University of Szeged
Faculty of Medicine
Doctoral School of Multidisciplinary Medicine

Compensatory evolution as a driver of morphological novelties

PhD Thesis

Zsuzsa Sarkadi

Supervisor:

Balázs Papp PhD, principal investigator
Biological Research Centre, Institute of Biochemistry
Lendület Laboratory of Computational Systems Biology

Szeged
2022

TABLE OF CONTENTS

TABLE OF CONTENTS 2
1. LIST OF PUBLICATIONS AND CONFERENCE ABSTRACTS 4
2. SUMMARY 6
3. INTRODUCTION 8
3.1. Compensatory evolution 8
3.2. Morphological evolution of microbes 11
3.2.1. Single cell morphology 11
3.2.2. Cell cycle dependency of yeast cell morphogenesis 11
3.2.3. Multicellular morphology 13
3.3. Molecular mechanism behind the transition to multicellularity 14
3.4. Ecological and clinical implications of multicellular structures 15
4. AIMS OF THE THESIS 17
5. RESULTS 18
5.1. Rapid evolution of cellular morphology in the laboratory 18
5.2. Comparable morphological diversity of compensated and natural strains 22
5.3. Rapid evolution of multicellular phenotypes 26
5.4. Synergistic epistasis at the level of morphology 30
6. DISCUSSION 33
7. METHODS 37
7.1. Yeast strains and media 37
7.2. High-throughput single-cell morphology measurements 38
7.3. Quality control of raw morphological data 39
7.4. Statistical analysis of morphological data 40
7.5. High-throughput DNA content analysis 42
7.6. Cell cycle analysis by flow cytometry 43
7.7. Strain construction for generating a deletion of *WHI2* in the wild-type and *Δrpb9* ancestor strain 43
7.8. Strain construction for introducing a point mutation into the wild type and Δbub3 ancestor strain 44
7.9. Measuring the activities of the filamentous response and *FLO11* promoter 44
7.10. Quantitative invasive growth assay 45
7.11. Imaging of agar-embedded filamentous structures upon invasive growth 47
7.12. Quantitative settling assay to detect multicellular aggregates 48

7.13. Quantitative biofilm formation assay 49
8. ACKNOWLEDGEMENTS 50
9. REFERENCES 51

1. LIST OF PUBLICATIONS AND CONFERENCE ABSTRACTS

Publication related to Ph.D. thesis:

Farkas, Z. †, Kovács, K. †, **Sarkadi, Z.** †, Kalapis, D., Fekete, G., Birtyik, F., Ayaydin, F., Molnár, C., Horváth, P., Pál, C. and Papp, B. (2022). Gene loss and compensatory evolution promotes the emergence of morphological novelties in budding yeast. Nat. Ecol. Evol. 1–11. https://doi.org/10.1038/s41559-022-01730-1.

† These authors contributed equally

Conference abstracts:

Z. Sarkadi*, Z. Farkas, K. Kovács, G. Fekete, C. Molnár, P. Horváth, D. Kalapis, Z. Bódi, C. Pál, B. Papp, Compensatory Mutations Drive Morphological Evolution, SMBE (Society for Molekular Biology and Evolution) in Manchester, UK, 23 July 2019.

Zsuzsa Sarkadi, Zoltán Farkas, Károly Kovács, Gergely Fekete, Dorottya Kalapis, Fanni Birtyik, Csaba Molnár, Péter Horváth, Ferhan Ayaydin, Csaba Pál, Balázs Papp, Deleterious mutations as drivers of morphological evolution, EMBL Conference: Molecular Mechanisms in Evolution and Ecology, 30 Sep - 2 Oct 2020.

Publications unrelated to Ph.D. thesis:

E. T. S. van, M. Forn, I. Forne, **Z. Sarkadi**, M. Capella, L. M. Caballero, S. Fischer-Burkart, C. Broenner, M. Simonetta, D. Toczyski, M. Halic, A. Imhof, and S. Braun, “Shelterin and subtelomeric DNA sequences control nucleosome maintenance and genome stability,” EMBO REPORTS, vol. 20, no. 1, 2019.

R. R. Barrales, M. Forn, P. R. Georgescu, **Z. Sarkadi**, and S. Braun, “Control of heterochromatin localization and silencing by the nuclear membrane protein Lem2,” GENES & DEVELOPMENT, vol. 30, no. 2, pp. 133–148, 2016.

Z. Sarkadi and M. Babits, “A megtermékenyülés és a spermiumalagút,” TERMÉSZET VILÁGA, vol. 135, no. 6, pp. 259–262, 2004.

2. SUMMARY

The morphology of microbes is very diverse in nature. What drives morphological evolution? According to the traditional view, organisms adapt to the changing environment. However, recent studies suggest that beside environmental change, there might be other selective pressures that shape the genetic makeup of organisms. Here we suggest an alternative scenario, where the environment is constant, and deleterious mutations induce selection pressure to compensate their harmful effects by specific mutations elsewhere in the genome, a phenomenon termed compensatory evolution. As a by-product, such genomic changes can lead to substantial divergence in morphological traits without direct selection acting on them. To test this hypothesis, we used ~150 originally slow-growing laboratory-evolved yeast strains that went through compensatory evolution.

First, we measured single cell morphology of the initial knock-out and evolved strains by quantifying ~150 traits by using high-throughput microscopy and automated image analysis. We found that new morphologies emerged rapidly; especially cell size, cell elongation and neck position angle showed the most extreme values. Furthermore, cell morphology reflects several processes throughout the cell cycle. By measuring the distribution of cell cycle phases, we found that compensated strains frequently had altered G1 and G2 phases. Cell size and cell elongation were earlier linked to G1 and G2 phases, respectively. We found that cell size negatively correlates with the fraction of cells in G1 phase. Furthermore, cell elongation positively correlates with the G2 phase, suggesting that cell elongation was caused by the expanded G2 phase.

Transition from unicellular to multicellular form is ecologically and clinically very important. Therefore, we investigated the emergence of three multicellular forms among the compensated strains. First, emergence of invasive growth (that is, the capability of fungi to penetrate into solid media or tissue) was frequent, and the level of the invasiveness was comparable with that of the measured natural isolates. Second, clump formation (that is, cell aggregation) in liquid culture was also frequently found. Moreover, we showed that these clumps might develop by incomplete cell separation from a single cell. Third, in a few cases we found increased biofilm production. The emergence of these multicellular structures is quite surprising, because they appeared without any direct selection. The present evidence suggests that these multicellular structures are by-products of compensatory evolution.

Finally, through two examples, we dissected the genetic mechanism of compensation and the linked morphological alterations. By reconstructing the single compensatory mutations, we detected synergistic interactions between deleterious and compensatory mutations.

Taken together, this work suggests that morphological evolution can be driven not only by environmental adaptation but also by compensation of deleterious mutations. Moreover, invasive growth and biofilm formation are virulence factors in pathogenic fungi, therefore our results imply that gene loss and subsequent compensatory evolution could have relevance for the evolution of pathogenicity.

3. INTRODUCTION

3.1. Compensatory evolution

During evolution, organisms adapt to a changing environment through mutations and natural selection. The traditional view is that the selected mutations are beneficial and accumulate step by step. However, the increasing numbers of sequenced genomes and comparative studies reveal that gene loss is frequent and has a substantial contribution in shaping evolution (Albalat and Cañestro, 2016). These loss-of-function mutations can become fixed through genetic drift, environmental change or hitchhiking with beneficial mutations (Albalat and Cañestro, 2016; Lang et al., 2013; Qian et al., 2012). Once fixed, such loss-of-function mutations generate selection pressure to further beneficial mutations elsewhere in the genome. This event called compensatory evolution (Figure 1).

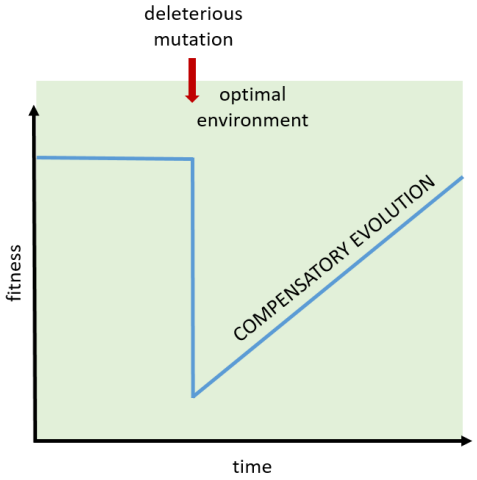


Figure 1. Scheme of compensatory evolution

Several pieces of evidence support the existence of compensatory evolution in the wild. One example is the thermostability maintenance of mitochondrial tRNA. Mitochondrial tRNAs harbor several mutations that reduce the thermostability of the secondary and tertiary structures of the molecule by causing muscular or neurological disorders in human (Taylor and Turnbull, 2005). Surprisingly, in other species, the same mutations do not cause any disease, because other base substitutions compensate the tRNA instability (Ivankov et al., 2014).

Another example for compensatory evolution is related to antibiotic resistance. Gain of resistance is usually coupled with a decrease in fitness. In the absence of antibiotics, the resistant bacteria are outcompeted by fitter susceptible strains. However, compensatory evolution can restore the fitness of the resistant strain (Andersson and Hughes, 2010).

The role of deleterious mutations in adaptive evolution has also been studied experimentally. Compensatory evolution was studied in laboratory experiments by several groups.

Their observations indicate some general conclusions. First, the fitness decrease caused by gene loss was rapidly restored in bacteria and yeast (Blank et al., 2014; Echenique et al., 2019; McCloskey et al., 2018; Szamecz et al., 2014). Second, the phenotype of the gene loss usually did not get restored, instead new phenotypes emerged. These new phenotypes were detected at the level of gene expression, metabolic networks, and viability under stressful conditions (Blank et al., 2014; Helsen et al., 2020; LaBar et al., 2020; McCloskey et al., 2018; Szamecz et al., 2014).

Suboptimal states in metabolite concentrations lead to alteration of regulatory networks whose function is subsequently changed and re-optimised during adaptive laboratory evolution (McCloskey et al., 2018). McCloskey's group performed adaptive laboratory evolution with *Escherichia coli* to study the alteration of regulatory network architectures by using metabolic gene knockout (KO) strains. In order to find the drive to increase fitness, they measured the changes in metabolite, transcript and flux levels. They found that mutations during adaptation rewired many existing metabolic pathways and/or introduced novel network functions that addressed the imbalances that the initial gene deletion created. Especially, flux distribution changed during the evolutionary experiment. For example, the flux was re-routed through the Entner–Doudoroff (ED) pathway in one of the KO strains, however the evolved strain rather used the higher energy producing glycolysis pathway instead of the ED pathway.

As mentioned before, despite the fitness compensation, gene expression state and viability under stressful conditions do not get restored during compensatory evolution. This is well demonstrated by the study with a large-scale experimental evolution performed in our laboratory few years ago, when over 180 haploid knockout yeast strains went through compensatory evolution (Szamecz et al., 2014). After ~400 generations, 68% of the initially slow-growing strains reached near wild-type fitness. Despite the fitness recovery, the wild-type gene expression pattern was generally not restored. Furthermore, to study the pleiotropic effects of fitness compensation, the evolved strains were grown at 14 environmental settings, including different nutrients and stress factors. The authors found that the evolved populations showed substantial fitness variation; both positive and negative fitness trade-offs were observed. Because the strains were not exposed to these environments during the laboratory evolution, the authors suggest that these phenotypic changes emerged as a byproduct through evolution. Similarly to McCloskey's work, they conclude that compensatory evolution can lead to changes in phenotypes. Important question is whether the

phenotypic changes found during compensatory evolution are advantageous in certain circumstances.

What kind of other phenotypes could evolve through gene loss and subsequent compensation? Single-cell morphology is influenced by thousands of genes (Ohya et al., 2005). Also, multicellular morphological features, such as invasive growth and biofilm formation are governed by several hundreds of genes (Ryan et al., 2012; Shively et al., 2013). This large mutational target size suggests that compensatory mutations could frequently occur in genes involved in morphological traits.

Single cell morphology can be measured with high accuracy and high-throughput manner (Ohya et al., 2005). Ohya and coworkers established a method that measures 254 parameters of a single cell by automated image analysis. Cells stained for cell wall, DNA, and actin are categorized into cell cycle stages (A, A1B, C), and specific area, distances and angles are measured by the program. Such parameters are, for example, the area of the mother cell and the bud, the lengths of the long and the short axis, the angle between the mother and the bud, etc. (Figure 2). The authors measured the single-cell morphology of the haploid deletion collection of *S. cerevisae*, and found that specific traits correlate with certain cellular functions. For example, polarisome mutants are enriched in mutants with altered bud morphology. Taken together, this method is ideal for detecting even small morphological changes in an unbiased way.

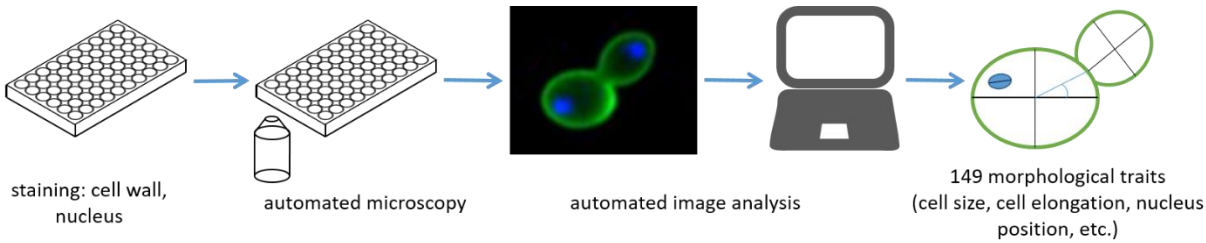


Figure 2. Workflow of single-cell morphological measurement in S. cerevisiae. *Figure shows a modified version of the original workflow that we used (without actin staining).*

3.2. Morphological evolution of microbes

3.2.1. Single cell morphology

Bacteria show a wide range of different shapes. What drives the evolution of morphological diversity? At low nutrient condition, the rod shape is more advantageous comparing to the spherical one, because the surface-volume ratio of the cell is higher, allowing higher nutrient uptake (Ojkic et al.). In contrast, the spherical form has an advantage when the cell is exposed to chemical or physical danger. According to the traditional view, organisms adapt to the changing environment during evolution. However, some recent studies show that the evolution of cell morphology is driven not just by adaptive but also by deleterious mutations. A single mutation (*mreB*) in the rod-shaped *Pseudomonas fluorescens* bacteria changed its shape to spherical form, and the same mutation decreased the fitness as well. Laboratory evolution restored the fitness, meanwhile the cell kept the spherical shape (Yulo et al., 2018). This observation suggests the existence of an alternative route for morphological evolution.

Not only bacterial cells, but also yeast cells show a high diversity of single cell morphology in nature. Natural isolates of *S. cerevisiae* show a wide range in such traits as cell size, cell elongation and bud growth direction (Yvert et al., 2013), indicating that cell morphology evolves rapidly. However, these morphologies are linked neither to the phylogenetic distance nor to the geographical or ecological environment. This observation suggests that morphological evolution is driven not only by the changing environment, but also by other forces.

3.2.2. Cell cycle dependency of yeast cell morphogenesis

Single cell morphology reflects several cellular events, such as cell size regulation, progression through the cell cycle and establishment of cell polarity (Ohya et al., 2005). Cells enter the S phase when they reach a critical cell size (Figure 3). Smaller cells need to spend more time to reach the critical size for entering the cell cycle, therefore a negative correlation can be observed between cell size and G1 phase length (Turner et al., 2012).

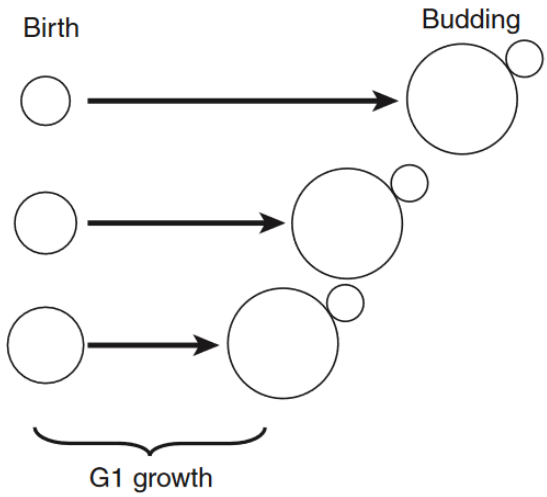


Figure 3. Cell size control in S. cerevisiae. *Negative correlation between cell size at birth and growth in G1. Source:* (Turner et al., 2012)

In budding yeast cell size can increase in two different ways depending on the direction of growth. In G1 phase the cell expands uniformly, increasing only the size, then switches to polarized growth during the S and G2 phases (Figure 4). In polarized growth, the cell or the bud grows in one direction, by transporting cell components toward the tip. Those cells that spend longer time in the polarized growth phase have elongated buds (Watanabe et al., 2009), indicating that cell cycle regulation contributes to the shape of the cell. Importantly, polarized growth is required for the yeast to hypha switch of the human pathogen *Candida albicans* (Sudbery, 2011), therefore understanding the evolution of polarized growth is clinically relevant.

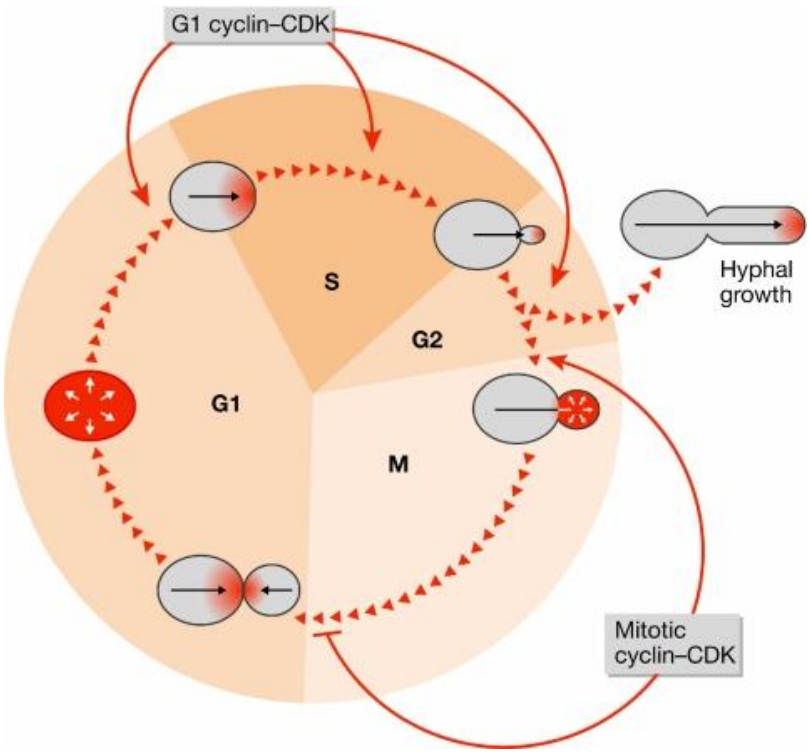


Figure 4. Cell-cycle regulation of cell polarity in budding yeast. *Source: (Yoshida and Pellman, 2008) CDK stands for cyclin dependent kinase*

3.2.3. Multicellular morphology

Unicellular organisms frequently form multicellular structures upon environmental change (Roberts and Fink, 1994). These structures are formed either by aggregation of many single cells or through a defect of cell separation. One example for cell aggregation is *biofilm* formation, which is observed in several microbes, including bacteria, archaea, fungi and microalgae (Figure 5a). Another aggregation form is *flocculation* (Soares, 2011), a characteristic of certain yeast species (Figure 5b). Structures formed by cell separation defects have several different names depending on the ploidity of the organism and the consistence of the medium. *Clumps* are cell communities that form in liquid media and were described by several studies (Kuzdzal-Fick et al., 2019; Ratcliff et al., 2012) (Figure 5 c-d). These clusters derive from a single cell through mother-daughter separation defect (Ratcliff et al., 2012). When elongated diploid yeast cells stay together, they form filamentous structures called *pseudohypha* (Madhani and Fink, 1998) (Figure 5c). When haploid yeast cells are under starvation, they can invade into solid media or tissue, a process called *invasive growth* (Figure 5e).

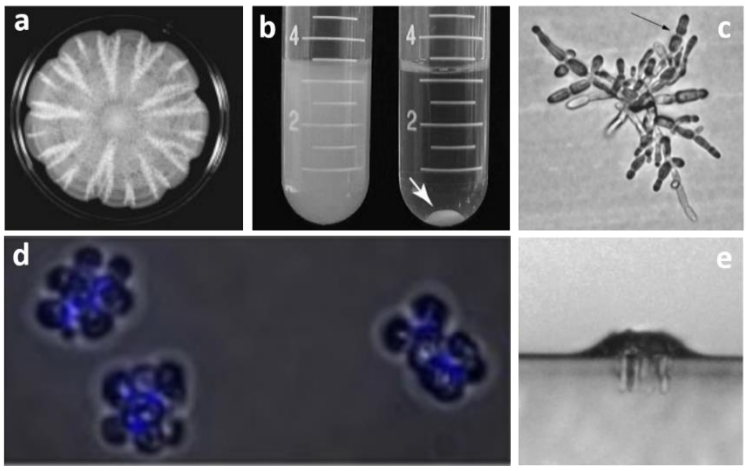


Figure 5. Multicellular structures in yeast. *a) Biofilm formation of S. cerevisae on soft-agar (0.3%) plates after 13 days. Source* (Reynolds and Fink, 2001) *b) Flocculating (right) and nonflocculating (left) S. cerevisiae strain. Source (Smukalla et al., 2008). c) Pseudohypha of diploid S. cerevisiae. Cells containing activated MAP kinase pathway and are growing in low-nitrogen medium. Arrow points to a constriction between two cells. Source:* (Madhani and Fink, 1998). *d) Clump formation by incomplete cell separation in a haploid yeast strain. Calcofluor white stained bud necks (blue) indicate the incomplete separation. Source:* (Kuzdzal-Fick et al., 2019). *e) Agar invasion of haploid S. cerevisiae S1278b cells on medium lacking glucose. Source:* (Cullen and Sprague, 2000).

How fast do these multicellular structures evolve in nature? How many independent routes of evolution are there? The levels of biofilm formation, flocculation and invasive growth show high diversity among *Saccharomyces cerevisiae* natural isolates, indicating that these multicellular traits evolve rapidly. Interestingly, the morphology of these strains does not reflect their phylogenetic distribution. Moreover, only few strains cluster phenotypically according to the niche of origin (Hope and Dunham, 2014). Some strains are not capable of adopting any of the abovementioned multicellular phenotypes. One of these is the S288C laboratory strain, therefore it is an optimal model organism for investigating the emergence of multicellular phenotypes. Rapid evolution of multicellularity was also proved experimentally, when a unicellular yeast evolved to clump-forming yeast in the laboratory by using sedimentation speed for selection (Ratcliff et al., 2012). In another laboratory evolution, yeast cells became adapted to nutrient deprivation by cell aggregation (Hope et al., 2017).

3.3. Molecular mechanism behind the transition to multicellularity

In yeast strains capable of switching from unicellular to multicellular form, the developmental transition is induced by environmental changes which lead to the activation of signaling pathways (Chavel et al., 2010; Gancedo, 2001; Ryan et al., 2012; Shively et al., 2013). Among the several multicellular structures, the filament formation mechanism was extensively studied in budding yeast. The major signaling pathway involved in filament formation is the filamentous MAPK pathway (fMAPK), which induces the transcription of several filamentation genes (Cullen and Sprague, 2012). For the transition from yeast to filamentous form, three major changes are required: (i) a change in budding pattern, (ii) cell elongation, and (iii) increased adhesion between cells (Figure 6.) (Cullen and Sprague, 2012; Roberts and Fink, 1994). In haploid yeast cells the buds normally emerge at the proximal pole (adjacent to the birth site). During filamentous growth the buds emerge at the distal pole (opposite to the birth site) in a unipolar manner (always at the distal pole). The second factor, cell elongation was shown to be the result of an alteration in the duration of cell cycle phases. For example, extended G2 resulted in elongation of the cells (Gladfelter et al., 2005). The third factor is cell-to-cell connection, which is mediated either by increased adhesin expression or by daughter separation defect (Vandermeulen and Cullen,

2020). The major adhesin involved in filamentous growth is Flo11, which is induced by glucose starvation in haploids and nitrogen starvation in diploids.

Enhanced expression of adhesins is required not only for filamentous growth, but also for biofilm formation and flocculation (Brückner and Mösch, 2012). In addition, biofilms consist of extracellular matrix, which prevents the cells from external dangers, such as desiccation, osmotic shock, UV-radiation, pH change, and exposure to antimicrobial agents (Pandit et al., 2020).

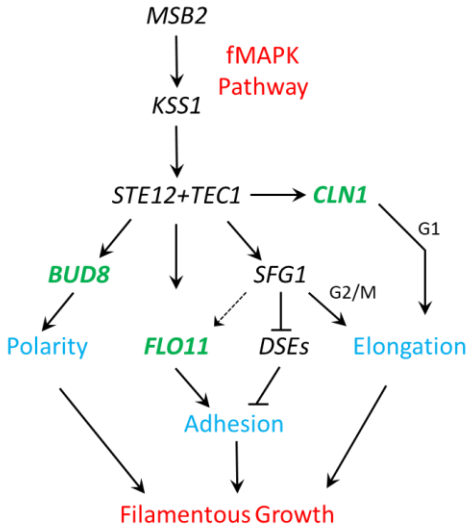


Figure 6. fMAPK pathway induces target genes (green) required for filamentous growth. *Source: (Vandermeulen and Cullen, 2020) Original figure was reconstructed with minor modifications*

3.4. Ecological and clinical implications of multicellular structures

Many fungi change their morphology when the environment changes. Dimorphic fungi can switch between yeast and hyphae (Gow et al., 2002).

In many pathogenic fungi hyphae are responsible for penetrating the host either through natural openings or by tissue invasion. Once they are in the host, they can go through morphogenesis, often under quorum-sensing regulation. This promotes the switch to yeast form, which can spread in the host more efficiently (Sharma and Jangid, 2016). By contrast, filamentous hyphae are better able to grow invasively and disseminate into tissues. At the same time, long hyphal cells are too large to be attacked by macrophages. Some species switch from filamentous hyphae to budding yeast as they enter the host, whereas others do the opposite way (Figure 7) (Min et al., 2020).

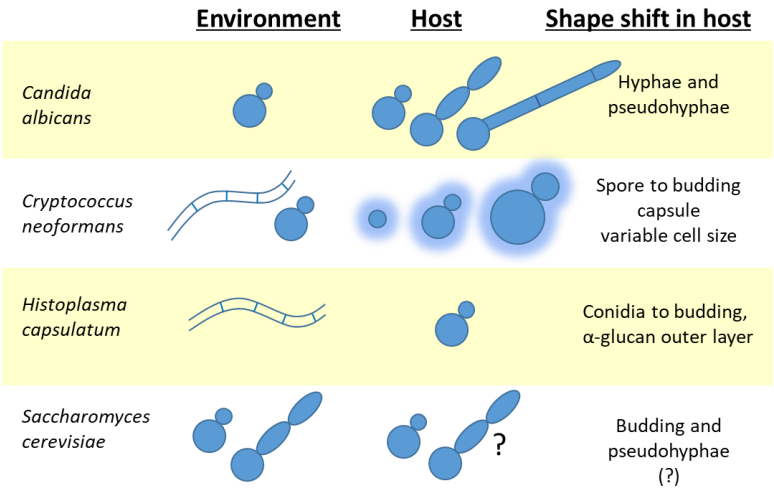


Figure 7. Morphological change in fungal pathogens. *Source (Min et al., 2020) Original figure was reconstructed and modified. S. cerevisiae was also added as an opportunistic pathogen. Question mark indicates the lack of sufficient evidence for the pseudohyphae form in host.*

Another clinically relevant multicellular structure is the biofilm. Biofilms can be resistant to antimicrobial agents and the host's immune system. Biofilms can develop on medical devices, such as cardiac pacemakers, artificial joints, catheters, endotracheal tubes etc. Furthermore, cells may separate from the biofilm and spread through the blood vessels, thereby aggravating the infection. Many medically important fungi produce biofilms, including *Candida*, *Aspergillus*, *Cryptococcus*, *Trichosporon*, *Coccidioides* and *Pneumocystis* (Fanning and Mitchell, 2012). Because biofilms are often resistant to antimicrobial agents, it is challenging to treat infections involving biofilm formation.

Understanding the underlying molecular diversity of filamentous growth and biofilm formation can help develop novel strategies for treatment.

4. AIMS OF THE THESIS

Here we hypothesize that deleterious mutations followed by compensatory evolution initiate cellular and multicellular morphological changes without direct selection on them. To test this hypothesis we aimed to systematically study the effect of compensatory evolution on the emergence of new morphologies.

