higher if the thresholds for neutrality and synthetic lethality were chosen more leniently.

Interestingly, one of the two candidate genes that emerged from our screen, *ABP1*, appears to have a functional connection to polarity establishment and Bem1. Abp1 promotes the nucleation of actin polymers from the Actin Related Protein (ARP) 2/3 complex by binding to acting through its actin depolymerization factor (ADF)-homology domain (Michelot et al., 2013; Quintero-Monzon et al., 2005). In addition to regulating actin dynamics, it links the actin network to the endocytosis network by binding proteins with its SH3 domain (Colwill et al., 1999; Lila & Drubin, 1997; Warren et al., 2002). Intriguingly, a domain analysis of the transposon insertion pattern of *BEM1* suggested that the fitness defects incurred by the deletion of *BEM1* are partially alleviated by protein truncations that allow expression of the two SH3 domains of Bem1 (Figure 5.4). This finding is supported by the study of Grinhagens et al. (2020) which showed that the central SH3 domain of Bem1 can rescue the phenotype of *bem1*Δ cells. However, it has been shown that the binding properties of SH3 domains are often not conserved between proteins and depend on their relative position to other domains within the same protein (Dionne et al., 2021), making it difficult to infer whether the domain of Abp1 could functionally replace an SH3 domain of Bem1. In addition, the precise mechanism that would make the beneficial effect of deleting *BEM3* in a *bem1*Δ background dependent on Abp1 remains unclear, although the findings that Bem3 depends on the endocytosis machinery for its localization to the bud site (Mukherjee et al., 2013) and may share common binding partners with Abp1 might provide (Drees et al., 2001) a starting point to elucidate such a mechanism. Second, Abp1 has been shown to exhibit conditional essentiality with other proteins. In wild-type cells, the deletion of *ABP1* shows no significant phenotype (Drubin et al., 1990), but becomes lethal when combined with the deletion of the endocytosis related genes such as *SLA1*, *SLA2* and *SAC6* (Garcia et al., 2012; Holtzman et al., 1993). This demonstrates that Abp1 is able to buffer for essential functions, and we hypothesize that this may allow it to compensate for the loss of *BEM1*, but only through a somewhat more complex relationship that depends on the presence of *BEM3*. Furthermore, the neutrality of an *ABP1* deletion in a wild-type strain creates a realistic likelihood that a genetic background in which the deletion of *bem3*Δ following the deletion of *bem1*Δ is not permitted may exist in natural populations.

Our results indicate that the accessibility of evolutionary pathways can be controlled through epistasis. In particular, our analysis implies that this control can be achieved through a single genetic change. Furthermore, considering that Abp1 and Bem1 both have an SH3 domain and that the two proteins have a role in overlapping cellular functions, our findings suggest that it may be possible to identify genes that allow this type of control from their functional annotations. However, this prompts the question why other proteins with an SH3 domain did not emerge as candidate genes in our screen. Indeed, it has been estimated that between 23-25 proteins with an SH3 domain exist in the yeast proteome (Mirey et al., 2005). While we do not provide direct evidence that makes Abp1 a more likely candidate than other SH3 containing proteins, we argue that the context in which the protein-protein interaction takes place likely plays an important role. This is supported by reports that have shown that, although the sequence of the SH3 domain is conserved, which interactions the domain participates in depends on the other domains present in the host protein (Dionne et al., 2021) and on cellular context (Zarrinpar et al., 2003). Thus, the functional relation between the protein that allows epistatic control of evolutionary trajectories and the initially induced genetic perturbation may be complex.

5.4. Methods

5.4.1. Strains

The auxin inducible degron (AID) system used in this study requires the introduction of two components: the AID-tag that targets a protein of interest for degradation and OsTIR1, which is required for ubiqitination of the AID-tag in the presence of auxin. The OsTIR1-3XMyc gene fragment was obtained from plasmid pOsTIR1w/oGFP (Addgene catalog # 102883) through PCR amplification with primers Seq4_Fwd and Seq4_Rev and integrated at the *HO* locus of strain yWT01a via homologous recombination, yielding strain yWT02a. Correct integration and sequence integrity of the integrated *OsTIR1* construct in yWT02a was verified by Sanger sequencing (Macrogen), which showed no deviations from the expected sequence. For the construction of a gene fragment that allows the conditional depletion of Bem1, plasmid pG23A (Addgene catalog # 102884) was modified using Gibson assembly to replace the sequences coding for the *CDC14* N-terminus and the downstream flanking site with those of *BEM1*. The *BEM1-mCherry-AID* gene fragment was amplified by PCR with primers Bem1-mCherry-FW-AID and Bem1-mCherry-FW-AID and transformed into strain yWT02a, resulting in strain yWT03a. The integrated *BEM1-mCherry-AID* in yWT03a was verified with Sanger sequencing (Macrogen), which revealed a single SNP (A to T conversion) at position 1433 of the coding sequence of *BEM1*. This variant occurs naturally in the S288C genomic background of *S. cerevisiae*, and it was therefore assumed to have no detrimental effects on the functionality of Bem1. Finally, the endogenous *BEM3* locus of yWT03a was replaced with the natMX4 marker to obtain strain yWT04a. Strains were stored at -80°C as frozen stocks in 40% (v/v) glycerol. Plasmids pOsTIR1w/oGFP (Addgene plasmid # 10288) and pG23A (Addgene plasmid # 102884) were a gift from Matthias Heinemann.

5.4.2. Media

Standard culturing and growth assays were performed in YPD (10g/L Yeast extract, 20 g/L Peptone, 20 g/L dextrose), SC (6.9 g/L Yeast nitrogen base, 0.75 g/L Complete supplement mixture, 20 g/L dextrose). For ade$^-$ strains, standard growth media was supplemented with 20 mg/L adenine just before incubation. Liquid and solid media for SATAY were prepared as described in section 4.4.2 and table 4.2. 0.1 M stock solutions of indole-3-acetic acid (IAA) were prepared by dissolving 175 mg IAA in 2 ml of 100% absolute ethanol. The final volume was adjusted to 10 ml, adding additional ethanol when necessary to prevent precipitation of IAA. The stock solution was filter sterilized using a 0.2 µm syringe filter, aliquoted and stored at -20°C.

5.4.3. Microscopy

Cells were grown overnight from a glycerol stock in 10 ml liquid SC media (using the low-fluorescence version of CSM). The next day, cells were diluted in fresh SC media. For Bem1 depletion experiments, IAA was added to a final concentration of 0.1 mM. 100 µm of the diluted cell suspension was transferred to a well of a glass bottom 96 well plate (Greiner Bio-One, #655097) that was pre-coated with Concanavalin A (20 minute incubation at room temperature with 200 µm of 0.1 mg/ml Concanavalin A solution). Microscopy images were acquired on a Nikon Ti eclipse inverted wide-field microscope equipped with a cage incubator (Okolab) to maintain the temperature at 30°C. Images were obtained every minute in brightfield (100 ms exposure time) and RFP (555 nm excitation, 100 ms exposure time, 15 % laser intensity, Z-stack) channels with a 60x oil objective. Because Bem1-mCherry-AID was not visible in the acquired images, data from the RFP channel is not shown.

5.4.4. Growth assays

Cells were grown overnight from a glycerol stock in 10 ml YPD media. The next day, the cells were diluted to an OD600 of 0.1 (measured on a Nanodrop 2000 spectrophotometer) and IAA was added from the 0.1 M stock solution to a final concentration of either 0 mM, 0.1 mM, 0.25 mM or 0.50 mM. Cells were vortexed and the cell suspension was distributed over multiple wells of a 96 well plate (Thermo Fisher, #267427), with each well containing 100 µl of cell suspension. The edge moats of the well plate were filled with 1.5 ml sterile MiliQ water and the plate was sealed with parafilm to minimize evaporation. The 96 well plates were incubated in a plate reader (Biotek Epoch 2) set to 30 °C with a 1 °C gradient between the top and bottom of the plate. Density measurements at 600 nm were taken every 7 minutes over a period of 4 days, with continuous linear shaking (567 cycles per minute, 3 mm amplitude) in between measurements. Doubling times were extracted from the growth curves using an in-house Matlab code.

5.4.5. SATAY library generation

SATAY libraries were generated based on the protocol developed by Michel et al. (2019). The procedures for transposition, DNA extraction and circularization were performed as described in section 4.4.3. Transposon-genome junctions were PCR amplified with barcoded primers 1 and 2 (table 4.3) when genomic DNA was digested with DpnII and primers 3 and 4 (table 4.3) when genomic DNA was digested with NlaIII. PCR amplification was performed on a thermal cycler (Bio-Rad C1000 Touch) with the block settings shown in table 4.4. PCR amplified samples were purified using the NucleoSpin Gel and PCR cleanup kit (Macherey-Nagel) and quantified on the NanoDrop 2000 spectrophotometer (Thermo Scientific). For each sample, equal ratios (w/w) of DpnIII and NlaII digested DNA were pooled. Library preparation and sample sequencing were performed by Novogene (UK) Company Limited. Sequencing libraries were prepared with the NEBNext Ultra II DNA Library Prep Kit, omitting the size selection and PCR enrichment steps. Libraries were sequenced on the NovaSeq 6000 platform (Illumina) set to Paired-End (PE) sequencing with a read length of 150 bp.

5.4.6. Volcano plots

The number of transposons mapping to a chromosome (both in coding and non-coding regions) was determined from the WIG output file of the transposonmapper Python package (Iñigo de la Cruz et al., 2022). This value was divided by the chromosome length (in bp) given by the S288C reference genome (release R64-3-1_20210421) available on SGD to obtain the insertion rate of each chromosome. The insertion rate per gene was calculated from the output data of the transposonmapper package (per_gene_file) and scaled to the insertion rate of the corresponding chromosome. The scaled insertion rates were used to determine changes in gene disruption tolerance between SATAY libraries through volcano plots. The $\log_2$-fold changes of each gene shown in the volcano plots were calculated with the average scaled insertion rate of replicate libraries. P-values were generated using the unequal variance independent t-test available from the SciPy library in Python (Virtanen et al., 2020). Gene deletions were defined as neutral if its scaled insertion rate was between 0.8 and 1.2.

5.4.7. Domain analysis

Bem1 domains were obtained from the PROSITE database (Sigrist et al., 2013). Domain coordinates were manually converted into a General Feature Format 3 (GFF3) file. This GFF3 file was

imported into the IGV genome viewer (version 2.12.3, Robinson et al. (2011)) together with a wiggle (WIG) file containing the SATAY data and visualized against the sacCer3 genome annotation.

5.5. Contributions & Acknowledgements

The strains used in this chapter were constructed by Wessel Teunisse, Enzo Kingma and Ramon van der Valk. Enzo Kingma conducted live cell microscopy of the Bem1-AID strains. Population growth assays were carried out by Wessel Teunisse. Enzo Kingma and Wessel Teunisse performed the SATAY experiments and conducted the data analysis. This chapter was written by Enzo Kingma.

Bibliography

Bateson, W., Mendel, G., & Leighton, A. G. (1909). *Mendel's principles of heredity, by W. Bateson*. University Press. https://www.biodiversitylibrary.org/item/15713

Carlborg, Ö., & Haley, C. S. (2004). Epistasis: too often neglected in complex trait studies? *Nature Reviews Genetics*, *5*(8), 618–625. https://doi.org/10.1038/nrg1407

Chenevert, J., Corrado, K., Bender, A., Pringle, J., & Herskowitz, I. (1992). A yeast gene (BEM1) necessary for cell polarization whose product contains two SH3 domains. *Nature*, *356*(6364), 77–79. https://doi.org/10.1038/356077a0

Chiou, J. G., Balasubramanian, M. K., & Lew, D. J. (2017). Cell Polarity in Yeast. *Annual review of cell and developmental biology*, *33*, 77. https://doi.org/10.1146/ANNUREV-CELLBIO-100616-060856

Colwill, K., Field, D., Moore, L., Friesen, J., & Andrews, B. (1999). In Vivo Analysis of the Domains of Yeast Rvs167p Suggests Rvs167p Function Is Mediated Through Multiple Protein Interactions. *Genetics*, *152*(3), 881–893. https://doi.org/10.1093/GENETICS/152.3.881

Cooper, T. F., Remold, S. K., Lenski, R. E., & Schneider, D. (2008). Expression Profiles Reveal Parallel Evolution of Epistatic Interactions Involving the CRP Regulon in Escherichia coli. *PLoS Genetics*, *4*(2). https://doi.org/10.1371/JOURNAL.PGEN.0040035

De Visser, J. A. G. M., & Krug, J. (2014). Empirical fitness landscapes and the predictability of evolution. https://doi.org/10.1038/nrg3744

de Vos, M. G. J., Dawid, A., Sunderlikova, V., & Tans, S. J. (2015). Breaking evolutionary constraint with a tradeoff ratchet. *Proceedings of the National Academy of Sciences*, *112*(48), 14906–14911. https://doi.org/10.1073/pnas.1510282112

Dionne, U., Bourgault, É., Dubé, A. K., Bradley, D., Chartier, F. J. M., Dandage, R., Dibyachintan, S., Després, P. C., Gish, G. D., Pham, N. T. H., Létourneau, M., Lambert, J.-P., Doucet, N., Bisson, N., & Landry, C. R. (2021). Protein context shapes the specificity of SH3 domain-mediated interactions in vivo. *Nature Communications*, *12*(1), 1597. https://doi.org/10.1038/s41467-021-21873-2

Domingo, J., Baeza-Centurion, P., & Lehner, B. (2019). The Causes and Consequences of Genetic Interactions (Epistasis). *Annual Review of Genomics and Human Genetics*, *20*(1), 433–460. https://doi.org/10.1146/annurev-genom-083118-014857

Drees, B. L., Sundin, B., Brazeau, E., Caviston, J. P., Chen, G. C., Guo, W., Kozminski, K. G., Lau, M. W., Moskow, J. J., Tong, A., Schenkman, L. R., McKenzie, A., Brennwald, P., Longtine, M., Bi, E., Chan, C., Novick, P., Boone, C., Pringle, J. R., … Drubin, D. G. (2001). A protein interaction map for cell polarity development. *Journal of Cell Biology*, *154*(3), 549–576. https://doi.org/10.1083/JCB.200104057

Drubin, D. G., Mulholland, J., Zhu, Z., & Botstein, D. (1990). Homology of a yeast actin-binding protein to signal transduction proteins and myosin-I. *Nature*, *343*(6255), 288–290. https://doi.org/10.1038/343288a0

Etienne-Manneville, S. (2004). Cdc42 - the centre of polarity. *Journal of Cell Science*, *117*(8), 1291–1300. https://doi.org/10.1242/JCS.01115

Fenster, C. B., Galloway, L. F., & Chao, L. (1997). Epistasis and its consequences for the evolution of natural populations. *Trends in Ecology & Evolution*, *12*(7), 282–286. https://doi.org/10.1016/S0169-5347(97)81027-0

Freisinger, T., Klünder, B., Johnson, J., Müller, N., Pichler, G., Beck, G., Costanzo, M., Boone, C., Cerione, R. A., Frey, E., & Wedlich-Söldner, R. (2013). Establishment of a robust single axis of cell polarity by coupling multiple positive feedback loops. *Nature Communications*, *4*(1), 1807. https://doi.org/10.1038/ncomms2795

Garcia, B., Stollar, E. J., & Davidson, A. R. (2012). The Importance of Conserved Features of Yeast Actin-Binding Protein 1 (Abp1p): The Conditional Nature of Essentiality. *Genetics*, *191*(4), 1199–1211. https://doi.org/10.1534/GENETICS.112.141739

Grinhagens, S., Dünkler, A., Wu, Y., Rieger, L., Brenner, P., Gronemeyer, T., Mulaw, M. A., & Johnsson, N. (2020). A time-resolved interaction analysis of Bem1 reconstructs the flow of Cdc42 during polar growth. *Life Science Alliance*, *3*(9). https://doi.org/10.26508/LSA.202000813

Guo, S., Sokolova, O. S., Chung, J., Padrick, S., Gelles, J., & Goode, B. L. (2018). Abp1 promotes Arp2/3 complex-dependent actin nucleation and stabilizes branch junctions by antagonizing GMF. *Nature Communications*, *9*(1), 2895. https://doi.org/10.1038/s41467-018-05260-y

Holtzman, D. A., Yang, S., & Drubin, D. G. (1993). Synthetic-lethal interactions identify two novel genes, SLA1 and SLA2, that control membrane cytoskeleton assembly in Saccharomyces cerevisiae. *The Journal of cell biology*, *122*(3), 635–644. https://doi.org/10.1083/JCB.122.3.635

Ida, Y., Furusawa, C., Hirasawa, T., & Shimizu, H. (2012). Stable disruption of ethanol production by deletion of the genes encoding alcohol dehydrogenase isozymes in Saccharomyces cerevisiae. *Journal of Bioscience and Bioengineering*, *113*(2), 192–195. https://doi.org/10.1016/J.JBIOSC.2011.09.019

Iñigo de la Cruz, L., van Beek, G., & Kok, M. (2022). transposonmapper. https://doi.org/10.5281/ZENODO.5903081

Irazoqui, J. E., Gladfelter, A. S., & Lew, D. J. (2003). Scaffold-mediated symmetry breaking by Cdc42p. *Nature Cell Biology*, *5*(12), 1062–1070. https://doi.org/10.1038/ncb1068

Kauffman, S. A., & Weinberger, E. D. (1989). The NK model of rugged fitness landscapes and its application to maturation of the immune response. *Journal of Theoretical Biology*, *141*(2), 211–245. https://doi.org/10.1016/S0022-5193(89)80019-0

Klünder, B., Freisinger, T., Wedlich-Söldner, R., & Frey, E. (2013). GDI-Mediated Cell Polarization in Yeast Provides Precise Spatial and Temporal Control of Cdc42 Signaling. *PLOS Computational Biology*, *9*(12), e1003396. https://doi.org/10.1371/JOURNAL.PCBI.1003396

Kogenaru, M., de Vos, M. G. J., & Tans, S. J. (2009). Revealing evolutionary pathways by fitness landscape reconstruction. *Critical Reviews in Biochemistry and Molecular Biology*, *44*(4), 169–174. https://doi.org/10.1080/10409230903039658

Laan, L., Koschwanez, J. H., & Murray, A. W. (2015). Evolutionary adaptation after crippling cell polarization follows reproducible trajectories. *eLife*, *4*. https://doi.org/10.7554/elife.09638

Lila, T., & Drubin, D. G. (1997). Evidence for physical and functional interactions among two Saccharomyces cerevisiae SH3 domain proteins, an adenylyl cyclase-associated protein and

the actin cytoskeleton. *Molecular Biology of the Cell*, *8*(2), 367–385. https://doi.org/10.1091/mbc.8.2.367

Martin, S. G. (2015). Spontaneous cell polarization: Feedback control of Cdc42 GTPase breaks cellular symmetry. *BioEssays*, *37*(11), 1193–1201. https://doi.org/10.1002/bies.201500077

Martin, S. G., & Arkowitz, R. A. (2014). Cell polarization in budding and fission yeasts. *FEMS Microbiology Reviews*, *38*(2), 228–253. https://doi.org/10.1111/1574-6976.12055

Michel, A. H., Hatakeyama, R., Kimmig, P., Arter, M., Peter, M., Matos, J., De Virgilio, C., & Kornmann, B. (2017). Functional mapping of yeast genomes by saturated transposition. *eLife*, *6*. https://doi.org/10.7554/eLife.23570

Michel, A. H., van Schie, S., Mosbach, A., Scalliet, G., & Kornmann, B. (2019). Exploiting homologous recombination increases SATAY efficiency for loss- and gain-of-function screening. *bioRxiv*, 866483. https://doi.org/https://doi.org/10.1101/866483

Michelot, A., Grassart, A., Okreglak, V., Costanzo, M., Boone, C., & Drubin, D. G. (2013). Actin Filament Elongation in Arp2/3-Derived Networks Is Controlled by Three Distinct Mechanisms. *Developmental Cell*, *24*(2), 182–195. https://doi.org/10.1016/J.DEVCEL.2012.12.008

Mirey, G., Soulard, A., Orange, C., Friant, S., & Winsor, B. (2005). SH3 domain-containing proteins and the actin cytoskeleton in yeast. *Biochemical Society Transactions*, *33*(6), 1247–1249. https://doi.org/10.1042/BST0331247

Mukherjee, D., Sen, A., Boettner, D. R., Fairn, G. D., Schlam, D., Bonilla Valentin, F. J., McCaffery, J. M., Hazbun, T., Staiger, C. J., Grinstein, S., Lemmon, S. K., & Aguilar, R. C. (2013). Bem3, a Cdc42 GTPase-Activating Protein, traffics to an intracellular compartment and recruits the secretory Rab GTPase Sec4 to endomembranes. *Journal of Cell Science*, *126*(Pt 20), 4560–71. https://doi.org/10.1242/jcs.117663

Nadeau, J. H. (2001). Modifier genes in mice and humans. *Nature Reviews Genetics*, *2*(3), 165–174. https://doi.org/10.1038/35056009

Neidhart, J., Szendro, I. G., & Krug, J. (2014). Adaptation in Tunably Rugged Fitness Landscapes: The Rough Mount Fuji Model. *Genetics*, *198*(2), 699–721. https://doi.org/10.1534/GENETICS.114.167668

Nghe, P., Kogenaru, M., & Tans, S. J. (2018). Sign epistasis caused by hierarchy within signalling cascades (2018/04/15). *Nature Communications*, *9*(1), 1451. https://doi.org/10.1038/s41467-018-03644-8

Nicastro, R., Raucci, S., Michel, A. H., Stumpe, M., Osuna, G. M. G., Jaquenoud, M., Kornmann, B., & de Virgilio, C. (2021). Indole-3-acetic acid is a physiological inhibitor of TORC1 in yeast. *PLOS Genetics*, *17*(3), e1009414. https://doi.org/10.1371/JOURNAL.PGEN.1009414

Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T., & Kanemaki, M. (2009). An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nature Methods*, *6*(12), 917–922. https://doi.org/10.1038/nmeth.1401

Papagiannakis, A., de Jonge, J. J., Zhang, Z., & Heinemann, M. (2017). Quantitative characterization of the auxin-inducible degron: a guide for dynamic protein depletion in single yeast cells. *Scientific Reports*, *7*(1), 4704. https://doi.org/10.1038/s41598-017-04791-6