To achieve our goals we used a set of isogenic haploid *Saccharomyces cerevisiae* strains that had gone through laboratory compensatory evolution (Szamecz et al., 2014). Each of these strains was initially slow-growing due to having one single gene deletion. These genes were involved in widespread biological processes. Here we measured the morphology of 86 slow-growing single-gene deletion strains (referred to as ancestor strains), and the corresponding 142 parallelly evolved lines (referred to as compensated strains). To control for potential adaptation to the media during the evolutionary experiment, 21 wild-type strains were also propagated. In the earlier work, all of the 142 compensated strains showed larger fitness increase than the control wild-type evolved lines (Szamecz et al., 2014). Our analyses addressed the following specific questions:

- By measuring single-cell morphology using high-throughput microscopy and automated image analysis, we systematically tested whether compensatory evolution generated novel cellular morphologies (morphotypes).
- Next, by cell cycle analysis of the ancestor and compensated strains using flow-cytometry we investigated whether the cell cycle phase alterations correlate with single-cell morphological traits, such as cell size and cell elongation.
- By measuring three ecologically important multicellular traits, we tested whether compensatory evolution leads to the emergence of invasive growth, clump formation or increased biofilm formation.
- Then, by measuring both single-cell morphology and multicellular morphology of yeast strains derived from nature, we tested whether the natural strains have comparable morphology to that of the laboratory-evolved strains.
- Finally, by measuring genetic interactions at the level of morphology using two examples, we tested the existence of epistasis between gene deletion and compensatory mutation.

I wish to emphasize that these analyses were jointly carried out with my co-authors. I was involved in all major analyses and played a leading role in the implementation and measurement of multicellular phenotypes and cell cycle phases.

5. RESULTS[1]

5.1. Rapid evolution of cellular morphology in the laboratory

We quantitatively measured single-cell morphology of four major classes of strains, including the wild-type, the evolved control (N=21), the ancestor (N=86) and the corresponding compensated strains (N=142). For the morphological measurements we isolated single clones from the compensated populations. We confirmed that all investigated strains had remained haploid (Methods). The haploid state of the strains is necessary for comparing their morphology, because diploid cellular morphology differs in many aspects from the haploid one. We applied an established protocol designed for quantitative morphological phenotyping (Ohya et al., 2005; Okada et al., 2015) that is well-suited for large-scale analyses (Yvert et al., 2013; Bauer et al., 2015). By using image analysis, we extracted 149 single-cell morphological traits that describe various morphological attributes of exponentially growing cell populations, including average cell size, bud size and bud growth direction (see Methods).

To allow unbiased comparison of different strains in the high-dimensional morphological space, we accounted for intrinsic correlations between traits using principal component analysis (PCA) (Suzuki et al., 2018; Yvert et al., 2013) (see Methods). This resulted in 8 morphological principal components (PCs), together explaining ~96% of the morphological variance. For an intuitive understanding of the morphological PCs, see Figure 8. We found that 71% of the 142 compensated strains and 66% of the 86 ancestors display significant changes in morphological PCs compared to the wild-type (Figure 9, Methods).

[1] Results text is reproduced with permission from Springer Nature

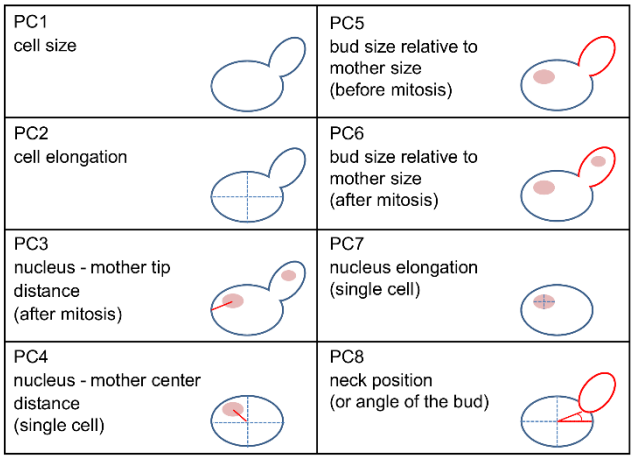


Figure 8. Compensatory evolution alters cellular morphology.
Schematic representation of the principal components. *The figure shows the intuitive meaning of the first 8 principal components (PC), based on the top contributing morphological traits. Separate morphological traits are defined for three cell cycle stages (single cell, budded cell before or after mitosis). PCs without an indication of a specific cell cycle stage can be linked to all cell cycle stages for a given type of trait.*

Compensatory evolution may shape cellular morphology in three different manners: it i) may restore the wild-type morphology initially perturbed by gene loss, ii) may retain the morphological change initiated by gene loss, without invoking any further changes or iii) may generate novel morphological alterations (Figure 9).

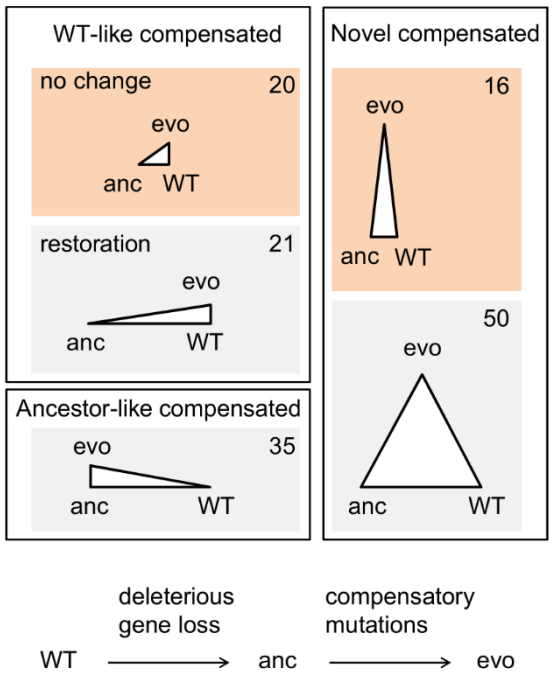


Figure 9. Main classes of evolutionary trajectories in morphological space and their prevalence. *Compensated strains may not differ significantly from the wild-type ('WT-like compensated'), differ from WT while being similar to the ancestor ('Ancestor-like compensated') or differ from both the WT and the ancestor ('Novel compensated'). Compensated strains with wild type-like ancestors are colored with orange background. For each triplet of genotypes involving a compensated strain (evo), morphological differences between genotypes are visualized by a triangle (long and short edges denote presence and absence, respectively, of significant difference in at least one principal component). Numbers indicate the number of compensated strains corresponding to each evolutionary trajectory. The bottom flowchart depicts the two steps of compensatory evolution: (i) deleterious gene loss that results in a low fitness state (anc) and (ii) accumulation of subsequent compensatory mutations that mitigate the fitness defect of gene loss and result in a fitness-compensated strain (evo).*

To investigate these possibilities, significant changes in morphological PCs were calculated between the wild-type, the ancestral and the compensated strain carrying the same gene deletion (Fig. 9, at a false discovery rate of 2.53%, see Methods). Generally, the cellular morphology of the compensated and the wild-type strains significantly differed from each other, hence restoration of the wild-type morphology was relatively rare (15% of the 142 cases, Fig. 9). In 25% of cases, the ancestral and the corresponding compensated strains showed similar cellular morphologies, but they substantially differed from that of the wild-type. Hence, in these cases, compensatory evolution improved fitness, but left the knock-out's cellular morphology unaltered. In 46% of the cases, the compensated strains displayed markedly different morphologies compared to that of the wild-type and the corresponding ancestral strain as well. Additionally, when only ancestral knock-out strains with wild-type morphology were considered, compensatory evolution generated novel morphological states in 16 out of the 36 cases (44%). Importantly, the morphological differences between the compensated and the ancestor strains are also apparent when specific traits representative of each PC were considered individually (Figure 10). Finally, hierarchical clustering of the morphological profiles coupled with a bootstrap analysis (Suzuki and Shimodaira, 2006) identified 11 morphologically distinct groups of compensated strains with statistical support, indicating the existence of multiple distinct morphotypes. Together, these results suggest that compensatory evolution following gene loss frequently promotes the evolutionary divergence of morphological traits despite yielding wild-type like fitness.

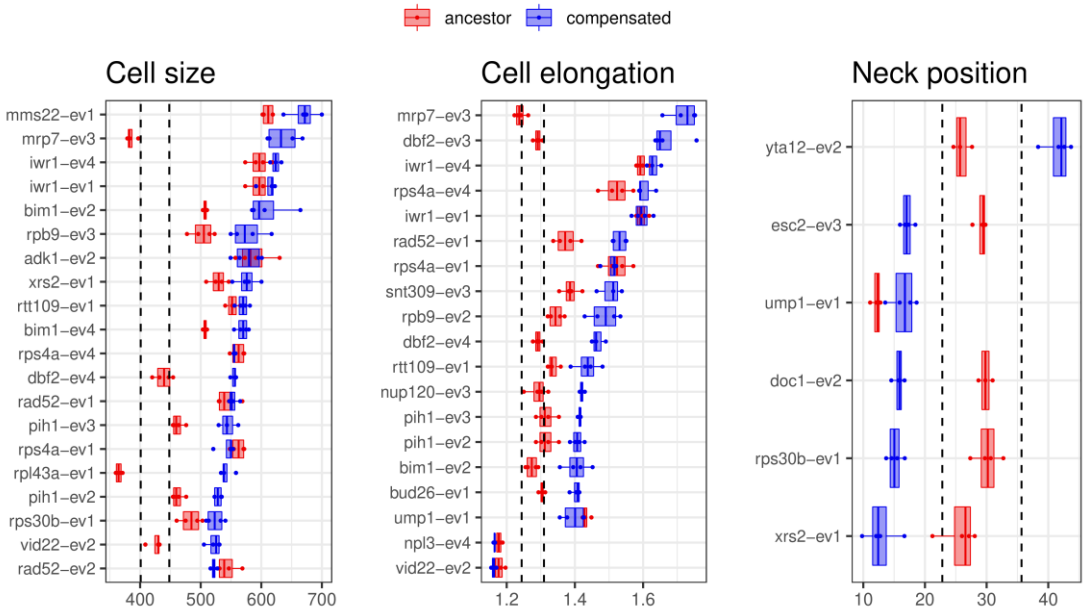


Figure 10. Changes in three morphological traits in compensated strains with the most extreme trait values. *Boxplots showing cell size, cell elongation and neck position values of compensated strains that differ most from the wild-type (based on 3 or 4 biological replicates per strain). Cell size (area in pixels of unbudded cell), cell elongation (ratio of the long and short axis length of the cell) and neck (i.e. boundary between mother and daughter cell) position (angle unit) correspond to CalMorph traits C11-1-A, C115-A and C105-A1B, respectively. Dashed lines indicate the range of the wild-type values (average ± 2 standard deviations). Boxplots show the median, first and third quartiles, with whiskers showing the 5th and 95th percentiles.*

Several lines of evidence indicate that the observed morphological changes are driven by specific deleterious and compensatory mutations. First, and foremost, the morphological changes in the evolved controls were limited, indicating that morphological divergence of compensated strains is not the result of adaptation to the physical environment. Second, morphological changes are not simply side-effects of growth rate increase for two reasons: (i) only one of the 149 single-cell morphological traits (D197-C, i.e. nuclei size ratio) shows a significant association with growth rate of the compensated strains (Spearman's correlation, FDR < 5%,); and (ii) the extent of changes in morphological traits are also not associated with the extent of growth rate increase during compensatory evolution (Spearman's correlation, FDR < 5%,). Third, morphological evolution in the laboratory may be driven by numerous mutations all of which have small effects individually. If it were so, one might expect a positive correlation between the number of mutations that have

accumulated during course of laboratory evolution and the extent of morphological change. In the ealier work (Szamecz et al., 2014) compensated strains were selected for whole genome sequencing. The list of sequenced strains included genotypes that showed large initial fitness defect, substantial fitness improvement and gradual fitness increase during evolution. However, analysis of available genome sequences of a set of 18 compensated strains (Szamecz et al., 2014) revealed no significant correlation between these two variables, indicating that certain compensatory mutations have exceptionally large effects on cellular morphology.

5.2. Comparable morphological diversity of compensated and natural strains

Next, we focused on three morphological traits; cell size, cell elongation and bud neck position, for three reasons. First, out of the 149 measured morphological traits, changes in cell size and cell elongation were especially high, yielding up to ~50% and ~35% increments respectively, in the compensated strains compared to that in the wild-type (Figure 11) and explaining most variation across strains (Figure 8). Second, compensated strains with extremely low bud neck positions form a distinct group (Figure 11), corresponding to a highly diverged and statistically significant cluster in the cluster analysis. Third and most importantly, these three traits are highly variable across natural *S. cerevisiae* isolates (Yvert et al., 2013).

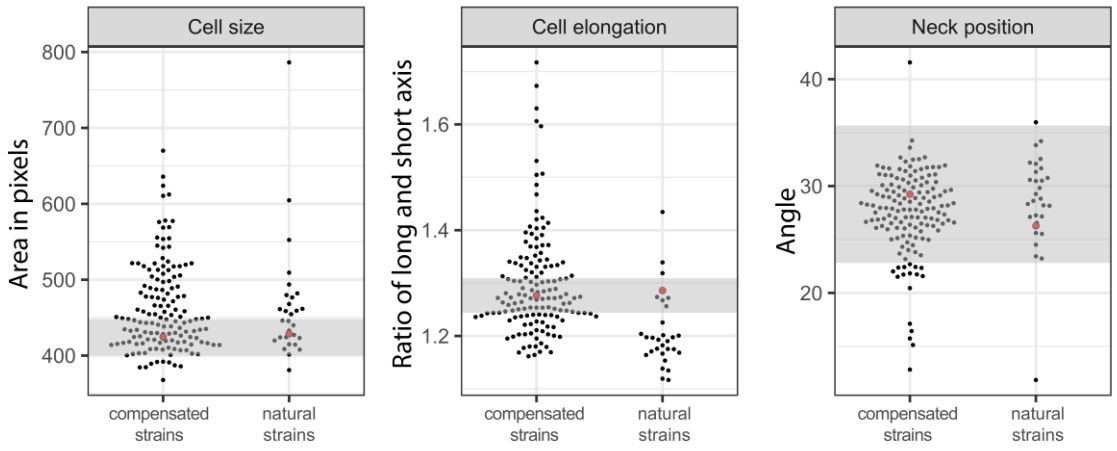


Figure 11. Comparable morphological diversity of compensated and natural strains. Distribution of three morphological traits across compensated and natural strains. *Black dots in swarm plots represent the average values of cell size, cell elongation and neck position (based on*

at least three biological replicates per genotype) for compensated and natural strains. Red dots indicate the average value of the wild-type biological replicates in the two sets of measurements (N = 88 and N = 16 for compensated strains and natural isolates, respectively). Grey area represents the average ± 2 standard deviations of the 88 wild-type replicates of the set of measurement involving the compensated strains.

We compared the extent of morphological diversity induced by laboratory evolution to that observed in nature. For this purpose, we measured these three morphological traits in 29 natural *S. cerevisiae* isolates with diverse ecological origins (Methods). The selected isolates represent several diverged clades of *S. cerevisiae* (Peter et al., 2018) and were confirmed to be haploids (Methods). While the natural isolates differ by ~10^2 – 10^5 point mutations from each other (Peter et al., 2018), the compensated strains typically differ from the wild-type by only ~6 point mutations in addition to the focal gene deletion (Szamecz et al., 2014). Despite these vast differences in the extent of genomic divergence, the overall extent of morphological variation in the compensated strains is comparable to that of natural strains (Figure 11, 12). This pattern is especially remarkable considering that we likely overestimate morphological variation of natural strains, as many of them are monosporic derivatives of isolates that are diploid in the wild (Peter et al., 2018). Surprisingly, certain compensated strains resemble specific natural isolates in their cell size and cell elongation values (Figure 13). We conclude that compensatory evolution in the laboratory generates morphological diversity in an exceptionally rapid manner (i.e. within ~400 generations).

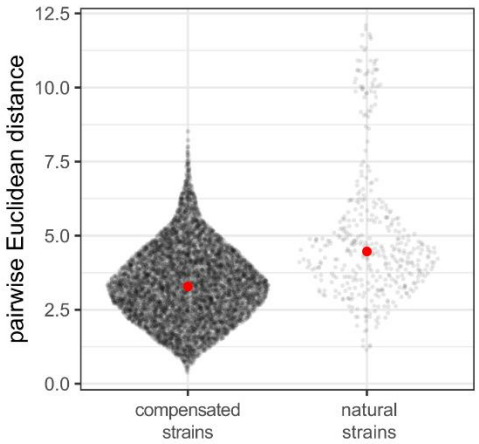


Figure 12. Distribution of pairwise Euclidean distances of the morphological profiles for all pairs of compensated strains (N=10,011) and natural isolates (N=406). *Each dot represents a pair of compensated or natural strains. Red dots indicate the median distance of all strain pairs in the two sets.*

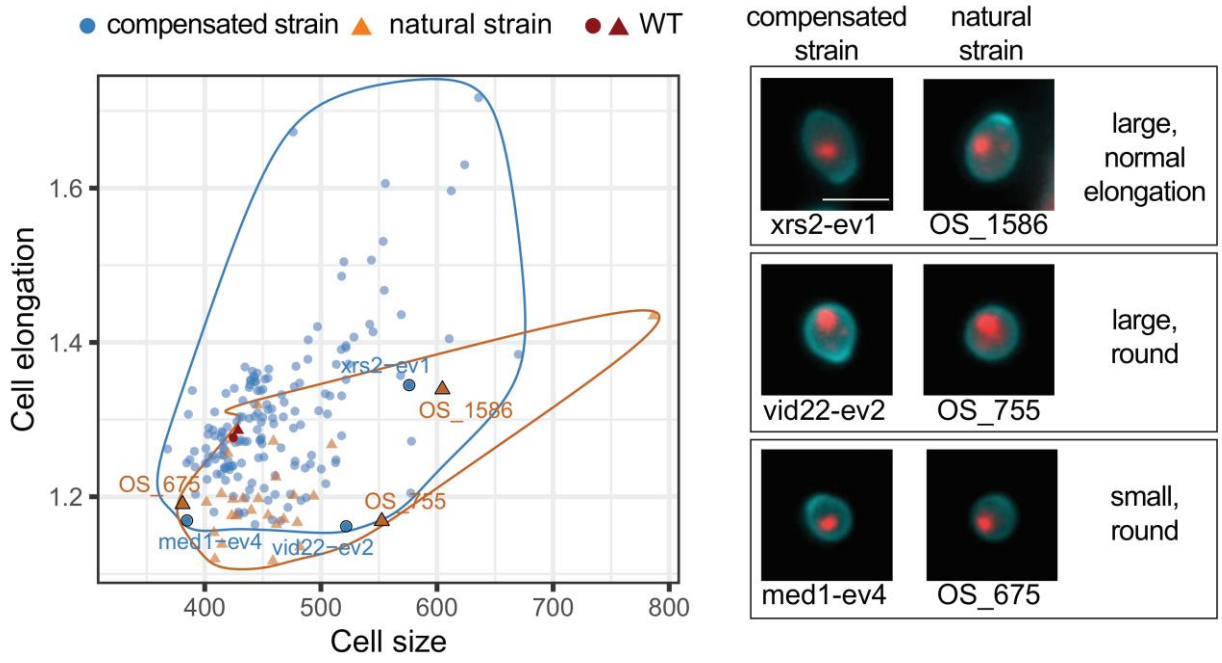


Figure 13. Similar cellular morphology of compensated strains and natural isolates. *Scatterplot showing the distribution of compensated strains (blue dots) and natural isolates (green triangles) according to average cell size and cell elongation values (based on at least 3 biological replicates). Red dot and triangle indicate the average value of the wild-type biological replicates in the two sets of measurements (N = 88 and N = 16 for compensated strains and natural isolates, respectively). Images show pairs of compensated and natural strains that display similar morphological trait combinations (cell wall and nuclei are colored with green and red, respectively): (i) large cells with normal, wild-type-like elongation: xrs2-ev1, OS_1586 isolate from tree leaves, (ii) large round cells: vid22-ev2, OS_755 wine yeast isolate, (iii) small round cells: med1-ev4, OS_675 isolate from human blood. Red dot and red triangle show the same wild-type genotype (WT) in the screens of the compensated and natural strains, respectively (N=88 and 16 replicates). Cell size and cell elongation in panels 2a, and 2c, and neck position in panel 2a correspond to the same CalMorph parameters as in Fig. 10. Scale bar (on image of xrs2-ev1) represents 5 μm distance.*

Yeast cell size and cell shape are intimately linked with progression through the cell cycle, therefore we next asked whether the rapid evolution of cell size and cell elongation is associated with changes in cell cycle phase distributions. To this end, we estimated the fraction of cells residing in G1 and G2 cell cycle phases (i.e. G1 and G2 percentage, respectively) using flow cytometry (Methods). Indeed, we found large changes in both G1 and G2 percentages in the

compensated strains compared to that in the wild-type (Figure 14). Specifically, 32.4% (N = 46) and 43.7% (N = 62) of the compensated strains show a marked alteration in G1 and G2 percentage, respectively. Prior works unveiled a control mechanism that regulates G1 duration based on cell size, resulting in a relatively shorter G1 for larger cells (Turner et al., 2012). As might be expected, we found a negative correlation between cell size and G1 percentage (Figure 14). More remarkably, bud elongation has been shown to be directly affected by the relative duration of G2 (Watanabe et al., 2009) and there was a positive correlation between G2 percentage and both bud and cell elongation across compensated strains (Figure 14). While deciphering the mechanistic underpinnings of these patterns goes beyond the scope of the current work, the latter result indicates that cell elongation has partly evolved through changes in cell cycle phase distributions.

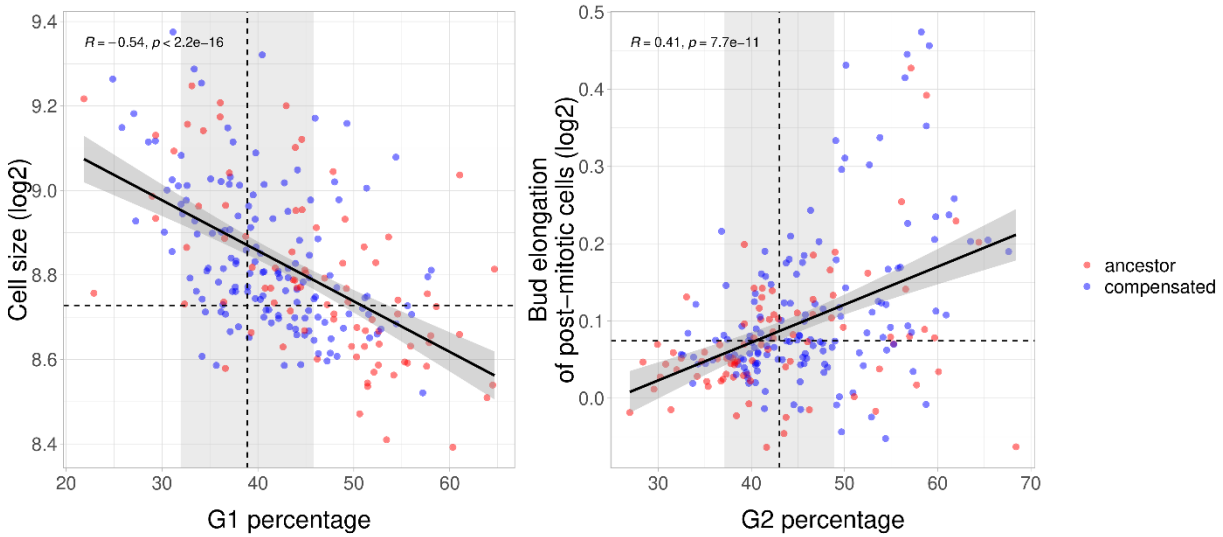


Figure 14. Compensatory evolution alters cell cycle. *Left panel shows the scatterplot of mean cell size and mean G1 percentage in both the ancestor (red) and compensated (blue) strains. Cell size shows a significant negative correlation with G1 percentage (i.e. fraction of cells in the G1 cell cycle phase in the population) across all genotypes (Pearson's correlation). Right panel shows the scatterplot of mean bud elongation and mean G2 percentage (i.e. fraction of cells in the G2 cell cycle phase) in both the ancestor (red) and compensated (blue) strains. Bud elongation shows a significant positive correlation with G2 percentage (Pearson's correlation). Cell size and bud elongation correspond to the CalMorph traits C11-1-A (i.e. the size of unbudded mother cell, as in Fig 13) and C114-C (i.e. the ratio of the long and short axis length of the bud in post-mitotic cells), respectively. Mean G1 and G2 percentage was estimated by performing flow-cytometry analysis of two biological replicates for each strain (Methods). Dashed line represents the average of the*

WT. Grey area represents the WT average ± 2 standard deviations. We estimated standard deviation using the pool of strain-wise centered replicate measurements of all investigated strains.

5.3. Rapid evolution of multicellular phenotypes

Alteration of cellular morphologies and formation of multicellular aggregates have been reported to be linked in yeast (Gimeno et al., 1992). Therefore, we next asked whether compensatory evolution alters the capacity to exhibit multicellular phenotypes. Different species of yeasts typically undergo a developmental transition from a single-cell form into multicellular forms upon environmental change (Madhani and Fink, 1998; Roberts and Fink, 1994). We studied three different forms of multicellular phenotypes: (i) invasive growth phenotype that permits penetration into solid agar (Cullen and Sprague, 2000), (ii) biofilm formation that allows adherence to semi-solid agar (Reynolds and Fink, 2001), and (iii) cellular aggregation in liquid medium via flocculation (Soares, 2011) or clump formation (Kuzdzal-Fick et al., 2019). These three traits enable survival under stressful conditions, aid nutrient acquisition (Cullen and Sprague, 2000) and contribute to virulence in pathogenic yeasts (Desai et al., 2014; Madhani and Fink, 1998). However, several natural *S. cerevisiae* strains have lost their capacity to display these multicellular traits (Hope and Dunham, 2014).

We screened the wild-type, the ancestor and the compensated strains for invasive growth, cell aggregation and biofilm formation using established methods (see Methods). Consistent with earlier works, the wild-type fails to show invasive growth and cell aggregation phenotype (Liu et al., 1996). It also forms 50% smaller biofilms (Reynolds and Fink, 2001) compared to the positive control sigma 1278b strain (Fig 15E). Strikingly, however, several compensated strains gained the capacity of invasive growth (13%), formed enlarged biofilms (2.8%) or displayed multicellular aggregates in liquid (12.4%), see Fig. 15. Multicellular aggregation in the compensated strains was achieved by incomplete daughter cell separation (i.e. clumping (Kuzdzal-Fick et al., 2019)), rather than by flocculation of previously separated cells (Figure 16).

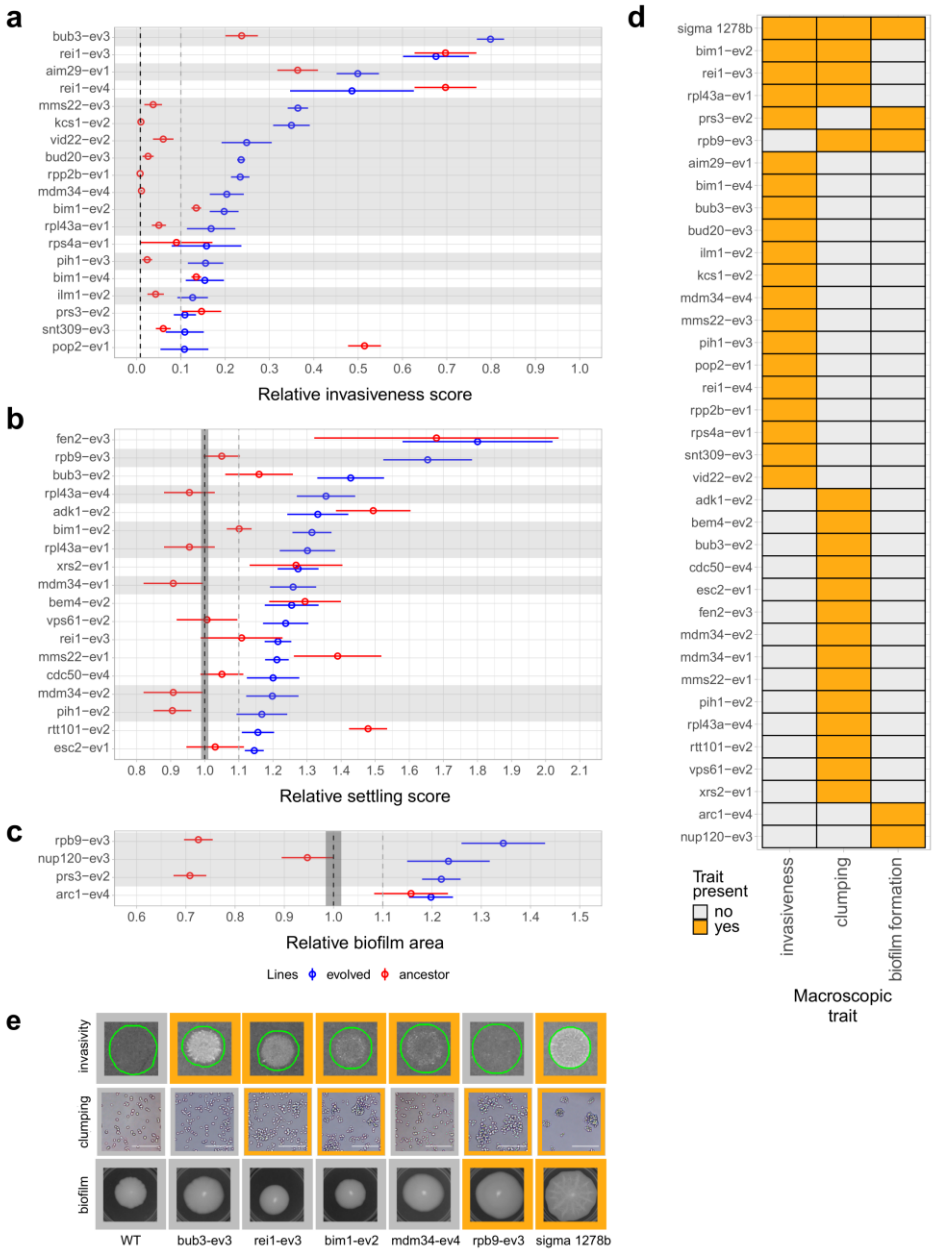


Figure 15. Compensatory evolution promotes the emergence of multicellular phenotypes.

(a-b-c) *Plots show trait values for three types of multicellular forms: invasiveness (panel a), multicellular aggregates as measured by settling (panel b) and biofilm area (panel c). Along with the corresponding ancestor strains (red), only those compensated strains (blue) are plotted that show a marked change in trait value compared to the wild-type based on both q-value and effect size thresholds. Mean (black dashed line) and standard error (dark-grey shaded area) of the WT for the invasive growth / settling / biofilm assay was calculated based on 156 / 106 / 42 biological replicates, respectively. Please note the small standard error of the WT on panel a. The horizontal*

light-grey shaded areas mark those ancestor-compensated pairs that differ significantly (Student's T-test, 10% false discovery rate). The empty circle and the point-range represent the mean and the standard error of the corresponding trait values, respectively, based on at least 3 biological replicates. ***(d)*** *Heatmap summarizes the presence (orange) or absence (light-grey) of alterations in the three investigated traits across 41 compensated strains, each of which display a marked change in at least one multicellular trait compared to WT.* ***(e)*** *Representative images of six compensated strains showing an increased level in at least one multicellular trait (with sigma 1278b as a positive control). Photos with orange frame mark significant trait changes. Scale bars on the microscopic images (middle row) represent 50 μm.*

Overall, 23.4% of the compensated strains showed an enhanced capacity to display at least one multicellular trait (Fig. 15d). This figure is striking, as the capacity to display these traits is unlikely to confer any benefit in the well-shaken liquid medium employed during the course of laboratory evolution. Consistent with this notion, none of the 21 control evolved strains exhibited any of these phenotypes (P = 0.0055, Fisher's exact test). Similarly, the corresponding ancestral strains typically displayed only mild or no changes in these traits (Fig. 15a-c). Together, these results indicate a prevalent role of compensatory evolution in generating multicellular phenotypes.

An important issue is whether multicellular phenotypes arise through convergent evolution at the molecular level. If it were so, one would expect an overlapping set of genes to be mutated in strains displaying the same evolved phenotype. This was clearly not the case: compensated strains that independently evolved the same multicellular trait showed no overlap in their sets of mutated genes (N=46 mutations in 9 sequenced strains). This indicates that morphological novelties arise through diverse mutational routes in the laboratory.

Prior studies demonstrated a substantial variation in the invasive growth phenotype among natural yeast strains (Hope and Dunham, 2014). Here we compared the extent of variation in this phenotype in the laboratory-evolved strains and natural isolates with different ecological origins. For this purpose, we measured the capacity for invasive growth phenotype in a set of natural *S. cerevisiae* isolates (see Methods). Unexpectedly, the extent of gain in invasive growth phenotype in the compensated strains reaches as high as ~50% of the range of invasiveness displayed by the natural isolates.

A compensated laboratory strain with *Δbub3* genetic background (bub3-ev3) shows an especially high invasiveness score, and it displays filament-like multicellular structures invading into the agar (Fig. 16a). Invasive growth in *S. cerevisiae* is generally mediated by Flo11p, a key cell surface protein whose regulation is impaired in the laboratory yeast background used as a wild-type here (Lo and Dranginis, 1998). We found that bub3-ev3 did not regain the ability to express *FLO11*, as evidenced by low activities of the *FLO11* promoter (Fig. 16b) and of a reporter protein of the filamentous growth pathway. Thus, our data suggest that invasiveness in this strain emerged without reactivating the canonical filamentous regulatory pathway.

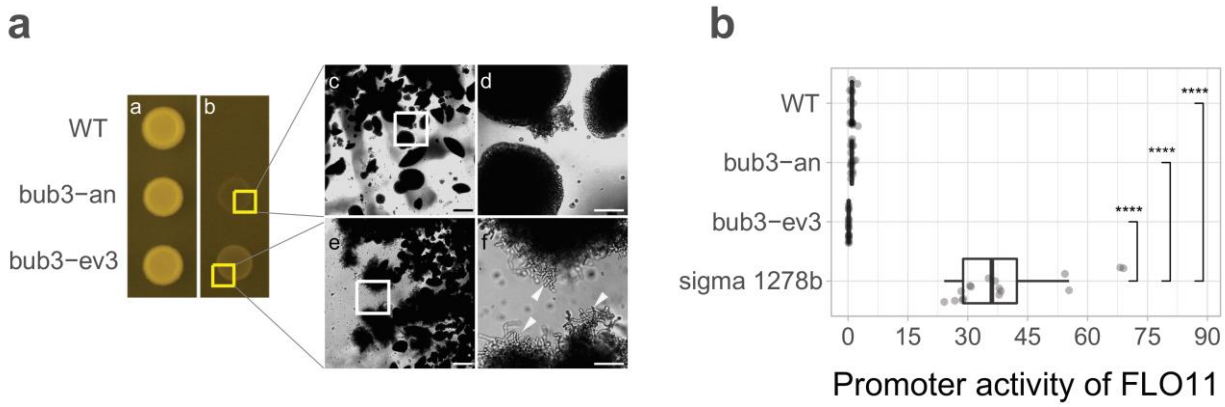


Figure 16. High invasiveness of a compensated strain. (a) *Filament formation appeared during compensatory evolution in a compensated strain of Δbub3. Subpanel a-b) The figure shows yeast spots (illuminated from the top) of wild-type, the ancestor (bub3-an) and a compensated strain (bub3-ev3) of Δbub3 on solid medium after 7 days of incubation (for further details, see Methods). Subpanel c-f) Transmitted light microscopy images show the agar-embedded structures after plate-washing. White triangles indicate extensive filament formation in bub3-ev3 line. Scale bars: 250 µm (subpanels c and e), 50 µm (subpanels d and f).* ***(b)*** *Promoter activity of FLO11 across four genotypes, including the wild-type (WT), the ancestor (bub3-an) and a compensated strain (bub3-ev3) of Δbub3 and the positive control strain (sigma 1278b). Promoter activity was estimated after 7 days of incubation on solid medium by using a pFLO11-yEVenus-tADH1 reporter plasmid. Promoter activity of FLO11 (%) was calculated by normalizing the colony fluorescence of the genotypes to that of the wild type (WT). The promoter activity of FLO11 shows an 82-fold increase compared to the laboratory wild-type. Boxplots show the median, first and third quartiles, with whiskers showing the 5th and 95th percentiles of two independent measurements of eight biological replicates for each genotype. Significant differences were assessed by Student's T-tests (**** indicates P < 0.0001).*

5.4. Synergistic epistasis at the level of morphology

Finally, we conducted in-depth genetic analyses to decipher potential epistatic interactions between loss-of-function and compensatory mutations. We first focused on a compensated strain with *Δrpb9* genetic background (rpb9-ev2), displaying an elongated cell shape (Fig. 17a). Rpb9-ev2 carries a compensatory loss-of-function mutation in *WHI2* (Szamecz et al., 2014). The *RPB9* gene encodes an RNA polymerase II subunit, whose deletion reduces transcriptional fidelity and inhibits growth as a result of proteotoxic stress response (Vermulst et al., 2015) and subsequent downregulation of the major growth control (TOR, target of rapamycin) pathway. As noted earlier, a loss-of-function mutation in Whi2p, a negative regulator of the TOR pathway (Chen et al., 2018) mitigates the fitness defect of *Δrpb9* (Szamecz et al., 2014) via relief of growth inhibition. Intriguingly, deletion of *WHI2* not only improves fitness but also yields elongated cells in *Δrpb9*, while it is non-beneficial (Szamecz et al., 2014) and has no major effect on cell elongation in the wild-type (Fig. 17a-b). We conclude that the initial loss-of-function mutation (*Δrpb9*) and the compensatory mutation (*Δwhi2*) display synergistic epistasis on cell elongation (Fig. 17b).

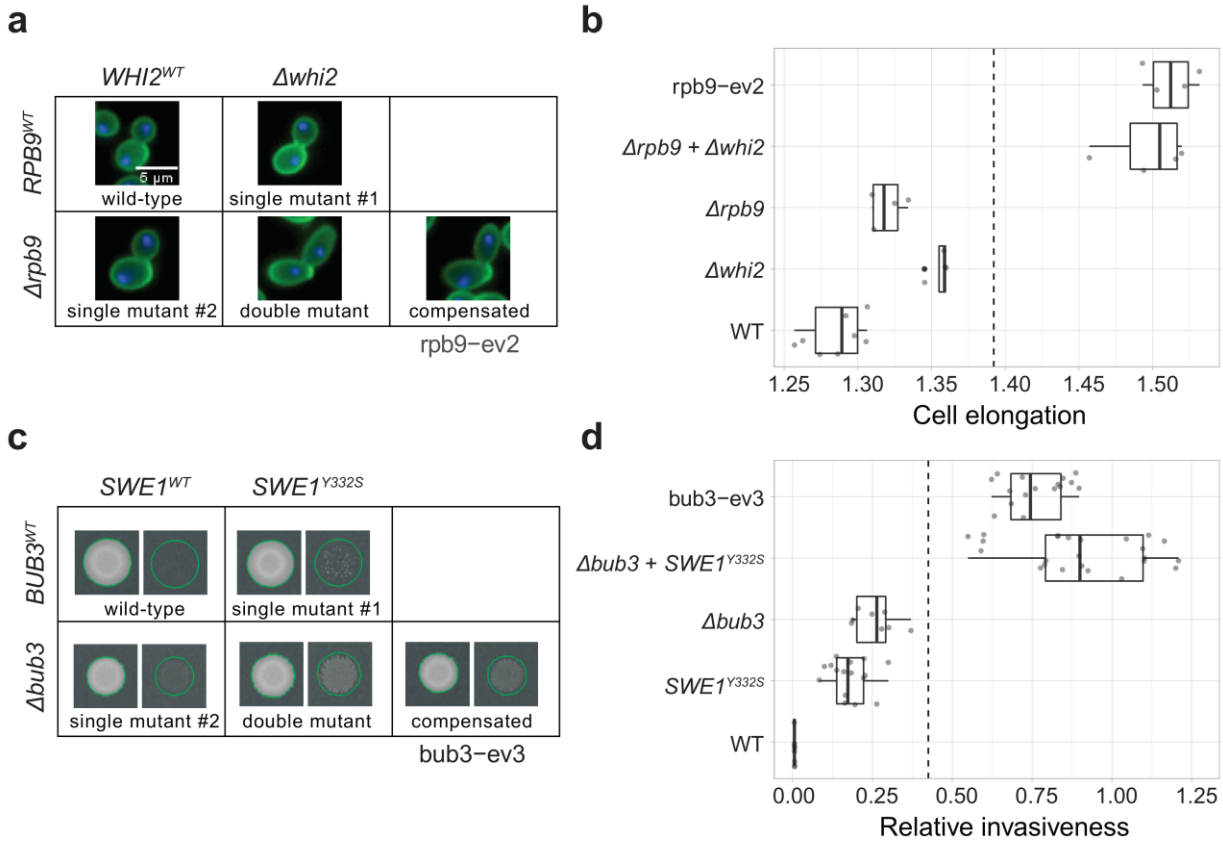


Figure 17. Synergistic epistasis underlying morphological changes in two compensated strains. ***(a)*** *Representative images of yeast cells of 5 selected genotypes, including the wild-type (WT), two single mutants (Δrpb9 and Δwhi2) and a reconstructed double mutant (Δrpb9 + Δwhi2). The fifth genotype is the compensated strain of Δrpb9 (rpb9-ev2) that harbors the whi2^{S133*} loss-of-function allele. Loss-of-function mutation of WHI2 recapitulates the elongated cell shape of the compensated strain when introduced into Δrpb9, but not when introduced into the WT. Cell wall and nuclei are colored with green and blue, respectively. Scale bar represents 5µm for all images.* ***(b)*** *Boxplots show cell elongation values of the same five genotypes (based on N=4 and N=8 replicates for the mutants and the WT, respectively) as in panel a. Vertical dashed line shows the expected value of the double mutant in the absence of epistasis (see Methods).* ***(c)*** *Representative images of yeast spots before and after plate washing (left and right images, respectively, in each grid) across 5 selected genotypes, including the wild-type (WT), two single mutants (Δbub3 and SWE1^{Y332S}) and a reconstructed double mutant (Δbub3 + SWE1^{Y332S}). The fifth genotype is the compensated strain of Δbub3 (bub3-ev3) that harbors the SWE1^{Y332S} mutant allele. SWE1^{Y332S} confers a high capacity for invasive growth when introduced into Δbub3, whereas it has only a minor effect in the WT background.* ***(d)*** *Boxplots show the relative invasiveness score of the same five genotypes (using at least eight biological replicates for each genotype) as in panel c. Vertical dashed line shows the expected value of the double mutant in the absence of epistasis (see*

Methods). Boxplots show the median, first and third quartiles, with whiskers showing the 5th and 95th percentiles.

Analysis of a compensated strain (bub3-ev3) exhibiting a novel multicellular phenotype reached a similar conclusion. Bub3-ev3 shows an exceptionally high capacity for invasive growth (Fig. 15a) compared to that of all other compensated, several natural and the corresponding *Δbub3* ancestor strains. The *BUB3* gene encodes a spindle assembly checkpoint protein involved in delaying the cell division until each chromosome is properly attached to the microtubules (Hardwick, 1998). Bub3-ev3 carries a non-synonymous point mutation (Y332S) in a morphogenesis checkpoint protein, Swe1p (Szamecz et al., 2014). Swe1p normally delays nuclear division until the bud has been formed (Lew, 2003). It is plausible that the mutation in Swe1p (*SWE1*^{Y332S}) alters cell cycle kinetics and thereby partly compensates for spindle checkpoint defects. We introduced the *SWE1*^{Y332S} mutation into both the wild-type and *Δbub3* genetic backgrounds. The mutation significantly improved fitness in *Δbub3*, but not in the wild-type, confirming its causal role in fitness compensation. Remarkably, the reconstructed double mutant *Δbub3 SWE1*^{Y332S} showed a strong invasive growth phenotype similar to that of the bub3-ev3 (Fig. 19c-d). In contrast, the *SWE1*^{Y332S} mutation had no major impact on invasiveness in the wild-type, indicating synergistic epistasis between gene loss and the associated compensatory mutation on invasiveness (Fig. 17d). (Farkas et al., 2022)

6. DISCUSSION

In this thesis, I experimentally test a new mechanism of morphological evolution. The traditional view of morphological evolution is that organisms adapt to changing environments. Here we propose an alternative scenario where the environment is constant and deleterious mutations play a role as a driving force for morphological evolution. Natural isolates of *Saccharomyces cerevisiae* show high diversity in morphology at both the single cell and the multicellular level, however, no significant association was observed between morphology and the ecological or geographical origin (Hope and Dunham, 2014; Yvert et al., 2013). This observation suggests that morphological evolution is driven not only by adaptation to the changing environment, but there should be an alternative route as well.

The aim of our work was to test whether compensatory evolution contributes to morphological novelties. Therefore we used a set of slow-growing knock-out *Saccharomyces cerevisiae* strains that had gone through laboratory compensatory evolution (Szamecz et al., 2014). We found that restoration of the wild-type morphology was relatively rare (15%), whereas in 46% of the cases single-cell morphology of the compensated strains markedly differed from that of the wild-type and the corresponding ancestral strain. Changes in cell size and cell elongation were especially high. In addition, a distinct group of compensated strains showed very low bud neck position, indicating altered bud site selection in these cells. Similarly, in *S. cerevisae* natural isolates the same three traits were the most diverse within species (Yvert et al., 2013). The extent of morphological diversity in the laboratory-evolved strains was similar to that of natural isolates. This is quite surprising, because natural strains differ by ~10^2 – 10^5 point mutations from each other, while the compensated strains typically differ by only ~6 point mutations.

At the level of multicellular morphology, we also found substantial alterations compared to the wild type. By examining invasive growth, we found that 13% of the compensated strains gained this ability. This result is quite surprising, because the evolutionary experiment was performed in liquid medium and invasive growth was not selected for during its course. What causes this rapid and frequent gain of invasive growth? It is known that laboratory strain (S288C) – that is, the ancestor strain background – lost the invasive growth phenotype due to a mutation in the *FLO8*

gene (Liu et al., 1996). Thus, at an earlier time this strain had carried all properties to grow invasive, and – probably by compensatory evolution – it could find an alternative route to restore invasiveness. Moreover, yeasts evolved from multicellular fungi multiple times during fungal evolution. Intriguingly, these parallel morphological changes involve parallelism at the genetic level by repeated diversification of Zn-cluster transcription factors involved in regulating yeast-filamentous switches (Nagy et al., 2014). Thus, during this regulatory evolution they could retained gene sets required for the multicellular life cycle.

It was reported that the wild-type (S288C) strain – which is the initial strain background – is impaired in the regulation of Flo11p (due to the mutation in the *FLO8* gene) that is a key component involved in cell adhesion during invasive growth (Lo and Dranginis, 1998). One can ask the question, whether fitness compensation accompanied by the restoration of *FLO11* expression. By examining the most invasive compensated strain (BUB3-ev3), we found that *FLO11* expression stayed at a low level, indicating that the canonical filamentous pathway has not changed, and cell adhesion is achieved through an alternative pathway. One may speculate that such an alternative pathway might involve the Sfg1 protein. A recent work showed that Sfg1 promoted cell adhesion independently from the fMAPK pathway and the Flo11 adhesin by repressing proteins involved in cell-wall degrading enzymes (Vandermeulen and Cullen, 2020). Clearly, this hypothesis deserves experimental testing in the future.

Invasive growth and biofilm formation are virulence factors in pathogenic fungi, therefore, it is important to study the evolution of these traits. A future evolutionary experiment can be done in the model organism *Candida albicans*, to test whether compensatory evolution shapes virulence in a pathogenic species.

The second multicellular phenotype was the clump-forming ability. There are two ways how a multicellular structure can be formed from a unicellular yeast: either through aggregation of different pre-existing cells (coming together) or by daughter separation defect, resulting in irreversible clonal clusters (staying together) (Driscoll and Travisano, 2017). We found that compensatory evolution resulted in the formation of irreversible clonal clusters, indicating that these cells have failure in daughter separation during mitosis. Similar multicellular clusters emerged during laboratory evolution of diploid yeast cells when selection was done by gravity (Ratcliff et al., 2012). Sequence analysis of the evolved strains revealed that the mutation of the Ace2 transcription factor caused the emergence of multicellularity by downregulating genes that

normally degrade the bud septum (Ratcliff et al., 2015). In the future, sequencing of the compensated strains that show the clumping phenotype could uncover the molecular basis of the daughter separation defects.

Compensatory evolution is observed not only in microbes but also in human cells. A recent phylogenic analysis revealed that most of the driver mutations in cancer cells are deleterious, and for many of these, compensators are identifiable (Rochman et al., 2020).

In the development of most tumor cells, due to the lack of the G1/S DNA repair checkpoint, the key checkpoint becomes the one at the G2/M transition. This imbalance in checkpoint usage is similar to that observed in the bub3-ev3 strain. This strain carries a deletion in the *BUB3* spindle assembly checkpoint protein and a compensatory mutation (Y332S) in the *SWE1* morphogenesis checkpoint kinase. This specific amino acid (Y332) locates at the binding site of Hsl7p, a negative inhibitor of Swe1p (McMillan et al., 2002). Such a change in the binding site of Hsl7p might interfere with the degradation and could thereby lead to the accumulation of the Swe1 protein. Thus, the lack of the spindle assembly checkpoint is compensated by the enhanced morphogenesis checkpoint. Similarly to our example, enhanced expression of Wee1 (human homolog of Swe1) was observed in several cancer types, including melanoma (Magnussen et al., 2012) and lung cancer (Yoshida et al., 2004), suggesting that dynamic alteration of checkpoint usage might be a general phenomenon.

In our study, synergistic genetic interaction between initial gene deletion and compensatory mutation was responsible for the morphological alterations in both cases investigated. Other studies on compensatory evolution also performed large-scale genetic interaction analysis, where they measured the fitness effects and found a general trend toward positive epistasis (i.e. compensatory mutations are more beneficial in the background in which they arose) (Echenique et al., 2019). Here we showed that, besides the fitness effect, there is synergistic epistasis on morphology between deleterious and compensatory mutations. One of our examples showed synergism between the lack of spindle assembly protein Bub3 and a mutation in the Swe1 protein on the level of invasive growth. Similarly, in another study, stable filamentous growth was induced by the synergism between septin defects and altered MAP-kinase pathway (Kim and Rose, 2015). These findings indicate that the emergence of novel morphology often requires both the initial gene deletion and the compensatory mutation.

One important question is how compensatory evolution takes place in natural microbial populations. Deleterious loss-of-function mutations could be fixed in the population through bottlenecks, genetic drifts, hitchhiking or environmental change. Then subsequent compensatory mutations increase the fitness. However, the order of the mutational events is not clear. It is possible that nearly neutral compensatory mutations emerge first in the population that allow the fixation of loss-of-function mutations without any serious deleterious effect. A recent work is consistent with the latter scenario, where pre-existing natural genetic variants suppress deleterious mutations (Parts et al., 2021).

Future studies should dissect the connection between the unicellular and multicellular morphologies. For example, it is possible that strains in our collection simultaneously harboring elongated cells and increased clump formation have a higher chance to develop filamentous growth.

Taken together, our work demonstrates that novel morphologies can evolve not only through adaptation to a changing environment but also by compensatory evolution. In the latter scenario, compensatory evolution generates rudimentary morphological changes as a byproduct (Figure 18), that later could be fine-tuned by direct selection.

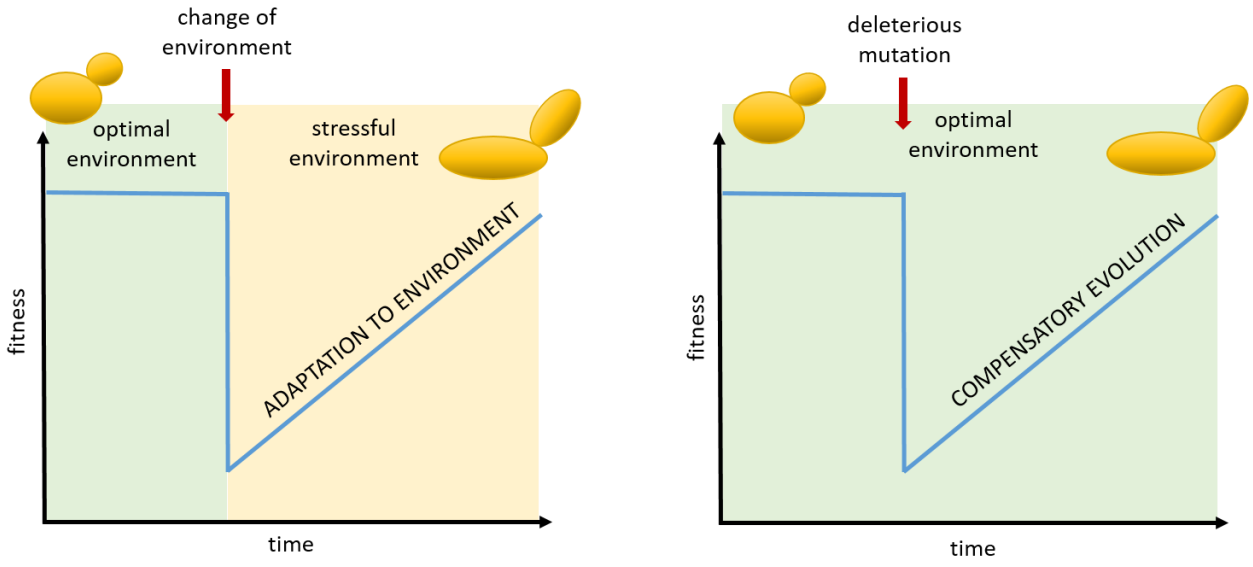


Figure 18. Model of morphological evolution. *Left: Classical view of evolution, where cells adapt to a changing environment. Right: Emergence of novel morphology by compensatory evolution, where deleterious mutations induce selection pressure.*

7. METHODS[2]

7.1. Yeast strains and media

All strains used in this study were derived from the BY4741 *S. cerevisiae* parental strain. A set of slow-growing non-essential single-gene deletion strains from the haploid yeast deletion collection (MATa; *his3Δ 1; leu2Δ 0; met15Δ 0; ura3Δ 0; xxx::KanMX4*) were used in our previous study to investigate compensatory evolution in the laboratory (Szamecz et al., 2014). In this prior batch selection experiment, each slow-growing deletion strain (i.e. ancestor strain) was propagated for ~400 generations under standard laboratory conditions (YPD medium) that resulted in fitness-compensated evolved lines (referred to as compensated strains throughout the text). As a proxy for fitness, we previously measured growth rate (Szamecz et al., 2014). Growth rate was estimated by monitoring the growth curves of yeast populations in liquid medium, using a modified version of an established procedure (Warringer and Blomberg, 2003; Warringer et al., 2003). To investigate the effect of the initial deleterious and subsequent compensatory mutations on single-cell and multicellular morphological features, we here used isolated clones of both the slow-growing ancestor strains (N = 86, impaired in diverse cellular functions) and 1-4 parallel evolved compensated strains (N = 145) per ancestor that showed fitness compensation and remained haploid during the laboratory evolution (see below for details). Note that all strains, including ancestor, compensated and control evolved strains, were saved and stored as glycerol stock until further experiments. All experiments throughout this study were also performed on a set of control evolved lines (N = 21) initiated from the wild-type strain (*his3::KanMX*, designated as WT throughout the manuscript) in our previous study (Szamecz et al., 2014) to control for morphological changes unrelated to compensatory evolution. To investigate the diversity of morphological traits across natural baker's yeast isolates, we selected 29 haploid and euploid strains from a collection of 1,011 *S. cerevisiae* isolates (Peter et al., 2018). The selected isolates represent several phylogenetically diverged clades of *S. cerevisiae* with diverse ecological origins. For the quantitative invasive growth / settling / biofilm formation / filamentous response assays (see below) we used the sigma 1278b parental strain (L5684, matA, *ura3-52, leu2::hisG*) as a positive control (Cullen and Sprague, 2000). To distinguish between calcium-dependent

[2] Methods text is reproduced with permission from Springer Nature

aggregation (i.e. flocculation) and mother-daughter separation defect (i.e. clump formation) as a mechanism underlying enhanced settling (see below), we used a highly flocculating natural strain (OS_1189) as a positive control (Peter et al., 2018).

For single-cell morphological phenotyping, ploidy analysis and fitness measurements, all strains were cultured in rich liquid medium (YPD, 1% yeast extract, 2% peptone, 2% glucose). Quantitative biofilm formation assays, colony size-based fitness measurement and quantitative invasive growth assays were performed on solid YPD medium with low-density (0.3%), medium-density (2%), and high density (3%) agar, respectively.

7.2. High-throughput single-cell morphology measurements

To quantitatively measure single-cell morphology, we applied a previously established protocol (Ohya et al., 2005; Okada et al., 2015) with minor modifications. Single colonies were selected from yeast populations in order to establish single clones. Yeast strains were inoculated into standard 96-well plates in 4 biological replicates as follows. Random layouts were generated by a custom R script to avoid plate effects biasing the genotype-specific morphological trait values. This protocol resulted in layouts where each biological replicates position is randomly chosen with two constraints. First, each biological replicate of a given genotype is being placed on a separate plate. Second, each plate contains four control wells (wild-type) in random positions within different quadrants, in distinct rows and columns. Overall this resulted in 88 replicates for wild-type. After reaching the saturation in cell density, each culture was diluted into 500 µl fresh YPD medium in a 96 deep-well plate including 0.5 mm glass beads in each well and grown until mid-exponential phase. After that, the cells were fixed in phosphate buffer (pH 6.5) containing 3.7 % formaldehyde (Sigma-Aldrich). Fixed cells were washed with PBS and P buffer (10 mM Na_2HPO_4, pH 7.2, 150 mM NaCl) and then we performed fluorescent immunostaining of the cell wall and the nucleus (actin-staining was omitted due to low reproducibility as in a previous study (Bauer et al., 2015)). Staining of the cell wall was performed by incubating the fixed cells with 1 mg/mL Alexa Fluor 488 conjugate of concanavalin A (Alexa-488-conA, Thermo-Fisher) solution for 2 hours at 4°C. After washing with P buffer, the cells were kept in PBS buffer at 4°C until microscopy. Prior to microscopy, the nuclei were stained with 350 ng/mL 4', 6-diamidino-2-phenylindole (DAPI, Thermo-Fisher) in PBS buffer supplemented with 0.1% Triton X-100 (Molar Chemicals) for 30 min at room temperature. After repeated pipetting (at least 30 times to minimize cell aggregates),

the cells were diluted and transferred into black clear bottom 96 well plates (Greiner) coated with 1 mg/mL concanavalin A solution (conA, Santa-Cruz Biotechnology). Fluorescently stained cells were sedimented by centrifugation (1500 rpm for 4 min). Microscopy screening was performed by an Operetta High-Content Imaging System (PerkinElmer Inc.) using a 63x high numerical aperture dry objective. During the imaging, 13 fields were captured from each well, with two channels configured for ConA and DAPI, in 4 layers of z-stack. Raw tiff images were processed using custom a MATLAB script to select the optimal z-stack layer for each cell and to produce 696 × 520 8-bit jpeg images (4 images/field of view), which were then used as inputs for the CalMorph software (Ohya et al., 2005). The same procedures were applied in the cases of reconstructed mutants and natural isolates.