Phillips, P. C. (1998). The Language of Gene Interaction. *Genetics*, *149*(3), 1167–1171. https://doi.org/10.1093/genetics/149.3.1167

Phillips, P. C. (2008). Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Reviews Genetics*, *9*(11), 855–867. https://doi.org/10.1038/nrg2452

Poelwijk, F. J., Socolich, M., & Ranganathan, R. (2019). Learning the pattern of epistasis linking genotype and phenotype in a protein. *Nature Communications*, *10*(1), 4213. https://doi.org/10.1038/s41467-019-12130-8

Poelwijk, F. J., Tănase-Nicola, S., Kiviet, D. J., & Tans, S. J. (2011). Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. *Journal of Theoretical Biology*, *272*(1), 141–144. https://doi.org/10.1016/j.jtbi.2010.12.015

Quintero-Monzon, O., Rodal, A. A., Strokopytov, B., Almo, S. C., & Goode, B. L. (2005). Structural and Functional Dissection of the Abp1 ADFH Actin-binding Domain Reveals Versatile In Vivo Adapter Functions. *Molecular Biology of the Cell*, *16*(7), 3128–3139. https://doi.org/10.1091/mbc.e05-01-0059

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, *29*(1), 24–26. https://doi.org/10.1038/nbt.1754

Sailer, Z. R., & Harms, M. J. (2017). High-order epistasis shapes evolutionary trajectories. *PLOS Computational Biology*, *13*(5), e1005541. https://doi.org/10.1371/JOURNAL.PCBI.1005541

Shetty, A., Reim, N. I., & Winston, F. (2019). Auxin-Inducible Degron System for Depletion of Proteins in Saccharomyces cerevisiae. *Current Protocols in Molecular Biology*, *128*(1), e104. https://doi.org/10.1002/CPMB.104

Sigrist, C. J. A., De Castro, E., Cerutti, L., Cuche, B. A., Hulo, N., Bridge, A., Bougueleret, L., & Xenarios, I. (2013). New and continuing developments at PROSITE. *Nucleic Acids Research*, *41*(D1), D344–D347. https://doi.org/10.1093/NAR/GKS1067

Smidt, O., du Preez, J. C., & Albertyn, J. (2012). Molecular and physiological aspects of alcohol dehydrogenases in the ethanol metabolism of Saccharomyces cerevisiae. *FEMS Yeast Research*, *12*(1), 33–47. https://doi.org/10.1111/j.1567-1364.2011.00760.x

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … Vázquez-Baeza, Y. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Warren, D. T., Andrews, P. D., Gourlay, C. W., & Ayscough, K. R. (2002). Sla1p couples the yeast endocytic machinery to proteins regulating actin dynamics. *Journal of Cell Science*, *115*(8), 1703–1715. https://doi.org/10.1242/JCS.115.8.1703

Weinreich, D. M. (2006). Darwinian Evolution Can Follow Only Very Few Mutational Paths to Fitter Proteins. *Science*, *312*(5770), 111–114. https://doi.org/10.1126/science.1123539

Weinreich, D. M., Watson, R. A., & Chao, L. (2005). PERSPECTIVE:SIGN EPISTASIS AND GENETIC CONSTRAINT ON EVOLUTIONARY TRAJECTORIES. *Evolution*, *59*(6), 1165–1174. https://doi.org/10.1554/04-272

Woods, B., Kuo, C.-C., Wu, C.-F., Zyla, T. R., & Lew, D. J. (2015). Polarity establishment requires localized activation of Cdc42. *Journal of Cell Biology*, *211*(1), 19–26. https://doi.org/10.1083/jcb.201506108

Woods, R., Schneider, D., Winkworth, C. L., Riley, M. A., & Lenski, R. E. (2006). Tests of parallel molecular evolution in a long-term experiment with Escherichia coli. *Proceedings of the National Academy of Sciences*, *103*(24), 9107–9112. https://doi.org/10.1073/pnas.0602917103

Wright, S. (1932). The Roles of Mutation, Inbreeding, crossbreeding and Selection in Evolution. *Proceedings of the XI International Congress of Genetics*, 209–222.

Zarrinpar, A., Park, S.-H., & Lim, W. A. (2003). Optimization of specificity in a cellular protein interaction network by negative selection. *Nature*, *426*(6967), 676–680. https://doi.org/10.1038/nature02178

5.6. Supplement

5.6.1. Supplementary figures

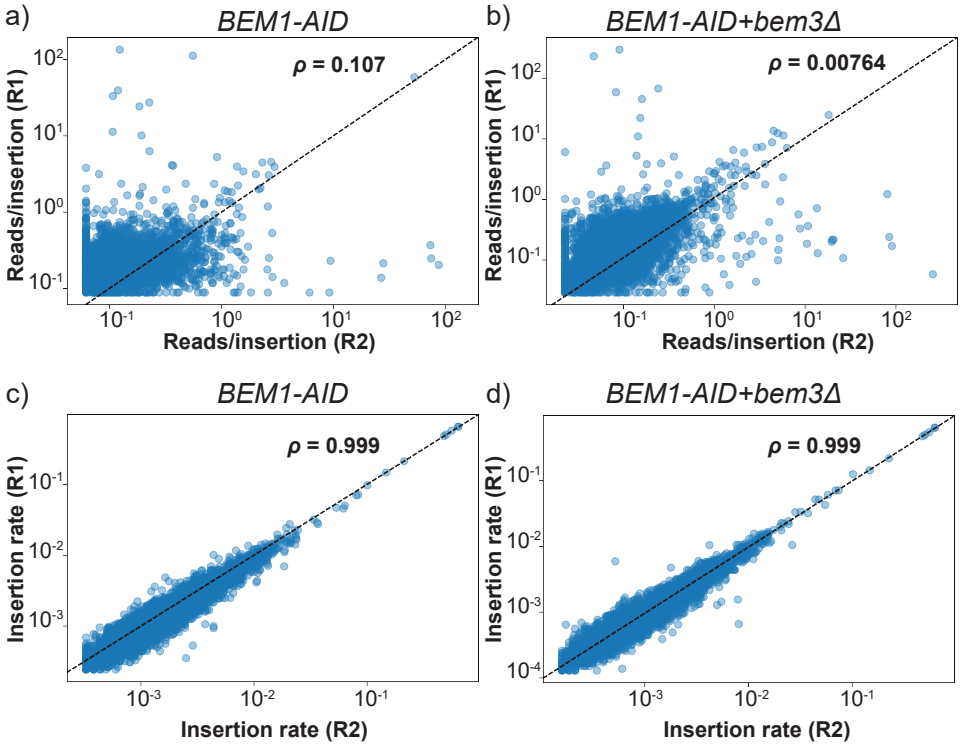


Figure S5.1. Transposon density strongly correlates between replicates while the read counts per transposons insertion site is weakly correlated. (**a-b**) The relation in transposon density of all annotated genes between different replicates of the (**a**) *BEM1-AID* and (**b**) *BEM1-AIDbem3*Δ strains. The plots show a strong correlation between replicates of the same genetic background. (**c-d**) The relation in the average read count per insertion site of all annotated genes between different replicates of the (**c**) *BEM1-AID* and (**d**) *BEM1-AIDbem3*Δ strains. The plots show that the average read count per insertion site is highly variable between replicates of the same strain. Dashed lines represent the identity line. ρ: Pearson's correlation coefficient.

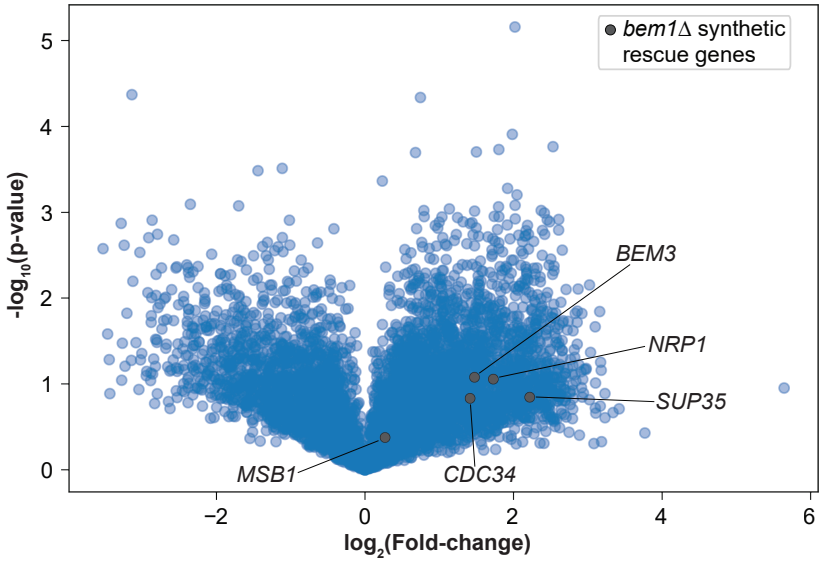


Figure S5.2. Genes that are annotated as synthetic rescue genes are enriched for transposon insertions in *BEM1-AID* strain relative to the wild-type strain. Volcano plot of the $\log_2$ fold-change of gene insertion density (ratio Bem1-AID:WT) against the significance value of the fold change (p-value, student's t-test). The plot shows that genes that have been annotated to rescue the *bem1*Δ phenotype have a higher insertion density in Bem1-AID strains than in wild-type strains. Gene annotations were retrieved from the saccharomyces genome database (SGD).

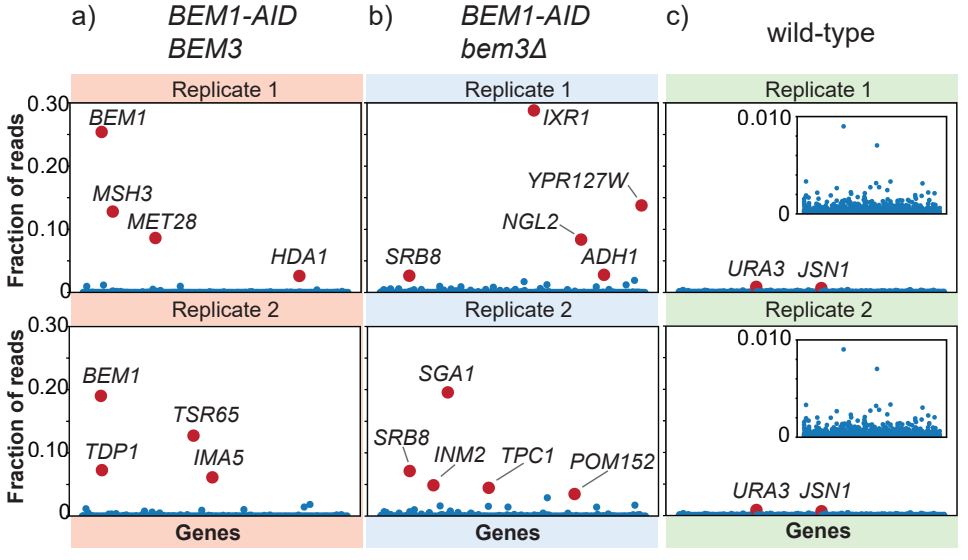


Figure S5.3. Using the AID system to conditionally knockdown Bem1 reduces the resolution of the SATAY screen. The fraction of reads mapping to each annotated gene is shown for two replicates of (**a**) a *BEM1-AID+BEM3* strain, (**b**) a *BEM1-AID+bem3*Δ strain and (**c**) a wild-type strain. Data points that stand out above the average are shown in red. The plots show that in the *BEM1-AID+BEM3* and the *BEM1-AID+bem3*Δ strains a large fraction of the reads is mapped to a small set of 4-5 genes. These genes likely suppress the fitness defects caused by knockdown of Bem1 with the AID system. As a consequence, there is a smaller number of reads available to distinguish between the fitness level of the remaining genes. In the wild-type strains, reads are spread more uniformly across genes.

5.6.2. Supplementary tables

Table 5.1. List of primers used in this chapter

#	Name	Sequence (5'-3')
1	Seq4_Fwd	AAT TAT CCT GGG CAC GAG
2	Seq4_Rev	ACT GTA AGA TTC CGC CAC
3	Bem1-mCherry-FW-AID	ACC TAG TGA ATC TTC CCG ATT ATA TAT CTC GCT C
4	Bem1-mCherry-FW-AID	AGG GAG CCA CAT TAT CCT TTGACA CAT ATG

6

The idea that a few people have about the gene being the target of selection is completely impractical; a gene is never visible to natural selection, and in the genotype, it is always in the context with other genes, and the interaction with those other genes make a particular gene either more favorable or less favorable. [...] In the 30s and 40s, it was widely accepted that genes were the target of selection, because that was the only way they could be made accessible to mathematics, but now we know that it is really the whole genotype of the individual, not the gene. Except for that slight revision, the basic Darwinian theory hasn't changed in the last 50 years.

— Ernst Mayr

Compensatory evolution changes the global structure of the genotype-to-fitness map

Abstract Polarity establishment plays an important role in biological functions that are observed throughout the tree of life. In the budding yeast *Saccharomyces cerevisiae* the formation of a polarized spot of the protein Cdc42 is an essential part of the cell cycle, as it marks the position of new bud formation. Disruption of this process due to the loss of a core polarity protein (Bem1) leads to a severe decrease in viability. Interestingly, these detrimental effects can be almost completely compensated during evolution by the subsequent deletion of two other proteins (Bem3 and Nrp1). The fact that gene deletions are sufficient to restore fitness suggests that the interaction pattern of the remaining proteins has changed in such a way that an alternative pathway for Cdc42 polarization is established. However, how gene deletions mediate these changes and whether it has local or global consequences for the protein interaction network is unknown. In this chapter, we use SATAY, a recently developed transposon mutagenesis screen developed for *S. cerevisiae*, to determine the consequences of compensatory evolution for gene importance on a genome-wide scale. Our results show that the deletion of Bem1, Bem3 and Nrp1 causes changes in the gene disruption tolerance for genes throughout the genome. In addition, these changes affect multiple distinct cellular processes, in some cases making genes related to seemingly vital functions become dispensable. We conclude that compensatory evolution can cause drastic rearrangements in cellular physiology that will likely impact the adaptive response to future genetic perturbations.

6.1. Introduction

Cellular functions are properties that emerge from a dense network of interactions between proteins, DNA, RNA and other small molecules. The advent of genomics and proteomics techniques that facilitate the mapping of these interactions has sparked several efforts towards generating complete maps of biological interaction networks (the interactome) (Barabási & Oltvai, 2004; Fromont-Racine et al., 1997; Yu et al., 2008), with the aim of unraveling how sub-structures within this network are related to cellular functions (Dunn et al., 2005; Gavin et al., 2002; Girvan & Newman, 2002; Han, 2008; Hartwell et al., 1999). However, biological networks are dynamic entities that can dramatically change their structure and functional relations over evolutionary timescales (Ghadie et al., 2018). For example, a study comparing the interactomes of two related yeast species found that the interaction partners of orthologous proteins were often not conserved across species, suggesting that each had adapted to their own ecological niche by rewiring their protein interaction network (Das et al., 2013). On shorter evolutionary timescales it has been shown that network rewiring plays a role in the acquired resistance of tumor cells to therapeutic intervention (Komurov et al., 2012), allowing cells to proliferate despite the inhibition of a protein that was deemed to be essential for survival.

Despite the importance of network rewiring for adaptive evolution, relatively little is known about how adaptive mutations change the structure of biological networks. Specifically, it is

unclear to what degree the effects of network rewiring remain contained within (sub-)modules of the network or if it changes the global topology. In addition, whether the network topology or functional relations between genes dominates the dynamics of rewiring still needs to be determined. An attractive cellular function to study the dynamics of biological networks during evolution is cell polarity. Nearly all cells possess the ability to polarize (Arkowitz & Iglesias, 2008; Thompson, 2013), which demonstrates that it is a strongly conserved function throughout evolution. At the same time, the composition of the protein network that regulates cell polarity has been found to be highly dynamic on evolutionary timescales (Diepeveen et al., 2018; Thompson, 2013). This combination of functional conservation and structural variability indicates that polarity establishment can be achieved through several degenerate pathways (Goryachev & Leda, 2017). Interestingly, a recent study in *Saccharomyces cerevisiae* has shown that the fitness effects of deleting a core polarity protein can be nearly completely compensated during evolution through a pathway consisting of only gene deletions (Laan et al., 2015). The fact that compensation can be achieved through gene deletions only and does not require mutations that change protein function suggests that the gene deletions promote the manifestation of an alternative pathway for polarity establishment by inducing changes in the interactome.

Because *S. cerevisiae* proliferates exclusively through asymmetric cell division (budding), polarity establishment is an essential part of the cell cycle and failure to polarize leads to cell death (Chiou et al., 2017; Laan et al., 2015; Martin & Arkowitz, 2014). As in nearly all eukaryotic species, its polarity network is centered around the cycling of the small GTPase Cdc42 between an inactive GDP-bound state and an active GTP-bound state (Chiou et al., 2017; Etienne-Manneville, 2004; Martin & Arkowitz, 2014). Only the active state of Cdc42 is able to signal to downstream effectors to induce an assymetric distribution of cellular constituents and the formation of a single polarized cap of active Cdc42 on the plasma membrane is therefore a prerequisite for bud formation. Mathematical models have proposed that a mechanistic requirement for the formation of a single polar cap is the existence of a non-linear positive feedback loop that drives Cdc42 activation (Freisinger et al., 2013; Goryachev & Leda, 2017; Klünder et al., 2013). Currently, the pathway considered to be the major contributor to this positive feedback loop relies on the scaffold protein Bem1 (Freisinger et al., 2013; Klünder et al., 2013), which simultaneously binds Cdc42 and its activator, the guanine nucleotide exchange factor (GEF) Cdc24 (Bose et al., 2001; Zheng et al., 1995). This causes clusters of active Cdc42 to contain enriched concentrations of Cdc24, locally increasing the rate of Cdc42 activation in a manner that correlates with cluster size (Irazoqui et al., 2003; Klünder et al., 2013; Kozubowski et al., 2008). Although the axis of polarity generated by this mechanism is normally directed by spatial landmarks, it is sufficient to drive polarity establishment at a random location when spatial cues are absent (symmetry breaking). The central role of Bem1 for Cdc42 polarization is supported by observations that the deletion of Bem1 strongly diminishes the ability of cells to initiate bud formation (Bendert & Pringle, 1991; Chenevert et al., 1992). Furthermore, local recruitment of Bem1 to a predetermined site on the plasma membrane using optogenetics is sufficient to direct the position of bud formation (Witte et al., 2017). In addition to Bem1-mediated positive feedback, a second pathway based on actin based delivery of vesicles carrying Cdc42-GTP to the bud site has been suggested to act in parallel and contribute to cell polarization (Slaughter et al., 2009; Slaughter et al., 2013; Wedlich-Soldner et al., 2003). In this alternative pathway, clusters of active Cdc42 would increase the local rate of Cdc42-GTP delivery via actin-based vesicle transport by promoting the nucleation of actin filaments (Wedlich-Soldner et al., 2003). However, recent theoretical models have proposed that

actin based delivery may antagonize rather than enhance the formation of the polar cap using the argument that the concentration of Cdc42-GTP on transport vesicles is too low compared to the concentration at the polar cap (Layton et al., 2011; Savage et al., 2012; Woods et al., 2016).

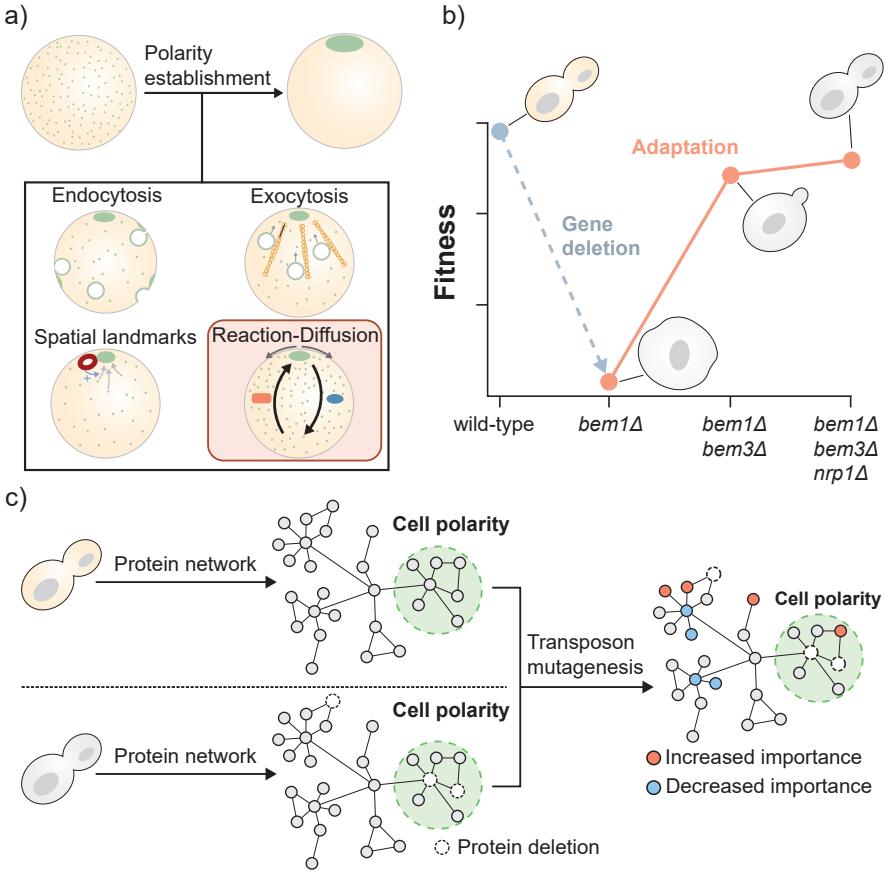


Figure 6.1. Investigating the consequences of compensatory evolution for the genotype-fitness map. (**a**) Polarity establishment involves the formation of an anisotropic distribution of Cdc42, which is shown here in green. In budding yeast, several pathways have been proposed to contribute to the polarization of Cdc42 (endo-/exocytosis, spatial landmarks and reaction diffusion). In this chapter, we study the effects of compensatory evolution in response to a strong genetic perturbation of the reaction-diffusion pathway. (**b**) Deletion of the polarity protein Bem1 can be compensated by additionally deleting Bem3 and Nrp1 in a conserved order. The result is a polarity mutant (*bem1*Δ*bem3*Δ*nrp1*Δ) that has a fitness nearly equal to that of a wild-type strain. (**c**) We determine the effects of compensatory evolution by comparing the disruption tolerance of all genes on a genome-wide scale in the wild-type strain and the polarity mutant using transposon mutagenesis. This allows us to identify the set of proteins that have an increased or decreased importance for fitness in the polarity mutant.