7.3. Quality control of raw morphological data

The output raw data files of CalMorph were further processed in the R programming environment. We applied several quality control steps to improve data quality as follows.

Visual inspection of the images with extremely high or low fluorescence values revealed contamination or absence of cells; therefore these images were removed from further analysis. Visual inspection of the images revealed erroneously detected buds in the cases of extremely large bud short axis (C108_A1B or C108_C > 50) or extremely small bud roundness (C114_C < 0.75, C114_A1B < 0.4), caused by two cells in close proximity, thus the corresponding cells were removed from further analysis. For traits with normally distributed cells, outlier cells above and below their estimated maximum and minimum values, respectively, were removed.

Only morphological traits representing mean values of cells were used (i.e. we excluded traits representing cell-to-cell variation). Following a previous study (Bauer et al., 2015), morphological traits with codes C127, D160, D164, D171, D188, and D189 were removed, because of annotation issues or because they measured distances on the order of one or two pixels. Traits related to the intensity of fluorescence were excluded. Any morphological traits that contained more than 10% missing values were removed. Any cells with missing values for any remaining traits were removed. Wells with less than 40 remaining cells per cell cycle stage (A, A1B, C), as defined by CalMorph, were removed from the dataset. This resulted in a minimum of 120 cells for each strain / cell cycle stage (median = 953 for laboratory strains and natural isolates).

Finally, wells in which the average morphological values of cells were extremely different from those of the other three biological replicates were removed. More specifically, we removed a well if its exclusion from the set of 4 biological replicates decreased the standard deviation of the well averages by half in at least 30 morphological traits. Note that each strain is represented by at least 3 biological replicates in the final dataset.

7.4. Statistical analysis of morphological data

All data analysis was performed by using the R language (R Core Team, 2017). First, average morphological trait values were calculated for each biological replicate (i.e. separate well). These average trait values (with the exception of traits representing angles) were log2-transformed and then the transformed values were weighted by the number of cells per well for each genotype. Next, these genotype-level averages were standardized to the wild-type: the average value of the wild-type was subtracted, then this value was divided by the standard deviation of the wild-type biological replicates (N = 88). Dimension reduction was performed on this wild-type standardized data by using principal component analysis resulting in 8 'principal component traits' explaining ~ 96% of all variance. We used the prcomp() function from R using default parameter values.

The morphological profile of each strain is defined by the 8 principal component values, transformed to have unit variance across strains. This ensures that each principal component (type of morphological trait) is equally considered in the analyses below, not only the ones with the highest variance.

To define 'significant' morphological differences between strains, biological replicates of all strains were projected to the principal component space calculated above. Average ± 2 standard deviation intervals were calculated from the biological replicates of each strain. Two strains are defined as being significantly different in a given principal component (PC) when their intervals are non-overlapping. We used standard deviations instead of statistical tests or confidence intervals because the results of the latter methods are highly influenced by the differences in sample sizes (3, 4 or 88). Two strains are defined as 'morphologically different' if they differ from each other in at least one PC.

We estimated the false discovery rate (FDR) of identifying significantly altered morphologies between two strains as follows. We first estimated the FDR of comparing 4-4 biological replicates, representing most of the comparisons between ancestral and compensated deletion strains (94 of

142 comparisons). To this end, we used the 88 biological replicates of the wild-type strain, randomly selecting two sets of 4 biological replicates for comparison. We reasoned that significant differences between such subsets of wild-type replicates represent false positives. Next, we calculated whether the two sets show a significant difference in their morphology (i.e. their mean +- 2 standard deviation intervals overlap for at least one PC trait). To estimate the FDR, we repeated the random sampling 10,000 times, and calculated the proportion of significant cases, resulting in a FDR of 1.3%, yielding ~1.3 expected false positive cases in our dataset. In a similar vein, we also estimated the FDR for comparing 3 vs 4 biological replicates (representing 42 out of 142 comparisons). This yielded ~1.8 false positive cases. Finally, comparison of 3 vs 3 biological replicates (6 out of 142 comparisons) yielded ~0.5 false positive cases. Thus, in total, we expect 3.6 false positives (2.53%) when comparing gene deletion ancestors and corresponding compensated strains. We note that comparisons between the wild-type versus ancestor or compensated strains imply comparing 88 vs. 4 biological replicates, therefore the FDR for those comparisons is expected to be negligible. Overall, these analyses show that the rate and number of falsely detected significantly different strains is very low and is unlikely to affect our conclusions.
Classes of evolutionary trajectories depicted in Fig. 2b are defined as follows. 'No change': neither the compensated strain nor its ancestor is morphologically different from the wild-type. 'Restoration': the ancestor is morphologically different from the wild-type, but the compensated strain is not. For simplicity, we also included in this category the two compensated strains in which all the PCs are at least partially restored. Partial restoration for a given PC is defined by both the compensated strain and its ancestor differing from the wild-type in the same direction but the compensated strain being closer to the wild-type while being significantly different from its ancestor. 'Ancestor-like compensated': both the compensated strain and its ancestor are morphologically different from the wild-type, but they are not different from each other (with the possible exception of partial or full restoration of some but not all PCs). The remaining strains are classified as 'Novel compensated' and can be further divided into two subclasses: i) the compensated strain is different from the wild-type, but its ancestor is not and ii) both the compensated strain and its ancestor are different from the wild-type and from each other.
For Fig. 10 we selected compensated strains with the most extreme trait values based on the following criteria. We considered only those compensated strains whose average ± 2 standard deviation interval is non-overlapping with the wild-type interval for a given trait (calculated on

log2 scale for cell size and cell elongation). From these sets, we display up to 20 strains with the largest absolute differences from the wild-type for each trait (based on log2 fold-change for cell size and cell elongation).

To quantify morphological distances between wild-type and control evolved strains, we calculated pairwise Euclidean distances between their morphological profiles. To compare the degree of morphological diversity of compensated strains and natural isolates, we calculated Euclidean distances between the morphological profiles of each pair of strains within each group. Morphological profiles were defined by the principal components as above. More specifically, morphological trait values were standardized for the two sets of strains separately using their corresponding replicate measurements of the wild-type strain. The two sets of standardized values were merged and PCA was carried out. The standardized scores of the first 8 PC (explaining ~ 96% of the total variation) were used to calculate all pairwise Euclidean distances within each set of strains.

To identify significantly different groups of strains with similar morphologies, we performed hierarchical clustering of their morphological profiles by using the R package pvclust (Suzuki and Shimodaira, 2006), with Euclidean distance and average linkage. Significant clusters were defined as having higher than 0.95 approximately unbiased (AU) bootstrap value after 10000-time bootstrap sampling. To define a minimal set of significant clusters, we selected the largest significant clusters above the height of the node where the wild-type and all the control evolved lines are merged (‘wild-type like cluster’), not including those that contain the ‘wild-type like cluster’. From this list, we considered only those clusters which contained compensated strains initiated from at least three different gene deletion strains.

7.5. High-throughput DNA content analysis

Changes in ploidy level could dramatically change the cellular morphology of the strains. Thus, inclusion of strains that became diploid during laboratory evolution could distort any morphological analysis. To exclude this possibility, we systematically measured the ploidy level of our strain set (Szamecz et al., 2014) by flow cytometry (Delobel and Tesnière, 2014). Briefly, yeast cells were cultured in YPD liquid media in 96 deep-well plates including 0.5 mm glass beads in each well. Saturated cultures were diluted to OD_{600} = 0.5 in 750 µl fresh YPD medium and then incubated for two additional rounds of cell division. Cells in exponential phase (OD_{600} ~ $2x10^7$

cells/mL) were harvested by centrifugation and fixed with 3.7% formaldehyde (Sigma-Aldrich) containing phosphate buffer (pH 6.5) for 45 min at 30°C. Cells were then washed with PBS and treated with PBS with 0.1% Triton X-100 (Molar Chemicals) for 30 min at 30°C. After washing the cells with 50 mM citrate buffer, RNA species were removed with an overnight RNase treatment (0.25 mg/mL RNase in citrate buffer, 37°C with rotation). The samples were then washed with a citrate buffer and stained with 1 μM SYTOX™ Green nucleic acid stain (Thermo-Fisher) solution in citrate buffer for 30 min at room temperature. Samples were then analyzed by flow cytometry (CytoFLEX S, Beckman Coulter Diagnostic, CA, USA) and ploidy was determined by comparing the signal distribution of each investigated strain to those of haploid and diploid reference strains. All diploid strains were excluded from our dataset and only the haploid strains were retained for all analyses in the paper. The ploidy level of selected natural isolates was also confirmed by the same method.

7.6. Cell cycle analysis by flow cytometry

To investigate potential changes in cell cycle phase distributions as a result of compensatory evolution, we analysed the DNA content of the compensated strains and their corresponding ancestors in two biological replicates using SYTOX™ Green nucleic acid stain (Delobel and Tesnière, 2014). We assigned cells to specific cell cycle phases (G1, S and G2/M) and quantified the relative fraction of these cell cycle phases using the Michael H. Fox algorithm of the Kaluza Analysis 2.1 software from Beckman Coulter (Beckman Coulter Diagnostic, CA, USA).

7.7. Strain construction for generating a deletion of *WHI2* in the wild-type and *Δrpb9* ancestor strain

To investigate epistasis between *Δrpb9* and *Δwhi2* at the level of cell elongation, we first deleted the *WHI2* gene both in the wild-type and the rpb9-an strains by using a NatMX resistance marker (providing resistance to nourseothricin selection drug). Briefly, the *whi2::NatMX* deletion cassette was amplified using the genomic DNA of the corresponding deletion strain from the SGA (Synthetic Genetic Array) query collection (Tong and Boone, 2006). The deletion cassette was transformed both into the wild-type and rpb9-an strains using the standard lithium acetate method (Gietz and Schiestl, 2007). Successful transformants were selected on YPD containing nourseothricin (clonNAT, WERNER BioAgents) selection drug at 100 mg/L final concentration.

Next, we measured cell elongation values (as described above) of the wild-type (WT), the two single mutants (*Δrpb9* and *Δwhi2*) and the reconstructed double mutant (*Δrpb9* + *Δwhi2*) using N=4 and N=8 replicates for the mutants and the WT, respectively. Cell elongation relative to the wild-type average was calculated. With the presumption of multiplicative epistasis on cell elongation, the expected value for the double mutant was estimated by multiplying the averages of the two single mutants. Please note that in Fig. 19b, we used the original cell elongation values, hence we multiplied the expected value with the wild-type average.

7.8. Strain construction for introducing a point mutation into the wild type and Δbub3 ancestor strain

Using the *delitto perfetto* method (Stuckey et al., 2011), a non-synonymous point mutation identified in the *SWE1* gene (*Y332S*) in a compensated strain of *Δbub3* (bub3-ev3) was introduced into both the wild-type and the corresponding ancestor *Δbub3* background. For yeast transformation, we used the standard lithium acetate procedure (Gietz and Schiestl, 2007). First, the CORE-UH cassette was inserted into the *SWE1* gene at the desired position of the SNP. Successful transformants were selected on YPD medium supplemented with 400 µg/mL hygromycin (PAA Laboratories). Then, two complementary oligonucleotides of 81 bp length with the sequence of the region of interest and the SNP in the 41st position were transformed. The replacement of the KlURA3 marker with the missense SNP was counter-selected using 5-Fluoroorotic Acid (5-FOA, Thermo Scientific Inc.) containing medium, loss of hygromycin resistance was confirmed, and the result of the site-directed mutagenesis was verified by capillary sequencing.

7.9. Measuring the activities of the filamentous response and *FLO11* promoter

To investigate the expression of a Tec1p-dependent filamentation response element (FRE) in a compensated strain of *Δbub3* (bub3-ev3), we used a previously established protocol (Cook et al., 1997). Briefly, we introduced the FRE-LacZ plasmid (YEpU-FTyZ, a kind gift from Prof. Jeremy Thorner) into the *bub3* ev-3 strain, the wild-type and the positive control strain (sigma 1278b) with the standard lithium acetate procedure. After selecting for successful transformants, the strains were inoculated into liquid selective medium, and then saturated cultures were spotted onto solid YPD medium with high density (3%) agar. After 3 days of incubation, the cells were scraped off

from the plates and washed with chilled Z-buffer (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, 50 mM 2-mercaptoethanol, pH 7.0). The cells were diluted to get 2×10^7 cells in 1 mL chilled Z-buffer and then 20 µl of 0.1% SDS and 50 µl of chloroform were added to permeabilize the cells. After incubating the samples at 28°C for 5 min, 200 µl of o-nitrophenyl-β-D-galactoside substrate (ONPG; 4 mg/mL, Sigma-Aldrich) was added. The samples were incubated at 30°C until a sufficiently yellow color had been developed and the time required was documented. The reaction was stopped by adding 500 µl 1M Na_2CO_3. After centrifugation, to remove cellular debris and the chloroform, the optical density of the samples at 420 nm and at 550 nm were recorded and the activity of the beta-galactosidase was determined by calculating Miller units using the following formula: $U = 1000 \times [(OD420) - (1.75 \times OD550)] / [(Time) \times (Vol) \times OD_{600}]$, where *Vol* and OD_{600} are the volume (in mL) and the optical density of culture used in the assay, respectively, and *Time* is the length of incubation until the yellow color has developed. Relative beta-galactosidase activity was calculated by normalizing the absolute Miller unit of the experimental strains to that of the wild type strain.

To investigate the expression of the major flocculin (*FLO11*) required for invasive growth (Lo and Dranginis, 1998), we measured the activity of the *FLO11* promoter by using a high-copy reporter plasmid harboring the *pFLO11-yEVenus-tADH1* gene. For this purpose, we transformed the reporter plasmid into the wild-type, the ancestor (bub3-an) and a compensated strain (bub3-ev3) of *Δbub3* and the sigma 1278b strain (positive control). Briefly, we measured the fluorescence intensity of the yeast spots expressing the fluorescent reporter gene after 7 days of incubation on solid YPD medium with high-density (3%) agar. For this purpose, we used a Synergy H1 (Biotek) plate reader with a monochromator module using YFP excitation/emission settings. The fluorescent intensity was used as a proxy for *pFLO11* promoter activity. Relative *pFLO11* activity was calculated by normalizing the absolute fluorescence of the investigated strains to that of the wild-type strain.

7.10. Quantitative invasive growth assay

To screen the ancestor and compensated strains of deletion mutants (along with control evolved strains) for invasive growth phenotype, we used an established plate washing assay (Roberts and Fink, 1994) with minor modifications. Briefly, strains were inoculated into randomly selected positions of a standard 96-well plate, excluding the borders of the plates that were inoculated with

a “filler” strain that was not analysed afterwards. The positive control strain sigma 1278b was also inoculated into the plates. After reaching saturation in cell density, each 96-well plate was spotted onto a solid YPD medium with high density (3%) agar. After 7 days of incubation at 30°C, all plates were photographed to obtain digital images of spot growth and then 10 mL water was pipetted onto the plates. After 10 min of incubation, the non-invasive cells were washed off with a gentle stream of water and gentle rubbing with a finger (wearing a latex glove). Dried plates were re-grown for 24 h at 30°C to enhance the growth of the agar invaded cells. Images of post-washed spots were also obtained by photographing the plates. Spots were illuminated from the bottom throughout the study, unless otherwise indicated (i.e. on the representative images of Fig. 17a, in which case illumination from the top was used). The pixel intensity of each spot pre- and post-wash was estimated with a custom-developed image analysis pipeline in MATLAB programming environment as follows.

As a first step, pairs of images taken before and after washing were registered (i.e. geometrically aligned with each other) to correct the possible misalignment errors between the two measurements. We fit a grid parameterized by the plate type on the reference image (image before plate washing), and we identified and segmented the spots at grid points. Then the intensities of the pixels were averaged over the extracted area of each spot. At the same time, we also measured background intensities for each spot that is required to calculate the normalized spot intensities. To achieve this, we averaged intensities in a fixed-sized region around each spot. Using the detected original position of the spots, the intensities after washing were calculated based on the segmented areas on the reference image and the image after the washing step. The same background intensity measurement process was applied to the images after washing. To measure the real “optical density” of spots, background correction was performed by subtracting the background intensity from the measured spot intensity for each individual spot. The level of invasive growth, i.e. absolute invasiveness, was determined as the ratio of the intensity after washing to the intensity before washing (based on 4 biological replicates of each). Relative invasiveness was calculated by normalizing the invasiveness of the investigated strains to that of the positive control strain (sigma 1278b). Strains that met the following criteria were considered to display an invasive growth phenotype: i) the relative invasiveness value was significantly higher than that of the wild-type (Wilcoxon rank-sum test, 10% false discovery rate); ii) the relative invasiveness value was higher than 10% of the corresponding value of the positive control strain (sigma 1278b, note that positive

hits were confirmed by visual inspection of the plates after washing). In order to assess significant differences between the compensated strains and their corresponding ancestors, Student's T-tests were used (10% false discovery rate) because these comparisons involved low sample sizes.

To investigate epistasis between *Δbub3* and *SWE1*^{Y332S} at the level of invasiveness, we performed an agar invasion assay using the wild-type (WT), the two single mutants (*Δbub3* and *SWE1*^{Y332S}) and the reconstructed double mutant (*Δbub3* + *SWE1*^{Y332S}) using at least 8 biological replicates. With the presumption of additive epistasis on invasiveness, we calculated the expected relative invasiveness value of the double mutant as follows: first, we subtracted the wild-type value from those of the two single mutants and then performed addition of the subtracted values.

7.11. Imaging of agar-embedded filamentous structures upon invasive growth

To investigate the structures of agar-embedded cells upon invasive growth, we cultivated the wild-type, the ancestor (bub3-an) and a compensated strain (bub3-ev3) of *Δbub3* in YPD liquid medium overnight. Saturated cultures were spotted onto solid YPD medium in 10 cm Petri dishes and incubated for 6 days at 30°C. Upper gel surface and bottom of the dishes were monitored every day for cell growth and presence of agar-embedded structures using Olympus SZX12 stereo zoom microscope (Olympus Life Science Europa GmbH, Hamburg, Germany). On the 7th day, digital images of the spots were taken with a Canon 200D camera and non-embedded, surface-growing cells were washed out with gentle stream of water and the agar-embedded structures were visualized by Leica SP5 AOBS laser scanning confocal microscope (Leica Microsystems GmbH, Wetzlar, Germany). The 10 cm Petri dishes without the lid were placed on Leica DMI 6000 inverted microscope stage upside down. To bring the surface of the gel into objective focal plane, custom-made aluminum adapters were used (i.e. two strips of 1.5 cm-wide aluminum sheet were cut and bent to form a "top hat profile"). The native stage insert of the microscope was replaced with two bent metal strips and the Petri dish was placed on top of them upside down with gel surface facing the objective. The embedded structures were captured with the transmitted light detector using 10x (HCX PL FLUOTAR, numerical aperture: 0.3) and 20x dry objectives (HC PL APO, numerical aperture: 0.7) using a 488 nm laser line and Leica LAS software.

7.12. Quantitative settling assay to detect multicellular aggregates

To systematically investigate the ability of the ancestor and compensated strains to form multicellular clumps or flocs, we first performed a well-established quantitative sedimentation assay (Hope and Dunham, 2014) with minor modification. Briefly, strains were inoculated into culture tubes filled with 5 mL YPD liquid medium and grown for 24 hours at 30°C with 200 rpm shaking. Strains were vortexed vigorously in the culture tubes and then were incubated without agitation for 60 minutes to help sedimentation of the cells. After the incubation time, the culture tubes were photographed then the fraction of the settled cells (absolute settling score) was estimated using an image analysis pipeline described previously (Hope and Dunham, 2014). Relative settling score was calculated by normalizing the absolute settling score of the investigated strains (based on 4 biological replicates of each) to that of the wild-type strain. Strains showing at least a 10% increase in relative settling score and a significant difference (Wilcoxon rank-sum test, 10% false discovery rate) compared to the wild-type were considered to display an increased settling capacity. To assess the statistical significance of the differences between the compensated strains and their corresponding ancestors, Student's T-tests were used (10% false discovery rate) because these comparisons involved low sample sizes.

Formation of multicellular aggregates of strains with increased settling capacity was further confirmed by microscopy analysis (Zeiss Laser Capture Microdissection (LCM) microscope) with 20x magnification. To distinguish between calcium-dependent aggregation (i.e. flocculation) and mother-daughter separation defect (i.e. clump formation) as a mechanism underlying multicellular structures, we performed a deflocculation assay as described in an earlier work (Liu et al., 1996) with minor modifications. Briefly, we treated cells with a chelating agent (EDTA) that sequesters calcium ions and thereby prevents cell adhesion mediated by calcium-dependent flocculins. Cells were incubated in 3ml YPD medium in 24 deep well plates for 24 hours with 220 rpm agitation. Cultures were settled for 60 minutes without agitation then 200 µl of sedimented culture was taken from the bottom of the wells, resuspended, split into two parts, and centrifuged. One part was resuspended in 100 µl water and the other part was resuspended in deflocculation solution (100 µl 4 mM EDTA). After 50 minutes of incubation at room temperature without agitation, cells were taken from the bottom of the centrifuge tube, mixed with mounting medium (9:1 mix of glycerol and 1 ×PBS) and examined by microscope with 20x magnification (Zeiss LCM and Operetta High-Content Imaging System (PerkinElmer Inc.)). For the deflocculation assay, we used a positive

control strain (natural strain (OS_1189) isolated from soil in the Netherlands, described in a previous study (Peter et al., 2018)) that flocculated well in minimal synthetic defined (SD) medium after two days and deflocculated completely after the EDTA treatment.

7.13. Quantitative biofilm formation assay

To screen the ancestor and compensated strains for increased biofilm production, we used an established protocol with minor modifications (Reynolds and Fink, 2001). Briefly, cells growing on solid YPD with medium density (2%) agar plates were picked and spotted by toothpick onto solid YPD with low-density (0.3%) agar plates and incubated for 5 days at 25°C. The plates were subsequently photographed to obtain digital images of the biofilms and then the area of the biofilm was estimated by using an image analysis pipeline described previously (Hope and Dunham, 2014). Relative biofilm area was calculated by normalizing the biofilm area of the investigated strains (based on 3 biological replicates of each) to that of the wild-type. Strains having at least a 10% increase and a significant difference (Wilcoxon rank-sum test, 10% false discovery rate) in biofilm area compared to the wild-type were considered to display an increased biofilm formation capacity. To assess the statistical significance of the differences between the compensated strains and their corresponding ancestors, Student's T-tests were used (10% false discovery rate) because these comparisons involved low sample sizes. (Farkas et al., 2022)

8. ACKNOWLEDGEMENTS

First of all, I would like to thank Balázs Papp for supervision and for introducing me to the science of evolution.

I would like to thank Zoltán Farkas for the joint labwork and the exciting discussions.

I am grateful to Károly Kovács for bioinformatic support and fruitful discussions.

I would like to thank Gergely Fekete for teaching me to write R scripts.

I would like to express my thanks to Fanni Birtyik for her help in the laboratory screenings.

I am grateful to Ferhan Ayaydin and Ildikó Kelemen-Várkony for support in microscopy analysis.

I would like to thank Krisztina Ambrus for her general technical assistance.

I am grateful for the scientific questions and advises of Csaba Pál.

Finally, I would like to thank my family for their support.

9. REFERENCES

Albalat, R., and Cañestro, C. (2016). Evolution by gene loss. Nat. Rev. Genet. *17*, 379–391. https://doi.org/10.1038/nrg.2016.39.

Andersson, D.I., and Hughes, D. (2010). Antibiotic resistance and its cost: is it possible to reverse resistance? Nat. Rev. Microbiol. *8*, 260–271. https://doi.org/10.1038/nrmicro2319.

Bauer, C.R., Li, S., and Siegal, M.L. (2015). Essential gene disruptions reveal complex relationships between phenotypic robustness, pleiotropy, and fitness. Mol. Syst. Biol. *11*, 773. https://doi.org/10.15252/msb.20145264.

Blank, D., Wolf, L., Ackermann, M., and Silander, O.K. (2014). The predictability of molecular evolution during functional innovation. Proc. Natl. Acad. Sci. *111*, 3044–3049. https://doi.org/10.1073/pnas.1318797111.

Brückner, S., and Mösch, H.-U. (2012). Choosing the right lifestyle: adhesion and development in Saccharomyces cerevisiae. FEMS Microbiol. Rev. *36*, 25–58. https://doi.org/10.1111/j.1574-6976.2011.00275.x.

Chavel, C.A., Dionne, H.M., Birkaya, B., Joshi, J., and Cullen, P.J. (2010). Multiple Signals Converge on a Differentiation MAPK Pathway. PLoS Genet. *6*. https://doi.org/10.1371/journal.pgen.1000883.

Chen, X., Wang, G., Zhang, Y., Dayhoff-Brannigan, M., Diny, N.L., Zhao, M., He, G., Sing, C.N., Metz, K.A., Stolp, Z.D., et al. (2018). Whi2 is a conserved negative regulator of TORC1 in response to low amino acids. PLoS Genet. *14*, e1007592. https://doi.org/10.1371/journal.pgen.1007592.

Cook, J.G., Bardwell, L., and Thorner, J. (1997). Inhibitory and activating functions for MAPK Kss1 in the S. cerevisiae filamentous- growth signalling pathway. Nature *390*, 85–88. https://doi.org/10.1038/36355.

Cullen, P.J., and Sprague, G.F. (2000). Glucose depletion causes haploid invasive growth in yeast. Proc. Natl. Acad. Sci. *97*, 13619–13624. https://doi.org/10.1073/pnas.240345197.

Cullen, P.J., and Sprague, G.F. (2012). The Regulation of Filamentous Growth in Yeast. Genetics *190*, 23–49. https://doi.org/10.1534/genetics.111.127456.

Delobel, P., and Tesnière, C. (2014). A Simple FCM Method to Avoid Misinterpretation in Saccharomyces cerevisiae Cell Cycle Assessment between G0 and Sub-G1. PLOS ONE *9*, e84645. https://doi.org/10.1371/journal.pone.0084645.

Desai, J.V., Mitchell, A.P., and Andes, D.R. (2014). Fungal Biofilms, Drug Resistance, and Recurrent Infection. Cold Spring Harb. Perspect. Med. *4*, a019729. https://doi.org/10.1101/cshperspect.a019729.

Driscoll, W.W., and Travisano, M. (2017). Synergistic cooperation promotes multicellular performance and unicellular free-rider persistence. Nat. Commun. *8*, 15707. https://doi.org/10.1038/ncomms15707.

Echenique, J.I.R., Kryazhimskiy, S., Ba, A.N.N., and Desai, M.M. (2019). Modular epistasis and the compensatory evolution of gene deletion mutants. PLOS Genet. *15*, e1007958. https://doi.org/10.1371/journal.pgen.1007958.

Fanning, S., and Mitchell, A.P. (2012). Fungal Biofilms. PLOS Pathog. *8*, e1002585. https://doi.org/10.1371/journal.ppat.1002585.

Farkas, Z., Kovács, K., Sarkadi, Z., Kalapis, D., Fekete, G., Birtyik, F., Ayaydin, F., Molnár, C., Horváth, P., Pál, C., et al. (2022). Gene loss and compensatory evolution promotes the emergence of morphological novelties in budding yeast. Nat. Ecol. Evol. 1–11. https://doi.org/10.1038/s41559-022-01730-1.

Gancedo, J.M. (2001). Control of pseudohyphae formation in Saccharomyces cerevisiae. FEMS Microbiol. Rev. *25*, 107–123. https://doi.org/10.1111/j.1574-6976.2001.tb00573.x.

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

Gimeno, C.J., Ljungdahl, P.O., Styles, C.A., and Fink, G.R. (1992). Unipolar cell divisions in the yeast S. cerevisiae lead to filamentous growth: Regulation by starvation and RAS. Cell *68*, 1077–1090. https://doi.org/10.1016/0092-8674(92)90079-R.

Gladfelter, A.S., Kozubowski, L., Zyla, T.R., and Lew, D.J. (2005). Interplay between septin organization, cell cycle and cell shape in yeast. J. Cell Sci. *118*, 1617–1628. https://doi.org/10.1242/jcs.02286.

Gow, N.A.R., Brown, A.J.P., and Odds, F.C. (2002). Fungal morphogenesis and host invasion. Curr. Opin. Microbiol. *5*, 366–371. https://doi.org/10.1016/s1369-5274(02)00338-7.

Hardwick, K.G. (1998). The spindle checkpoint. Trends Genet. *14*, 1–4. https://doi.org/10.1016/S0168-9525(97)01340-1.

Helsen, J., Voordeckers, K., Vanderwaeren, L., Santermans, T., Tsontaki, M., Verstrepen, K.J., and Jelier, R. (2020). Gene Loss Predictably Drives Evolutionary Adaptation. Mol. Biol. Evol. *37*, 2989–3002. https://doi.org/10.1093/molbev/msaa172.

Hope, E.A., and Dunham, M.J. (2014). Ploidy-Regulated Variation in Biofilm-Related Phenotypes in Natural Isolates of Saccharomyces cerevisiae. G3 Genes Genomes Genet. *4*, 1773–1786. https://doi.org/10.1534/g3.114.013250.

Hope, E.A., Amorosi, C.J., Miller, A.W., Dang, K., Heil, C.S., and Dunham, M.J. (2017). Experimental Evolution Reveals Favored Adaptive Routes to Cell Aggregation in Yeast. Genetics *206*, 1153–1167. https://doi.org/10.1534/genetics.116.198895.

Ivankov, D.N., Finkelstein, A.V., and Kondrashov, F.A. (2014). A structural perspective of compensatory evolution. Curr. Opin. Struct. Biol. *26*, 104–112. https://doi.org/10.1016/j.sbi.2014.05.004.

Kim, J., and Rose, M.D. (2015). Stable Pseudohyphal Growth in Budding Yeast Induced by Synergism between Septin Defects and Altered MAP-kinase Signaling. PLoS Genet. *11*. https://doi.org/10.1371/journal.pgen.1005684.

Kuzdzal-Fick, J.J., Chen, L., and Balázsi, G. (2019). Disadvantages and benefits of evolved unicellularity versus multicellularity in budding yeast. Ecol. Evol. *9*, 8509–8523. https://doi.org/10.1002/ece3.5322.

LaBar, T., Hsieh, Y.-Y.P., Fumasoni, M., and Murray, A. (2020). Evolutionary repair experiments as a window to the molecular diversity of life. Curr. Biol. *30*, R565. .

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

Lew, D.J. (2003). The morphogenesis checkpoint: how yeast cells watch their figures. Curr. Opin. Cell Biol. *15*, 648–653. .

Liu, H., Styles, C.A., and Fink, G.R. (1996). Saccharomyces Cerevisiae S288c Has a Mutation in Flo8, a Gene Required for Filamentous Growth. Genetics *144*, 967–978. .

Lo, W.-S., and Dranginis, A.M. (1998). The Cell Surface Flocculin Flo11 Is Required for Pseudohyphae Formation and Invasion by Saccharomyces cerevisiae. Mol. Biol. Cell *9*, 161–171. .

Madhani, H.D., and Fink, G.R. (1998). The control of filamentous differentiation and virulence in fungi. Trends Cell Biol. *8*, 348–353. https://doi.org/10.1016/S0962-8924(98)01298-7.

Magnussen, G.I., Holm, R., Emilsen, E., Rosnes, A.K.R., Slipicevic, A., and Flørenes, V.A. (2012). High Expression of Wee1 Is Associated with Poor Disease-Free Survival in Malignant Melanoma: Potential for Targeted Therapy. PLoS ONE *7*. https://doi.org/10.1371/journal.pone.0038254.

McCloskey, D., Xu, S., Sandberg, T.E., Brunk, E., Hefner, Y., Szubin, R., Feist, A.M., and Palsson, B.O. (2018). Evolution of gene knockout strains of E. coli reveal regulatory architectures governed by metabolism. Nat. Commun. *9*, 3796. https://doi.org/10.1038/s41467-018-06219-9.

McMillan, J.N., Theesfeld, C.L., Harrison, J.C., Bardes, E.S.G., and Lew, D.J. (2002). Determinants of Swe1p Degradation in Saccharomyces cerevisiae. Mol. Biol. Cell *13*, 3560–3575. https://doi.org/10.1091/mbc.E02-05-0283.

Min, K., Neiman, A.M., and Konopka, J.B. (2020). Fungal Pathogens: Shape-Shifting Invaders. Trends Microbiol. *28*, 922–933. https://doi.org/10.1016/j.tim.2020.05.001.

Nagy, L.G., Ohm, R.A., Kovács, G.M., Floudas, D., Riley, R., Gácser, A., Sipiczki, M., Davis, J.M., Doty, S.L., de Hoog, G.S., et al. (2014). Latent homology and convergent regulatory evolution underlies the repeated emergence of yeasts. Nat. Commun. *5*, 4471. https://doi.org/10.1038/ncomms5471.

Ohya, Y., Sese, J., Yukawa, M., Sano, F., Nakatani, Y., Saito, T.L., Saka, A., Fukuda, T., Ishihara, S., Oka, S., et al. (2005). High-dimensional and large-scale phenotyping of yeast mutants. Proc. Natl. Acad. Sci. *102*, 19015–19020. https://doi.org/10.1073/pnas.0509436102.

Ojkic, N., Serbanescu, D., and Banerjee, S. Surface-to-volume scaling and aspect ratio preservation in rod-shaped bacteria. ELife *8*. https://doi.org/10.7554/eLife.47033.

Okada, H., Ohnuki, S., and Ohya, Y. (2015). Quantification of Cell, Actin, and Nuclear DNA Morphology with High-Throughput Microscopy and CalMorph. Cold Spring Harb. Protoc. *2015*, pdb.prot078667. https://doi.org/10.1101/pdb.prot078667.

Pandit, A., Adholeya, A., Cahill, D., Brau, L., and Kochar, M. (2020). Microbial biofilms in nature: unlocking their potential for agricultural applications. J. Appl. Microbiol. *129*, 199–211. https://doi.org/10.1111/jam.14609.

Parts, L., Batté, A., Lopes, M., Yuen, M.W., Laver, M., San Luis, B.-J., Yue, J.-X., Pons, C., Eray, E., Aloy, P., et al. (2021). Natural variants suppress mutations in hundreds of essential genes. Mol. Syst. Biol. *17*, e10138. https://doi.org/10.15252/msb.202010138.

Peter, J., Chiara, M.D., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freel, K., Llored, A., et al. (2018). Genome evolution across 1,011 Saccharomyces cerevisiae isolates. Nature 1. https://doi.org/10.1038/s41586-018-0030-5.

Qian, W., Ma, D., Xiao, C., Wang, Z., and Zhang, J. (2012). The genomic landscape and evolutionary resolution of antagonistic pleiotropy in yeast. Cell Rep. *2*, 1399–1410. https://doi.org/10.1016/j.celrep.2012.09.017.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ratcliff, W.C., Denison, R.F., Borrello, M., and Travisano, M. (2012). Experimental evolution of multicellularity. Proc. Natl. Acad. Sci. *109*, 1595–1600. https://doi.org/10.1073/pnas.1115323109.

Ratcliff, W.C., Fankhauser, J.D., Rogers, D.W., Greig, D., and Travisano, M. (2015). Origins of multicellular evolvability in snowflake yeast. Nat. Commun. *6*, 6102. https://doi.org/10.1038/ncomms7102.

Reynolds, T.B., and Fink, G.R. (2001). Bakers' Yeast, a Model for Fungal Biofilm Formation. Science *291*, 878–881. https://doi.org/10.1126/science.291.5505.878.

Roberts, R.L., and Fink, G.R. (1994). Elements of a single MAP kinase cascade in Saccharomyces cerevisiae mediate two developmental programs in the same cell type: mating and invasive growth. Genes Dev. *8*, 2974–2985. https://doi.org/10.1101/gad.8.24.2974.

Rochman, N.D., Wolf, Y.I., and Koonin, E.V. (2020). Deep phylogeny of cancer drivers and compensatory mutations. Commun. Biol. *3*, 1–11. https://doi.org/10.1038/s42003-020-01276-7.

Ryan, O., Shapiro, R.S., Kurat, C.F., Mayhew, D., Baryshnikova, A., Chin, B., Lin, Z.-Y., Cox, M.J., Vizeacoumar, F., Cheung, D., et al. (2012). Global Gene Deletion Analysis Exploring Yeast Filamentous Growth. Science *337*, 1353–1356. https://doi.org/10.1126/science.1224339.

Sharma, R., and Jangid, K. (2016). Role of Quorum Sensing in Fungal Morphogenesis and Pathogenesis. In Fungal Metabolites, J.-M. Mérillon, and K.G. Ramawat, eds. (Cham: Springer International Publishing), pp. 1–28.

Shively, C.A., Eckwahl, M.J., Dobry, C.J., Mellacheruvu, D., Nesvizhskii, A., and Kumar, A. (2013). Genetic Networks Inducing Invasive Growth in Saccharomyces cerevisiae Identified Through Systematic Genome-Wide Overexpression. Genetics *193*, 1297–1310. https://doi.org/10.1534/genetics.112.147876.

Smukalla, S., Caldara, M., Pochet, N., Beauvais, A., Guadagnini, S., Yan, C., Vinces, M.D., Jansen, A., Prevost, M.C., Latgé, J.-P., et al. (2008). FLO1 Is a Variable Green Beard Gene that Drives Biofilm-like Cooperation in Budding Yeast. Cell *135*, 726–737. https://doi.org/10.1016/j.cell.2008.09.037.

Soares, E. v. (2011). Flocculation in Saccharomyces cerevisiae: a review. J. Appl. Microbiol. *110*, 1–18. https://doi.org/10.1111/j.1365-2672.2010.04897.x.

Stuckey, S., Mukherjee, K., and Storici, F. (2011). In Vivo Site-Specific Mutagenesis and Gene Collage Using the Delitto Perfetto System in Yeast Saccharomyces cerevisiae. Methods Mol. Biol. Clifton NJ *745*, 173–191. https://doi.org/10.1007/978-1-61779-129-1_11.

Sudbery, P.E. (2011). Growth of Candida albicans hyphae. Nat. Rev. Microbiol. *9*, 737–748. https://doi.org/10.1038/nrmicro2636.

Suzuki, R., and Shimodaira, H. (2006). Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinforma. Oxf. Engl. *22*, 1540–1542. https://doi.org/10.1093/bioinformatics/btl117.

Suzuki, G., Wang, Y., Kubo, K., Hirata, E., Ohnuki, S., and Ohya, Y. (2018). Global study of holistic morphological effectors in the budding yeast Saccharomyces cerevisiae. BMC Genomics *19*. https://doi.org/10.1186/s12864-018-4526-z.

Szamecz, B., Boross, G., Kalapis, D., Kovács, K., Fekete, G., Farkas, Z., Lázár, V., Hrtyan, M., Kemmeren, P., Koerkamp, M.J.A.G., et al. (2014). The Genomic Landscape of Compensatory Evolution. PLOS Biol. *12*, e1001935. https://doi.org/10.1371/journal.pbio.1001935.

Taylor, R.W., and Turnbull, D.M. (2005). Mitochondrial DNA mutations in human disease. Nat. Rev. Genet. *6*, 389–402. https://doi.org/10.1038/nrg1606.

Tong, A.H.Y., and Boone, C. (2006). Synthetic genetic array analysis in Saccharomyces cerevisiae. Methods Mol. Biol. Clifton NJ *313*, 171–192. .

Turner, J.J., Ewald, J.C., and Skotheim, J.M. (2012). Cell size control in yeast. Curr. Biol. CB *22*, R350-359. https://doi.org/10.1016/j.cub.2012.02.041.

Vandermeulen, M.D., and Cullen, P.J. (2020). New Aspects of Invasive Growth Regulation Identified by Functional Profiling of MAPK Pathway Targets in Saccharomyces cerevisiae. Genetics *216*, 95–116. https://doi.org/10.1534/genetics.120.303369.

Vermulst, M., Denney, A.S., Lang, M.J., Hung, C.-W., Moore, S., Moseley, M.A., Thompson, J.W., Madden, V., Gauer, J., Wolfe, K.J., et al. (2015). Transcription errors induce proteotoxic stress and shorten cellular lifespan. Nat. Commun. *6*, 8065. https://doi.org/10.1038/ncomms9065.

Warringer, J., and Blomberg, A. (2003). Automated screening in environmental arrays allows analysis of quantitative phenotypic profiles in Saccharomyces cerevisiae. Yeast Chichester Engl. *20*, 53–67. https://doi.org/10.1002/yea.931.

Warringer, J., Ericson, E., Fernandez, L., Nerman, O., and Blomberg, A. (2003). High-resolution yeast phenomics resolves different physiological features in the saline response. Proc. Natl. Acad. Sci. U. S. A. *100*, 15724–15729. https://doi.org/10.1073/pnas.2435976100.

Watanabe, M., Watanabe, D., Nogami, S., Morishita, S., and Ohya, Y. (2009). Comprehensive and quantitative analysis of yeast deletion mutants defective in apical and isotropic bud growth. Curr. Genet. *55*, 365–380. https://doi.org/10.1007/s00294-009-0251-0.

Yoshida, S., and Pellman, D. (2008). Plugging the GAP between cell polarity and cell cycle. EMBO Rep. *9*, 39–41. https://doi.org/10.1038/sj.embor.7401142.

Yoshida, T., Tanaka, S., Mogi, A., Shitara, Y., and Kuwano, H. (2004). The clinical significance of Cyclin B1 and Wee1 expression innon-small-cell lung cancer. Ann. Oncol. *15*, 252–256. https://doi.org/10.1093/annonc/mdh073.

Yulo, P.R.J., Desprat, N., Gerth, M.L., Liu, Y., Zhang, X.X., Rainey, P.B., and Hendrickson, H.L. (2018). Experimental Evolution of Cell Shape in Bacteria. BioRxiv 263681. https://doi.org/10.1101/263681.

Yvert, G., Ohnuki, S., Nogami, S., Imanaga, Y., Fehrmann, S., Schacherer, J., and Ohya, Y. (2013). Single-cell phenomics reveals intra-species variation of phenotypic noise in yeast. BMC Syst. Biol. *7*, 54. https://doi.org/10.1186/1752-0509-7-54.

nature ecology & evolution

ARTICLES

https://doi.org/10.1038/s41559-022-01730-1

Check for updates

Gene loss and compensatory evolution promotes the emergence of morphological novelties in budding yeast

Zoltán Farkas[1,10], Károly Kovács[1,2,10], Zsuzsa Sarkadi[1,2,3,10], Dorottya Kalapis[1,2], Gergely Fekete[1,2], Fanni Birtyik[1], Ferhan Ayaydin[4,5,6], Csaba Molnár[1,7], Péter Horváth[1,8,9], Csaba Pál[1]✉ and Balázs Papp[1,2]✉

Deleterious mutations are generally considered to be irrelevant for morphological evolution. However, they could be compensated by conditionally beneficial mutations, thereby providing access to new adaptive paths. Here we use high-dimensional phenotyping of laboratory-evolved budding yeast lineages to demonstrate that new cellular morphologies emerge exceptionally rapidly as a by-product of gene loss and subsequent compensatory evolution. Unexpectedly, the capacities for invasive growth, multicellular aggregation and biofilm formation also spontaneously evolve in response to gene loss. These multicellular phenotypes can be achieved by diverse mutational routes and without reactivating the canonical regulatory pathways. These ecologically and clinically relevant traits originate as pleiotropic side effects of compensatory evolution and have no obvious utility in the laboratory environment. The extent of morphological diversity in the evolved lineages is comparable to that of natural yeast isolates with diverse genetic backgrounds and lifestyles. Finally, we show that both the initial gene loss and subsequent compensatory mutations contribute to new morphologies, with their synergistic effects underlying specific morphological changes. We conclude that compensatory evolution is a previously unrecognized source of morphological diversity and phenotypic novelties.

It is generally believed that phenotypic innovations driven by selection arise as a result of step-by-step accumulation of multiple beneficial mutations. Accordingly, the contribution of deleterious mutations to the evolution of phenotypic innovations is generally disregarded[1]. However, slightly deleterious mutations are far more common than adaptive mutations and can reach high frequencies in natural populations as a result of genetic drift[2], hitchhiking with adaptive mutations[3] and environmental change[4]. Indeed, the genomes of unicellular eukaryotic and human populations carry a wide range of loss-of-function alleles[5,6]. While gene loss can be beneficial in special circumstances[7], the overwhelming majority of loss-of-function mutations are non-adaptive[2]. Long-term accumulation of deleterious loss-of-function mutations would have disastrous consequences for organism survival unless the corresponding harmful effects were later mitigated by conditionally beneficial mutations elsewhere in the genome. This process, dubbed compensatory evolution, could also explain how core cellular processes can remain conserved despite substantial changes in the underlying genetic networks during evolution[8,9] (refs. [10,11] provide alternative explanations).

Signs of compensatory evolution are apparent at many organizational levels, including RNA and protein secondary structures[12], antibiotic resistance[13], genomic expression[14] and mitochondrial–nuclear co-adaptation[15]. The phenotypic effects of gene inactivation substantially vary across isolates of the same unicellular species[16], indicating the widespread occurrence of compensatory mutations that can buffer gene loss in some, but not all, populations. In the laboratory, deleterious gene loss in yeast and bacteria initiates adaptive genomic changes that rapidly restore fitness[8,17–19]. This process not only promotes genomic divergence of parallel evolving laboratory populations but also has substantial pleiotropic side effects on genomic expression, cellular physiology and viability under stressful conditions[8,18].

Here we propose that compensatory evolution following fixation of deleterious loss-of-function mutations initiate major changes in cellular and macroscopic morphological traits without direct selection on these traits. We systematically study the effect of compensatory evolution on the emergence of new morphologies by focusing on the unicellular yeast *Saccharomyces cerevisiae*.

There are several reasons to focus on cellular morphology and use baker's yeast as a model organism for the study. First, yeast cell morphology reflects a wide range of cellular events, including regulation of cell size, progression through the cell cycle and establishment of cell polarity[20]. Accordingly, activity of nearly half of the non-essential genes in the yeast genome shapes at least one cell morphology trait[20]. As the mutational target size for morphological variation is enormous, it is feasible that a large fraction of adaptive mutations has unselected pleiotropic side effects on cell morphology. Second, comprehensive analysis of yeast cell morphology is

[1]Synthetic and Systems Biology Unit, Institute of Biochemistry, Biological Research Centre, Eötvös Loránd Research Network, Szeged, Hungary. [2]HCEMM-BRC Metabolic Systems Biology Lab, Szeged, Hungary. [3]Doctoral School of Multidisciplinary Medical Science, University of Szeged, Szeged, Hungary. [4]Functional Cell Biology and Immunology Advanced Core Facility (FCBI), Hungarian Centre of Excellence for Molecular Medicine (HCEMM), Szeged, Hungary. [5]Faculty of Medicine, Albert Szent-Györgyi Health Centre, Interdisciplinary R&D and Innovation Centre of Excellence, University of Szeged, Szeged, Hungary. [6]Laboratory of Cellular Imaging, Biological Research Centre, Eötvös Loránd Research Network, Szeged, Hungary. [7]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [8]Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Helsinki, Finland. [9]Single-Cell Technologies Ltd., Szeged, Hungary. [10]These authors contributed equally: Zoltán Farkas, Károly Kovács, Zsuzsa Sarkadi. ✉e-mail: cpal@brc.hu; pappb@brc.hu

feasible through powerful high-throughput fluorescent microscopy imaging analysis[20,21]. Third, specific morphological traits, such as cell size, have direct physiological relevance as they affect the cell's surface-to-volume ratio and hence nutrient uptake rate[22,23]. Finally, cell morphological traits of natural baker's yeast isolates do not correlate with their ecological origin, indicating that they are not simply the result of adaptation to different ecological niches[24].

In a prior work, we initiated laboratory evolutionary experiments with isogenic haploid *S. cerevisiae* strains all of which initially displayed slow growth owing to the inactivation of a single gene[8]. The corresponding genes cover a wide range of molecular processes and functions. Here we focus on 86 slow-growing isogenic single-gene deletion strains (referred to as ancestor strains) and the corresponding 142 parallel evolved lines (referred to as compensated strains) from our previous study[8]. All starting and compensated strains were subjected to high-throughput fitness measurements by monitoring growth rates in liquid cultures (Methods). The compensated strains rapidly restored wild type-like fitness during the course of laboratory evolution[8] (Extended Data Fig. 1). Genomic analysis revealed that the accumulated mutations were not beneficial in the wild-type genetic background and were generally specific to the functional defect incurred[8]. To control for potential adaptation unrelated to compensatory evolution, we also established 21 populations starting from the isogenic wild-type genotype (referred to as evolved controls).

Results

Rapid evolution of cellular morphology in the laboratory. We quantitatively measured single-cell morphology of four major classes of strains, including the wild type, the evolved control ($N=21$), the ancestor ($N=86$) and the corresponding compensated strains ($N=142$). We confirmed that all investigated strains have remained haploid (Methods). We applied an established protocol designed for quantitative morphological phenotyping[20,25] that is well suited for large-scale analyses[21,24]. By using image analysis, we extracted 149 single-cell morphological traits (Supplementary Data 1) that describe various morphological attributes of exponentially growing cell populations, including average cell size, bud size and bud growth direction (Methods).

To allow unbiased comparison of different strains in the high-dimensional morphological space, we accounted for intrinsic correlations between traits using principal component analysis (PCA)[24,26] (Methods). This resulted in eight morphological principal components, together explaining ~96% of the morphological variance (Supplementary Data 2 and Extended Data Fig. 2a). For an intuitive understanding of the morphological principal components, see Fig. 1a and Extended Data Fig. 2b,c. We found that 71% of the 142 compensated strains and 66% of the 86 ancestors display significant changes in morphological principal components compared with the wild type (Fig. 1b and Methods).

Compensatory evolution may shape cellular morphology in three different manners: it (1) may restore the wild-type morphology initially perturbed by gene loss, (2) may retain the morphological change initiated by gene loss without invoking any further changes or (3) may generate novel morphological alterations (Fig. 1b).

To investigate these possibilities, significant changes in morphological principal components were calculated between the wild type, the ancestral and the compensated strain carrying the same gene deletion (Fig. 1b at a false discovery rate (FDR) of 2.53%; Methods). Generally, the cellular morphology of the compensated and the wild-type strains significantly differed from each other, hence restoration of the wild-type morphology was relatively rare (15% of the 142 cases, Fig. 1b). In 25% of cases, the ancestral and the corresponding compensated strains showed similar cellular morphologies, but they substantially differed from that of the wild type. Hence, in these cases, compensatory evolution improved fitness but left the knockout's cellular morphology unaltered. In 46% of the cases, the compensated strains displayed markedly different morphology compared with that of the wild type and the corresponding ancestral strain as well. Additionally, when only ancestral knockout strains with wild-type morphology were considered, compensatory evolution generated novel morphological states in 16 out of the 36 cases (44%). Importantly, the morphological differences between the compensated and the ancestor strains are also apparent when specific traits representative of each principal component were considered individually (Fig. 1c and Extended Data Fig. 3). Finally, hierarchical clustering of the morphological profiles coupled with a bootstrap analysis[27] identified 11 morphologically distinct groups of compensated strains with statistical support, indicating the existence of multiple distinct morphotypes (Extended Data Fig. 4a,b and Supplementary Data 3). Together, these results suggest that compensatory evolution following gene loss frequently promotes the evolutionary divergence of morphological traits despite yielding wild type-like fitness.

Several lines of evidence indicate that the observed morphological changes are driven by specific deleterious and compensatory mutations. First and foremost, the morphological changes in the evolved controls were limited, indicating that morphological divergence of compensated strains is not the result of adaptation to the physical environment (Extended Data Fig. 5a–d). Second, morphological changes are not simply side effects of growth rate increase for two reasons: (1) only one of the 149 single-cell morphological traits (D197-C; that is, nuclei size ratio) shows a significant association with growth rate of the compensated strains (Spearman's correlation, FDR<5%, Supplementary Data 4); and (2) the extent of changes in morphological traits are also not associated with the extent of growth rate increase during compensatory evolution (Spearman's correlation, FDR<5%, Supplementary Data 4). Third, morphological evolution in the laboratory may be driven by numerous mutations, all of which have small effects individually. If it were so, one might expect a positive correlation between the number of mutations that have accumulated during the course of laboratory evolution and the extent of morphological change. However, analysis of available genome sequences of a set of 18 compensated strains[8] revealed no significant correlation between these two variables (Extended Data Fig. 5e), indicating that certain compensatory mutations have exceptionally large effects on cellular morphology.

Comparable diversity of compensated and natural strains. Next, we focused on three morphological traits: cell size, cell elongation and bud neck position for three reasons. First, out of the 149 measured morphological traits, changes in cell size and cell elongation were especially high, yielding up to ~50% and ~35% increments, respectively, in the compensated strains compared with that in the wild type (Fig. 2a) and explaining most variation across strains (Fig. 1a). Second, compensated strains with extremely low bud neck positions form a distinct group (Fig. 2a), corresponding to a highly diverged and statistically significant cluster in the cluster analysis (Extended Data Fig. 4a,b). Third, and most importantly, these three traits are highly variable across natural *S. cerevisiae* isolates[24].

We compared the extent of morphological diversity induced by laboratory evolution to that observed in nature. For this purpose, we measured these three morphological traits in 29 natural *S. cerevisiae* isolates with diverse ecological origins (Supplementary Data 5 and Methods). The selected isolates represent several diverged clades of *S. cerevisiae*[28] and were confirmed to be haploids (Methods). While the natural isolates differ by ~10^2–10^5 mutations from each other[28], the compensated strains typically differ from the wild type by only ~6 mutations in addition to the focal gene deletion[8]. Despite these vast differences in the extent of genomic divergence, the overall extent of morphological variation in the compensated strains is comparable to that of natural strains (Fig. 2a,b). This pattern

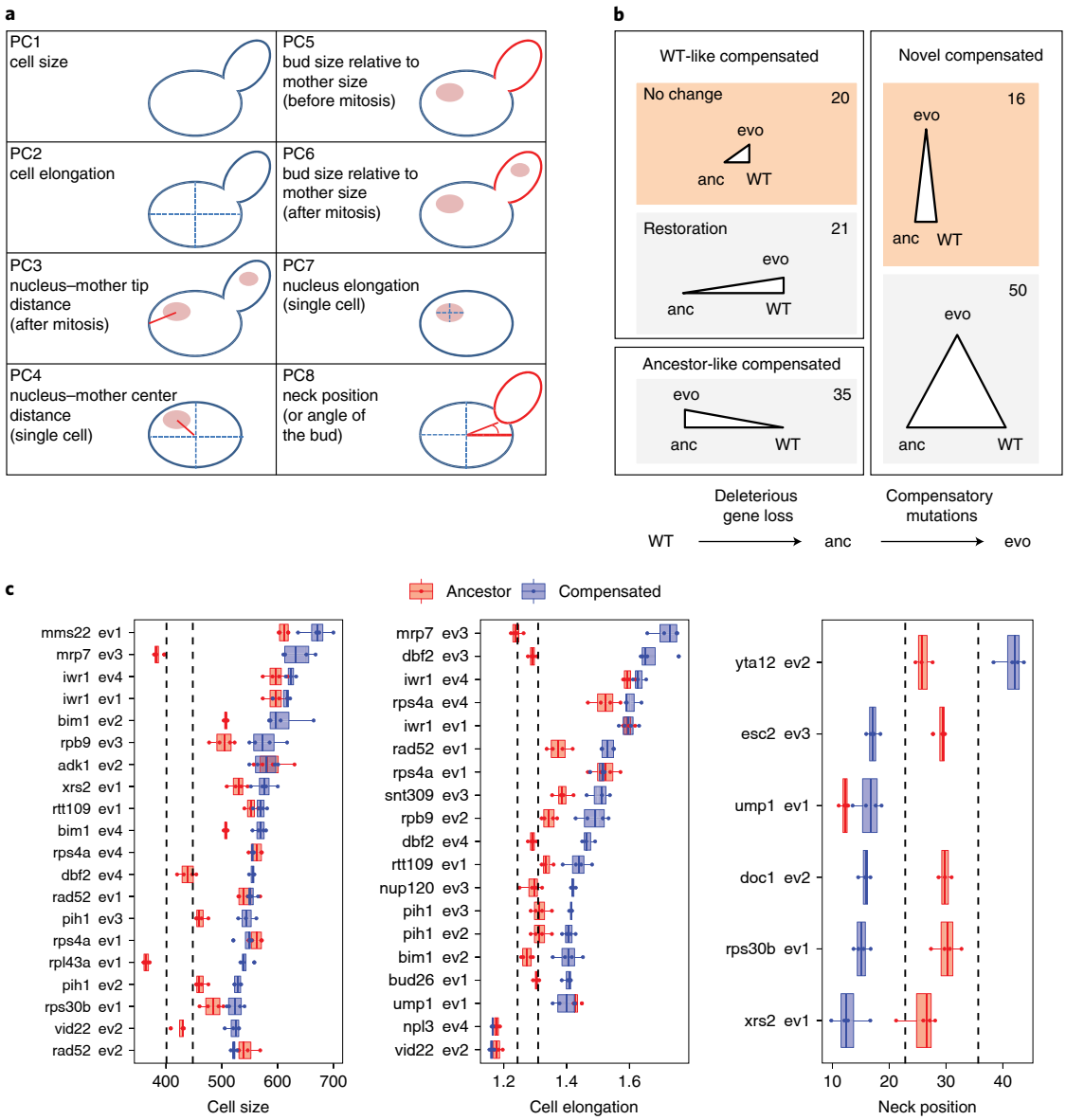


Fig. 1 | Compensatory evolution alters cellular morphology. a, Schematic representation of the principal components (PC1–PC8). The figure shows the intuitive meaning of the first eight principal components based on the top contributing morphological traits (Extended Data Fig. 2b). Separate morphological traits are defined for three cell cycle stages (single cell, budded cell before or after mitosis). Principal components without an indication of a specific cell cycle stage can be linked to all cell cycle stages for a given type of trait. **b**, Main classes of evolutionary trajectories in morphological space and their prevalence. Compensated strains may not differ significantly from the wild type ('WT-like compensated'), differ from wild type while being similar to the ancestor ('Ancestor-like compensated') or differ from both the wild type and the ancestor ('Novel compensated'). Compensated strains with wild type-like ancestors are coloured with orange background. For each triplet of genotypes involving a compensated strain (evo), morphological differences between genotypes are visualized by a triangle (long and short edges denote presence and absence, respectively, of significant difference in at least one principal component; Methods). Numbers indicate the number of compensated strains corresponding to each evolutionary trajectory. Bottom flowchart depicts the two steps of compensatory evolution: (1) deleterious gene loss that results in a low fitness state (anc) and (2) accumulation of subsequent compensatory mutations that mitigate the fitness defect of gene loss and result in a fitness-compensated strain (evo). **c**, Changes in three morphological traits in compensated strains with the most extreme trait values. Box plots showing cell size, cell elongation and neck position values of compensated strains that differ most from the wild type (based on $n=3$ or $n=4$ biological replicates per strain). Cell size (area in pixels of unbudded cell), cell elongation (ratio of the long- and short-axis length of the cell) and neck (that is, boundary between mother and daughter cell) position (angle unit) correspond to CalMorph traits C11-1-A, C115-A and C105-A1B, respectively. Dashed lines indicate the range of the wild-type values (average ±2 standard deviations). Box plots show the median, first and third quartiles, with whiskers showing the 5th and 95th percentiles.

is especially remarkable considering that we probably overestimate morphological variation of natural strains as many of them are monosporic derivatives of isolates that are diploid in the wild[28]. Surprisingly, certain compensated strains resemble specific natural isolates in their cell size and cell elongation values (Fig. 2c and Extended Data Fig. 6a). We conclude that compensatory evolution in the laboratory generates morphological diversity in an exceptionally rapid manner (that is, within ~400 generations).