In this chapter, we investigate the genome-wide consequences of compensatory evolution of the polarity network (Figure 6.1). To do so, we focus on the reproducible evolutionary trajectory mentioned above, in which the loss of Bem1 is compensated by loss-of-function mutations in Bem3 and Nrp1 (Laan et al., 2015). Using the saturated transposon analysis in yeast (SATAY) (Michel et al., 2017), we determine which genes display an altered tolerance to transposon insertions after all three proteins (Bem1, Bem3 and Nrp1) are deleted. Surprisingly, we find that

the compensatory mutations impact a large number of genes and that these genes are related to a variety of cellular processes, suggesting extensive interactome rewiring. We conclude that mutations that compensate the defects of a single function can have global consequences for cellular physiology.

6.2. Results

6.2.1. Compensatory evolution of the polarity module leads to global changes in gene disruption tolerance

The deletion of Bem3 and Nrp1 is sufficient for the nearly complete compensation of the defects in polarity establishment caused by the loss of Bem1 (Laan et al., 2015). The fact that this compensation can be achieved through gene deletions only suggests that the remaining proteins have an altered interaction pattern such that it allows cell polarity to be established through a degenerate pathway. How the gene deletions mediate these changes in the interaction pattern and whether it induces local or global changes in the interaction network is unknown.

To objectively search for genes that have changed their physiological role after compensatory evolution, we performed a SATAY screen in the wild-type and *bem1*Δ*bem3*Δ*nrp1*Δ genetic backgrounds. We used the rationale that genes that become associated with a new function after the deletion of Bem1, Bem3 and Nrp1 are likely to have a different effect on fitness when deleted in the two genetic backgrounds. Thus, changes in the transposon insertion pattern should allow us to identify genes that play a role in the degenerate pathway for polarity establishment. In order to resolve smaller changes in fitness, we used the total read count rather than insertion density. However, we found that the raw read counts show substantially more variability between replicate experiments than the insertion density (see S6.1), an effect that has been described to be caused by artificial spikes in the read counts (DeJesus & Ioerger, 2013, 2016; Parekh et al., 2016). To reduce the impact of these spikes on our analysis, we implemented the non-linear normalization method proposed by DeJesus and Ioerger (2016). This normalization substantially improved the correlation between replicate datasets (see section 6.6.1).

A pairwise comparison between six wild-type SATAY libraries and six *bem1*Δ*bem3*Δ*nrp1*Δ libraries revealed significant ($p_{adj} < 0.05$) differences in the fitness impact of 882 genes between the two genetic backgrounds (Figure 6.2 and Figure S6.3). Importantly, applying the same significance threshold to compare datasets within the same genetic background did not yield any significantly different genes. Interestingly, among the identified genes, 468 exhibited increased tolerance to disruptions in the polarity mutant, while 414 displayed reduced tolerance relative to the wild-type strain (Figure 6.2c).

Collectively, these findings demonstrate that compensatory evolution within the polarity network affects the fitness impact of genes throughout the genome. Notably, while some of these genes (such as *SPA2*, *BNI1*, and *AXL2*) are associated with polarity establishment, others (such as *DYN1*, *CWP2*, and *MEC1*) lack clear connections to cell polarity, indicating global effects on cell physiology. Furthermore, the observation that nearly equal numbers of genes become less or more important for survival suggests that compensatory evolution reshapes the relative importance of cellular processes rather than causing to a general decline in mutational robustness.

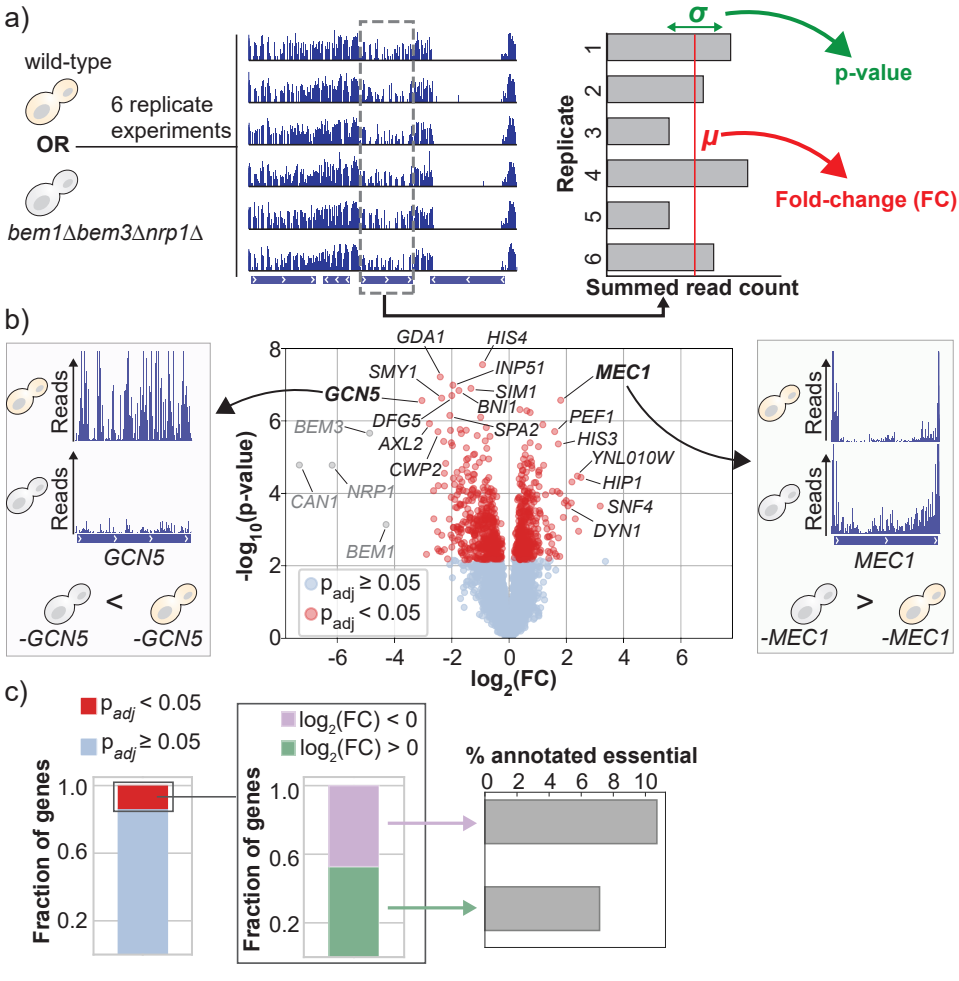


Figure 6.2. Adaptation of the polarity network globally changes the fitness contribution of genes. (**a**) To determine changes in the effect of gene disruptions on fitness between a wild-type and polarity mutant, we generated six replicate SATAY datasets per strain. For each strain, reads that mapped to the same gene were summed and the mean and variance of this summed read count across replicates was determined. The mean and variance were respectively used to calculate a fold change and p-value for the difference in fitness effect between the two genetic backgrounds. (**b**) Volcano plot of the p-value and log fold-change (FC) for all annotated genes in the yeast genome when comparing the wild-type to the polarity mutant. A positive log fold-change indicates that disruption of the gene is less detrimental in the wild-type than in the polarity mutant, a negative log fold-change indicates the reverse. P-values were determined using Welch's t-test and corrected for multiple hypothesis testing with the Benjamini-Hochberg procedure. Genes with a p-value below the significance threshold ($p_{adj} < 0.05$) are annotated in red. Grayed out data points correspond to genes that were deleted in the polarity mutant genes are the differences between the genetic backgrounds of the two strains. (**c**) Approximately 14% of the genes have a significantly altered tolerance to transposon disruptions between the wild-type strain and the polarity mutant. Of these significant genes, a nearly equal ratio has a negative and positive log fold-change. Slightly more genes with a negative log-fold change have been annotated as essential than genes with a positive log-fold change.

6.2.2. Network centralities do not correlate with the probability of altered gene importance

The large number of genes that contribute differently to fitness between the wild-type strain and the polarity mutant suggests that the deletion of Bem1, Bem3 and Nrp1 causes extensive rewiring of the protein interaction network. Interestingly, the ability of a protein to participate in new interactions during evolution has been linked to its topological role within the protein interaction (PPI) network (Alvarez-Ponce et al., 2017; Fraser & Hirsh, 2004; Fraser et al., 2002; Joy et al., 2005). For example, proteins that interact with many other proteins have been found to be more likely to be essential for survival than those with fewer interactions, a phenomenon known as the *lethality-centrality* rule (Jeong et al., 2001). As essential genes are often conserved throughout evolution (Giaever et al., 2002; Winzeler et al., 1999), this implies a negative correlation between the rate of protein evolution and the number interactions the protein engages in (Fraser & Hirsh, 2004; Fraser et al., 2002). Based on this proposed link between the adaptive potential of a protein and its position within the PPI network, one hypothesis is that a similar relation might exist between PPI structure and the probability that a gene deletion has a different fitness effect in the two genetic backgrounds. The existence of such a relation would indicate that it is possible to build a predictive model of interactome rewiring during compensatory evolution based on mapped protein interactions.

To identify a possible relation, we constructed a PPI network for the proteome of *S. cerevisiae* based on the physical interactions annotated in the BioGRID database. This resulted in an undirected network containing 3799 proteins (nodes) and 17205 interactions (edges). Although this network is incomplete (only about 65% of an estimated total of 5800 proteins are included in the network (Mackiewicz et al., 2002)), we were able to validate the lethality-centrality rule in this network based on a positive correlation between the number of interactions a protein has and the increased likelihood that the protein is essential (Figure S6.4). The existence of the centrality-lethality rule in the PPI network indicates that important topological properties of the network are preserved, despite that it contains only a subset of all known proteins in *S. cerevisiae*.

There are several other measures that can be used to describe the role of a protein with respect to different structural features of the PPI network. These measures are generically known as *centralities*, as they describe how central (important) each node is for a given structural component of the network (Koutrouli et al., 2020). For example, nodes with a high degree are often described to be important for the connectivity of the network and based on this feature one may expect that the loss of a high degree node has a strong detrimental effect on fitness. Alternatively, their ability to interact with many other proteins may facilitate their interaction with novel partners during network rewiring.

Here, we focus on three widely used centralities: degree, betweenness and closeness. *Degree centrality* ranks nodes based on their number of direct physical interactors. *Betweenness centrality* describes whether a node acts as a bridge between groups of nodes and therefore mediates the flow of information between them. The loss of nodes with a high betweenness disrupts this flow of information between communities in the network. Finally, *closeness centrality* relates to the ability of a node to quickly communicate with many other nodes in the network. In contrast to degree, closeness is not limited to the immediate neighbours of a node, but is instead based on the average distance from a given node to all other nodes in the network. Here, distance is defined as the number of edges that need to be traversed to in order to reach a second node in the network. The distribution of these centralities in our PPI network is shown in Figure 6.3

We looked for a relation between these three parameters and the likelihood that a protein is encoded by a gene that has a different importance between the wild-type strain and the polarity mutant. None of the three centralities showed a correlation with the likelihood that a gene has a changed effect on fitness. Instead, we found that this likelihood was more or less uniformly distributed across degree, betweenness and closeness centrality values (figure 6.4). Similarly, we did not find a relationship between the values of these centralities and the effect size of changes in gene disruption tolerance.

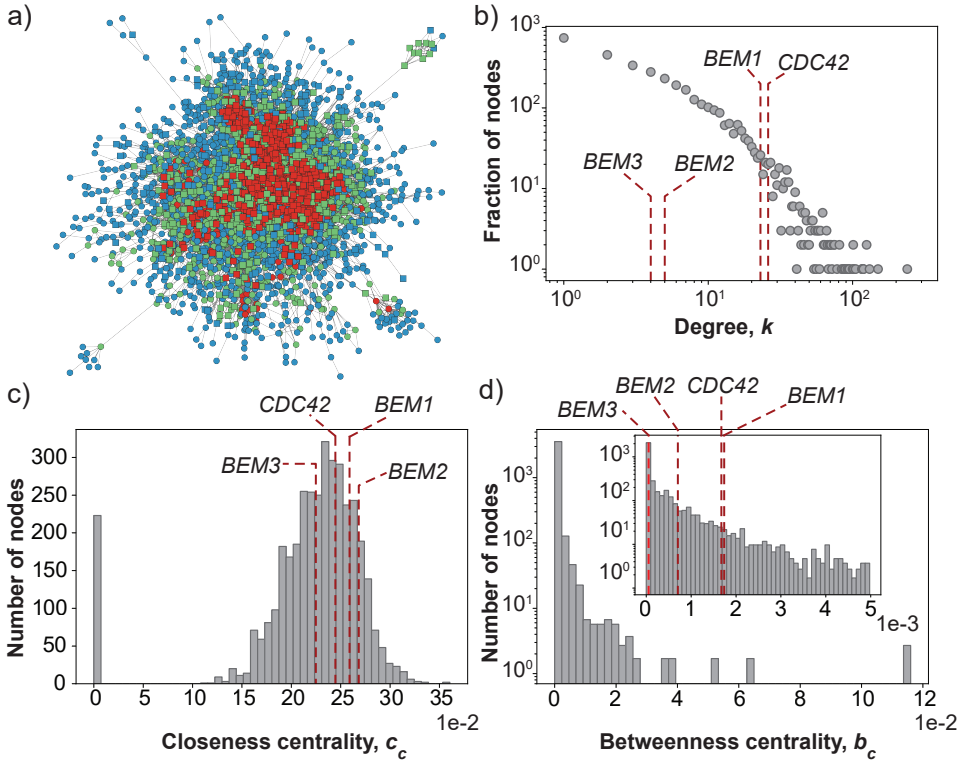


Figure 6.3. Centralities in the protein-protein interaction network of *S. cerevisiae*. (**a**) Depiction of the multi-validated protein-protein interaction network (PPI). Nodes are colored according to their degree k (blue: $k \leq 3$, green: $4 \leq k \leq 10$, red: $k > 10$). Essential genes are shown as square nodes, non-essential genes are shown as circles. For clarity, only nodes that are connected to the main network are shown. The complete network can be found in Figure S6.4. (**b-d**) Distributions of the (**b**) degree, (**c**) closeness and (**d**) betweenness centralities of the PPI network. The centrality values of the two deleted proteins, Bem1 and Bem3, and the master regulator for cell polarity, Cdc42, are indicated by red dashed lines. The value of Nrp1 is not shown because Nrp1 is absent in the multi-validated dataset, which is likely due to the fact that Nrp1 is an understudied protein.

These results suggest that, at least for the interaction network used here, centrality measures of the PPI are uninformative for the prediction whether a gene will attain an altered role after compensatory evolution.

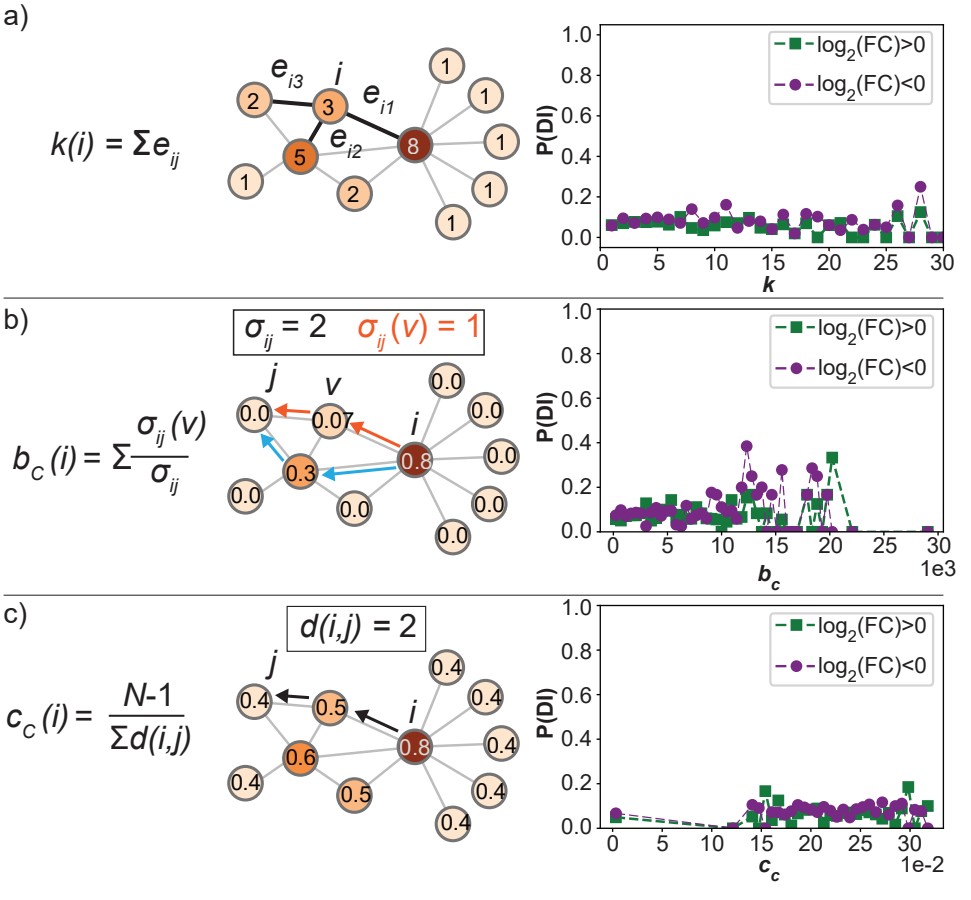


Figure 6.4. Centrality measures in the protein-protein interaction network do not correlate with an altered tolerance to gene disruption. (**a**) The degree centrality k is determined by counting the number of edges e_{ij} a node has connecting it to other nodes in the network. The example graph shows a node i with a degree of three. The plot shows the conditional likelihood P(DI) that a node with degree k has an altered disruption tolerance. (**b**) The betweenness centrality b_c provides a measure for the importance of a node for the information flow in the network based on the number of shortest paths that pass through that node. In the example there are two shortest paths σ_{ij} from node i to node j, but only one of these paths passes through node v. The plot shows the conditional likelihood P(DI) that a node with betweenness b_c has an altered disruption tolerance. (**c**) The closeness centrality reflects the distance of a node to all other N nodes in the network based on the average shortest path $d(i,j)$. The example graph shows a node i with a shortest pathlength of two to node j. The plot shows the conditional likelihood P(DI) that a node with closeness c_c has an altered disruption tolerance. For all centralities shown in (a-c), the distribution of P(DI) is approximately uniform.

6.2.3. Compensatory evolution affects a broad range of cellular processes

We had already found that compensatory evolution of the polarity network affects genes related to other processes than polarity establishment (figure 6.2). However, it is still unclear whether the observed changes reflect a shift in importance at the level of a gene's individual function or rather at the level of the pathway in which the gene product acts. If the latter case is true, we expect to find clusters of functionally related genes in our dataset, as the loss of any of the proteins that act cooperatively in the same pathway will perturb the same process. Alternatively, in the former case most genes should be functionally unrelated.

To distinguish between these two cases, we constructed a functional interaction map of the genes that were found to have a significantly different tolerance to disruption between the wild-type strain and the polarity mutant. To construct this functional map, we used the interactions available from the STRING database (Doncheva et al., 2019; Szklarczyk et al., 2019; Szklarczyk et al., 2021). The STRING database infers functional associations between proteins by integrating information from several sources, which includes text mining, annotated molecular complexes and computationally predicted interactions. As such, the interactions obtained from this database are based on the notion that functionally related proteins do not necessarily interact physically, making it distinct from a PPI network. Moreover, the database scores each interaction based on the quality of the evidence, allowing more weight to be given to high-confidence interactions during downstream processing steps.

We found that the functional interaction map was represented by a single connected network with only few disconnected nodes. This demonstrates that the vast majority of the genes that are affected by compensatory evolution are functionally related. Interestingly, plotting the degree distribution of the connected part of the network revealed that the distribution approximately follows a power law, suggesting that the network is scale free. A hallmark of scale-free networks is that they consist of smaller sub-graphs that are connected to each other through hubs. We therefore expected that the connected network could be fragmented into smaller clusters that may reveal the relevant cellular process association of each gene.

To identify these clusters, we ran the Markov Cluster (MCL) algorithm (Van Dongen, 2008) on the functional association network using the STRING interaction score as weights. This resulted in 120 clusters, of which 49 contained three or more genes. Because we are unable to discuss the biological meaning of all these clusters, we focus on the seven largest clusters as these likely represent the major pathways affected by compensatory evolution. We performed an enrichment analysis of the biological process gene ontology (GO) terms to see if we could identify the process related to each cluster. All seven clusters showed a a clear enrichment and none of the clusters were enriched for the same GO term. Five clusters consisted of genes related to cellular homeostasis (translation, metabolism, signalling, chromatin remodeling and ribosome biogeneses), while two consisted of genes related to specific cellular functions (cell polarity and microtubule dynamics). Four of the five clusters related to cellular homeostasis (the signalling cluster was the exception) predominantly contained genes that were less tolerant to disruption in the polarity mutant. This could indicate that the loss of Bem1, Bem3 and Nrp1 causes an intolerance to fluctuations in metabolites and protein copy numbers. The same predominance was present in the cluster related to cell polarity, which suggests that compensatory evolution through gene deletions causes an increased dependence on pathways that may act in parallel to the one perturbed (discussed in the next section). Surprisingly, all genes in the cluster enriched for microtubule dynamics are more tolerant to disruption in the polarity mutant than in the

wild-type strain. Thus, dynamic microtubules appear to be relatively dispensable in the polarity mutant, which is unexpected considering their role in vital processes such as organelle mobility and spindle positioning.

In conclusion, these results show that many of the genes affected by compensatory evolution after the loss of Bem1 are functionally associated, indicating that changes are induced on the level of pathways rather than individual genes. However, only a minority of the affected genes can be directly related to polarity establishment, which demonstrates that these changes go beyond the originally perturbed pathway.

6.2.4. Identification of an alternative pathway for polarity establishment

The discovery that the set of genes affected by the compensatory evolution of the polarity network contained a large cluster associated with polarity establishment (figure 6.5c) prompted us to analyze this cluster in more detail. In particular, we considered it likely that the genes that were originally associated with polarity establishment play an important role in forming an alternative pathway for polarity after the loss of Bem1, Bem3 and Nrp1. We therefore manually annotated the molecular complex or pathway of each gene and found that most could be assigned to one of the following: the axial landmark, the polarisome and the endocytic pathway. In the sections below, we give an overview of how each of these complexes and pathways relate to the genes affected by compensatory evolution and how they might promote polarity establishment in the absence of a Bem1-mediated feedback loop.

The axial landmark

In wild-type haploid yeast cells, budding occurs next to, but not at, the previous division site. The structure that directs bud formation towards this location is a molecular complex consisting of four proteins (Bud3, Bud4, Axl1 and Axl2) known as the axial landmark (Chant, 1995). These four proteins assemble into a ring-like structure on the membrane during the M phase of the cell cycle, guided by the remnants of the septin ring of the previous budding event. The majority of this ring (Bud3, Bud4 and Axl1) forms only transiently during the M phase and early G1, but a patch of Axl2 remains present on the membrane throughout the cell cycle (Gao et al., 2007; Kang et al., 2012). Axl2 has been proposed to play a key role in providing the spatial cue for axial budding based on observations that the presence of Axl2, but not Bud3, Bud4 or Axl1, is a prerequisite for an axial budding pattern (Fujita et al., 2004; Kang et al., 2004). The other three proteins of the axial landmark complex have instead been implied to serve as regulators for Axl2 localization and activity (Gao et al., 2007; Kang et al., 2012).

Recent work has shown that Cdc42 polarization induced by landmark proteins occurs in two phases of the cell cycle: in early G1 before START (as marked by the nuclear exit of Whi5) and in late G1 after START (Miller et al., 2020; Moran et al., 2019). These two phases of Cdc42 polarization appear to be required for the complete assembly of the components necessary for cell division (Kang et al., 2018; Lai et al., 2018; Miller et al., 2020). Interestingly, it has been suggested that distinct positive feedback loops drive Cdc42 polarization towards the axial landmark before and after START (Moran et al., 2019). In both phases of polarity establishment, Axl2 connects the axial landmark complex to the polarization machinery by recruiting Bud5, a GEF which activates the small GTPase Rsr1. Active Rsr1-GTP in turn induces polarity establishment through either of two pathways depending on whether the signal for START has been passed. During pre-START polarization, Rsr1-GTP binds to Cdc42-GDP. This interaction mediates the conversion of Cdc42-

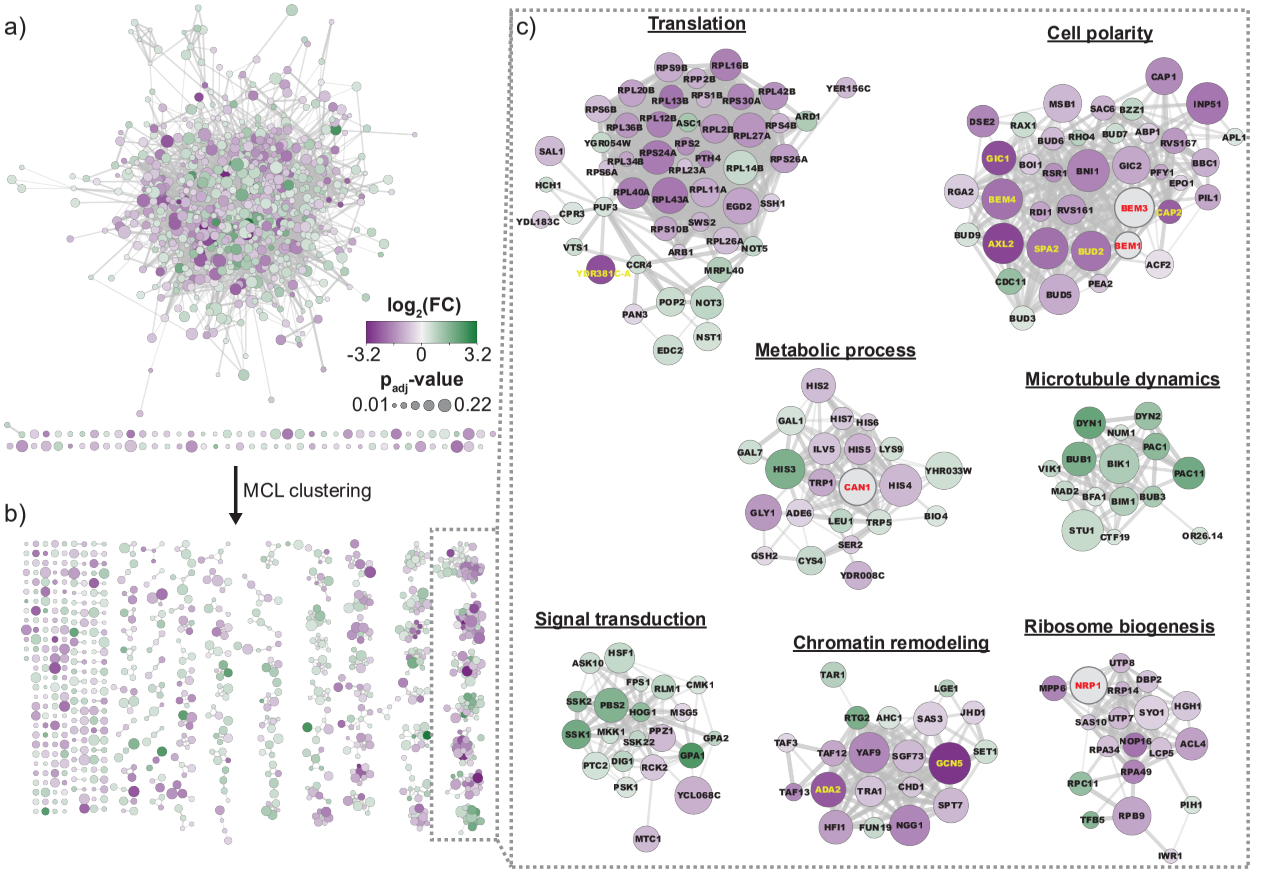


Figure 6.5. Genes with an altered fitness contribution can be clustered according to their biological process. Analysis of the STRING functional associations between genes that have a significantly different tolerance to gene disruption in the *bem1*Δ*bem3*Δ*nrp1*Δ and wild-type strain. (**a**) Graph of the functional associations between genes that have a different effect on fitness when deleted in the wild-type strain and the polarity mutant. Node color scales with the magnitude and sign of the fitness effect, node size scales according to the significance level (adjusted p-value). The graph shows that most genes are contained in a single, densely connected network. (**b**) The clusters formed from the graph shown in (a) by the Markov clustering algorithm (MCL). The clusters are arranged from large (top right) to small (bottom left). The clustering result reveals that the functional association network contains several sub-structures. (**c**) Gene ontology enrichment for the seven largest clusters identified by the Markov clustering algorithm. The second largest cluster is enriched for genes related to polarity establishment, which suggests that known pathways play a role in the functional recovery of polarity establishment in the polarity mutant.

GDP into active Cdc42-GTP by the landmark component Bud3, which has been reported to show GEF activity for Cdc42 in addition to its role in landmark assembly (Kang et al., 2014). The requirement for Bud3 to act as a GEF for Cdc42 appears to be supported by reports that Cdc24 is sequestered in the nucleus prior to START and is therefore unable to participate in Cdc42 activation (Nern & Arkowitz, 2000; Shimada et al., 2000). Notice that the functionality of this pathway would not require Bem1 (Miller et al., 2020; Witte et al., 2017). Alternatively, after START has been passed and Cdc24 is released from the nucleus, Rsr1-GTP binds both Cdc42-GDP and Cdc24 to initialize the Bem1-mediated feedback loop for Cdc42 polarization and direct the polarity axis towards the axial landmark.

Our data shows that the deletion of Axl2 in the polarity mutant has a profoundly more negative effect on fitness than in a wild-type strain, indicating that the communication between this landmark protein and the Cdc42 polarity module plays an important role during bud formation in the polarity mutant. This is further supported by the similar trend for components of the Rsr1-GTPase module (the GAP Bud2, the GEF Bud5 and the GTPase Rsr1), which functionally connects Axl2 to Cdc42 polarity. In addition, we find that the deletion of Gic1 and Gic2 have an increased negative effect on fitness. Gic1 and Gic2 were initially considered to be Cdc42 effectors, but have lately also been implied to play a role upstream of Cdc42 during polarity establishment by reducing the diffusion rate of Cdc42-GTP on the membrane (Kang et al., 2018; Miller et al., 2020), thereby stabilizing the polar cap. It has been proposed that this stabilizing effect of Gic1 and Gic2 relies on their interaction with PIP2 on the plasma membrane (Kang et al., 2018). Consistent with this, we find that Inp51, which regulates the PIP2 levels, becomes more important in the *bem1*Δ*bem3*Δ*nrp1*Δ background.

At first sight, this pattern appears to suggest that pre-START polarization, which is mediated by the GEF activity of Bud3, is capable of promoting post-START polarization and bud emergence in the polarity mutant. This would allow the cells to surpass the requirement of Bem1 activity for efficient polarity establishment. However, if polarity establishment in *bem1*Δ*bem3*Δ*nrp1*Δ cells would be purely based on the redundancy of the pre- and post-START polarization mechanisms, we would expect to see a strong switch in the essentiality of Bud3 (from non-essential to essential) and Cdc24 (from essential to non-essential). Instead, we find that the tolerance to disruptions in the *BUD3* is even slightly increased, while *CDC24* appears to be essential in both the wild-type strain and the polarity mutant. Thus, the mechanism for polarity establishment in *bem1*Δ*bem3*Δ*nrp1*Δ cells remains dependent on Cdc24 to act as a GEF for Cdc42.

The polarisome

The polarisome is a molecular complex that forms at the incipient bud site. It is generally considered to consist of seven components (Spa2, Pea2, Bni1, Bud6, Aip5, Msb3 and Msb4), although the role of Msb3 and Msb4 in the formation of this complex is still unknown (Xie & Miao, 2021). Localization of the polarisome during bud formation is dependent on active Cd42 as the Cdc42-effector Gic2 recruits the components Bud6, Bni1 and Spa2 to the incipient bud site (Chen et al., 2012; Jaquenoud & Peter, 2000). At the bud site, Spa2 acts as a scaffold and mediates the assembly of the complex by binding different components of the polarisome (Sheu et al., 1998). Within the polarisome, the proteins Bni1, Bud6 and Aip5 form the nucleation-core that promotes the assembly of F-actin cables towards the bud site (Xie & Miao, 2021). Thus, formation of the polarisome is crucial for the directed transport of membrane components towards the bud site and hence for polarized growth.

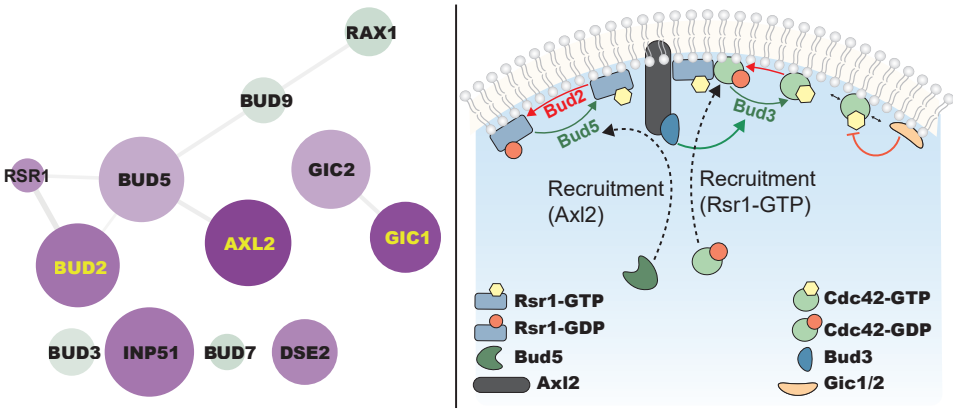


Figure 6.6. Genes coding for proteins related to the spatial landmark proteins become important for fitness in the polarity mutant. (Left) The physical subnetwork of proteins related to the spatial landmark for which the corresponding gene was found to have an altered contribution to fitness between the wild-type strain and polarity mutant. As in figure 6.5, node color scales with the magnitude and sign of the fitness effect, node size scales according to the significance level (adjusted p-value). The graph shows that most genes coding for axial landmark proteins (Axl2), proteins of the Rsr1-GTPase module (Bud2, Bud5, Rsr1) and interacting proteins (Gic1, Gic1) become less tolerant to disruption in the polarity mutant. In contrast, components of the bipolar landmark (Rax1, Bud9, Bud7) become more tolerant to disruption. (Right) Illustration of the currently proposed mechanism through which the axial landmark can promote Cdc42 polarization.

We find that several core components of the polarisome (Spa2, Bni1, Pea2, Bud6) appear as having a decreased tolerance to disruption of their corresponding gene in our dataset. As the deletion of Bem3 and Nrp1 was originally found to restore the defect in polarity establishment in *bem1*Δ cells (Laan et al., 2015), we speculate that the increased importance of the polarisome in *bem1*Δ*bem3*Δ*nrp1*Δ cells is due to a contribution of actin-mediated transport to Cdc42 polarization rather than downstream processes. Thus, while the role of actin-mediated transport has been disputed, our results suggest that vesicle delivery of active Cdc42 to the polarity cap makes a significant contribution to polarity establishment in the absence of Bem1, Bem3 and Nrp1.

The endocytic pathway

Endocytosis occurs through a dense network of short actin cables that generate the force required for invagination of the plasma membrane (Goode et al., 2015). The formation and dynamics of these actin patches is regulated by the Arp2/3 complex (D'Agostino & Goode, 2005; Mishra et al., 2014). During polarity establishment, active Cdc42 initiates a cascade that recruits and activates the Arp2/3 complex, resulting in the increased formation of actin patches near the site of polarized growth (Lechler et al., 2001). The role of endocytosis in polarity establishment is currently unclear, as it has been proposed to be involved in both positive and negative feedback loops (Harris & Tepass, 2010; Irazoqui et al., 2005; Jose et al., 2013).

Many of the genes that belong to the cell polarity cluster are related to the regulation of actin, but several proteins stand out as specific components of actin patches and the endocytic machinery. First of all, we find that the two subunits of the capping protein heterodimer (Cap1 and Cap2) become increasingly important in the *bem1*Δ*bem3*Δ*nrp1*Δ mutant. Capping proteins are required for inhibiting the growth of actin filaments to allow the formation of a dense structure of short actin filaments that make up the actin patch (K. Kim et al., 2004; Young et al., 2004).

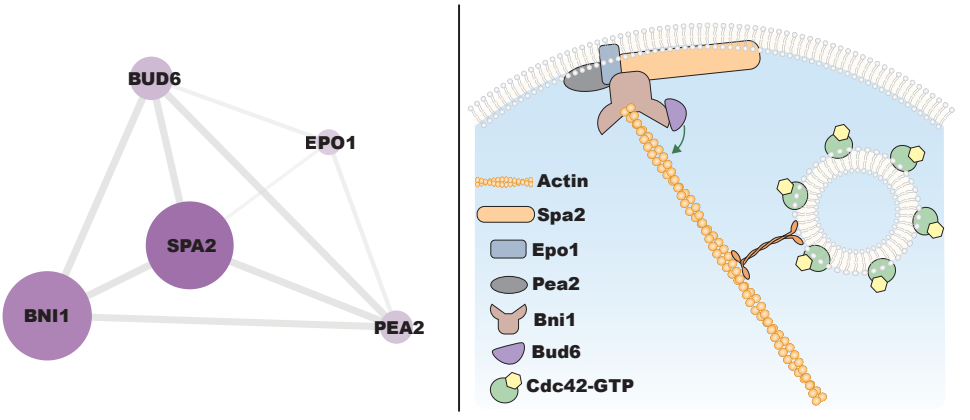


Figure 6.7. Proteins of the polarisome become more important for fitness in the polarity mutant. (left) The physical subnetwork of polarisome proteins for which the corresponding gene was found to have an altered contribution to fitness between the wild-type strain and polarity mutant. As in figure 6.5, node color scales with the magnitude and sign of the fitness effect, node size scales according to the significance level (adjusted p-value). (Right) Illustration of the interactions between the polarisome proteins shown in the figure on the left. Actin cables and myosin have been added to emphasize the role of the polarisome in directing actin-mediated transport towards the polar site.

Second, we see an increased importance of the BAR-domain containing amphiphysin proteins Rvs161 and Rvs167. These proteins stabilize the curvature of invaginated membranes and promote their scission during endocytosis. The deletion of Rvs161 or Rvs167 has been shown to increase the rate of failure to create endocytic veiscles (Kaksonen et al., 2005). Lastly, we see the same effect for the adapter protein Abp1 that connects actin patches to the endocytic machinery (Quintero-Monzon et al., 2005). Taken together, this pattern of increased essentiality of proteins involved in actin patch formation and endocytosis indicates that endocytosis makes an important contribution to polarity establishment. Whether this is due to its involvement in a negative or positive feedback loop for Cdc42 polarization remains to be determined.

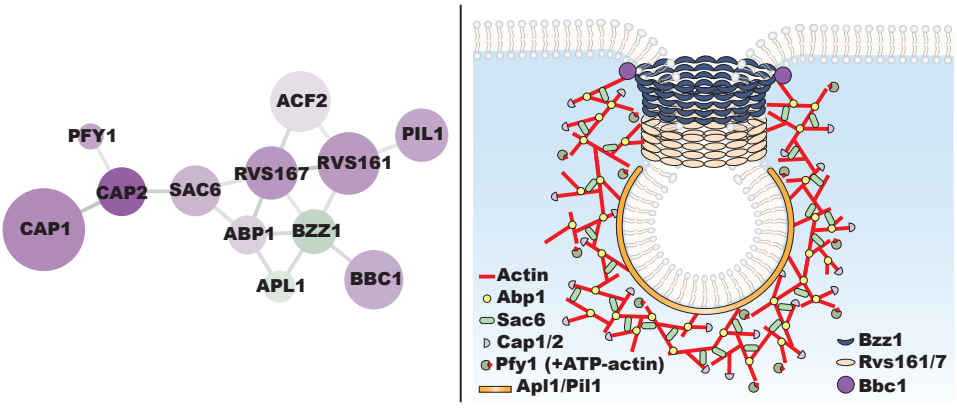


Figure 6.8. Endocytic proteins become more important for fitness in the polarity mutant. (left) The physical subnetwork of the proteins involved in endocytosis for which the corresponding gene was found to have an altered contribution to fitness between the wild-type strain and polarity mutant. As in figure 6.5, node color scales with the magnitude and sign of the fitness effect, node size scales according to the significance level (adjusted p-value). (Right) Illustration of the interactions between the proteins related to endocytosis shown in the figure on the left.

6.3. Conclusions

The ability of cells to compensate for the loss of proteins by changing their interactome plays an important role in their ability to quickly adapt to genetic perturbations (Ding et al., 2018; Halaoui & McCaffrey, 2015; J. Kim et al., 2012). Understanding how cellular functions can be restored without restoring the perturbed pathway has important applications in healthcare. For example, microbial cells can become resistant to antibiotics by rewiring their interaction network in such a way that an alternative pathway for the perturbed function is formed (Grézal et al., 2023; Taylor et al., 2022).

In this chapter, we studied the genome-wide changes that occur at the end of an adaptive pathway that compensates for the loss of the polarity protein Bem1 through additional gene deletions only. We find that a surprisingly large number of genes (882) are significantly affected by the gene deletions, which demonstrates that compensatory evolution causes global changes in the importance of genes to fitness. Importantly, these changes cause an approximately equal number of genes to increase and decrease their importance after compensatory evolution. Adaptation by gene deletions therefore does not appear to directly affect the mutational robustness of the strain. Instead, compensatory evolution appears to merely cause a shift in gene importance. These shifts in gene importance are global and span several, seemingly unrelated, cellular functions. This has the unexpected result that processes that appear to be important for viability in the wild-type cells can become dispensable in the polarity mutant. A particularly notable example is that genes related to microtubule dynamics become less important in the polarity mutant, raising the question of how processes such as spindle positioning is regulated in these cells (Carminati & Stearns, 1997). We predict that these changes can profoundly alter the response of these cells when faced with a new genetic perturbation.

Encouraged by studies that indicate a relation between the evolutionary rate of a protein and its role in maintaining the structural features of the PPI (Alvarez-Ponce et al., 2017; Fraser & Hirsh, 2004; Fraser et al., 2002; Helsen et al., 2019; Koubkova-Yu et al., 2018), we searched

for a possible correlation between the genes that we found to be affected by compensatory evolution and several network centralities. Our analysis did not reveal any correlation between the likelihood that a gene is affected and any of the three major centralities in the PPI network (degree, betweenness and connectedness) of the corresponding protein. Thus, the predictive power of the PPI network appears to be limited. Instead, our finding that most affected genes can be functionally associated with each other suggests the functional information may be more useful. However, we cannot exclude that this result is a consequence of low PPI network quality. Although we have chosen to use a multi-validated dataset to construct our PPI network, the structure of this network may still significantly deviate from the true interactome that exists in cells. It has been noted by others that interaction network quality is not only determined by a high confidence that an interaction between two components exists, but also by a high confidence that an interaction between two components is truly absent (Van Dam & Snel, 2008). In addition, many interaction networks that are available from integrated databases suffer from research bias (Gillis et al., 2014; von Mering et al., 2002). Both these factors will affect the metrics that we use to determine the position of a protein within the network with respect to any of the three centralities. Aside from the quality of the used protein interaction network, a final possibility is that the transition from the wild-type strain to the polarity mutant induces such large changes in the interactome that the previously mapped interactions are no longer valid. This implies that interaction networks can be highly context dependent and would have far reaching consequences for the ability to draw conclusions based on interaction maps generated with high-throughput methods. These methods frequently determine the presence of a protein-protein interaction based on an assay that takes the components out of the normal cellular context (such as two-hybrid screenings).