Yeast cell size and cell shape are intimately linked with progression through the cell cycle, therefore we next asked whether the rapid

evolution of cell size and cell elongation is associated with changes in cell cycle phase distributions. To this end, we estimated the fraction of cells residing in G1 and G2 cell cycle phases (that is, G1 and G2 percentages, respectively) using flow cytometry (Methods and Supplementary Data 6). Indeed, we found large changes in both G1 and G2 percentages in the compensated strains compared with that in the wild type (Fig. 2d). Specifically, 32.4% ($n=46$) and 43.7% ($n=62$) of the compensated strains show a marked alteration in G1 and G2 percentages, respectively. Prior works unveiled a control mechanism that regulates G1 duration based on cell size, resulting in a relatively shorter G1 for larger cells[23]. As might be expected, we found a negative correlation between cell size and G1 percentage (Fig. 2d). More remarkably, bud elongation has been shown to be directly affected by the relative duration of G2 (ref. [29]), and there was a positive correlation between G2 percentage and both bud and cell elongation across compensated strains (Fig. 2d and Extended Data Fig. 7). While deciphering the mechanistic underpinnings of these patterns goes beyond the scope of the current paper, the latter result indicates that cell elongation has partly evolved through changes in cell cycle phase distributions.

Rapid evolution of multicellular phenotypes. Alteration of cellular morphologies and formation of multicellular structures have been reported to be linked in yeast[30]. Therefore, we next asked whether compensatory evolution alters the capacity to exhibit multicellular phenotypes. Different species of yeasts typically undergo a developmental transition from a single-cell form into multicellular forms upon environmental change[31,32]. We studied three different forms of multicellular phenotypes: (1) invasive growth phenotype that permits penetration into solid agar[33], (2) biofilm formation that allows adherence to semi-solid agar[34] and (3) cellular aggregation in liquid medium via flocculation[35] or clump formation[36]. These three traits enable survival under stressful conditions, aid nutrient acquisition[33] and contribute to virulence in pathogenic yeasts[32,37]. However, several natural *S. cerevisiae* strains have lost their capacity to display these multicellular traits[38].

We screened the wild type, the ancestor and the compensated strains for invasive growth, cell aggregation and biofilm formation using established methods (Methods). Consistent with earlier works, the wild type fails to show invasive growth and cell aggregation phenotype[39]. It also forms 50% smaller biofilms[34] compared with the positive control sigma 1278b strain (Supplementary Data 7 and Fig. 3e). Strikingly, however, several compensated strains gained the capacity of invasive growth (13%), formed enlarged biofilms (2.8%) or displayed multicellular aggregates in liquid (12.4%) (Fig. 3 and Extended Data Fig. 8a–d). Multicellular aggregation in the compensated strains was achieved by incomplete daughter cell separation (that is, clumping[36]) rather than by flocculation of previously separated cells (Extended Data Fig. 8e).

Overall, 23.4% of the compensated strains showed an enhanced capacity to display at least one multicellular trait (Fig. 3d). This figure is striking as the capacity to display these traits are unlikely to confer any benefit in the well-shaken liquid medium employed during the course of laboratory evolution. Consistent with this notion, none of the 21 control evolved strains exhibited any of these phenotypes ($P=0.0055$, Fisher's exact test; Extended Data Fig. 8a–c and Supplementary Data 7). Similarly, the corresponding ancestral strains typically displayed only mild or no changes in these traits (Fig. 3a–c and Supplementary Data 7). Together, these results indicate a prevalent role of compensatory evolution in generating multicellular phenotypes.

An important issue is whether multicellular phenotypes arise through convergent evolution at the molecular level. If it were so, one would expect an overlapping set of genes to be mutated in strains displaying the same evolved phenotype. This was clearly not the case: compensated strains that independently evolved the same multicellular trait showed no overlap in their sets of mutated genes ($N=46$ mutations in 9 sequenced strains, Supplementary Data 8). This indicates that morphological novelties arise through diverse mutational routes in the laboratory.

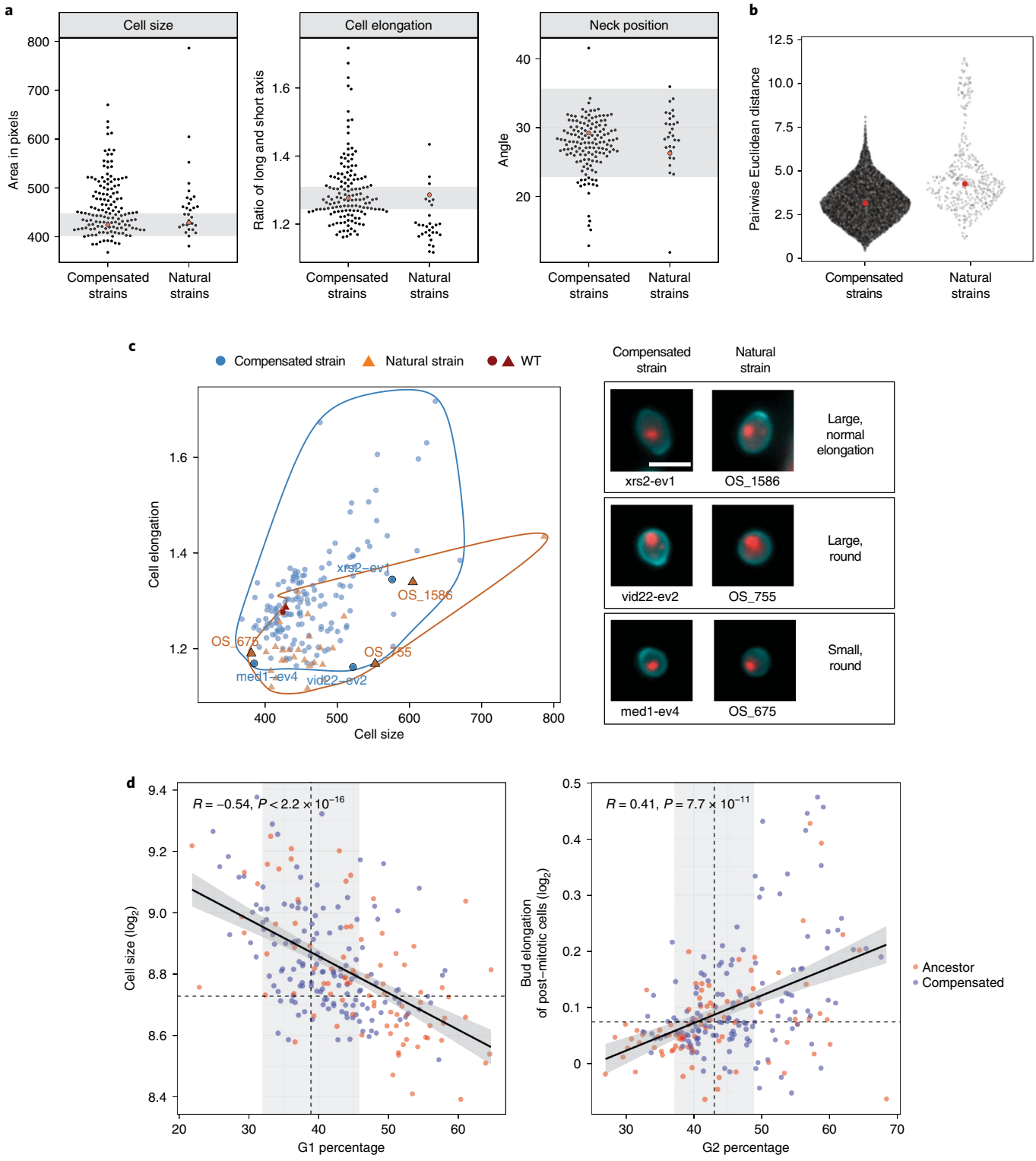
Prior studies demonstrated a substantial variation in invasive growth phenotype among natural yeast strains[38]. Here we compared the extent of variation in this phenotype in the laboratory-evolved strains and natural isolates with different ecological origins. For this purpose, we measured the capacity for invasive growth phenotype in a set of natural *S. cerevisiae* isolates (Methods and Supplementary Data 9). Unexpectedly, the extent of gain in invasive growth phenotype in the compensated strains reaches as high as ~50% of the range of invasiveness displayed by the natural isolates (Extended Data Fig. 9a).

A compensated laboratory strain with *Δbub3* genetic background (bub3-ev3) shows an especially high invasiveness score and it displays filament-like multicellular structures invading into the

Fig. 2 | Comparable morphological diversity of compensated and natural strains. a, Distribution of three morphological traits across compensated and natural strains. Black dots in swarm plots represent the average values of cell size, cell elongation and neck position (based on at least three biological replicates per genotype) for compensated and natural strains. Red dots indicate the average value of the wild-type biological replicates in the two sets of measurements ($n=88$ and $n=16$ for compensated strains and natural isolates, respectively). Grey area represents the average ±2 standard deviation of the 88 wild-type replicates of the set of measurements involving the compensated strains. **b**, Distribution of pairwise Euclidean distances of the morphological profiles for all pairs of compensated strains ($N=10{,}011$) and natural isolates ($N=406$). Each dot represents a pair of compensated or natural strains. Red dots indicate the median distance of all strain pairs in the two sets. **c**, Similar cellular morphology of compensated strains and natural isolates. Scatter plot showing the distribution of compensated strains (blue dots) and natural isolates (orange triangles) according to average cell size and cell elongation values (based on at least three biological replicates). Red dots and red triangles show the same wild-type genotype (WT) in the screens of the compensated and natural strains ($n=88$ and $n=16$ replicates, respectively). Images show pairs of compensated and natural strains that display similar morphological trait combinations (cell wall and nuclei are coloured with green and red, respectively): (1) large cells with normal, wild-type-like elongation: xrs2-ev1, OS_1586 isolate from tree leaves, (2) large round cells: vid22-ev2, OS_755 wine yeast isolate and (3) small round cells: med1-ev4, OS_675 isolate from human blood. Cell size and cell elongation in panels **a** and **c** and neck position in panel **a** correspond to the same CalMorph parameters as in Fig. 1c. Scale bar (on image of xrs2-ev1) represents 5 µm. For larger fields of view, see Extended Data Fig. 6a. **d**, Compensatory evolution alters cell cycle. Left panel shows the scatter plot of mean cell size and mean G1 percentage in both the ancestor (orange) and compensated (blue) strains. Cell size shows a significant negative correlation with G1 percentage (that is, fraction of cells in the G1 cell cycle phase in the population) across all genotypes (Pearson's correlation). Right panel shows the scatter plot of mean bud elongation and mean G2 percentage (that is, fraction of cells in the G2 cell cycle phase) in both the ancestor (orange) and compensated (blue) strains. Bud elongation shows a significant positive correlation with G2 percentage (Pearson's correlation). Cell size and bud elongation correspond to the CalMorph traits C11-1-A (that is, the size of unbudded mother cell as in Fig. 2c) and C114-C (that is, the ratio of the long- and short-axis length of the bud in post-mitotic cells), respectively. Mean G1 and G2 percentages were estimated by performing flow-cytometry analysis of two biological replicates for each strain (Methods). Dashed lines represent the average of the wild type. Grey area represents the wild type average ±2 standard deviations. We estimated standard deviation using the pool of strain-wise centred replicate measurements of all investigated strains.

agar (Fig. 4a). Invasive growth in *S. cerevisiae* is generally mediated by Flo11p, a key cell surface protein whose regulation is impaired in the laboratory yeast background used as a wild type here[40]. We found that bub3-ev3 did not regain the ability to express *FLO11*, as evidenced by low activities of the *FLO11* promoter (Fig. 4b) and of a reporter protein of the filamentous growth pathway (Extended Data Fig. 9b and Supplementary Data 10 and 11). Thus, our data suggest that invasiveness in this strain emerged without reactivating the canonical filamentous regulatory pathway.

Synergistic epistasis at the level of morphology. Finally, we conducted in-depth genetic analyses to decipher potential epistatic interactions between loss-of-function and compensatory mutations. We first focused on a compensated strain with *Δrpb9* genetic background (rpb9-ev2) displaying an elongated cell shape (Fig. 5a). Rpb9-ev2 carries a compensatory loss-of-function mutation in *WHI2* (ref. [8]). The *RPB9* gene encodes an RNA polymerase II subunit whose deletion reduces transcriptional fidelity and inhibits growth as a result of proteotoxic stress response[41] and subsequent



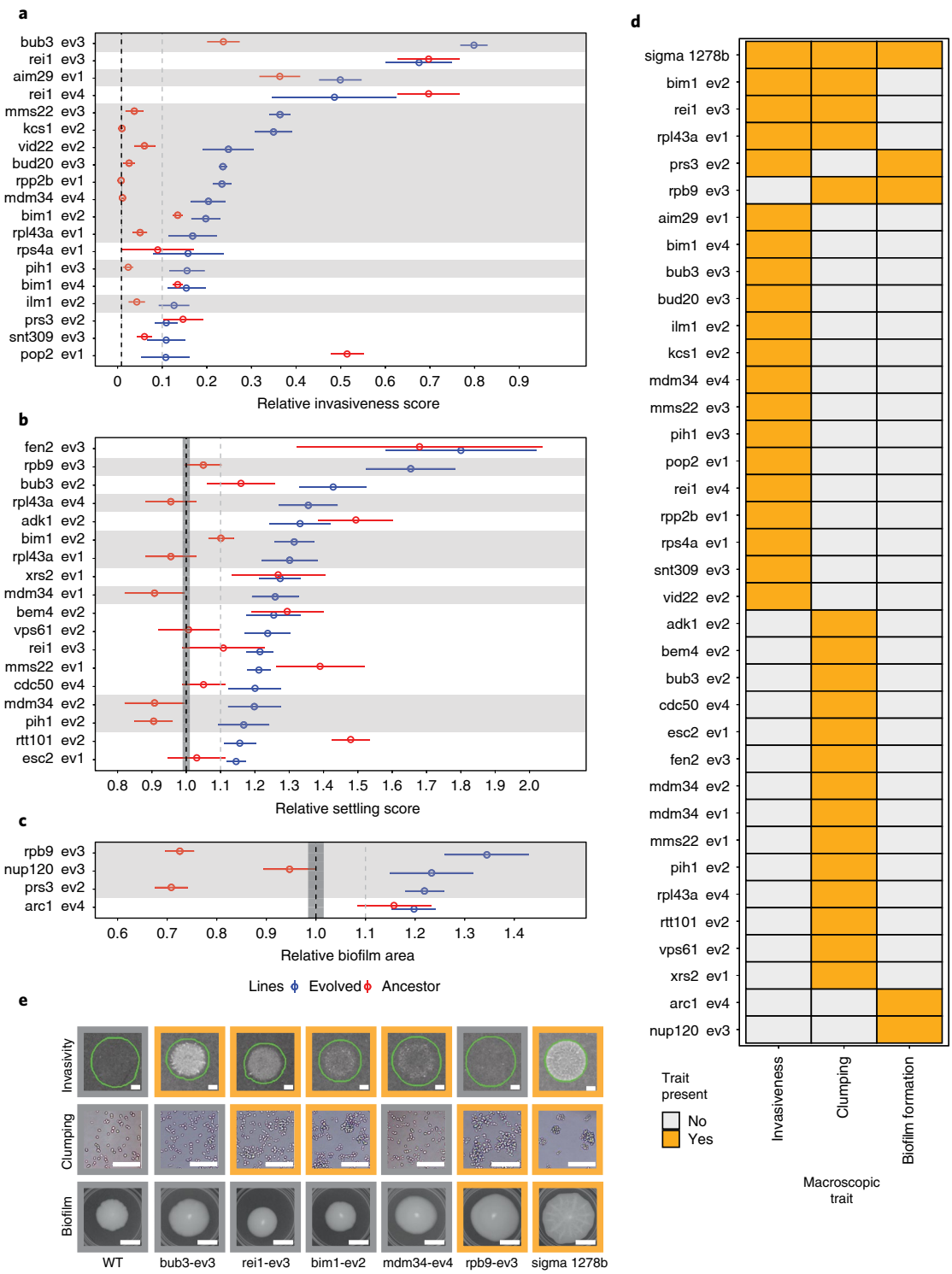


Fig. 3 | Compensatory evolution promotes the emergence of multicellular phenotypes. a,**b**,**c**, Plots show trait values for three types of multicellular form: invasiveness (**a**), multicellular aggregates as measured by settling (**b**) and biofilm area (**c**). Along with the corresponding ancestor strains (red), only those compensated strains (blue) are plotted that show a marked change in trait value compared with the wild type based on both *q*-value and effect size thresholds (grey dashed line) (Methods). Mean (black dashed line) and standard error (dark grey shaded area) of the wild type for the invasive growth/settling/biofilm assay was calculated based on 156/106/42 biological replicates, respectively. Please note the small standard error of the wild type in **a**. The horizontal light grey shaded areas mark those pairs of strains where the compensated strain displays a significantly higher value than that of the corresponding ancestor strain (one-sided Student's *t*-test, 10% FDR). The empty circle and the point range represent the mean and the standard error of the corresponding trait values, respectively, based on at least three biological replicates. **d**, Heat map summarizes the presence (orange) or absence (light grey) of alterations in the three investigated traits across 41 compensated strains, each of which display a marked change in at least one multicellular trait compared with the wild type. **e**, Representative images of six compensated strains showing an increased level in at least one multicellular trait (with sigma 1278b as a positive control). Photos with orange frame mark significant trait changes. Scale bars represent 1 mm, 50 μm and 1 cm on the images in rows invasivity, clumping and biofilm, respectively. For larger fields of view of microscopic images, see Extended Data Fig. 8d.

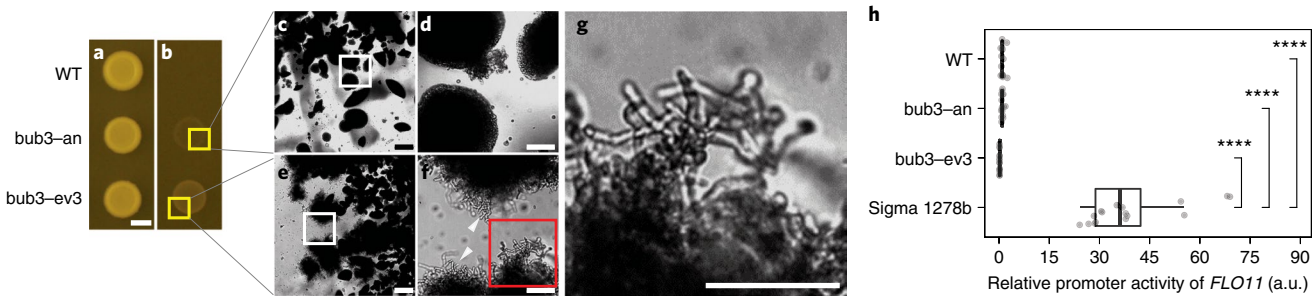


Fig. 4 | High invasiveness of a compensated strain. a–f, Filament formation appeared in a compensated strain of *Δbub3*. **a**,**b**, The figure shows yeast spots (illuminated from the top) of wild type, the ancestor (bub3-an) and a compensated strain (bub3-ev3) of *Δbub3* on solid medium after seven days of incubation, before (**a**) and after (**b**) plate washing (Methods). **c**–**g**, Transmitted light microscopy images show the agar-embedded structures after plate washing with different magnifications. Yellow boxes in **b** mark the areas that are magnified in **c** and **e**, respectively. White boxes in **c** and **e** mark the areas that are magnified in **d** and **f**, respectively. White triangles and red box in **f** indicate extensive filament formation in bub3-ev3 line. Red box in **f** marks the area that is magnified on **g**. Scale bars: 3 mm (**b**), 250 μm (**c** and **e**) and 50 μm (**d**, **f** and **g**). **h**, Promoter activity of *FLO11* across four genotypes, including the wild type, the ancestor (bub3-an) and a compensated strain (bub3-ev3) of *Δbub3* and the positive control strain (sigma 1278b). Absolute promoter activity was estimated by measuring the colony fluorescence (that is, the optical density normalized fluorescent level) of yeast spots after seven days of incubation on solid medium, using a pFLO11-yEVenus-tADH1 reporter plasmid (Supplementary Protocols). Relative promoter activity of *FLO11* was calculated by normalizing the absolute promoter activity of the genotypes to that of the wild type. The relative promoter activity of *FLO11* of the sigma 1278b genotype shows a median 36-fold increase compared with the laboratory wild type (median relative promoter activity of *FLO11*=1). Box plots show the median, first and third quartiles, with whiskers showing the 5th and 95th percentiles of two independent measurements of eight biological replicates for each genotype. Significant differences were assessed by two-sided Student's *t*-tests (**** indicates $P<0.0001$). The *P* values are 2.1×10^{-8}, 2.1×10^{-8} and 1.6×10^{-8} for comparing sigma 1278b with wild type, bub3-an and bub3-ev3, respectively. a.u., arbitrary units.

downregulation of the major growth control (target of rapamycin) pathway. As noted earlier, a loss-of-function mutation in Whi2p, a negative regulator of the target of rapamycin pathway[42], mitigates the fitness defect of *Δrpb9* (ref. [8]) via relief of growth inhibition. Intriguingly, deletion of *WHI2* not only improves fitness but also yields elongated cells in *Δrpb9*, while it is non-beneficial[8] and has no major effect on cell elongation in the wild type (Fig. 5a,b and Extended Data Fig. 6b.). We conclude that the initial loss-of-function mutation (*Δrpb9*) and the compensatory mutation (*Δwhi2*) display synergistic epistasis on cell elongation (Fig. 5b).

Analysis of a compensated strain (bub3-ev3) exhibiting a novel multicellular phenotype resulted in a similar conclusion: bub3-ev3 shows an exceptionally high capacity for invasive growth (Fig. 3a) compared with that of all other compensated, several natural (Extended Data Fig. 9a) and the corresponding *Δbub3* ancestor strains. The *BUB3* gene encodes a spindle assembly checkpoint protein involved in delaying the cell division until each chromosome is properly attached to the microtubules[43]. Bub3-ev3 carries a non-synonymous point mutation (Y332S) in a morphogenesis checkpoint protein Swe1p[8]. Swe1p normally delays nuclear division until the bud has been formed[44]. It is plausible that the mutation in Swe1p ($SWE1^{Y332}$) alters cell cycle kinetics and thereby partly compensates for spindle checkpoint defects. We introduced the $SWE1^{Y332}$ mutation into both the wild-type and *Δbub3* genetic backgrounds. The mutation significantly improved fitness in *Δbub3* but not in the wild type (Extended Data Fig. 10 and Supplementary Data 12), confirming its causal role in fitness compensation. Remarkably, the reconstructed double mutant *Δbub3* $SWE1^{Y332S}$ showed a strong invasive growth phenotype similar to that of bub3-ev3 (Fig. 5c,d). In contrast, the $SWE1^{Y332S}$ mutation had no major impact on invasiveness in the wild type, indicating synergistic epistasis between gene loss and the associated compensatory mutation on invasiveness (Fig. 5d).

Discussion

In this work, we addressed a long-standing debate on the role of non-adaptive mutations in generating evolutionary novelties. Compensatory evolution following gene loss generates hidden genetic and physiological variation across parallel evolving lines[8]. As a consequence, gene loss initiates adaptive genomic changes that rapidly restore fitness, but this process has pleiotropic side effects on cellular physiology and evolvability[8,45]. Here we hypothesized that organisms undergo major changes in morphological traits not simply to adapt to external conditions but also as a by-product of compensating for previously accumulated deleterious mutations. Using laboratory evolution, we found an exceptionally rapid diversification of morphological traits in response to gene loss. Furthermore, in-depth genetic analyses demonstrated that the loss-of-function and the compensatory mutations are jointly required for specific morphological alterations (Fig. 5).

Both the set of disrupted genes and the mutations that have accumulated during the course of laboratory evolution are very diverse in the compensated strains[8]. Therefore, we propose that a broad range of molecular processes associated with compensatory evolution can initiate major changes in cellular morphology. Indeed, compensated strains that independently evolved the same multicellular trait display no overlap in their sets of mutated genes, suggesting that morphological novelties can arise through diverse mutational routes. While these analyses focused on point mutations, structural variants such as chromosome aneuploidies might also contribute to morphological changes[46]. Clearly, future systematic works should reveal the relative contribution of different types of compensatory mutation to morphological novelties.

Our work indicates that compensatory evolution generates rudimentary forms of ecologically and clinically relevant traits, which could be later fine-tuned by direct selection. Some of these traits, including invasive growth and biofilm formation, are virulence factors in several pathogenic fungi[32,37], suggesting that gene loss and subsequent compensatory mutations could be relevant for the emergence of pathogenicity.