Our data allows us to formulate a hypothesis about a possible mechanism for polarity establishment that does not require Bem1, Bem3 and Nrp1. A cluster analysis of the genes that show a significant change in dispensability between the wild-type strain and the polarity mutant revealed the existence of a relatively large cluster of proteins that are annotated to be involved in polarity establishment. This suggests that proteins that participate in polarity establishment in our wild-type strain are also frequently involved in the pathway for polarity establishment in the polarity mutant, although their role may have changed. Further inspection of the cluster enriched for genes related to cell polarity indicated that there is an increased dependence on the axial landmark complex in the polarity mutant. Interestingly, a mechanism for Cdc42 polarization that does not depend on Bem1 has previously proposed to drive polarity establishment before START (Kang et al., 2014; Miller et al., 2020; Moran et al., 2019). It is tempting to speculate that this mechanism has become the main driver of the positive feedback loop that generates the polarized cap of Cdc42. However, our finding that Cdc24 remains essential and Bud3 does not become essential in the polarity mutant is inconsistent with the proposed mechanism for Bem1-independent polarization of Cdc42 before START (Kang et al., 2014). This suggests that in polarity mutants it is not simply the pre-START polarization mechanism that takes over when the Bem1-mediated feedback loop is lost, but rather a pathway that may be a mixture of the pre- and post-START mechanisms. An alternative possibility is that Cdc24 does act as the main GEF for Cdc42 during pre-START polarization, but does not require Bem1 to initiate the positive feedback loop. A study that has shown that recruitment of Cdc24 to a predetermined site on the membrane using optogenetics can initiate pre-START polarization supports this possibility (Witte et al., 2017). However, it remains to be determined how this can be reconciled with the se-

questration of Cdc24 in the nucleus before the onset of START. In addition to the axial landmark complex, we see a dependency on actin-mediated transport, both in the form of exocytosis and endocytosis, of the polarity mutant, suggesting that they are part of the alternative pathway for cell polarization. The role of these processes for the establishment of polarity in wild-type cells has been under debate (Irazoqui et al., 2005; Layton et al., 2011; Savage et al., 2012; Slaughter et al., 2009; Woods et al., 2016). While we currently cannot identify whether these actin-based processes contribute to the establishment or robustness and singularity of the polarity cap, we propose that their role will likely be important for efficient polarization of wild-type cells.

In conclusion, we have shown that compensatory evolution can lead to changes in the fitness contribution of a large number of genes. Importantly, these changes go beyond the initially perturbed and subsequently compensated pathway and instead spans across several, seemingly distinct, cellular processes. However, our data does suggest that the alternative pathway that allows a near wild-type efficiency for polarity establishment after the deletion of Bem1, Bem3 and Nrp1 relies on the axial landmark, the polarisome and endocytosis.

6.4. Methods

6.4.1. Strains

All strains used in this study are from the W303 genetic background. The cas9 cassette was obtained from plasmid p414-TEF1p-cas9-CYC1t (DiCarlo et al., 2013) and fused to the up- and downstream genomic sequences of the *HO*-locus and the *ScURA3* marker using overlap-extension PCR. The resulting genetic construct was transformed into a wild-type strain and *bem1*Δ*bem3*Δ*nrp1*Δ mutant using according to a lithium-acetate transformation protocol (Gietz & Schiestl, 2007). Correct integration was verified with colony PCR. Endogenous expression of Cas9 from the *HO*-locus had no significant effects on growth. Strains were made compatible with SATAY by removing the *ScURA3* marker and the endogenous *ADE2* locus using the CRISPR/Cas9 system according to the double guide-RNA (gRNA) method of Mans et al. (2015). gRNA sequences targeting the *ScURA3* marker and *ADE2* locus were designed using the online toolbox CHOPCHOP (Labun et al., 2019). The repair fragment for removal of the *ScURA3* marker was constructed by PCR amplification of the genomic sequence upstream and downstream of the *ScURA3* marker and fusing the two fragments using overlap-extension PCR. Similarly, the repair fragment for removal of the *ADE2* locus was constructed by fusing the up- and downstream genomic sequences with overlap-extension PCR. Removal of *ScURA3* marker and the *ADE2* was verified by colony PCR using primers 5 and 6 and primers 7 and 8, respectively. Correct clones were cured from their gRNA plasmids by inoculating them in non-selective medium (YPD) and growing them until saturation (~1.5 days) at 30 °C. After saturation was reached, the cultures were plated to single colonies on YPD agar plates and screened for the loss of growth on media selecting for the marker of the gRNA plasmid. Strains were stored at -80 °C as frozen stocks in 40% (v/v) glycerol. The double gRNA plasmids that were used as a template were a kind gift from Pascale Daran-Lapujade.

6.4.2. Media

Standard culturing and growth assays were performed in YPD (10g/L Yeast extract, 20 g/L Peptone, 20 g/L dextrose), SC (6.9 g/L Yeast nitrogen base, 0.75 g/L Complete supplement mixture, 20 g/L dextrose). For ade$^-$ strains, standard growth media was supplemented with 20 mg/L adenine just before incubation. Liquid and solid media for SATAY were prepared as described in section 4.4.2 and table 4.2.

6.4.3. SATAY Library generation

SATAY libraries were generated based on the protocol developed by Michel et al. (2019). The procedures for transposition, DNA extraction and circularization were performed as described in section 4.4.3. Transposon-genome junctions were amplified using the barcoded primers 1 and 2 (table 4.3) for DpnII digested DNA or primers 3 and 4 (table 4.3) for NlaIII digested DNA on a thermal cycler (Bio-Rad C1000 Touch) with the block settings shown in table 4.4. PCR amplified samples were purified using the NucleoSpin Gel and PCR cleanup kit (Macherey-Nagel) and quantified on the NanoDrop 2000 spectrophotometer (Thermo Scientific). For each sample, equal ratios (w/w) of DpnIII and NlaII digested DNA were pooled. Library preparation and sample sequencing were performed by Novogene (UK) Company Limited. Sequencing libraries were prepared with the NEBNext Ultra II DNA Library Prep Kit, omitting the size selection and PCR enrichment steps. Libraries were sequenced on the NovaSeq 6000 platform (Illumina) set to Paired-End (PE) sequencing with a read length of 150 bp.

6.4.4. Volcano plots

Read count distributions were corrected for spikes using the Beta-Geometric correction method and subsequently normalized for differences in transposon density and sequencing depth with the median of ratios normalization (see section 6.6.1). p-values were generated using the unequal variance independent t-test available from the SciPy library in Python (Virtanen et al., 2020) and corrected for multiple hypothesis testing with the Benjamini-Hochberg procedure implemented in the TRANSIT software tool (DeJesus et al. (2015), version 3.2.6).

6.4.5. Physical interaction network analysis

The Multi-Validated (MV) dataset of physical interactions (release BIOGRID-4.4.214) was downloaded from the BioGRID in Tab 3.0 format, which contains all physical interactions that have been validated by at least two different experimental systems or publication sources. The original set of physical interactions was sliced to obtain only those that correspond to interactions between proteins of *S. cerevisiae*. Interactions inferred from Affinity Capture-RNA or Protein-RNA were excluded from the dataset as these interactions do not represent direct physical interactions between proteins. Network visualizations were made with Cytoscape (Shannon et al. (2003), version 3.9.1). Network centralities (degree, betweenness and closeness) were calculated using the NetworkX package in Python (Hagberg et al., 2008).

6.4.6. Functional interaction network analysis

Functional associations between gene products were retrieved from the STRING database (Szklarczyk et al., 2021) through the stringApp plugin (Doncheva et al. (2019), version 1.7.1) of Cytoscape (Shannon et al. (2003), version 3.9.1). The confidence threshold of the imported interactions from STRING was set to medium (confidence level $\leq$0.40). Markov Clustering (MCL) was performed with the clusterMaker2 (version 2.2) plugin of Cytoscape using the confidence level as edge weight and setting the granularity parameter to 2.5.

6.5. Contributions & Acknowledgements

Enzo Kingma constructed the strains, performed the SATAY experiments, analyzed the data and wrote this chapter. We are grateful to Melanie Wijsman and Nicole Bennis for their assistance in establishing the CRISPR/Cas9 system for genome editing.

Bibliography

Alvarez-Ponce, D., Feyertag, F., & Chakraborty, S. (2017). Position Matters: Network Centrality Considerably Impacts Rates of Protein Evolution in the Human Protein–Protein Interaction Network. *Genome Biology and Evolution*, *9*(6), 1742–1756. https://doi.org/10.1093/GBE/EVX117

Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology*, *11*(10), R106. https://doi.org/10.1186/gb-2010-11-10-r106

Arkowitz, R. A., & Iglesias, P. A. (2008). Basic principles of polarity establishment and maintenance. *EMBO reports*, *9*(9), 847–852. https://doi.org/10.1038/EMBOR.2008.164

Barabási, A.-L., & Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, *5*(2), 101–113. https://doi.org/10.1038/nrg1272

Bendert, A., & Pringle, J. R. (1991). Use of a screen for synthetic lethal and multicopy suppressee mutants to identify two new genes involved in morphogenesis in Saccharomyces cerevisiae. *Molecular and Cellular Biology*, *11*(3), 1295–1305. https://doi.org/10.1128/MCB.11.3.1295-1305.1991

Bose, I., Irazoqui, J. E., Moskow, J. J., Bardes, E. S. G., Zyla, T. R., & Lew, D. J. (2001). Assembly of Scaffold-mediated Complexes Containing Cdc42p, the Exchange Factor Cdc24p, and the Effector Cla4p Required for Cell Cycle-regulated Phosphorylation of Cdc24p. *Journal of Biological Chemistry*, *276*(10), 7176–7186. https://doi.org/10.1074/JBC.M010546200

Carminati, J. L., & Stearns, T. (1997). Microtubules Orient the Mitotic Spindle in Yeast through Dynein-dependent Interactions with the Cell Cortex. *Journal of Cell Biology*, *138*(3), 629–641. https://doi.org/10.1083/jcb.138.3.629

Chant, J. (1995). Control of cell polarity by internal programs and external signals in yeast. *Seminars in Developmental Biology*, *6*(1), 13–23. https://doi.org/10.1016/S1044-5781(06)80081-2

Chen, H., Kuo, C.-C., Kang, H., Howell, A. S., Zyla, T. R., Jin, M., & Lew, D. J. (2012). Cdc42p regulation of the yeast formin Bni1p mediated by the effector Gic2p. *Molecular biology of the cell*, *23*(19), 3814–26. https://doi.org/10.1091/mbc.E12-05-0400

Chenevert, J., Corrado, K., Bender, A., Pringle, J., & Herskowitz, I. (1992). A yeast gene (BEM1) necessary for cell polarization whose product contains two SH3 domains. *Nature*, *356*(6364), 77–79. https://doi.org/10.1038/356077a0

Chiou, J. G., Balasubramanian, M. K., & Lew, D. J. (2017). Cell Polarity in Yeast. *Annual review of cell and developmental biology*, *33*, 77. https://doi.org/10.1146/ANNUREV-CELLBIO-100616-060856

D'Agostino, J. L., & Goode, B. L. (2005). Dissection of Arp2/3 Complex Actin Nucleation Mechanism and Distinct Roles for Its Nucleation-Promoting Factors in Saccharomyces cerevisiae. *Genetics*, *171*(1), 35–47. https://doi.org/10.1534/genetics.105.040634

Das, J., Vo, T. V., Wei, X., Mellor, J. C., Tong, V., Degatano, A. G., Wang, X., Wang, L., Cordero, N. A., Kruer-Zerhusen, N., Matsuyama, A., Pleiss, J. A., Lipkin, S. M., Yoshida, M., Roth, F. P., & Yu, H. (2013). Cross-Species Protein Interactome Mapping Reveals Species-Specific Wiring

of Stress Response Pathways. *Science Signaling*, *6*(276), ra38. https://doi.org/10.1126/scisignal.2003350

DeJesus, M. A., Ambadipudi, C., Baker, R., Sassetti, C., & Ioerger, T. R. (2015). TRANSIT - A Software Tool for Himar1 TnSeq Analysis (P. P. Gardner, Ed.). *PLOS Computational Biology*, *11*(10), e1004401. https://doi.org/10.1371/journal.pcbi.1004401

DeJesus, M. A., & Ioerger, T. R. (2013). A Hidden Markov Model for identifying essential and growth-defect regions in bacterial genomes from transposon insertion sequencing data. *BMC Bioinformatics*, *14*(1), 303. https://doi.org/10.1186/1471-2105-14-303

DeJesus, M. A., & Ioerger, T. R. (2016). Normalization of transposon-mutant library sequencing datasets to improve identification of conditionally essential genes. *Journal of Bioinformatics and Computational Biology*, *14*(03), 1642004. https://doi.org/10.1142/S021972001642004X

DiCarlo, J. E., Norville, J. E., Mali, P., Rios, X., Aach, J., & Church, G. M. (2013). Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic Acids Research*, *41*(7), 4336–4343. https://doi.org/10.1093/nar/gkt135

Diepeveen, E. T., Gehrmann, T., Pourquié, V., Abeel, T., & Laan, L. (2018). Patterns of Conservation and Diversification in the Fungal Polarization Network (B. Papp, Ed.). *Genome Biology and Evolution*, *10*(7), 1765–1782. https://doi.org/10.1093/gbe/evy121

Ding, K.-F., Finlay, D., Yin, H., Hendricks, W. P. D., Sereduk, C., Kiefer, J., Sekulic, A., LoRusso, P. M., Vuori, K., Trent, J. M., & Schork, N. J. (2018). Network Rewiring in Cancer: Applications to Melanoma Cell Lines and the Cancer Genome Atlas Patients. *Frontiers in Genetics*, *9*(JUL), 228. https://doi.org/10.3389/fgene.2018.00228

Doncheva, N. T., Morris, J. H., Gorodkin, J., & Jensen, L. J. (2019). Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. *Journal of Proteome Research*, *18*(2), 623–632. https://doi.org/10.1021/acs.jproteome.8b00702

Dunn, R., Dudbridge, F., & Sanderson, C. M. (2005). The use of edge-betweenness clustering to investigate biological function in protein interaction networks. *BMC bioinformatics*, *6*(1), 39. https://doi.org/10.1186/1471-2105-6-39

Etienne-Manneville, S. (2004). Cdc42 - the centre of polarity. *Journal of Cell Science*, *117*(8), 1291–1300. https://doi.org/10.1242/JCS.01115

Fraser, H. B., & Hirsh, A. E. (2004). Evolutionary rate depends on number of protein-protein interactions independently of gene expression level. *BMC evolutionary biology*, *4*(1), 13. https://doi.org/10.1186/1471-2148-4-13

Fraser, H. B., Hirsh, A. E., Steinmetz, L. M., Scharfe, C., & Feldman, M. W. (2002). Evolutionary Rate in the Protein Interaction Network. *Science*, *296*(5568), 750–752. https://doi.org/10.1126/science.1068696

Freisinger, T., Klünder, B., Johnson, J., Müller, N., Pichler, G., Beck, G., Costanzo, M., Boone, C., Cerione, R. A., Frey, E., & Wedlich-Söldner, R. (2013). Establishment of a robust single axis of cell polarity by coupling multiple positive feedback loops. *Nature Communications*, *4*(1), 1807. https://doi.org/10.1038/ncomms2795

Fromont-Racine, M., Rain, J.-C., & Legrain, P. (1997). Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nature Genetics*, *16*(3), 277–282. https://doi.org/10.1038/ng0797-277

Fujita, A., Lord, M., Hiroko, T., Hiroko, F., Chen, T., Oka, C., Misumi, Y., & Chant, J. (2004). Rax1, a protein required for the establishment of the bipolar budding pattern in yeast. *Gene*, *327*(2), 161–169. https://doi.org/10.1016/J.GENE.2003.11.021

Gao, X.-D., Sperber, L. M., Kane, S. A., Tong, Z., Tong, A. H. Y., Boone, C., & Bi, E. (2007). Sequential and Distinct Roles of the Cadherin Domain-containing Protein Axl2p in Cell Polarization in Yeast Cell Cycle (T. Stearns, Ed.). *Molecular Biology of the Cell*, *18*(7), 2542–2560. https://doi.org/10.1091/mbc.e06-09-0822

Gavin, A.-C., Bösche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A.-M., Cruciat, C.-M., Remor, M., Höfert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., … Superti-Furga, G. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, *415*(6868), 141–147. https://doi.org/10.1038/415141a

Ghadie, M. A., Coulombe-Huntington, J., & Xia, Y. (2018). Interactome evolution: insights from genome-wide analyses of protein–protein interactions. *Current Opinion in Structural Biology*, *50*, 42–48. https://doi.org/10.1016/J.SBI.2017.10.012

Giaever, G., Chu, A. M., Ni, L., Connelly, C., Riles, L., Véronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., André, B., Arkin, A. P., Astromoff, A., El Bakkoury, M., Bangham, R., Benito, R., Brachat, S., Campanaro, S., Curtiss, M., Davis, K., … Johnston, M. (2002). Functional profiling of the Saccharomyces cerevisiae genome. *Nature*, *418*(6896), 387–391. https://doi.org/10.1038/nature00935

Gietz, R. D., & Schiestl, R. H. (2007). High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nature Protocols*, *2*(1), 31–34. https://doi.org/10.1038/nprot.2007.13

Gillis, J., Ballouz, S., & Pavlidis, P. (2014). Bias tradeoffs in the creation and analysis of protein–protein interaction networks. *Journal of Proteomics*, *100*, 44–54. https://doi.org/10.1016/j.jprot.2014.01.020

Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, *99*(12), 7821–7826. https://doi.org/10.1073/pnas.122653799

Goode, B. L., Eskin, J. A., & Wendland, B. (2015). Actin and Endocytosis in Budding Yeast. *Genetics*, *199*(2), 315–358. https://doi.org/10.1534/genetics.112.145540

Goryachev, A. B., & Leda, M. (2017). Many roads to symmetry breaking: molecular mechanisms and theoretical models of yeast cell polarity (W. Bement, Ed.). *Molecular Biology of the Cell*, *28*(3), 370–380. https://doi.org/10.1091/mbc.e16-10-0739

Grézal, G., Spohn, R., Méhi, O., Dunai, A., Lázár, V., Bálint, B., Nagy, I., Pál, C., & Papp, B. (2023). Plasticity and Stereotypic Rewiring of the Transcriptome Upon Bacterial Evolution of Antibiotic Resistance (M. Barlow, Ed.). *Molecular Biology and Evolution*, *40*(2). https://doi.org/10.1093/molbev/msad020

Hagberg, A. A., Schult, D. A., & Swart, P. J. (2008). Exploring Network Structure, Dynamics, and Function using NetworkX. In G. Varoquaux, T. Vaught, & J. Millman (Eds.), *Proceedings of the 7th python in science conference* (pp. 11–15).

Halaoui, R., & McCaffrey, L. (2015). Rewiring cell polarity signaling in cancer. *Oncogene*, *34*(8), 939–950. https://doi.org/10.1038/onc.2014.59

Han, J.-D. J. (2008). Understanding biological functions through molecular networks. *Cell Research*, *18*(2), 224–237. https://doi.org/10.1038/cr.2008.16

Harris, K. P., & Tepass, U. (2010). Cdc42 and Vesicle Trafficking in Polarized Cells. *Traffic*, *11*(10), 1272–1279. https://doi.org/10.1111/j.1600-0854.2010.01102.x

Hartwell, L. H., Hopfield, J. J., Leibler, S., & Murray, A. W. (1999). From molecular to modular cell biology. *Nature*, *402*(S6761), C47–C52. https://doi.org/10.1038/35011540

Helsen, J., Frickel, J., Jelier, R., & Verstrepen, K. J. (2019). Network hubs affect evolvability. *PLOS Biology*, *17*(1), e3000111. https://doi.org/10.1371/JOURNAL.PBIO.3000111

Irazoqui, J. E., Gladfelter, A. S., & Lew, D. J. (2003). Scaffold-mediated symmetry breaking by Cdc42p. *Nature Cell Biology*, *5*(12), 1062–1070. https://doi.org/10.1038/ncb1068

Irazoqui, J. E., Howell, A. S., Theesfeld, C. L., & Lew, D. J. (2005). Opposing Roles for Actin in Cdc42p Polarization. *Molecular Biology of the Cell*, *16*(3), 1296–1304. https://doi.org/10.1091/mbc.e04-05-0430

Jaquenoud, M., & Peter, M. (2000). Gic2p May Link Activated Cdc42p to Components Involved in Actin Polarization, Including Bni1p and Bud6p (Aip3p). *Molecular and Cellular Biology*, *20*(17), 6244–6258. https://doi.org/10.1128/MCB.20.17.6244-6258.2000

Jeong, H., Mason, S. P., Barabási, A.-L., & Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature*, *411*(6833), 41–42. https://doi.org/10.1038/35075138

Jose, M., Tollis, S., Nair, D., Sibarita, J.-B., & McCusker, D. (2013). Robust polarity establishment occurs via an endocytosis-based cortical corralling mechanism. *Journal of Cell Biology*, *200*(4), 407–418. https://doi.org/10.1083/jcb.201206081

Joy, M. P., Brock, A., Ingber, D. E., & Huang, S. (2005). High-Betweenness Proteins in the Yeast Protein Interaction Network. *BioMed Research International*, *2005*(2), 96–103. https://doi.org/10.1155/JBB.2005.96

Kaksonen, M., Toret, C. P., & Drubin, D. G. (2005). A Modular Design for the Clathrin- and Actin-Mediated Endocytosis Machinery. *Cell*, *123*(2), 305–320. https://doi.org/10.1016/J.CELL.2005.09.024

Kang, P. J., Angerman, E., Jung, C.-H., & Park, H.-O. (2012). Bud4 mediates the cell-type-specific assembly of the axial landmark in budding yeast. *Journal of Cell Science*, *125*(16), 3840–3849. https://doi.org/10.1242/jcs.103697

Kang, P. J., Angerman, E., Nakashima, K., Pringle, J. R., & Park, H.-O. (2004). Interactions among Rax1p, Rax2p, Bud8p, and Bud9p in Marking Cortical Sites for Bipolar Bud-site Selection in Yeast. *Molecular Biology of the Cell*, *15*(11), 5145–5157. https://doi.org/10.1091/mbc.e04-07-0600

Kang, P. J., Lee, M. E., & Park, H.-O. (2014). Bud3 activates Cdc42 to establish a proper growth site in budding yeast. *Journal of Cell Biology*, *206*(1), 19–28. https://doi.org/10.1083/jcb.201402040

Kang, P. J., Miller, K. E., Guegueniat, J., Beven, L., & Park, H.-O. (2018). The shared role of the Rsr1 GTPase and Gic1/Gic2 in Cdc42 polarization. *Molecular Biology of the Cell*, *29*(20), 2359–2369. https://doi.org/10.1091/mbc.e18-02-0145

Kim, J., Kim, I., Han, S. K., Bowie, J. U., & Kim, S. (2012). Network rewiring is an important mechanism of gene essentiality change. *Scientific Reports*, *2*(1), 900. https://doi.org/10.1038/srep00900

Kim, K., Yamashita, A., Wear, M. A., Maéda, Y., & Cooper, J. A. (2004). Capping protein binding to actin in yeast biochemical mechanism and physiological relevance. *Journal of Cell Biology*, *164*(4), 567–580. https://doi.org/10.1083/JCB.200308061

Klünder, B., Freisinger, T., Wedlich-Söldner, R., & Frey, E. (2013). GDI-Mediated Cell Polarization in Yeast Provides Precise Spatial and Temporal Control of Cdc42 Signaling. *PLOS Computational Biology*, *9*(12), e1003396. https://doi.org/10.1371/JOURNAL.PCBI.1003396

Komurov, K., Tseng, J. T., Muller, M., Seviour, E. G., Moss, T. J., Yang, L., Nagrath, D., & Ram, P. T. (2012). The glucose-deprivation network counteracts lapatinib-induced toxicity in resis-

tant ErbB2-positive breast cancer cells. *Molecular Systems Biology*, *8*(1), 596. https://doi.org/10.1038/MSB.2012.25

Koubkova-Yu, T. C.-T., Chao, J.-C., & Leu, J.-Y. (2018). Heterologous Hsp90 promotes phenotypic diversity through network evolution (J. Gore, Ed.). *PLOS Biology*, *16*(11), e2006450. https://doi.org/10.1371/journal.pbio.2006450

Koutrouli, M., Karatzas, E., Paez-Espino, D., & Pavlopoulos, G. A. (2020). A Guide to Conquer the Biological Network Era Using Graph Theory. *Frontiers in Bioengineering and Biotechnology*, *8*, 34. https://doi.org/10.3389/fbioe.2020.00034

Kozubowski, L., Saito, K., Johnson, J. M., Howell, A. S., Zyla, T. R., & Lew, D. J. (2008). Symmetry-Breaking Polarization Driven by a Cdc42p GEF-PAK Complex. *Current Biology*, *18*(22), 1719–1726. https://doi.org/10.1016/j.cub.2008.09.060

Laan, L., Koschwanez, J. H., & Murray, A. W. (2015). Evolutionary adaptation after crippling cell polarization follows reproducible trajectories. *eLife*, *4*. https://doi.org/10.7554/elife.09638

Labun, K., Montague, T. G., Krause, M., Torres Cleuren, Y. N., Tjeldnes, H., & Valen, E. (2019). CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Research*, *47*(W1), W171–W174. https://doi.org/10.1093/nar/gkz365

Lai, H., Chiou, J.-G., Zhurikhina, A., Zyla, T. R., Tsygankov, D., & Lew, D. J. (2018). Temporal regulation of morphogenetic events in Saccharomyces cerevisiae (R. Li, Ed.). *Molecular Biology of the Cell*, *29*(17), 2069–2083. https://doi.org/10.1091/mbc.E18-03-0188

Layton, A. T., Savage, N. S., Howell, A. S., Carroll, S. Y., Drubin, D. G., & Lew, D. J. (2011). Modeling Vesicle Traffic Reveals Unexpected Consequences for Cdc42p-Mediated Polarity Establishment. *Current Biology*, *21*(3), 184–194. https://doi.org/10.1016/J.CUB.2011.01.012

Lechler, T., Jonsdottir, G. A., Klee, S. K., Pellman, D., & Li, R. (2001). A two-tiered mechanism by which Cdc42 controls the localization and activation of an Arp2/3-activating motor complex in yeast. *Journal of Cell Biology*, *155*(2), 261–270. https://doi.org/10.1083/JCB.200104094

Mackiewicz, P., Kowalczuk, M., Mackiewicz, D., Nowicka, A., Dudkiewicz, M., Laszkiewicz, A., Dudek, M. R., & Cebrat, S. (2002). How many protein-coding genes are there in the Saccharomyces cerevisiae genome? *Yeast*, *19*(7), 619–629. https://doi.org/10.1002/YEA.865

Mans, R., van Rossum, H. M., Wijsman, M., Backx, A., Kuijpers, N. G., van den Broek, M., Daran-Lapujade, P., Pronk, J. T., van Maris, A. J., & Daran, J.-M. G. (2015). CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in Saccharomyces cerevisiae. *FEMS Yeast Research*, *15*(2), 1–21. https://doi.org/10.1093/femsyr/fov004

Martin, S. G., & Arkowitz, R. A. (2014). Cell polarization in budding and fission yeasts. *FEMS Microbiology Reviews*, *38*(2), 228–253. https://doi.org/10.1111/1574-6976.12055

McGill, B. J., Etienne, R. S., Gray, J. S., Alonso, D., Anderson, M. J., Benecha, H. K., Dornelas, M., Enquist, B. J., Green, J. L., He, F., Hurlbert, A. H., Magurran, A. E., Marquet, P. A., Maurer, B. A., Ostling, A., Soykan, C. U., Ugland, K. I., & White, E. P. (2007). Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecology Letters*, *10*(10), 995–1015. https://doi.org/10.1111/J.1461-0248.2007.01094.X

Michel, A. H., Hatakeyama, R., Kimmig, P., Arter, M., Peter, M., Matos, J., De Virgilio, C., & Kornmann, B. (2017). Functional mapping of yeast genomes by saturated transposition. *eLife*, *6*. https://doi.org/10.7554/eLife.23570

Michel, A. H., van Schie, S., Mosbach, A., Scalliet, G., & Kornmann, B. (2019). Exploiting homologous recombination increases SATAY efficiency for loss- and gain-of-function screening. *bioRxiv*, 866483. https://doi.org/https://doi.org/10.1101/866483

Miller, K. E., Kang, P. J., & Park, H. O. (2020). Regulation of Cdc42 for polarized growth in budding yeast. *Microbial Cell*, *7*(7), 175. https://doi.org/10.15698/MIC2020.07.722

Mishra, M., Huang, J., & Balasubramanian, M. K. (2014). The yeast actin cytoskeleton. *FEMS Microbiology Reviews*, *38*(2), 213–227. https://doi.org/10.1111/1574-6976.12064

Moran, K. D., Kang, H., Araujo, A. V., Zyla, T. R., Saito, K., Tsygankov, D., & Lew, D. J. (2019). Cell-cycle control of cell polarity in yeast. *Journal of Cell Biology*, *218*(1), 171–189. https://doi.org/10.1083/JCB.201806196

Nern, A., & Arkowitz, R. A. (2000). Nucleocytoplasmic Shuttling of the Cdc42p Exchange Factor Cdc24p. *Journal of Cell Biology*, *148*(6), 1115–1122. https://doi.org/10.1083/JCB.148.6.1115

Parekh, S., Ziegenhain, C., Vieth, B., Enard, W., & Hellmann, I. (2016). The impact of amplification on differential expression analyses by RNA-seq. *Scientific Reports 2016 6:1*, *6*(1), 1–11. https://doi.org/10.1038/srep25533

Quintero-Monzon, O., Rodal, A. A., Strokopytov, B., Almo, S. C., & Goode, B. L. (2005). Structural and Functional Dissection of the Abp1 ADFH Actin-binding Domain Reveals Versatile In Vivo Adapter Functions. *Molecular Biology of the Cell*, *16*(7), 3128–3139. https://doi.org/10.1091/mbc.e05-01-0059

Robinson, M. D., & Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, *11*(3), R25. https://doi.org/10.1186/gb-2010-11-3-r25

Savage, N. S., Layton, A. T., & Lew, D. J. (2012). Mechanistic mathematical model of polarity in yeast (L. Edelstein-Keshet, Ed.). *Molecular Biology of the Cell*, *23*(10), 1998–2013. https://doi.org/10.1091/mbc.e11-10-0837

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Research*, *13*(11), 2498–2504. https://doi.org/10.1101/gr.1239303

Sheu, Y.-J., Santos, B., Fortin, N., Costigan, C., & Snyder, M. (1998). Spa2p Interacts with Cell Polarity Proteins and Signaling Components Involved in Yeast Cell Morphogenesis. *Molecular and Cellular Biology*, *18*(7), 4053–4069. https://doi.org/10.1128/MCB.18.7.4053

Shimada, Y., Gulli, M. P., & Peter, M. (2000). Nuclear sequestration of the exchange factor Cdc24 by Far1 regulates cell polarity during yeast mating. *Nature Cell Biology 2000 2:2*, *2*(2), 117–124. https://doi.org/10.1038/35000073

Slaughter, B. D., Das, A., Schwartz, J. W., Rubinstein, B., & Li, R. (2009). Dual Modes of Cdc42 Recycling Fine-Tune Polarized Morphogenesis. *Developmental Cell*, *17*(6), 823–835. https://doi.org/10.1016/j.devcel.2009.10.022

Slaughter, B. D., Unruh, J. R., Das, A., Smith, S. E., Rubinstein, B., & Li, R. (2013). Non-uniform membrane diffusion enables steady-state cell polarization via vesicular trafficking. *Nature Communications*, *4*(1), 1380. https://doi.org/10.1038/ncomms2370

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., & Mering, C. v. (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, *47*(D1), D607–D613. https://doi.org/10.1093/nar/gky1131

Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N. T., Legeay, M., Fang, T., Bork, P., Jensen, L. J., & von Mering, C. (2021). The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Research*, *49*(D1), D605–D612. https://doi.org/10.1093/NAR/GKAA1074

Taylor, T. B., Shepherd, M. J., Jackson, R. W., & Silby, M. W. (2022). Natural selection on crosstalk between gene regulatory networks facilitates bacterial adaptation to novel environments. *Current Opinion in Microbiology*, *67*, 102140. https://doi.org/10.1016/j.mib.2022.02.002

Thompson, B. J. (2013). Cell polarity: models and mechanisms from yeast, worms and flies. *Development*, *140*(1), 13–21. https://doi.org/10.1242/DEV.083634

Van Dam, T. J. P., & Snel, B. (2008). Protein Complex Evolution Does Not Involve Extensive Network Rewiring. *PLOS Computational Biology*, *4*(7), e1000132. https://doi.org/10.1371/JOURNAL.PCBI.1000132

Van Dongen, S. (2008). Graph Clustering Via a Discrete Uncoupling Process. *SIAM Journal on Matrix Analysis and Applications*, *30*(1), 121–141. https://doi.org/10.1137/040608635

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … Vázquez-Baeza, Y. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S. G., Fields, S., & Bork, P. (2002). Comparative assessment of large-scale data sets of protein–protein interactions. *Nature*, *417*(6887), 399–403. https://doi.org/10.1038/nature750

Wedlich-Soldner, R., Altschuler, S., Wu, L., & Li, R. (2003). Spontaneous Cell Polarization Through Actomyosin-Based Delivery of the Cdc42 GTPase. *Science*, *299*(5610), 1231–1235. https://doi.org/10.1126/science.1080944

Winzeler, E. A., Shoemaker, D. D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J. D., Bussey, H., Chu, A. M., Connelly, C., Davis, K., Dietrich, F., Dow, S. W., El Bakkoury, M., Foury, F., Friend, S. H., Gentalen, E., … Davis, R. W. (1999). Functional Characterization of the S. cerevisiae Genome by Gene Deletion and Parallel Analysis. *Science*, *285*(5429), 901–906. https://doi.org/10.1126/science.285.5429.901

Witte, K., Strickland, D., & Glotzer, M. (2017). Cell cycle entry triggers a switch between two modes of Cdc42 activation during yeast polarization. *eLife*, *6*. https://doi.org/10.7554/ELIFE.26722

Woods, B., Lai, H., Wu, C. F., Zyla, T. R., Savage, N. S., & Lew, D. J. (2016). Parallel Actin-Independent Recycling Pathways Polarize Cdc42 in Budding Yeast. *Current Biology*, *26*(16), 2114–2126. https://doi.org/10.1016/j.cub.2016.06.047

Xie, Y., & Miao, Y. (2021). Polarisome assembly mediates actin remodeling during polarized yeast and fungal growth. *Journal of Cell Science*, *134*(1). https://doi.org/10.1242/JCS.247916/237100

Young, M. E., Cooper, J. A., & Bridgman, P. C. (2004). Yeast actin patches are networks of branched actin filaments. *Journal of Cell Biology*, *166*(5), 629–635. https://doi.org/10.1083/JCB.200404159

Yu, H., Braun, P., Yıldırım, M. A., Lemmens, I., Venkatesan, K., Sahalie, J., Hirozane-Kishikawa, T., Gebreab, F., Li, N., Simonis, N., Hao, T., Rual, J.-F., Dricot, A., Vazquez, A., Murray, R. R., Simon, C., Tardivo, L., Tam, S., Svrzikapa, N., … Vidal, M. (2008). High-Quality Binary Protein Interaction Map of the Yeast Interactome Network. *Science*, *322*(5898), 104–110. https://doi.org/10.1126/science.1158684

Zheng, Y., Bender, A., & Cerione, R. A. (1995). Interactions among Proteins Involved in Bud-site Selection and Bud-site Assembly in Saccharomyces cerevisiae. *Journal of Biological Chemistry*, *270*(2), 626–630. https://doi.org/10.1074/JBC.270.2.626

6.6. Supplement

6.6.1. Read count correction and normalization procedure

The benefit of including read counts in our metric to identify changes in the fitness effect of gene disruptions across genetic backgrounds is that it allows for a more sensitive analysis than those based on insertion counts. For example, a gene may have a negligible difference in insertion counts between two genetic backgrounds if the change in gene disruption tolerance is not so drastic that it causes a large number of mutants to fall below the detection level in one of the two backgrounds. However, if a significant difference in gene disruption tolerance truly exists, we would expect that the read counts (which should reflect mutant abundance) associated with each insertion site will be lower in one of the two strains. Thus, extending the analysis from a binary assessment of the occupancy of insertion sites to a gradual scale based on read counts can uncover differences that would otherwise remain undetected. However, a drawback is that read counts tend to be more variable between replicate experiments than insertion counts, often owing to read count spikes caused by technical and biological artifacts (DeJesus & Ioerger, 2016). This increased noise level can hinder the desired increase of sensitivity of the analysis. In addition, spikes in the read counts can significantly skew the read count distribution. This means that the small subset of insertion sites with an artificially high read count will strongly affect methods that aim to correct differences in sequencing depth between libraries by a linear transformation of the total read count.

To address these issues, we implemented the Beta-Geometric correction (BGC) method developed by DeJesus and Ioerger (2013). This method is based on the observation that read count distributions obtained from transposon insertion sequencing (TIS) experiments resemble a geometric distribution where most insertion sites have a low read count, while only few sites contain many reads[1]. The quantiles of the empirical read count distribution are adjusted to match the quantiles of a fitted 'ideal' geometric distribution. To allow for greater flexibility of the fit, this ideal distribution is implemented as a mixture of geometric distributions with a Beta prior on the probability parameter. In practice, this procedure corrects for the 'skew' in the empirical read count distribution by suppressing spikes while inflating the insertion sites that have a low read count. The results in figure S6.1 show that the application of this procedure greatly improves the correlation between replicate datasets (Figure S6.1b) compared to uncorrected datasets (Figure S6.1a). Lastly, after applying the BGC correction, we normalized for differences in sequencing depth and library complexity across datasets using the median of ratios normalization (Anders & Huber, 2010; Robinson & Oshlack, 2010). In summary, the geometric mean across all samples was calculated for each gene:

$$\bar{X}_g = \exp \frac{1}{n} \cdot \sum_{a}^{n} \ln (R_{g,a})). \tag{6.1}$$

Where $\bar{X}_g$ is the geometric mean of the read counts mapping to gene g, n is the total number of datasets and $R_{g,a}$ is the number of reads that map to gene g in dataset a. Next, the ratio of the total read count to the geometric mean $\bar{X}_g$ was determined for each sample and the sample-specific normalization factor was taken to be the median of these ratios across all genes:

[1] For a mathematical reasoning of why read counts obtained from a TIS experiment would be expected to follow a geometric distribution, see DeJesus and Ioerger (2016) and McGill et al. (2007).

$$S_a = \text{median}\left(\frac{R_{g,a}}{\bar{X}_g}\right) \text{ , for all } a \in G \tag{6.2}$$

With S_a the normalization factor for dataset a and G is the complete set of annotated genes used in this chapter. Finally, the read counts were linearly scaled by the normalization factor S_a:

$$R_{g,a,\text{norm}} = \frac{R_{g,a}}{S_a}. \tag{6.3}$$

6.6.2. Supplementary figures

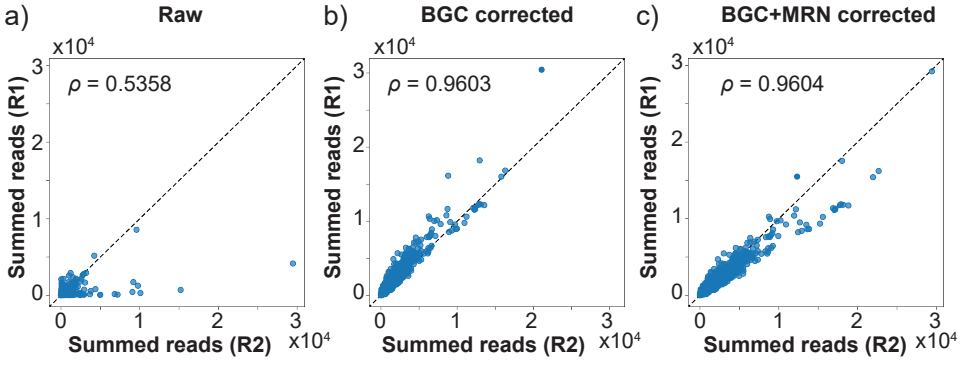


Figure S6.1. Normalization using the beta-geometric correction (BGC) increases the correlation in summed read counts between replicate experiments. The correlation between the summed read counts of each gene in two replicate strains (R1 and R2) is shown for (**a**) raw data (**b**) after BGC and (**c**) after both BGC and median of ratios normalization (MRN). The dashed line represents the identity line. ρ: Pearson's correlation coefficient.

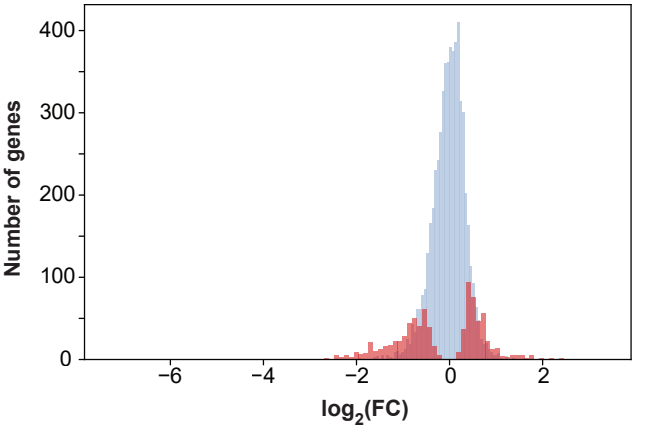


Figure S6.2. Significant genes have a larger effect size. Histogram of the $\log_2$-fold changes shown in figure 6.2. The plot shows that genes that are flagged to have a statistically significant difference between the two genetic backgrounds (red bars) typically occurs for larger fold-changes.

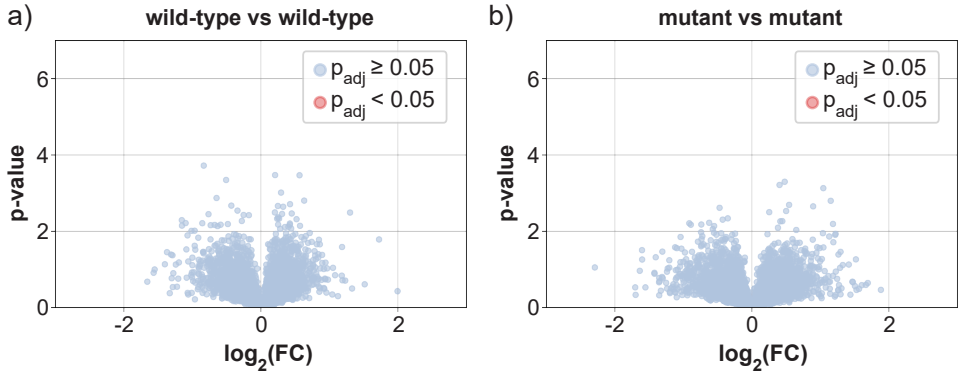


Figure S6.3. Volcano plots of samples of the same genetic background to check for false discoveries. The replicate datasets of each genetic background (6 in total) were split and compared 3 vs. 3. The plot shows that for both our wild-type samples and *bem1*Δ*bem3*Δ*nrp1*Δ samples, we do not find any false positives at a significance threshold of $p_{adj} < 0.05$.