Despite the short time scale of laboratory evolution, the extent of morphological diversity was comparable to that of haploid natural yeast isolates with diverse genetic backgrounds and lifestyles. These results could be relevant for understanding the evolution of cellular morphology in the wild. Systematic analyses of morphological traits in natural yeast isolates have revealed abundant trait variation

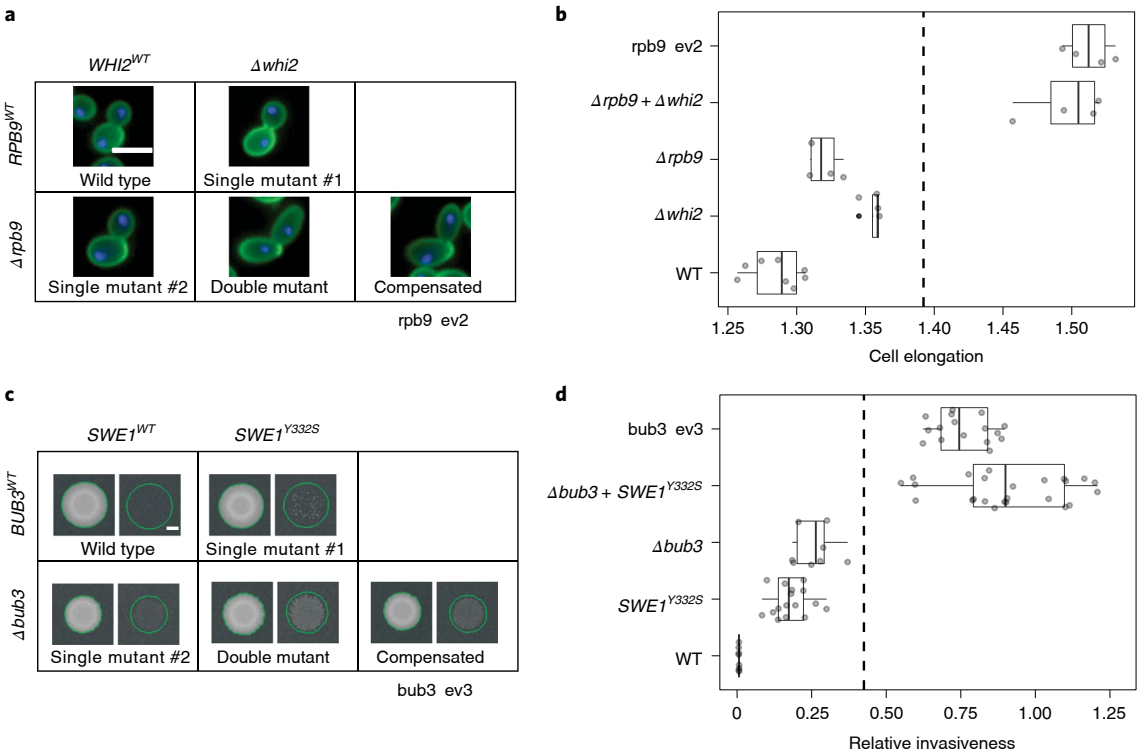


Fig. 5 | Synergistic epistasis underlying morphological changes in two compensated strains. a, Representative images of yeast cells of five selected genotypes, including the wild type, two single mutants (*Δrpb9* and *Δwhi2*) and a reconstructed double mutant (*Δrpb9* + *Δwhi2*). The fifth genotype is the compensated strain of *Δrpb9* (rpb9-ev2) that harbours the *whi2*^{S133*} loss-of-function allele. Loss-of-function mutation of *WHI2* recapitulates the elongated cell shape of the compensated strain when introduced into *Δrpb9* but not when introduced into the WT. Cell wall and nuclei are coloured with green and blue, respectively. Scale bar represents 5 µm for all images. For larger fields of view, see Extended Data Fig 6b. **b**, Box plots show cell elongation values (CalMorph trait C115-A) of the same five genotypes (based on $N = 4$ and $N = 8$ replicates for the mutants and the wild type, respectively) as in **a**. Vertical dashed line shows the expected value of the double mutant in the absence of epistasis (Methods). **c**, Representative images of yeast spots before and after plate washing (left and right images, respectively, in each grid) across five selected genotypes, including the wild type, two single mutants (*Δbub3* and *SWE1*^{Y332S}) and a reconstructed double mutant (*Δbub3* + *SWE1*^{Y332S}). The fifth genotype is the compensated strain of *Δbub3* (bub3-ev3) that harbours the *SWE1*^{Y332S} mutant allele. *SWE1*^{Y332S} confers a high capacity for invasive growth when introduced into *Δbub3*, whereas it has only a minor effect in the wild type background. Scale bar represents 1 mm for all images. **d**, Box plots show the relative invasiveness score of the same five genotypes (using at least eight biological replicates for each genotype) as in **c**. Vertical dashed line shows the expected value of the double mutant in the absence of epistasis (Methods). Box plots show the median, first and third quartiles, with whiskers showing the 5th and 95th percentiles.

between strains with no correlation with their ecological origin or genetic relatedness[24,38]. However, these traits are also claimed to be shaped by positive selection rather than neutral evolution[47]. These seemingly contradictory findings are consistent with the hypothesis that morphological diversity is driven partly by compensatory evolution and not simply by adaptation to changing ecological settings. Indeed, it has been previously suggested that adaptive mutations at the genomic level may generally reflect compensatory rather than progressive changes in the phenotype[48]. For example, an ancient lineage of budding yeasts (genus *Hanseniaspora*) has diversified and thrived, despite lacking many cell-cycle checkpoint genes which are essential in other fungal lineages[49]. Perhaps importantly, these yeasts underwent a burst of accelerated molecular evolution and display markedly altered budding morphology. Further work should be done to explore whether compensatory evolution has played a role in the phenotypic evolution of the *Hanseniaspora* lineage.

Although putatively deleterious loss-of-function alleles are prevalent in eukaryotic genomes, the underlying evolutionary mechanisms are far from being clear. One can envisage at least three possibilities. Population bottlenecks and genetic drift promote the accumulation of deleterious mutations, but this scenario may be unfeasible in microbes with large effective population sizes. It is more likely that loss-of-function mutations are positively selected if they are beneficial in one environment and detrimental in another environment[8,9]. Indeed, a large fraction of gene deletant mutants display such antagonistic pleiotropy in baker's yeast[4]. It is an interesting and open issue how subsequent accumulation of compensatory mutations shape fitness across environments. Finally, the order of events may occur the other way around: nearly neutral compensatory mutations may arise first in the population that later permit the fixation of loss-of-function mutations without serious fitness consequences. This scenario is consistent with a recent systematic study that showed that pre-existing natural genetic variants frequently suppress deleterious mutations in specific strains of baker's yeast[50].

Our work has some general implications on the tempo and mode of phenotypic evolution. It is generally believed that phenotypic evolution proceeds in small discrete steps. The justification for this assumption is that mutations with large effects tend to have deleterious side consequences[51]. While this could be generally true, our work indicates that the fitness defects associated with mutations can be rapidly mitigated by compensatory evolution without restoring the original phenotype. Hence, laboratory evolution can generate 'hopeful monsters'[52], organisms with profoundly altered morphological phenotypes but normal fitness. In the future, it will be important to establish whether such hopeful monsters have the potential to establish new evolutionary lineages through

promoting reproductive isolation. In sum, our work provides direct experimental evidence that loss-of-function mutations can act as stepping stones[1] in morphological evolution.

Methods

Yeast strains and media. All strains used in this study were derived from the BY4741 *S. cerevisiae* parental strain. A set of slow-growing non-essential single-gene deletion strains from the haploid yeast deletion collection (MATa; *his3Δ 1*; *leu2Δ 0*; *met15Δ 0*; *ura3Δ 0*; *xxx::KanMX4*) were used in our previous study to investigate compensatory evolution in the laboratory[8]. In this prior batch selection experiment, each slow-growing deletion strain (that is, ancestor strain) was propagated for ~400 generations under standard laboratory conditions (YPD medium: 1% yeast extract, 2% peptone, 2% glucose) that resulted in fitness-compensated evolved lines (referred to as compensated strains throughout the text). As a proxy for fitness, we previously measured growth rate[8]. Growth rate was estimated by monitoring the growth curves of yeast populations in liquid medium using a modified version of an established procedure[53,54]. To investigate the effect of the initial deleterious and subsequent compensatory mutations on single-cell and multicellular morphological features, we here used isolated clones of both the slow-growing ancestor strains ($n=86$, impaired in diverse cellular functions) and 1–4 parallel evolved compensated strains ($n=145$) per ancestor that showed fitness compensation and remained haploid during the laboratory evolution (details below). Note that all strains, including ancestor, compensated and control evolved strains, were saved and stored as glycerol stock until further experiments. All experiments throughout this study were also performed on a set of control evolved lines ($n=21$) initiated from the wild-type strain (*his3::KanMX*, designated as wild type throughout the manuscript) in our previous study[8] to control for morphological changes unrelated to compensatory evolution. To investigate the diversity of morphological traits across natural baker's yeast isolates, we selected 29 haploid and euploid strains (Supplementary Data 5) from a collection of 1,011 *S. cerevisiae* isolates[28]. The selected isolates represent several phylogenetically diverged clades of *S. cerevisiae* with diverse ecological origins. For the quantitative invasive growth/settling/biofilm formation/filamentous response assays (see below), we used the sigma 1278b parental strain (L5684, matA, *ura3-52*, *leu2::hisG*) as a positive control[33]. To distinguish between calcium-dependent aggregation (that is, flocculation) and mother–daughter separation defect (that is, clump formation) as a mechanism underlying enhanced settling (see below), we used a highly flocculating natural strain (OS_1189) as a positive control[28].

For single-cell morphological phenotyping, ploidy analysis and fitness measurements, all strains were cultured in rich, YPD liquid medium. Quantitative biofilm-formation assays, colony size-based fitness measurement and quantitative invasive growth assays were performed on solid YPD medium with low-density (0.3%), medium-density (2%) and high-density (3%) agar, respectively.

High-throughput single-cell morphology measurements. To quantitatively measure single-cell morphology, we applied a previously established protocol[20,25] with minor modifications. Single colonies were selected from yeast populations to establish single clones. Yeast strains were inoculated into standard 96-well plates in four biological replicates as follows. Random layouts were generated by a custom R script to avoid plate effects biasing the genotype-specific morphological trait values. This protocol resulted in layouts where each biological replicate position is randomly chosen with two constraints. First, each biological replicate of a given genotype is placed on a separate plate to ensure that it is measured on different days. Second, each plate contains four control wells (wild type) in random positions within different quadrants in distinct rows and columns. Overall, this resulted in 88 replicates for wild type. After reaching the saturation in cell density, each culture was diluted into 500 µl fresh YPD medium in a 96 deep-well plate including 0.5 mm glass beads in each well and grown until mid-exponential phase. After that, the cells were fixed in phosphate buffer (pH 6.5) containing 3.7% formaldehyde (Sigma-Aldrich). Fixed cells were washed with phosphate-buffered saline (PBS) (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 1 mM $CaCl_2$, 0.5 mM $MgCl_2$) and P buffer (10 mM Na_2HPO_4, pH 7.2, 150 mM NaCl), and then we performed fluorescent staining of the cell wall and the nucleus (actin-staining was omitted due to low reproducibility as in a previous study[21]). Staining of the cell wall was performed by incubating the fixed cells with 1 mg ml^{-1} Alexa Fluor 488 conjugate of concanavalin A (Alexa-488-conA, Thermo-Fisher) solution for 2 h at 4 °C. After washing with P buffer, the cells were kept in PBS buffer at 4 °C until microscopy. Before microscopy, the nuclei were stained with 350 ng ml^{-1} 4′,6-diamidino-2-phenylindole (DAPI, Thermo-Fisher) in PBS buffer supplemented with 0.1% Triton X-100 (Molar Chemicals) for 30 min at room temperature. After repeated pipetting (at least 30 times to minimize cell aggregates), the cells were diluted and transferred into black clear-bottom 96-well plates (Greiner) coated with 1 mg ml^{-1} concanavalin A solution (conA, Santa-Cruz Biotechnology). Fluorescently stained cells were sedimented by centrifugation (1,750 × gravity for 4 min). Microscopy screening was performed by an Operetta High-Content Imaging System (PerkinElmer Inc.) using a 63× high numerical aperture dry objective. During the imaging, 13 fields were captured from each well with two channels configured for Alexa-488-conA and DAPI in four layers of z-stack. Raw tiff images were processed using custom a MATLAB script to select the optimal z-stack layer for each cell and to produce 696 × 520 8-bit jpeg images (four images per field of view), which were then used as inputs for the CalMorph software[20]. The same procedures were applied in the cases of reconstructed mutants and natural isolates.

Statistical analysis of morphological data. All data analysis was performed using the R language[55]. First, average morphological trait values were calculated for each biological replicate (that is, separate well). These average trait values (with the exception of traits representing angles) were log2 transformed, and then the transformed values were weighted by the number of cells per well for each genotype. Next, these genotype-level averages were standardized to the wild type: the average value of the wild type was subtracted, then this value was divided by the standard deviation of the wild-type biological replicates ($n=88$). Dimension reduction was performed on this wild-type standardized data by using principal component analysis resulting in eight 'principal component traits' explaining ~96% of all variance. We used the 'prcomp()' function from R using default parameter values.

The morphological profile of each strain is defined by the eight principal component values, transformed to have unit variance across strains. This ensures that each principal component (type of morphological trait) is equally considered in the analyses below, not only the ones with the highest variance.

To define 'significant' morphological differences between strains, biological replicates of all strains were projected to the principal component space calculated above. Average ±2 standard deviation intervals were calculated from the biological replicates of each strain. Two strains are defined as being significantly different in a given principal component when their intervals are non-overlapping. We used standard deviations instead of statistical tests or confidence intervals because the results of the latter methods are highly influenced by the differences in sample sizes (3, 4 or 88). Two strains are defined as 'morphologically different' if they differ from each other in at least one principal component.

We estimated the FDR of identifying significantly altered morphologies between two strains as follows. We first estimated the FDR of comparing 4–4 biological replicates, representing most of the comparisons between ancestral and compensated deletion strains (94 of 142 comparisons). To this end, we used the 88 biological replicates of the wild-type strain, randomly selecting two sets of four biological replicates for comparison. We reasoned that significant differences between such subsets of wild-type replicates represent false positives. Next, we calculated whether the two sets show a significant difference in their morphology (that is, their mean ±2 standard deviation intervals overlap for at least one principal component trait). To estimate the FDR, we repeated the random sampling 10,000 times and calculated the proportion of significant cases, resulting in a FDR of 1.3%, yielding ~1.3 expected false positive cases in our dataset. In a similar vein, we also estimated the FDR for comparing three versus four biological replicates (representing 42 out of 142 comparisons). This yielded ~1.8 false positive cases. Finally, comparison of three versus three biological replicates (6 out of 142 comparisons) yielded ~0.5 false positive cases. Thus, in total, we expect 3.6 false positives (2.53%) when comparing gene deletion ancestors and corresponding compensated strains. We note that comparisons between the wild type versus ancestor or compensated strains imply comparing 88 versus four biological replicates, therefore the FDR for those comparisons is expected to be negligible. Overall, these analyses show that the rate and number of falsely detected significantly different strains is very low and is unlikely to affect our conclusions.

Classes of evolutionary trajectories depicted in Fig. 1b are defined as follows. 'No change': neither the compensated strain nor its ancestor is morphologically different from the wild type. 'Restoration': the ancestor is morphologically different from the wild type, but the compensated strain is not. For simplicity, we also included in this category the two compensated strains in which all the principal components are at least partially restored. Partial restoration for a given principal component is defined by both the compensated strain and its ancestor differing from the wild type in the same direction but the compensated strain being closer to the wild type while being significantly different from its ancestor. 'Ancestor-like compensated': both the compensated strain and its ancestor are morphologically different from the wild type, but they are not different from each other (with the possible exception of partial or full restoration of some but not all principal components). The remaining strains are classified as 'Novel compensated' and can be further divided into two subclasses: (1) the compensated strain is different from the wild type, but its ancestor is not and (2) both the compensated strain and its ancestor are different from the wild type and from each other.

For Fig. 1c we selected compensated strains with the most extreme trait values based on the following criteria. We considered only those compensated strains whose average ±2 standard deviation interval is non-overlapping with the wild-type interval for a given trait (calculated on log2 scale for cell size and cell elongation). From these sets, we display up to 20 strains with the largest absolute differences from the wild type for each trait (based on log2 fold change for cell size and cell elongation).

To quantify morphological distances between wild type and control evolved strains, we calculated pairwise Euclidean distances between their morphological profiles. To compare the degree of morphological diversity of compensated strains and natural isolates, we calculated Euclidean distances between the morphological profiles of each pair of strains within each group. Morphological profiles were defined by the principal components as above. More specifically, morphological trait values were standardized for the two sets of strains separately using their

corresponding replicate measurements of the wild-type strain. The two sets of standardized values were merged and PCA was carried out. The standardized scores of the first eight principal component (explaining ~96% of the total variation) were used to calculate all pairwise Euclidean distances within each set of strains.

To identify significantly different groups of strains with similar morphologies, we performed hierarchical clustering of their morphological profiles by using the R package 'pvclust'[27], with Euclidean distance and average linkage. Significant clusters were defined as having higher than 0.95 approximately unbiased bootstrap value after 10,000-time bootstrap sampling. To define a minimal set of significant clusters, we selected the largest significant clusters above the height of the node where the wild type and all the control evolved lines are merged ('wild type-like cluster'), not including those that contain the 'wild type-like cluster'. From this list, we considered only those clusters which contained compensated strains initiated from at least three different gene deletion strains.

Quantitative invasive growth assay. To screen the ancestor and compensated strains of deletion mutants (along with control evolved strains) for invasive growth phenotype, we used an established plate-washing assay[31] with minor modifications. Briefly, strains were inoculated into randomly selected positions of a standard 96-well plate, excluding the borders of the plates that were inoculated with a 'filler' strain that was not analysed afterwards. The positive control strain, sigma 1278b, was also inoculated into the plates. After reaching saturation in cell density, each 96-well plate was spotted onto a solid YPD medium with high-density (3%) agar. After seven days of incubation at 30 °C, all plates were photographed to obtain digital images of spot growth and then 10 ml water was pipetted onto the plates. After 10 min of incubation, the non-invasive cells were washed off with a gentle stream of water and gentle rubbing with a finger (wearing a latex glove). Dried plates were re-grown for 24 h at 30 °C to enhance the growth of the agar-invaded cells. Images of post-washed spots were also obtained by photographing the plates. Spots were illuminated from the bottom throughout the study unless otherwise indicated (that is, on the representative images of Fig. 4a in which case illumination from the top was used). The pixel intensity of each spot pre- and post-wash was estimated with a custom-developed image analysis pipeline in MATLAB programming environment (Supplementary Protocols).

The level of invasive growth, that is, absolute invasiveness, was determined as the ratio of the intensity after washing to the intensity before washing (based on four biological replicates of each). Relative invasiveness was calculated by normalizing the invasiveness of the investigated strains to that of the positive control strain (sigma 1278b). Strains that met the following criteria were considered to display an invasive growth phenotype: (1) the relative invasiveness value was significantly higher than that of the wild type (one-sided Wilcoxon rank-sum test, 10% FDR) and (2) the relative invasiveness value was higher than 10% of the corresponding value of the positive control strain (sigma 1278b, note that positive hits were confirmed by visual inspection of the plates after washing). To assess whether the trait value of the compensated strains are significantly higher than their corresponding ancestors, one-sided Student's *t*-tests were used (10% FDR) because these comparisons involved low sample sizes.

To investigate epistasis between *Δbub3* and $SWE1^{Y332S}$ at the level of invasiveness, we performed an agar invasion assay using the wild type, the two single mutants (*Δbub3* and $SWE1^{Y332S}$) and the reconstructed double mutant (*Δbub3* + $SWE1^{Y332S}$) using at least eight biological replicates. With the presumption of additive epistasis on invasiveness, we calculated the expected relative invasiveness value of the double mutant as follows: first, we subtracted the wild-type value from those of the two single mutants and then performed addition of the subtracted values.

Quantitative settling assay. To systematically investigate the ability of the ancestor and compensated strains to form multicellular clumps or flocs, we first performed a well-established quantitative sedimentation assay[38] with minor modification. Briefly, strains were inoculated into culture tubes filled with 5 ml YPD liquid medium and grown for 24 h at 30 °C with 200 r.p.m. shaking. Strains were vortexed vigorously in the culture tubes and then were incubated without agitation for 60 min to help sedimentation of the cells. After the incubation time, the culture tubes were photographed then the fraction of the settled cells (absolute settling score) was estimated using an image analysis pipeline described previously[38]. Relative settling score was calculated by normalizing the absolute settling score of the investigated strains (based on four biological replicates of each) to that of the wild-type strain. Strains showing at least a 10% increase in relative settling score and a significant difference (one-sided Wilcoxon rank-sum test, 10% FDR) compared with the wild type were considered to display an increased settling capacity. To assess whether the trait value of the compensated strains are significantly higher than their corresponding ancestors, one-sided Student's *t*-tests were used (10% FDR) because these comparisons involved low sample sizes.

Detection of multicellular aggregates. Formation of multicellular aggregates of strains with increased settling capacity was further confirmed by microscopy analysis (Zeiss Laser Capture Microdissection microscope) with 20× magnification. To distinguish between calcium-dependent aggregation (that is, flocculation) and mother–daughter separation defect (that is, clump formation) as a mechanism underlying multicellular structures, we performed a deflocculation assay as described in an earlier work[39] with minor modifications. Briefly, we treated cells with a chelating agent (EDTA) that sequesters calcium ions and thereby prevents cell adhesion mediated by calcium-dependent flocculins. Cells were incubated in 3 ml YPD medium in 24 deep-well plates for 24 h with 220 r.p.m. agitation. Cultures were settled for 60 min without agitation, then 200 µl of sedimented culture was taken from the bottom of the wells, resuspended, split into two parts and centrifuged. One part was resuspended in 100 µl water and the other part was resuspended in deflocculation solution (100 µl 4 mM EDTA). After 50 min of incubation at room temperature without agitation, cells were taken from the bottom of the centrifuge tube, mixed with mounting medium (9:1 mix of glycerol and 1 × PBS) and examined by microscope with 20× magnification (Zeiss Laser Capture Microdissection and Operetta High-Content Imaging System (PerkinElmer Inc.)). For the deflocculation assay, we used a positive control strain (natural strain (OS_1189) isolated from soil in the Netherlands, described in a previous study[28]) that flocculated well in minimal synthetic defined medium after two days and deflocculated completely after the EDTA treatment.

Quantitative biofilm-formation assay. To screen the ancestor and compensated strains for increased biofilm production, we used an established protocol with minor modifications[34]. Briefly, cells growing on solid YPD with medium density (2%) agar plates were picked and spotted by toothpick onto solid YPD with low-density (0.3%) agar plates and incubated for 5 days at 25 °C. The plates were subsequently photographed to obtain digital images of the biofilms and then the area of the biofilm was estimated by using an image analysis pipeline described previously[38]. Relative biofilm area was calculated by normalizing the biofilm area of the investigated strains (based on three biological replicates of each) to that of the wild type. Strains having at least a 10% increase and a significant difference (one-sided Wilcoxon rank-sum test, 10% FDR) in biofilm area compared with the wild type were considered to display an increased biofilm-formation capacity. To assess whether the trait value of the compensated strains are significantly higher than their corresponding ancestors, one-sided Student's *t*-tests were used (10% FDR) because these comparisons involved low sample sizes.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All data are available in the main text, Methods or the Supplementary Information. A multi-page pdf containing the investigation of ploidy level of yeast strains is available at https://figshare.com/s/cc55743a3c97d927db59. High-resolution image of Extended Data Fig. 4 can be found at https://figshare.com/s/a5f1571eb8cc5bada89b.

Code availability

Scripts used in the analysis of microscopic images are available at https://github.com/pappb/Farkas-et-al-Compensatory-evolution. The MATLAB code used in the image analysis of invasive growth is available at https://github.com/csmolnar/invasivegrowth.

Received: 6 October 2021; Accepted: 10 March 2022;
Published online: 28 April 2022

References

1. Covert, A. W., Lenski, R. E., Wilke, C. O. & Ofria, C. Experiments on the role of deleterious mutations as stepping stones in adaptive evolution. *Proc. Natl Acad. Sci. USA* https://doi.org/10.1073/pnas.1313424110 (2013).
2. Albalat, R. & Cañestro, C. Evolution by gene loss. *Nat. Rev. Genet.* **17**, 379–391 (2016).
3. Lang, G. I. et al. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature* **500**, 571–574 (2013).
4. Qian, W., Ma, D., Xiao, C., Wang, Z. & Zhang, J. The genomic landscape and evolutionary resolution of antagonistic pleiotropy in yeast. *Cell Rep.* **2**, 1399–1410 (2012).
5. Doniger, S. W. et al. A catalog of neutral and deleterious polymorphism in yeast. *PLoS Genet.* **4**, e1000183 (2008).
6. MacArthur, D. G. et al. A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335**, 823–828 (2012).
7. Sharma, V. et al. A genomics approach reveals insights into the importance of gene losses for mammalian adaptations. *Nat. Commun.* **9**, 1215 (2018).
8. Szamecz, B. et al. The genomic landscape of compensatory evolution. *PLoS Biol.* **12**, e1001935 (2014).
9. LaBar, T., Phoebe Hsieh, Y.-Y., Fumasoni, M. & Murray, A. W. Evolutionary repair experiments as a window to the molecular diversity of life. *Curr. Biol.* **30**, R565–R574 (2020).
10. Wagner, A. Robustness, evolvability, and neutrality. *FEBS Lett.* **579**, 1772–1778 (2005).
11. Lynch, M. The evolution of genetic networks by non-adaptive processes. *Nat. Rev. Genet.* **8**, 803–813 (2007).

12. Ivankov, D. N., Finkelstein, A. V. & Kondrashov, F. A. A structural perspective of compensatory evolution. *Curr. Opin. Struct. Biol.* **26**, 104–112 (2014).
13. Andersson, D. I. & Hughes, D. Antibiotic resistance and its cost: is it possible to reverse resistance? *Nat. Rev. Microbiol.* **8**, 260–271 (2010).
14. Wittkopp, P. J., Haerum, B. K. & Clark, A. G. Evolutionary changes in cis and trans gene regulation. *Nature* **430**, 85–88 (2004).
15. Connallon, T., Camus, M. F., Morrow, E. H. & Dowling, D. K. Coadaptation of mitochondrial and nuclear genes, and the cost of mother's curse. *Proc. R. Soc. B* **285**, 20172257 (2018).
16. Galardini, M. et al. The impact of the genetic background on gene deletion phenotypes in *Saccharomyces cerevisiae*. *Mol. Syst. Biol.* **15**, e8831 (2019).
17. Blank, D., Wolf, L., Ackermann, M. & Silander, O. K. The predictability of molecular evolution during functional innovation. *Proc. Natl Acad. Sci. USA* **111**, 3044–3049 (2014).
18. McCloskey, D. et al. Evolution of gene knockout strains of *E. coli* reveal regulatory architectures governed by metabolism. *Nat. Commun.* **9**, 3796 (2018).
19. Rojas Echenique, J. I., Kryazhimskiy, S., Nguyen Ba, A. N. & Desai, M. M. Modular epistasis and the compensatory evolution of gene deletion mutants. *PLoS Genet.* **15**, e1007958 (2019).
20. Ohya, Y. et al. High-dimensional and large-scale phenotyping of yeast mutants. *Proc. Natl Acad. Sci. USA* **102**, 19015–19020 (2005).
21. Bauer, C. R., Li, S. & Siegal, M. L. Essential gene disruptions reveal complex relationships between phenotypic robustness, pleiotropy, and fitness. *Mol. Syst. Biol.* **11**, 773 (2015).
22. Spor, A., Wang, S., Dillmann, C., Vienne, Dde & Sicard, D. "Ant" and "grasshopper" life-history strategies in *Saccharomyces cerevisiae*. *PLoS ONE* **3**, e1579 (2008).
23. Turner, J. J., Ewald, J. C. & Skotheim, J. M. Cell size control in yeast. *Curr. Biol.* **22**, R350–R359 (2012).
24. Yvert, G. et al. Single-cell phenomics reveals intra-species variation of phenotypic noise in yeast. *BMC Syst. Biol.* **7**, 54 (2013).
25. Okada, H., Ohnuki, S. & Ohya, Y. Quantification of cell, actin, and nuclear DNA morphology with high-throughput microscopy and CalMorph. *Cold Spring Harb. Protoc.* **4**, 408–412 (2015).
26. Suzuki, G. et al. Global study of holistic morphological effectors in the budding yeast *Saccharomyces cerevisiae*. *BMC Genom.* **19**, 149 (2018).
27. Suzuki, R. & Shimodaira, H. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540–1542 (2006).
28. Peter, J. et al. Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* **556**, 339–344 (2018).
29. Watanabe, M., Watanabe, D., Nogami, S., Morishita, S. & Ohya, Y. Comprehensive and quantitative analysis of yeast deletion mutants defective in apical and isotropic bud growth. *Curr. Genet.* **55**, 365–380 (2009).
30. Gimeno, C. J., Ljungdahl, P. O., Styles, C. A. & Fink, G. R. Unipolar cell divisions in the yeast *S. cerevisiae* lead to filamentous growth: regulation by starvation and RAS. *Cell* **68**, 1077–1090 (1992).
31. Roberts, R. L. & Fink, G. R. Elements of a single MAP kinase cascade in *Saccharomyces cerevisiae* mediate two developmental programs in the same cell type: mating and invasive growth. *Genes Dev.* **8**, 2974–2985 (1994).
32. Madhani, H. D. & Fink, G. R. The control of filamentous differentiation and virulence in fungi. *Trends Cell Biol.* **8**, 348–353 (1998).
33. Cullen, P. J. & Sprague, G. F. Glucose depletion causes haploid invasive growth in yeast. *Proc. Natl Acad. Sci. USA* **97**, 13619–13624 (2000).
34. Reynolds, T. B. & Fink, G. R. Bakers' yeast, a model for fungal biofilm formation. *Science* **291**, 878–881 (2001).
35. Soares, E. V. Flocculation in *Saccharomyces cerevisiae*: a review. *J. Appl. Microbiol.* **110**, 1–18 (2011).
36. Kuzdzal-Fick, J. J., Chen, L. & Balázsi, G. Disadvantages and benefits of evolved unicellularity versus multicellularity in budding yeast. *Ecol. Evol.* **9**, 8509–8523 (2019).
37. Desai, J. V., Mitchell, A. P. & Andes, D. R. Fungal biofilms, drug resistance, and recurrent infection. *Cold Spring Harb. Perspect. Med.* **4**, a019729 (2014).
38. Hope, E. A. & Dunham, M. J. Ploidy-regulated variation in biofilm-related phenotypes in natural isolates of *Saccharomyces cerevisiae*. *G3* **4**, 1773–1786 (2014).
39. Liu, H., Styles, C. A. & Fink, G. R. *Saccharomyces cerevisiae* S288c has a mutation in *Flo8*, a gene required for filamentous growth. *Genetics* **144**, 967–978 (1996).
40. Lo, W.-S. & Dranginis, A. M. The cell surface flocculin *Flo11* is required for pseudohyphae formation and invasion by *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **9**, 161–171 (1998).
41. Vermulst, M. et al. Transcription errors induce proteotoxic stress and shorten cellular lifespan. *Nat. Commun.* **6**, 8065 (2015).
42. Chen, X. et al. Whi2 is a conserved negative regulator of TORC1 in response to low amino acids. *PLoS Genet.* **14**, e1007592 (2018).
43. Hardwick, K. G. The spindle checkpoint. *Trends Genet.* **14**, 1–4 (1998).
44. Lew, D. J. The morphogenesis checkpoint: how yeast cells watch their figures. *Curr. Opin. Cell Biol.* **15**, 648–653 (2003).
45. Helsen, J. et al. Gene loss predictably drives evolutionary adaptation. *Mol. Biol. Evol.* **37**, 2989–3002 (2020).
46. Tan, Z. et al. Aneuploidy underlies a multicellular phenotypic switch. *Proc. Natl Acad. Sci. USA* **110**, 12367–12372 (2013).
47. Ho, W.-C., Ohya, Y. & Zhang, J. Testing the neutral hypothesis of phenotypic evolution. *Proc. Natl Acad. Sci. USA* **114**, 12219–12224 (2017).
48. Pavlicev, M. & Wagner, G. P. A model of developmental evolution: selection, pleiotropy and compensation. *Trends Ecol. Evol.* **27**, 316–322 (2012).
49. Steenwyk, J. L. et al. Extensive loss of cell-cycle and DNA repair genes in an ancient lineage of bipolar budding yeasts. *PLoS Biol.* **17**, e3000255 (2019).
50. Parts, L. Natural variants suppress mutations in hundreds of essential genes. *Mol. Syst. Biol.* **17**, e10138 (2021).
51. Orr, H. A. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution* **52**, 935–949 (1998).
52. Goldschmidt, R. *The Material Basis of Evolution* (Yale Univ. Press, 1940).
53. Warringer, J., Ericson, E., Fernandez, L., Nerman, O. & Blomberg, A. High-resolution yeast phenomics resolves different physiological features in the saline response. *Proc. Natl Acad. Sci. USA* **100**, 15724–15729 (2003).
54. Warringer, J. & Blomberg, A. Automated screening in environmental arrays allows analysis of quantitative phenotypic profiles in *Saccharomyces cerevisiae*. *Yeast* **20**, 53–67 (2003).
55. R Core Team *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2019).