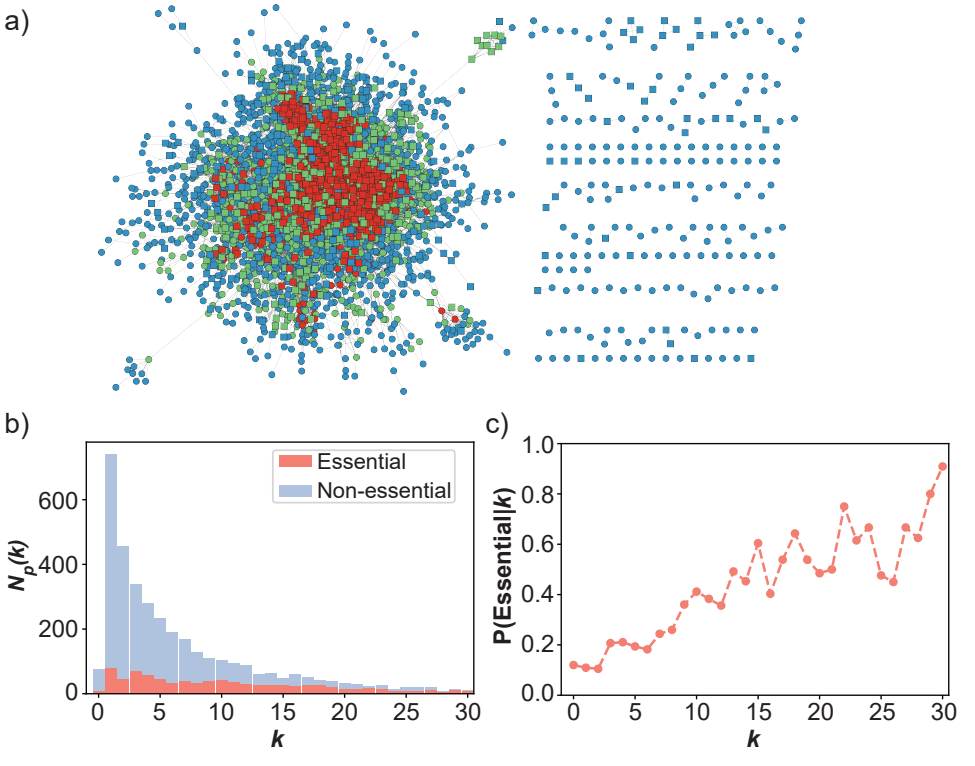


Figure S6.4. Gene essentiality correlates with degree in the protein-protein interaction network. (**a**) Visualization of the protein-protein interaction network constructed from the interactions annotated by the BioGRID database. Nodes are colored according to their degree k (blue: $k \leq 3$, green: $4 \leq k \leq 10$, red: $k > 10$). Essential genes are shown as square nodes, non-essential genes are shown as circles. (**b**) Stacked histogram of the number of proteins $N_p(k)$ with degree k. proteins annotated as essential are shown in red, non-essential proteins are shown in blue. The histogram shows that the distribution for k is roughly uniform for essential genes, while it is skewed towards lower values of k for non-essential genes. (**c**) The probability that a node is essential as a function of k. The graph displays an increasing trend for the probability as k increases. Combined with the histogram shown in (b), this trend demonstrates that genes with a high degree are more likely to be essential than genes with a low degree.

7

The four stages of acceptance:

1. This is worthless nonsense.
2. This is an interesting, but perverse, point of view.
3. This is true, but quite unimportant.
4. I always said so.

— J.B.S. Haldane

Conclusion

Living systems possess the remarkable ability to adapt to their environment through the process of evolution. The question of whether we will ever be able to understand this process to a degree that allows prediction has been asked repeatably in the past decades (Gould, 1989; Johnson et al., 2023; Lobkovsky & Koonin, 2012; Orgogozo, 2015). Initially, the overwhelming complexity of living systems combined with the unpredictable forces that drive evolution made prediction seem like a daunting, if not impossible, task (Gould, 1989). However, experiments demonstrating that adaptive evolution can follow reproducible paths have provided insights into the conditions that can lead to predictable evolutionary trajectories (Laan et al., 2015; Lind et al., 2019). Crucially, these experiments identified epistatic interactions between mutations as a source of evolutionary constraint (de Visser et al., 2018; Szendro et al., 2013). If these constraints are severe enough to render the majority of possible evolutionary trajectories inaccessible, they open up an opportunity for predicting evolution. Nevertheless, epistatic constraints have mostly been used retrospectively to explain observed adaptive pathways (Gorter et al., 2018; Laan et al., 2015).

To employ epistasis as a tool for prediction, either one of the following two approaches must be undertaken: (1) an extensive mapping of all possible combinations of mutations to empirically determine the structure of the fitness landscape (Poelwijk et al., 2007) or (2) deriving epistasis from the physiochemical properties of the cell. These approaches are inherently linked, as understanding the molecular basis of epistatic interactions requires an in-depth analysis of observed epistatic patterns (Poelwijk et al., 2019). This link has been underscored by studies that successfully inferred functional relationships between genes based on the similarity of their epistatic interaction profile (Costanzo et al., 2016; van Leeuwen et al., 2016).

In this thesis, we studied how molecular interactions can facilitate or constrain evolution using the polarity network of *Saccharomyces cerevisiae* as a model system. We focus on a specific evolutionary pathway where cells adapt to the loss of Bem1, a key polarity protein, through the sequential deletion of Bem3 and Nrp1 (Laan et al., 2015). Under standard laboratory growth conditions, the phenotype of the resulting polarity mutant is surprisingly similar to that of a wild-type strain, indicating that these gene deletions effectively suppress the detrimental effects of losing Bem1. The ability to mask the detrimental effects of a mutation through genetic changes elsewhere in the genome has been an active area of research, in part due to its importance for understanding the inheritance pattern of Mendelian diseases (Chen et al., 2016). However, these studies traditionally focus on the potential functional relation between the allele responsible for a specific phenotype and the gene capable of mitigating these effects through secondary mutations. In this work, we go one step further and explore the impact of suppressor mutations on the overall genetic wiring of the cell (Tong et al., 2004; Usaj et al., 2017). By doing so, we take a holistic approach to unravel the complex mechanisms that govern adaptive pathways. Our results demonstrate that such an approach may be instrumental for the development of models that can make evolutionary predictions.

Mapping the fitness effect of gene disruptions on a genome-wide scale using SATAY.
Fitness landscapes are at the heart of many predictive models for adaptive evolution. Specifically, knowledge of the topological features of the fitness landscape, such as its ruggedness, provides information on the accessibility and likelihood of evolutionary trajectories, which can in turn be

used to make predictions. Empirical fitness landscapes, constructed by systematically assessing all possible allele combinations, have indeed been shown capable to explain observed patterns of evolution (Gorter et al., 2018; Weinreich et al., 2005). However, what complicates their application in predictive models is that their structure can be dynamic across environments and genetic backgrounds. For example, Bajić et al. (2018) found that an empirical fitness landscape of the metabolic network of *Escherichia coli* may only be valid for short mutational distances due to the effect evolving cells have on their environment. Similarly, van Leeuwen et al. (2020) demonstrated that the lethality of gene loss in *S. cerevisiae* can vary substantially between genetic backgrounds. These findings highlight the need for techniques that facilitate the construction of fitness maps in different environmental and genetic contexts —a feature lacking in most currently used methods (Butland et al., 2008; Kogenaru et al., 2009; Tong & Boone, 2006; Winzeler et al., 1999).

Transposon insertion sequencing (TIS) methods enable the rapid construction of complete libraries containing single gene deletion mutants within specific genetic backgrounds or environments. Initially, these methods were primarily used to distinguish genes based on their essentiality (Gawronski et al., 2009; Goodman et al., 2009; Langridge et al., 2009), but the relation between mutant abundance and observed read counts has prompted their application towards the construction of genome-wide fitness maps of gene disruption mutants (van Opijnen et al., 2009). In *chapter 4*, we describe a method to estimate the fitness of gene deletion mutants from data generated using the newly developed TIS screen for *S. cerevisiae* called SATAY (Michel et al., 2017; Michel & Kornmann, 2022). Our approach is based on the average read count per transposon insertion site and we demonstrate that this proxy provides fitness estimates that are robust across replicate SATAY experiments. However, despite this robustness, the distribution of fitness effects (DFE) differs significantly from those published by other studies. These inconsistencies between fitness distributions generated with different high-throughput techniques are not uncommon. Notably, it appears that the major conserved feature of the DFE is that the most gene deletions have a near-neutral effect on fitness (Baryshnikova et al., 2010).

The absence of a definitive 'gold standard' dataset for gene fitness values prevents us from conclusively determining accuracy of fitness values obtained from SATAY in comparison to those found using other methods. While it is uncertain whether this gold standard exists –and if it does, whether it can practically be constructed–, we speculate about its desired characteristics. In order for a gold standard to serve as a calibration for other datasets, it must accurately represent the conserved features of the fitness landscape. This requirement presents a challenge, given the notorious context-dependency of fitness (Bajić et al., 2018; Kinsler et al., 2020; van Leeuwen et al., 2020). However, genetic interaction maps, which are derived from fitness landscapes (Costanzo et al., 2016), do appear to have conserved properties. For instance, research in various yeast species and humans has shown that the impact of a gene deletion on fitness correlates with the number of genetic interactors of the gene (Koch et al., 2012; van Leeuwen et al., 2017). Moreover, studies have indicated that specific genetic interactions between certain pairs of genes remain conserved during evolution (Tosti et al., 2014). Therefore, using the conserved structures of the genetic interaction map, rather than directly comparing fitness values, may be a more reliable approach to assess different fitness estimation techniques.

Functional restoration is aided by redundant pathways.

The deleterious effects of losing an important gene can be compensated during evolution without re-establishing the pathway that was perturbed (Helsen et al., 2020; Szamecz et al., 2014).

This remarkable property demonstrates that organisms can exploit genetic redundancies to quickly restore functions affected by the gene deletion. However, these redundancies within biological networks often remain concealed in unperturbed systems and it is unclear how the compensatory mutations promote their role in restoring the perturbed function. In *chapter 6*, we performed a genome-wide mutagenesis screen to identify genes that recover the defects in polarity establishment induced by the deletion of *BEM1* through the subsequent deletion of *BEM3* and *NRP1*. The ability to restore the defects solely through gene deletions implies the existence of a latent pathway for polarity establishment within the cell. Identifying the structure of this pathway may provide a mechanistic explanation for the constrained order in which the compensatory mutations occur during experimental evolution.

Interestingly, we were able to identify a subset of genes associated with cell polarity that predominantly exhibit a reduced tolerance to disruptions in the polarity mutant when compared to the same genes in a wild-type strain. Considering the known molecular function of these genes, we propose the axial landmark complex, endocytosis and exocytosis as the main structures and pathways that contribute to polarity establishment after the loss of Bem1, Bem3 and Nrp1. This finding supports the previous observation that redundancies at the level of pathways, rather than individual genes, play an important role in compensatory evolution (Szamecz et al., 2014).

The fact that none of these processes are required for polarity establishment in the wild-type genetic background demonstrates that their redundant function remains hidden when a stronger feedback loop for Cdc42 polarization is active. Furthermore, each of these three pathways alone may be too weak to drive cell polarization independently, which would explain why genes related to all three pathways display heightened sensitivity to disruptions in *bem1*Δ*bem3*Δ*nrp1*Δ cells. Despite their apparent involvement in an underlying pathway for polarity establishment, the existing knowledge about the interactions among the molecular components of these processes does not fully explain how they would contribute to a positive feedback loop for Cdc42 polarization. For example, even in the presence of the axial landmark, the primary positive feedback loop driving Cdc42 polarization during bud formation is still generally believed to depend on Bem1 (Jost & Weiner, 2015; Miller et al., 2020; Woods et al., 2015). Consequently, the involvement of the axial landmark in establishing a Bem1-independent mechanism for Cdc42 polarization may depend on molecular interactions that have not ye been identified.

Additional research is therefore necessary to validate the mechanisms through which the proposed processes create an alternative pathway for polarity establishment. For instance, fluorescence microscopy using labelled Cdc42 could reveal whether *bem1*Δ*bem3*Δ*nrp1*Δ cells contain higher levels of Cdc42 on their internal membranes compared to wild-type cells. This observation would support the hypothesis that these cells have an increased reliance on endo- and exocytic pathways in the absence of cytoplasmic recycling of Cdc42. In addition, creating conditional knock-outs of the proteins indicated by our data to have an increased importance in *bem1*Δ*bem3*Δ*nrp1*Δ cells could elucidate how their loss affects Cdc42 polarization dynamics. For this purpose, light-induced sequestration experiments may prove to be particularly useful (Jost & Weiner, 2015). These experiments allow the reversible release and sequestration of proteins at different phases of the cell cycle and could be used to identify the temporal sequence of events leading to polarity establishment (Kang et al., 2018; Miller et al., 2020; Miller et al., 2019; Moran et al., 2019).

Understanding the alternative pathway for polarity establishment that acts after the deletion of Bem1, Bem3 and Nrp1 could theoretically allow control over the evolutionary trajectory. A key

element of the ability to exert control lies in the finding that this alternative pathway consists of redundant functions within existing cellular processes. Mutations in these processes might have limited effects on fitness in a wild-type genetic background but can be highly detrimental when combined with the gene deletions that compensate for the loss of Bem1. Such mutations would then act as an epistatic switch that controls the accessibility of adaptive pathways.

Interestingly, our search for gene deletions that could render the pathway from a *bem1*Δ to a *bem1*Δ*bem3*Δ genetic background inaccessible identified *ABP1* as a gene with the necessary epistatic properties (see chapter 5). Notably, Abp1 belongs to a conserved family of actin-binding proteins and connects the cortical actin network to the endocytic machinery (Michelot et al., 2013; Quintero-Monzon et al., 2005), suggesting a functional association with cellular processes involved in polarity establishment. This supports the notion that understanding the molecular mechanism underlying the evolutionary recovery of polarity establishment would allow for control over the evolutionary process.

While further experimental validation is required, we propose the following mechanistic explanation for how the deletion of Abp1 would prevent the deletion of Bem3 from acting as a suppressor in a *bem1*Δ background. Because Bem3 acts as a GAP for Cdc42, its deletion marks the transition from a reaction-diffusion type mechanism to an actin-mediated mechanism for Cdc42 recycling. This transition increases the dependence of polarity establishment on endocytosis for the extraction of Cdc42 from the membrane, a process that normally acts in parallel with the faster GDI-mediated pathway (Slaughter et al., 2009). The deletion of *ABP1* disrupts the ability to use the endocytic pathway for Cdc42 recycling, rendering the transition from a *bem1*Δ*abp1*Δ to a *bem1*Δ*bem3*Δ*abp1*Δ mutant deleterious.

In addition to its functional relation to polarity establishment, the role of Abp1 is dispensable in wild-type cells (Drubin et al., 1990). Consequently, we anticipate that a population *abp1*Δ cells can persist in natural environments. This feature makes the deletion of Abp1 a viable strategy for designing cells with a distinct adaptive response to the absence of Bem1, while maintaining the same fitness as wild-type cells under standard conditions. Overall, our findings support the hypothesis that inhibiting pathways redundant with the disrupted function can be used to control the accessibility of evolutionary pathways.

Compensatory evolution has genome-wide consequences on genetic interactions.
Mutations that suppress the effects of gene loss are frequently found in genes that are functionally related to the disrupted gene (Helsen et al., 2020; Szamecz et al., 2014). However, the highly interconnected nature of genetic interaction maps implies that the consequences of compensatory mutations can have effects that go beyond the initially perturbed module. In *chapter 6*, our genome-wide mutagenesis screen revealed a substantial number of genes with an altered fitness contribution following the evolutionary repair of cell polarity. Based on this finding we propose that, while suppressor mutations might be associated with the induced functional defect, compensatory evolution induces a widespread rewiring of the interaction network. This is in agreement with the observation that compensatory mutations rarely result in the restoration of the ancestral gene expression pattern (Szamecz et al., 2014).

Moreover, our findings underscore the concept of global epistasis, where the fitness of individual gene relies on a large number of other loci (Kryazhimskiy et al., 2014; Reddy & Desai, 2021). A notable implication of global epistasis is that the epistatic constraints associated with compensatory mutations will tend to be complex. For example, our results indicate that the ability of loss-of-function mutations in Bem3 and Nrp1 to suppress the *bem1*Δ phenotype depends

on genes from pathways unrelated to cell polarity. Furthermore, given that a gene's relevance for fitness can vary based on environmental conditions, the epistatic constraints observed during adaptive evolution might differ considerably across environments (Kinsler et al., 2020). For the same reason, although the polarity mutant closely resembles a wild-type strain under standard laboratory growth conditions, differences between the two strains may become apparent in alternative environments (Kinsler et al., 2020).

An important question is how the individual mutations along the evolutionary trajectory contribute to this complexity of the epistatic interactions. For example, the deletion of Bem3, which has a clear functional connection to polarity establishment, may only locally alter the fitness impact of genes within the polarity module. Alternatively, the loss of Nrp1, a protein proposed to be capable of binding RNA (Hogan et al., 2008), may induce more global changes through general effects on translation. Finally, it is possible that the gene deletions exhibit higher-order epistasis (Ferretti et al., 2016; Weinreich et al., 2018; Yang et al., 2019; Zhou et al., 2022), where the loss of one gene changes the epistatic interaction pattern of the other, shifting it from local to global, or vice versa. The ability to predict whether gene loss causes local or global changes in the genotype-fitness map based on the gene's function might allow us to distinguish between different levels of evolutionary predictability (Bank et al., 2016; Zhou et al., 2022). Performing SATAY screens in mutants from different steps of the observed evolutionary trajectory should reveal the extent to which the different gene deletions contribute locally or globally to changes in the epistatic interaction pattern.

The discovery that several genes crucial for survival in a wild-type strain become increasingly dispensable in the polarity mutant implies that gene deletions provide opportunities for evolutionary innovations. Indeed, several studies have proposed gene loss and the subsequent compensatory evolution as a driver of evolutionary novelty (Cañestro & Postlethwait, 2007; Farkas et al., 2022; Guijarro-Clarke et al., 2020; Helsen et al., 2020; Murray, 2020). Our work provides a compelling example of how compensatory evolution may lead to novelty by releasing evolutionary constraints. In particular, the finding that multiple genes related to microtubule dynamics become increasingly dispensable in the polarity mutants indicates that these released constraints affect not only individual genes, but entire biological processes. Interestingly, it has been suggested that gene loss can, in some cases, also promote evolvability by allowing a population to escape a local fitness peak. For example, a study by Helsen et al. (2020) discovered that some gene deletion mutants adapted faster and achieved higher fitness than a wild-type strain evolved under the same conditions. Based on this finding, we anticipate that the altered genotype-fitness relationship in the polarity mutant may allow it to adapt better to some conditions than a wild-type strain.

Pleiotropic interactions can cause similar phenotypes to emerge during evolution in different selective environments.

Cells must coordinate their behaviour with the dynamics of their environment. Achieving this coordination requires the integration of interaction networks governing cellular traits with networks responsible for sensing and responding to environmental changes. Such a coupling between networks often leads to environmental pleiotropy, which causes the effect of a mutation to differ across environments. Traditionally, it is believed that the number of evolutionary constraints increases as interaction networks become more pleiotropic, as pleiotropy increases the likelihood that the overall effect a mutation has on fitness is negative (Fisher, 1930; Orr, 2000; Wagner & Zhang, 2011). However, this concept should only be true for dynamic environments

where several of the traits affected by pleiotropy are under directional selection.

In contrast, in static environments, antagonistic pleiotropic effects between two traits would lead to the deterioration of the trait not under directional selection, but without imposing additional evolutionary constraints. Based on this understanding of the consequences of pleiotropy, the evolution of coupled traits can follow paths in static environments that are inaccessible in dynamic environments. In *chapter 3*, we explored the impact of environmental dynamics on the observed adaptive pathways using a *bem3*Δ*nrp1*Δ mutant. We demonstrated that while the deletion of Bem3 and Nrp1 causes mild detrimental effects during vegetative growth, these defects significantly aggravate upon entry into the diauxic shift. Surprisingly, we found that the recovery of diauxic growth was not dependent on evolution in a dynamic environment, as we frequently saw these phenotypes emerge in populations evolving in a static environment. Strikingly, we could link the emergence of these phenotypes in both environments to mutations in the same metabolic pathway.

These results contrast with the proposed strong connection between environmental selective pressures and the observed phenotypes after evolution. Specifically, these findings illustrate that pleiotropy could be an important structural feature of biological networks that maintains traits under directional selection regardless of the specific environmental context. Therefore, while global epistasis may make adaptive trajectories increasingly contingent on environmental conditions, pleiotropy can drive adaptation toward similar solutions across diverse environments.

Bibliography

Bajić, D., Vila, J. C. C., Blount, Z. D., & Sánchez, A. (2018). On the deformability of an empirical fitness landscape by microbial evolution. *Proceedings of the National Academy of Sciences*, *115*(44), 11286–11291. https://doi.org/10.1073/pnas.1808485115

Bank, C., Matuszewski, S., Hietpas, R. T., & Jensen, J. D. (2016). On the (un)predictability of a large intragenic fitness landscape. *Proceedings of the National Academy of Sciences*, *113*(49), 14085–14090. https://doi.org/10.1073/pnas.1612676113

Baryshnikova, A., Costanzo, M., Kim, Y., Ding, H., Koh, J., Toufighi, K., Youn, J.-Y., Ou, J., San Luis, B.-J., Bandyopadhyay, S., Hibbs, M., Hess, D., Gingras, A.-C., Bader, G. D., Troyanskaya, O. G., Brown, G. W., Andrews, B., Boone, C., & Myers, C. L. (2010). Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nature Methods*, *7*(12), 1017–1024. https://doi.org/10.1038/nmeth.1534

Butland, G., Babu, M., Díaz-Mejía, J. J., Bohdana, F., Phanse, S., Gold, B., Yang, W., Li, J., Gagarinova, A. G., Pogoutse, O., Mori, H., Wanner, B. L., Lo, H., Wasniewski, J., Christopoulos, C., Ali, M., Venn, P., Safavi-Naini, A., Sourour, N., … Emili, A. (2008). eSGA: E. coli synthetic genetic array analysis. *Nature Methods*, *5*(9), 789–795. https://doi.org/10.1038/nmeth.1239

Cañestro, C., & Postlethwait, J. H. (2007). Development of a chordate anterior–posterior axis without classical retinoic acid signaling. *Developmental Biology*, *305*(2), 522–538. https://doi.org/10.1016/j.ydbio.2007.02.032

Chen, R., Shi, L., Hakenberg, J., Naughton, B., Sklar, P., Zhang, J., Zhou, H., Tian, L., Prakash, O., Lemire, M., Sleiman, P., Cheng, W.-y., Chen, W., Shah, H., Shen, Y., Fromer, M., Omberg, L., Deardorff, M. A., Zackai, E., … Friend, S. H. (2016). Analysis of 589,306 genomes identifies individuals resilient to severe Mendelian childhood diseases. *Nature Biotechnology*, *34*(5), 531–538. https://doi.org/10.1038/nbt.3514

Costanzo, M., VanderSluis, B., Koch, E. N., Baryshnikova, A., Pons, C., Tan, G., Wang, W., Usaj, M., Hanchard, J., Lee, S. D., Pelechano, V., Styles, E. B., Billmann, M., van Leeuwen, J., van Dyk, N., Lin, Z.-Y., Kuzmin, E., Nelson, J., Piotrowski, J. S., … Boone, C. (2016). A global genetic interaction network maps a wiring diagram of cellular function. *Science*, *353*(6306), aaf1420–aaf1420. https://doi.org/10.1126/science.aaf1420

de Visser, J. A. G. M., Elena, S. F., Fragata, I., & Matuszewski, S. (2018). The utility of fitness landscapes and big data for predicting evolution. *Heredity*, *121*(5), 401–405. https://doi.org/10.1038/s41437-018-0128-4

Drubin, D. G., Mulholland, J., Zhu, Z., & Botstein, D. (1990). Homology of a yeast actin-binding protein to signal transduction proteins and myosin-I. *Nature*, *343*(6255), 288–290. https://doi.org/10.1038/343288a0

Farkas, Z., Kovács, K., Sarkadi, Z., Kalapis, D., Fekete, G., Birtyik, F., Ayaydin, F., Molnár, C., Horváth, P., Pál, C., & Papp, B. (2022). Gene loss and compensatory evolution promotes the emergence of morphological novelties in budding yeast. *Nature Ecology & Evolution*, *6*(6), 763–773. https://doi.org/10.1038/s41559-022-01730-1

Ferretti, L., Schmiegelt, B., Weinreich, D., Yamauchi, A., Kobayashi, Y., Tajima, F., & Achaz, G. (2016). Measuring epistasis in fitness landscapes: The correlation of fitness effects of mutations. *Journal of Theoretical Biology*, *396*, 132–143. https://doi.org/10.1016/j.jtbi.2016.01.037

Fisher, R. A. (1930). *The genetical theory of natural selection*. Clarendon Press. https://doi.org/10.5962/bhl.title.27468

Gawronski, J. D., Wong, S. M. S., Giannoukos, G., Ward, D. V., & Akerley, B. J. (2009). Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for Haemophilus genes required in the lung. *Proc Natl Acad Sci USA*, *106*(38), 16422–16427. https://doi.org/10.1073/pnas.0906627106

Goodman, A. L., McNulty, N. P., Zhao, Y., Leip, D., Mitra, R. D., Lozupone, C. A., Knight, R., & Gordon, J. I. (2009). Identifying Genetic Determinants Needed to Establish a Human Gut Symbiont in Its Habitat. *Cell Host & Microbe*, *6*(3), 279–289. https://doi.org/10.1016/j.chom.2009.08.003

Gorter, F. A., Aarts, M. G. M., Zwaan, B. J., & De Visser, J. A. G. M. (2018). Local fitness landscapes predict yeast evolutionary dynamics in directionally changing environments. *Genetics*, *208*(1), 307–322. https://doi.org/10.1534/GENETICS.117.300519/-/DC1

Gould, S. J. (1989). *Wonderful life: the Burgess Shale and the nature of history*. WW Norton & Company.

Guijarro-Clarke, C., Holland, P. W. H., & Paps, J. (2020). Widespread patterns of gene loss in the evolution of the animal kingdom. *Nature Ecology & Evolution*, *4*(4), 519–523. https://doi.org/10.1038/s41559-020-1129-2

Helsen, J., Voordeckers, K., Vanderwaeren, L., Santermans, T., Tsontaki, M., Verstrepen, K. J., & Jelier, R. (2020). Gene Loss Predictably Drives Evolutionary Adaptation. *Molecular Biology and Evolution*, *37*(10), 2989–3002. https://doi.org/10.1093/MOLBEV/MSAA172

Hogan, D. J., Riordan, D. P., Gerber, A. P., Herschlag, D., & Brown, P. O. (2008). Diverse RNA-Binding Proteins Interact with Functionally Related Sets of RNAs, Suggesting an Extensive Regulatory System (Sean R. Eddy, Ed.). *PLoS Biology*, *6*(10), e255. https://doi.org/10.1371/journal.pbio.0060255

Johnson, M. S., Reddy, G., & Desai, M. M. (2023). Epistasis and evolution: recent advances and an outlook for prediction. *BMC Biology*, *21*(1), 120. https://doi.org/10.1186/s12915-023-01585-3

Jost, A. P.-T., & Weiner, O. D. (2015). Probing Yeast Polarity with Acute, Reversible, Optogenetic Inhibition of Protein Function. *ACS Synthetic Biology*, *4*(10), 1077–1085. https://doi.org/10.1021/acssynbio.5b00053

Kang, P. J., Miller, K. E., Guegueniat, J., Beven, L., & Park, H.-O. (2018). The shared role of the Rsr1 GTPase and Gic1/Gic2 in Cdc42 polarization. *Molecular Biology of the Cell*, *29*(20), 2359–2369. https://doi.org/10.1091/mbc.e18-02-0145

Kinsler, G., Geiler-Samerotte, K., & Petrov, D. A. (2020). Fitness variation across subtle environmental perturbations reveals local modularity and global pleiotropy of adaptation. *eLife*, *9*, 1–52. https://doi.org/10.7554/eLife.61271

Koch, E. N., Costanzo, M., Bellay, J., Deshpande, R., Chatfield-Reed, K., Chua, G., D'Urso, G., Andrews, B. J., Boone, C., & Myers, C. L. (2012). Conserved rules govern genetic interaction degree across species. *Genome Biology*, *13*(7), R57. https://doi.org/10.1186/gb-2012-13-7-r57

Kogenaru, M., de Vos, M. G. J., & Tans, S. J. (2009). Revealing evolutionary pathways by fitness landscape reconstruction. *Critical Reviews in Biochemistry and Molecular Biology*, *44*(4), 169–174. https://doi.org/10.1080/10409230903039658

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

Laan, L., Koschwanez, J. H., & Murray, A. W. (2015). Evolutionary adaptation after crippling cell polarization follows reproducible trajectories. *eLife*, *4*. https://doi.org/10.7554/elife.09638

Langridge, G. C., Phan, M.-D. D., Turner, D. J., Perkins, T. T., Parts, L., Haase, J., Charles, I., Maskell, D. J., Peters, S. E., Dougan, G., Wain, J., Parkhill, J., & Turner, A. K. (2009). Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. *Genome Research*, *19*(12), 2308–2316. https://doi.org/10.1101/gr.097097.109

Lind, P. A., Libby, E., Herzog, J., & Rainey, P. B. (2019). Predicting mutational routes to new adaptive phenotypes (P. J. Wittkopp, Ed.). *eLife*, *8*, e38822. https://doi.org/10.7554/eLife.38822

Lobkovsky, A. E., & Koonin, E. V. (2012). Replaying the Tape of Life: Quantification of the Predictability of Evolution. *Frontiers in Genetics*, *3*, 246. https://doi.org/10.3389/fgene.2012.00246

Michel, A. H., Hatakeyama, R., Kimmig, P., Arter, M., Peter, M., Matos, J., De Virgilio, C., & Kornmann, B. (2017). Functional mapping of yeast genomes by saturated transposition. *eLife*, *6*. https://doi.org/10.7554/eLife.23570

Michel, A. H., & Kornmann, B. (2022). SAturated Transposon Analysis in Yeast (SATAY) for Deep Functional Mapping of Yeast Genomes. In *Methods in molecular biology (clifton, n.j.)* (pp. 349–379). Humana Press Inc. https://doi.org/10.1007/978-1-0716-2257-5{_}20

Michelot, A., Grassart, A., Okreglak, V., Costanzo, M., Boone, C., & Drubin, D. G. (2013). Actin Filament Elongation in Arp2/3-Derived Networks Is Controlled by Three Distinct Mechanisms. *Developmental Cell*, *24*(2), 182–195. https://doi.org/10.1016/J.DEVCEL.2012.12.008

Miller, K. E., Kang, P. J., & Park, H. O. (2020). Regulation of Cdc42 for polarized growth in budding yeast. *Microbial Cell*, *7*(7), 175. https://doi.org/10.15698/MIC2020.07.722

Miller, K. E., Lo, W.-C., Chou, C.-S., & Park, H.-O. (2019). Temporal regulation of cell polarity via the interaction of the Ras GTPase Rsr1 and the scaffold protein Bem1 (R. Li, Ed.). *Molecular Biology of the Cell*, *30*(20), 2543–2557. https://doi.org/10.1091/mbc.E19-02-0106

Moran, K. D., Kang, H., Araujo, A. V., Zyla, T. R., Saito, K., Tsygankov, D., & Lew, D. J. (2019). Cell-cycle control of cell polarity in yeast. *Journal of Cell Biology*, *218*(1), 171–189. https://doi.org/10.1083/JCB.201806196

Murray, A. W. (2020). Can gene-inactivating mutations lead to evolutionary novelty? *Current Biology*, *30*(10), R465–R471. https://doi.org/10.1016/J.CUB.2020.03.072

Orgogozo, V. (2015). Replaying the tape of life in the twenty-first century. *Interface Focus*, *5*(6), 20150057. https://doi.org/10.1098/rsfs.2015.0057

Orr, H. A. (2000). ADAPTATION AND THE COST OF COMPLEXITY. *Evolution*, *54*(1), 13–20. https://doi.org/10.1111/j.0014-3820.2000.tb00002.x

Poelwijk, F. J., Kiviet, D. J., Weinreich, D. M., & Tans, S. J. (2007). Empirical fitness landscapes reveal accessible evolutionary paths. *Nature*, *445*(7126), 383–386. https://doi.org/10.1038/nature05451

Poelwijk, F. J., Socolich, M., & Ranganathan, R. (2019). Learning the pattern of epistasis linking genotype and phenotype in a protein. *Nature Communications*, *10*(1), 4213. https://doi.org/10.1038/s41467-019-12130-8

Quintero-Monzon, O., Rodal, A. A., Strokopytov, B., Almo, S. C., & Goode, B. L. (2005). Structural and Functional Dissection of the Abp1 ADFH Actin-binding Domain Reveals Versatile In Vivo Adapter Functions. *Molecular Biology of the Cell*, *16*(7), 3128–3139. https://doi.org/10.1091/mbc.e05-01-0059

Reddy, G., & Desai, M. M. (2021). Global epistasis emerges from a generic model of a complex trait. *eLife*, *10*. https://doi.org/10.7554/eLife.64740

Slaughter, B. D., Das, A., Schwartz, J. W., Rubinstein, B., & Li, R. (2009). Dual Modes of Cdc42 Recycling Fine-Tune Polarized Morphogenesis. *Developmental Cell*, *17*(6), 823–835. https://doi.org/10.1016/j.devcel.2009.10.022

Szamecz, B., Boross, G., Kalapis, D., Kovács, K., Fekete, G., Farkas, Z., Lázár, V., Hrtyan, M., Kemmeren, P., Groot Koerkamp, M. J. A., Rutkai, E., Holstege, F. C. P., Papp, B., & Pál, C. (2014). The Genomic Landscape of Compensatory Evolution (N. H. Barton, Ed.). *PLoS Biology*, *12*(8), e1001935. https://doi.org/10.1371/journal.pbio.1001935

Szendro, I. G., Schenk, M. F., Franke, J., Krug, J., & de Visser, J. A. G. M. (2013). Quantitative analyses of empirical fitness landscapes. *Journal of Statistical Mechanics: Theory and Experiment*, *2013*(01), P01005. https://doi.org/10.1088/1742-5468/2013/01/P01005

Tong, A. H. Y., & Boone, C. (2006). Synthetic Genetic Array Analysis in <I>Saccharomyces cerevisiae</I>. In W. Xiao (Ed.), *Yeast protocols* (pp. 171–192). Humana Press. https://doi.org/10.1385/1-59259-958-3:171

Tong, A. H. Y., Lesage, G., Bader, G. D., Ding, H., Xu, H., Xin, X., Young, J., Berriz, G. F., Brost, R. L., Chang, M., Chen, Y., Cheng, X., Chua, G., Friesen, H., Goldberg, D. S., Haynes, J., Humphries, C., He, G., Hussein, S., … Boone, C. (2004). Global Mapping of the Yeast Genetic Interaction Network. *Science*, *303*(5659), 808–813. https://doi.org/10.1126/science.1091317

Tosti, E., Katakowski, J. A., Schaetzlein, S., Kim, H.-S., Ryan, C. J., Shales, M., Roguev, A., Krogan, N. J., Palliser, D., Keogh, M.-C., & Edelmann, W. (2014). Evolutionarily conserved genetic interactions with budding and fission yeast MutS identify orthologous relationships in mismatch repair-deficient cancer cells. *Genome Medicine*, *6*(9), 68. https://doi.org/10.1186/s13073-014-0068-4

Usaj, M., Tan, Y., Wang, W., VanderSluis, B., Zou, A., Myers, C. L., Costanzo, M., Andrews, B., & Boone, C. (2017). TheCellMap.org: A Web-Accessible Database for Visualizing and Mining the Global Yeast Genetic Interaction Network. *G3 Genes|Genomes|Genetics*, *7*(5), 1539–1549. https://doi.org/10.1534/g3.117.040220

van Leeuwen, J., Boone, C., & Andrews, B. J. (2017). Mapping a diversity of genetic interactions in yeast. *Current opinion in systems biology*, *6*, 14. https://doi.org/10.1016/J.COISB.2017.08.002

van Leeuwen, J., Pons, C., Mellor, J. C., Yamaguchi, T. N., Friesen, H., Koschwanez, J., Ušaj, M. M., Pechlaner, M., Takar, M., Ušaj, M., VanderSluis, B., Andrusiak, K., Bansal, P., Baryshnikova, A., Boone, C. E., Cao, J., Cote, A., Gebbia, M., Horecka, G., … Boone, C. (2016). Exploring genetic suppression interactions on a global scale. *Science*, *354*(6312). https://doi.org/10.1126/science.aag0839

van Leeuwen, J., Pons, C., Tan, G., Wang, J. Z., Hou, J., Weile, J., Gebbia, M., Liang, W., Shuteriqi, E., Li, Z., Lopes, M., Ušaj, M., Dos Santos Lopes, A., van Lieshout, N., Myers, C. L., Roth, F. P., Aloy, P., Andrews, B. J., & Boone, C. (2020). Systematic analysis of bypass suppression of essential genes. *Molecular Systems Biology*, *16*(9), e9828. https://doi.org/10.15252/msb.20209828

van Opijnen, T., Bodi, K. L., & Camilli, A. (2009). Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nature Methods*, *6*(10), 767–772. https://doi.org/10.1038/nmeth.1377

Wagner, G. P., & Zhang, J. (2011). The pleiotropic structure of the genotype–phenotype map: the evolvability of complex organisms. *Nature Reviews Genetics*, *12*(3), 204–213. https://doi.org/10.1038/nrg2949

Weinreich, D. M., Lan, Y., Jaffe, J., & Heckendorn, R. B. (2018). The Influence of Higher-Order Epistasis on Biological Fitness Landscape Topography. *Journal of Statistical Physics*, *172*(1), 208–225. https://doi.org/10.1007/s10955-018-1975-3

Weinreich, D. M., Watson, R. A., & Chao, L. (2005). PERSPECTIVE:SIGN EPISTASIS AND GENETIC CONSTRAINT ON EVOLUTIONARY TRAJECTORIES. *Evolution*, *59*(6), 1165–1174. https://doi.org/10.1554/04-272

Winzeler, E. A., Shoemaker, D. D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J. D., Bussey, H., Chu, A. M., Connelly, C., Davis, K., Dietrich, F., Dow, S. W., El Bakkoury, M., Foury, F., Friend, S. H., Gentalen, E., … Davis, R. W. (1999). Functional Characterization of the S. cerevisiae Genome by Gene Deletion and Parallel Analysis. *Science*, *285*(5429), 901–906. https://doi.org/10.1126/science.285.5429.901

Woods, B., Kuo, C.-C., Wu, C.-F., Zyla, T. R., & Lew, D. J. (2015). Polarity establishment requires localized activation of Cdc42. *Journal of Cell Biology*, *211*(1), 19–26. https://doi.org/10.1083/jcb.201506108

Yang, G., Anderson, D. W., Baier, F., Dohmen, E., Hong, N., Carr, P. D., Kamerlin, S. C. L., Jackson, C. J., Bornberg-Bauer, E., & Tokuriki, N. (2019). Higher-order epistasis shapes the fitness landscape of a xenobiotic-degrading enzyme. *Nature Chemical Biology*, *15*(11), 1120–1128. https://doi.org/10.1038/s41589-019-0386-3

Zhou, J., Wong, M. S., Chen, W.-C., Krainer, A. R., Kinney, J. B., & McCandlish, D. M. (2022). Higher-order epistasis and phenotypic prediction. *Proceedings of the National Academy of Sciences*, *119*(39). https://doi.org/10.1073/pnas.2204233119

Acknowledgements

This work is the result of the contributions made by many people, both directly and indirectly, and I am very grateful to all of them.

First of all, Liedewij, thank you for not only giving me the opportunity to work in your group, but also for your guidance, mentoring, trust and endless enthusiasm. I admire how you have been able to bring together a group of people that "just want to do science and have fun". The past years have been a great learning experience and I am grateful that I have had the opportunity to spend that time in your lab. I would also like to thank Timon for being my promoter and for his feedback on this thesis. I thank all the other members of my committee for their assessment of my work. I thank Marianne, Jasmijn, Benoît, Agnès and Johan for their help and discussions on the experiments and data analysis.

I thank my fellow lab members, Leila, Christine, Els, Sophie, Fayezeh, Ramon, Frank, Shazia, Marieke, Nynke, Esengül, Caspar, Werner, Reza, Mathia, Eveline and Eve for their help and companionship. Leila, as my fellow SATAY'er, thank you for your laughter, games of domino, your salsa?!? –yes, salsa–, your warmth and kindness. It has truly been a pleasure to work with you and I treasure all the fun moments we have had. Marieke, Nynke and Caspar, I will never forget our amazing trip through the US. Sophie, thank you for being the creative force of the lab, as is showcased by the interior design of this thesis, and for providing me with a unique cell division mug. Thank you Marieke for providing the cover of this thesis and continuing the *in vivo* part of the lab, I know you will do great. Nynke, thank you for never giving up on asking if anyone from "that office" will join for lunch. Caspar, thank you for your fun discussions and the best of luck with finding the PhD position of your dreams. Christine, thanks for introducing and quizzing us on classical music and for convincing Liedewij that we should make a trip to Lausanne ;). Els, thank you for keeping a strict organization of the lab meetings. Ramon, thank you for helping with strain constructions, keeping the lab organized and your trade-mark style of jokes which I somehow managed to survive for several years. Frank and Esengül, thank you for keeping the lab up and running. Shazia, thank you for the amazing dinners and creative poems. Eveline, when we started the evolution project together, I could never have imaged that it would lead to this! Thank you for your guidance, and I'm happy to see that you have realised your dream of having your own ecological farm. Werner, thank you for being a friend and sharing your knowledge and many talents with us.

I thank my former students –Wessel, Ingmar, Floor, Nina, Maaike, Pim, Albert, Antoine, Germain and Pam – for their contributions to this work and being brave enough to join me on this project. I thank Valentina, Thomas, Esther, Renske, Zhiheng, Miranda, Gregory, Felix, Zaïda, Thijs, Constant, Laurie, Vlad, Julia, Daphne, Rosanne and Bram, for contributing to the social atmosphere of the group.

I am grateful to Joyce for her care, kindness and hospitality. I thank Cátia, Fayezeh and Sadaf for the intense games of table football and being an amazing group of friends. Cátia, thank you for trying to teach me Portuguese. Fayezeh and Sadaf, thank you for introducing me to Nowruz and the other traditions of the Iranian culture. Nuo, thank you for all your help inside and outside

the lab. You are one of the most skilled people I know, and what I've learned from you has greatly contributed to this work. I'm happy to have had the privilege to work with you and have you as a friend. I thank Hugo and Evril for always making time for a chat when we meet in the hallways and maintaining the building in good condition. I thank Anke and Jan for their friendliness, keeping the labs organized and autoclaving the large amounts of liquid waste that I often produced.

I thank Reza, Diederik, Theo, Hirad, Milan, Lars Yiteng, Diego, Daniel, Aleksandre, Miloš, Federico, Ana, Hidde, Beatriz, Antoni and all others in BN for making this such a fun and active department. I have enjoyed taking part in all the social activities, often accompanied by live music from the in-house band, that have taken place in the past years.

I thank my family and friends for their unconditional love and support.

Curriculum Vitæ

Enzo Kingma

18-05-1993 Born in Willemstad, Curaçao.

Education

2017-2023 PhD. in Biophysics
Delft University of Technology, the Netherlands
Faculty of Applied Sciences, Department of Bionanoscience
Thesis: The path towards prediction: An exploration of the evolutionary dynamics of the yeast polarity network
Promotor: Dr. ir. L. Laan
Promotor: Dr. T. Idema

2014-2016 M. Sc. in Applied Physics
Delft University of Technology, the Netherlands

2011-2014 B. Sc. in Applied Physics
Delft University of Technology, the Netherlands

List of Publications

In this thesis

2. **E. Kingma**, E.T. Diepeveen, L. Iñigo de la Cruz, L. Laan, *Pleiotropy drives evolutionary repair of the responsiveness of polarized cell growth to environmental cues*, Frontiers in Microbiology **14**, (2023).

1. P. Nghe, M.G.J. de Vos, **E. Kingma**, M. Kogenaru, F.J. Poelwijk, Laan L, S.J. Tans,*Predicting Evolution Using Regulatory Architecture*, Annual Review of Biophysics **49**, (2020)

Other

1. F. Wu, J. Halatek, M. Reiter, **E. Kingma**, E. Frey, C. Dekker, *Multistability and dynamic transitions of intracellular Min protein patterns*, Molecular Systems Biology, **12(6)**, (2016)