Acknowledgements

The FRE-LacZ plasmid (YEpU-FTyZ) was a kind gift from J. Thorner. We thank Z. Bódi for informal discussions, K. Ambrus for her general technical assistance, E. Kotogány for her help in the flow-cytometry measurements and I. Kelemen-Valkony for her help in laser scanning confocal microscopy. Funding and grant sources are as follows: 'Lendület' program of the Hungarian Academy of Sciences LP2009-013/2012 (B.P.); 'Lendület' program of the Hungarian Academy of Sciences LP-2017-10/2020 (C.P.); LENDULET-BIOMAG grant 2018-342 (P.H.); Wellcome Trust WT 098016/Z/11/Z (B.P.); National Laboratory of Biotechnology grant NKFIH-871-3/2020 (C.P.); the European Research Council H2020-ERC-2014-CoG 648364- Resistance Evolution (C.P.); National Research, Development and Innovation Office Élvonal Program KKP 126506 (C.P.); National Research, Development and Innovation Office Élvonal Program KKP 129814 (B.P.); Economic Development and Innovation Operational Programme: European Regional Development Funds GINOP-2.3.2-15-2016-00006 (P.H.); Economic Development and Innovation Operational Programme: European Regional Development Funds GINOP-2.3.2-15-2016-00037 (P.H.); Economic Development and Innovation Operational Programme: European Regional Development Funds GINOP-2.3.2-15-2016-00014 (C.P., B.P.); Economic Development and Innovation Operational Programme: European Regional Development Funds GINOP-2.3.2-15-2016-00020 (C.P.); Economic Development and Innovation Operational Programme: European Regional Development Funds GINOP-2.3.2-15-2016-00026 (B.P., P.H.); the European Union's Horizon 2020 research and innovation program grant number 739593 (B.P., F.A.); COMPASS-ERA PerMed H2020 (P.H.); CZI Deep Visual Proteomics (P.H.); H2020-DiscovAir (P.H.); ELKH-Excellence grant (P.H.); Hungarian Academy of Sciences Postdoctoral Fellowship Program Postdoc2014-85 (K.K.); National Research, Development and Innovation Office FK 128775 (Z.F.); National Research, Development and Innovation Office FK 128916 (D.K.); Janos Bolyai Research Fellowship from the Hungarian Academy of Sciences BO/779/20 (Z.F.); New National Excellence Program of the Ministry of Human Capacities Bolyai+, UNKP-20-5-SZTE-646 (Z.F.); and New National Excellence Program of the Ministry of Human Capacities Bolyai+, UNKP-21-5-SZTE-562 (Z.F.).

Author contributions

Conceptualization: B.P. and C.P. Methodology: Z.F., K.K., Z.S., D.K., G.F., F.B., F.A., C.M. and P.H. Investigation: Z.F., K.K., Z.S., G.F. and C.M. Visualization: Z.F., K.K., Z.S. and G.F. Funding acquisition: B.P., C.P. and P.H. Supervision: B.P. and C.P. Writing—original draft: B.P., C.P., Z.F., K.K. and Z.S. Writing—review and editing: B.P., C.P., Z.F., K.K. and Z.S.

Competing interests

Authors declare no competing interests.

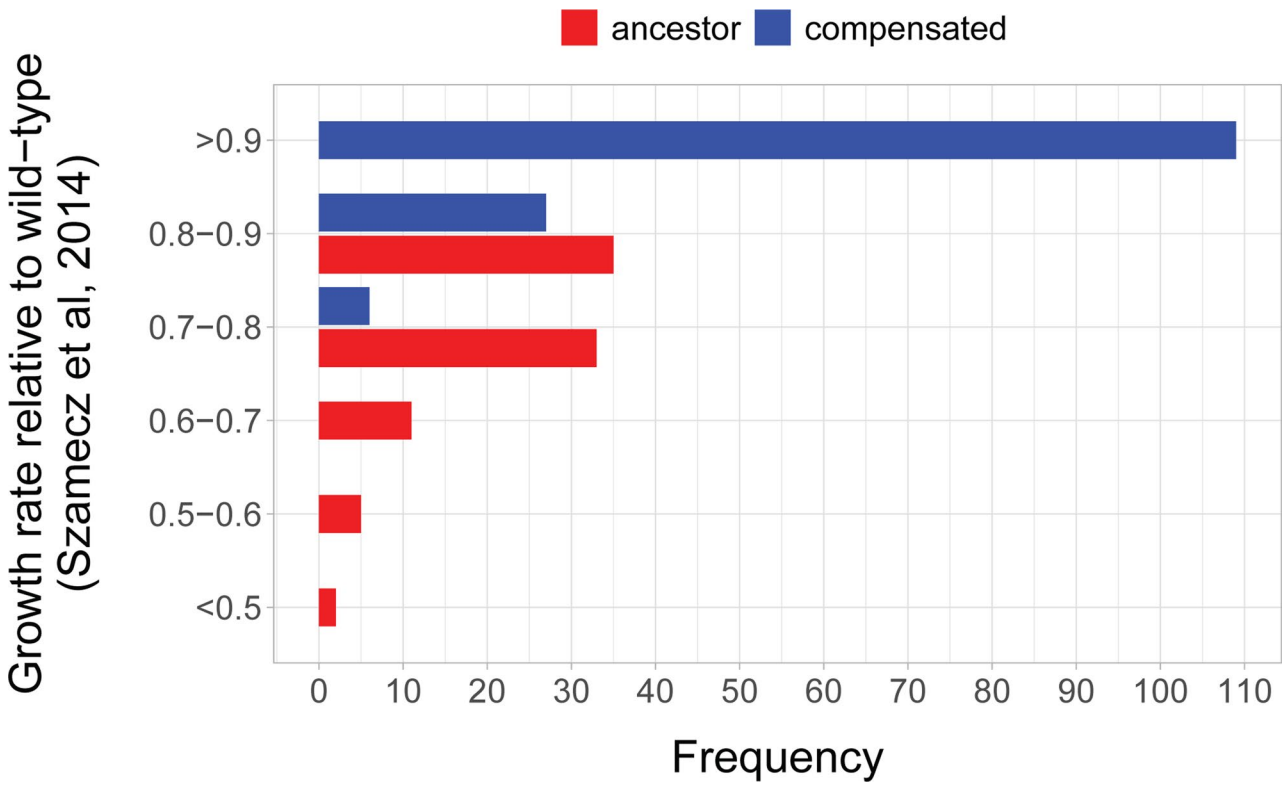
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41559-022-01730-1.

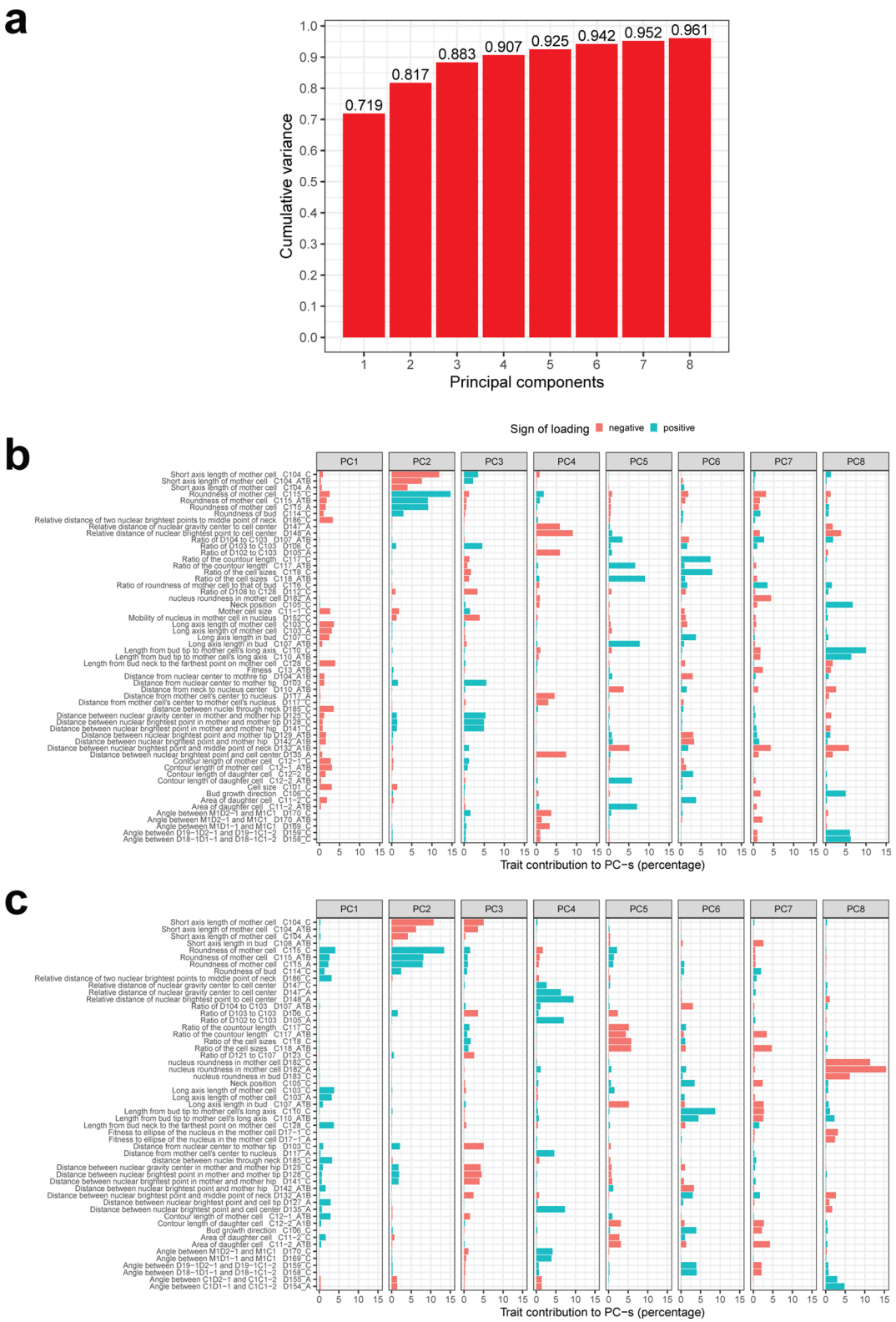
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41559-022-01730-1.

Correspondence and requests for materials should be addressed to Csaba Pál or Balázs Papp.

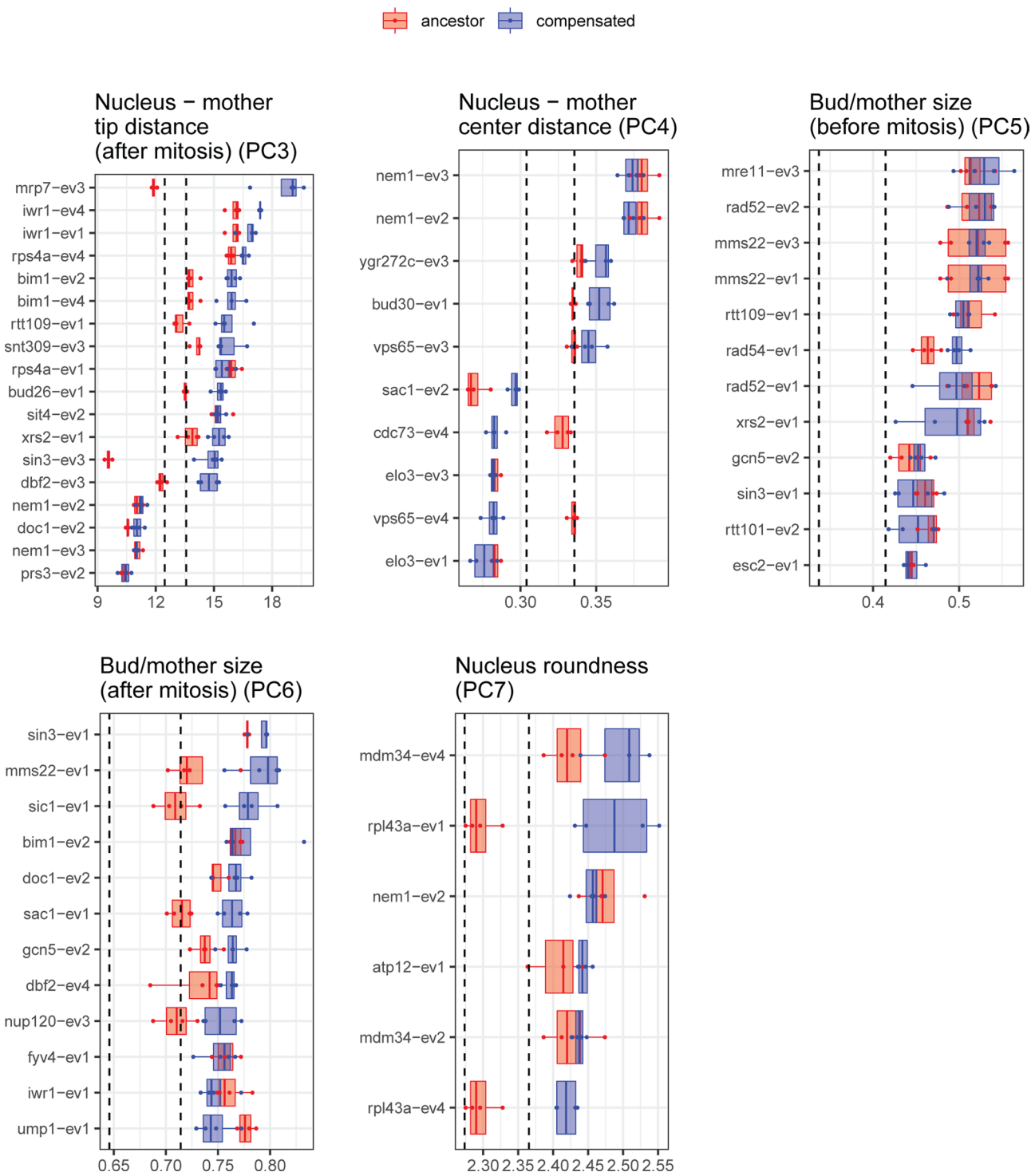
Peer review information *Nature Ecology & Evolution* thanks Yoshikazu Ohya, Alys Cheatle Jarvela and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



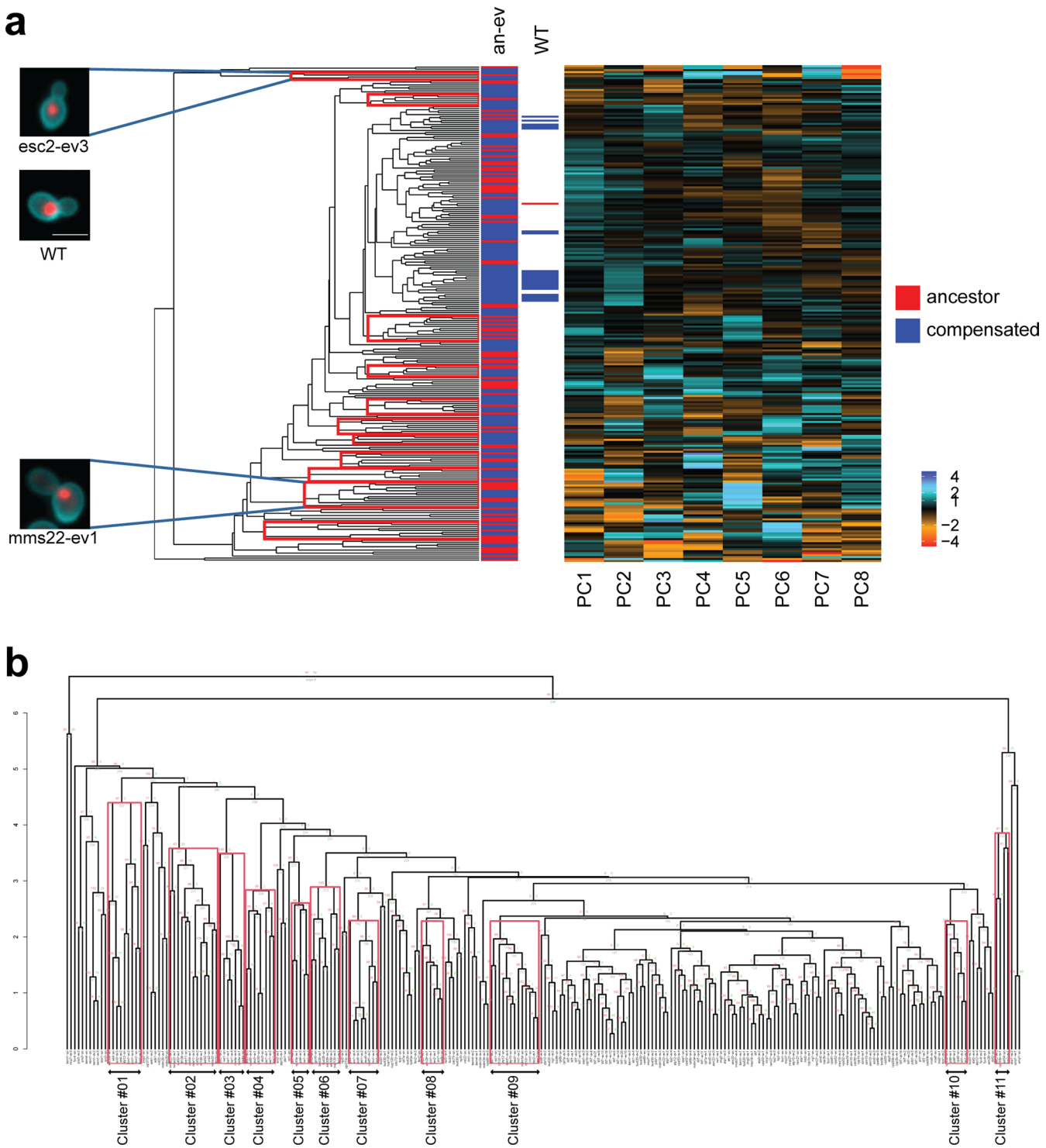
Extended Data Fig. 1 | Fitness distribution of the investigated strains. The barplot shows the distribution of relative fitness of initial knock-out mutant strains (i.e. ancestors, red) and the compensated strains (blue). Data from our previous study[8] is re-plotted here. Relative fitness was estimated by growth rate in liquid medium relative to the wild-type.



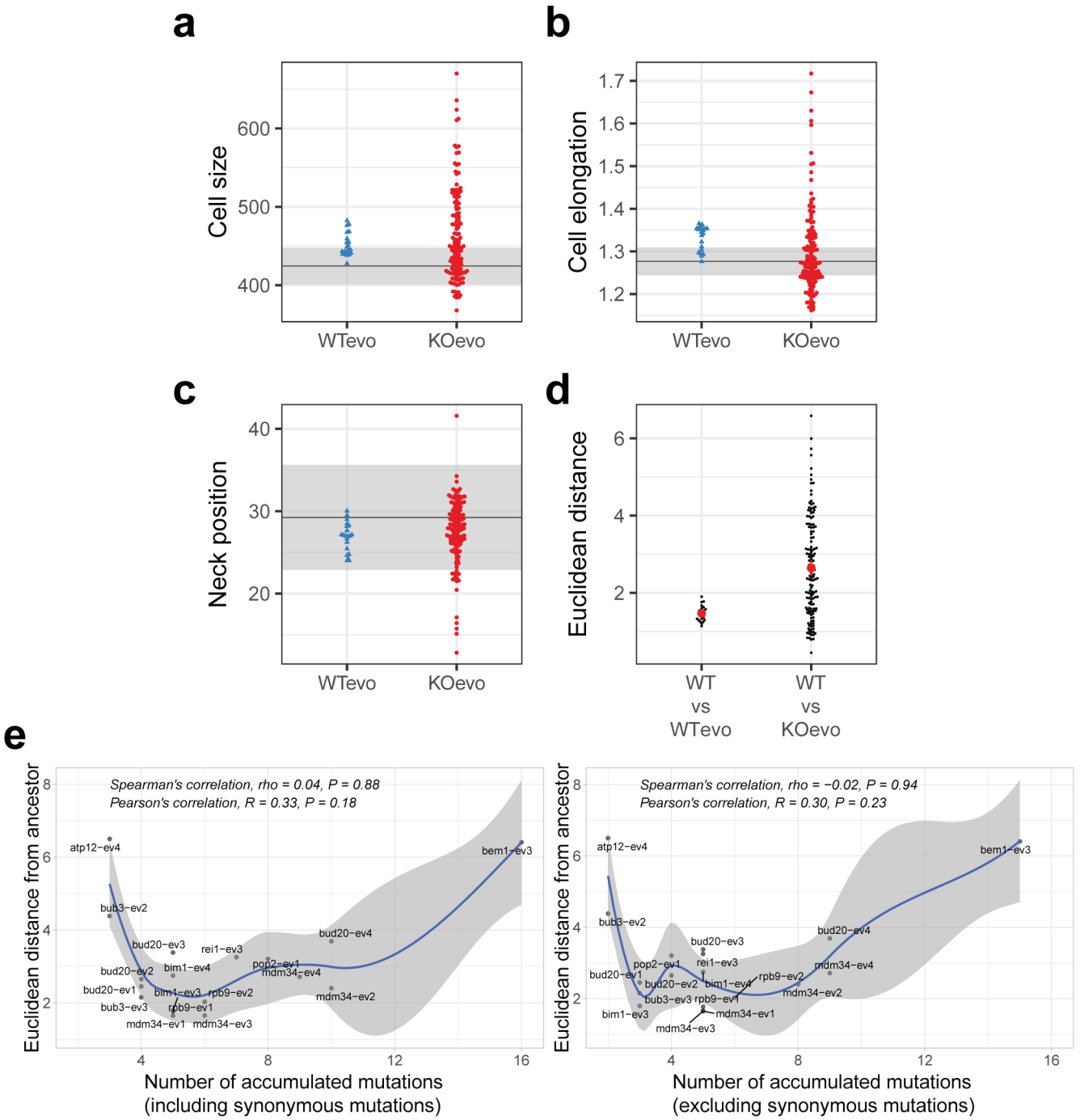
Extended Data Fig. 2 | Results of principal component analysis on single-cell morphology. (a) Cumulative variance of all single-cell morphological traits explained by the first 8 principal components in a principal component analysis (PCA). Note that PCA was performed on all genotypes, including wild-type and gene deletion ancestors. **(bc) Contribution of specific morphological traits to the first 8 principal components**. Panels (**b**) and (**c**) show the results of separate PCAs carried out for all strains and the subset of novel compensated strains (including the WT), respectively. Colors of the bars indicate the sign of the effects of specific traits on the given principal component (loading). Only traits providing the 8 largest contributions to the principal components are shown. Note that the traits contributing to PC1 to PC8 in panels (**b**) and (**c**) show substantial overlap with each other.



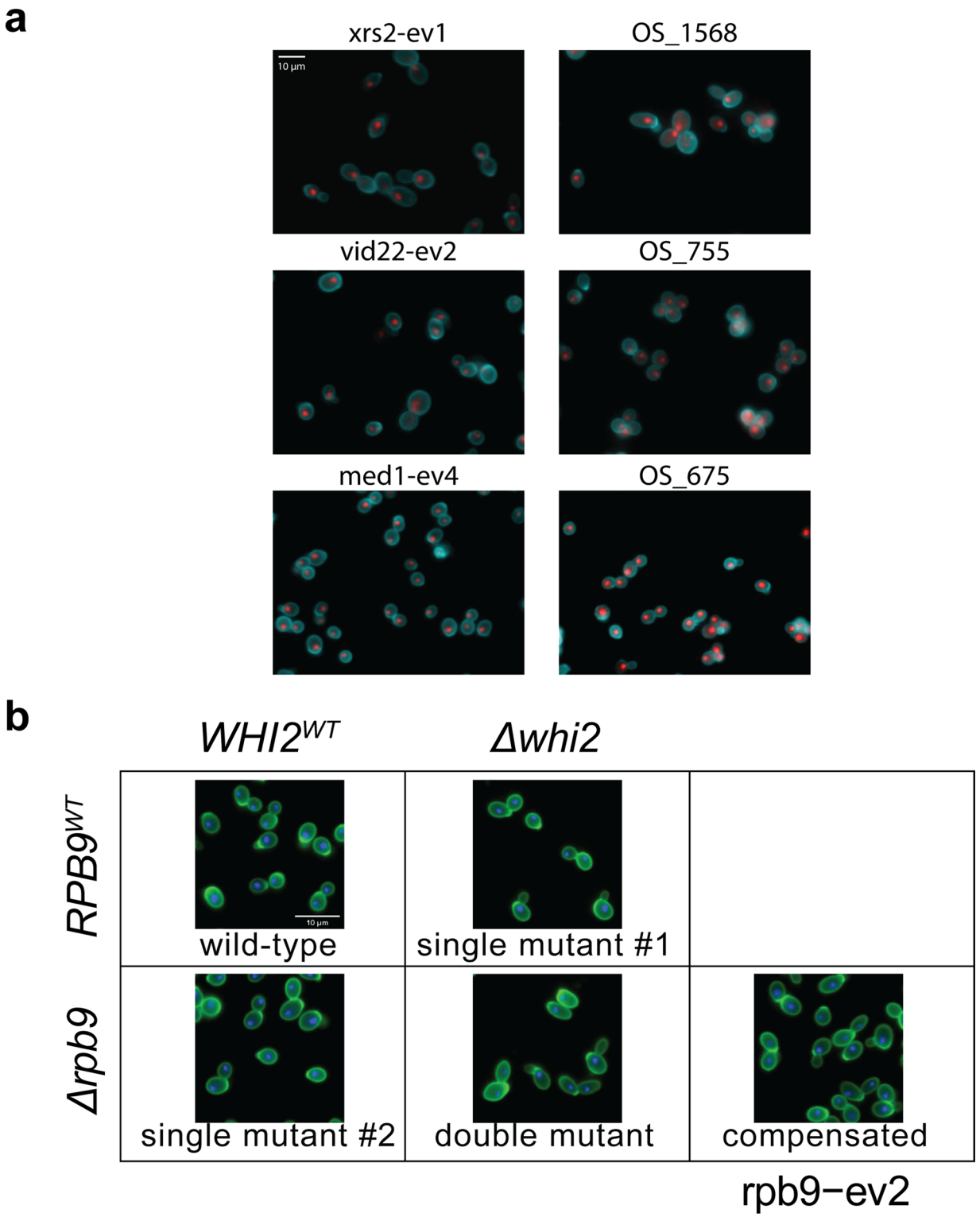
Extended Data Fig. 3 | Evolution of five representative morphological traits in compensated strains with the most extreme trait values. The boxplots display the trait values in the compensated strains and corresponding ancestors compared to that of the wild-type (based on n = 3 or n = 4 biological replicates each). Traits are representative traits of PCs 3-7, shown in the order of PCs (Fig. 1a). Note that for each trait, a subset of compensated strains displaying the most extreme trait values are displayed. The corresponding CalMorph traits are D103_C, D148_A, C118_A1B, C118_C, and D182_A respectively. Dashed lines indicate the range of the wild-type trait values (average ± 2 standard deviations). Boxplots show the median, first and third quartiles, with whiskers showing the 5th and 95th percentiles.



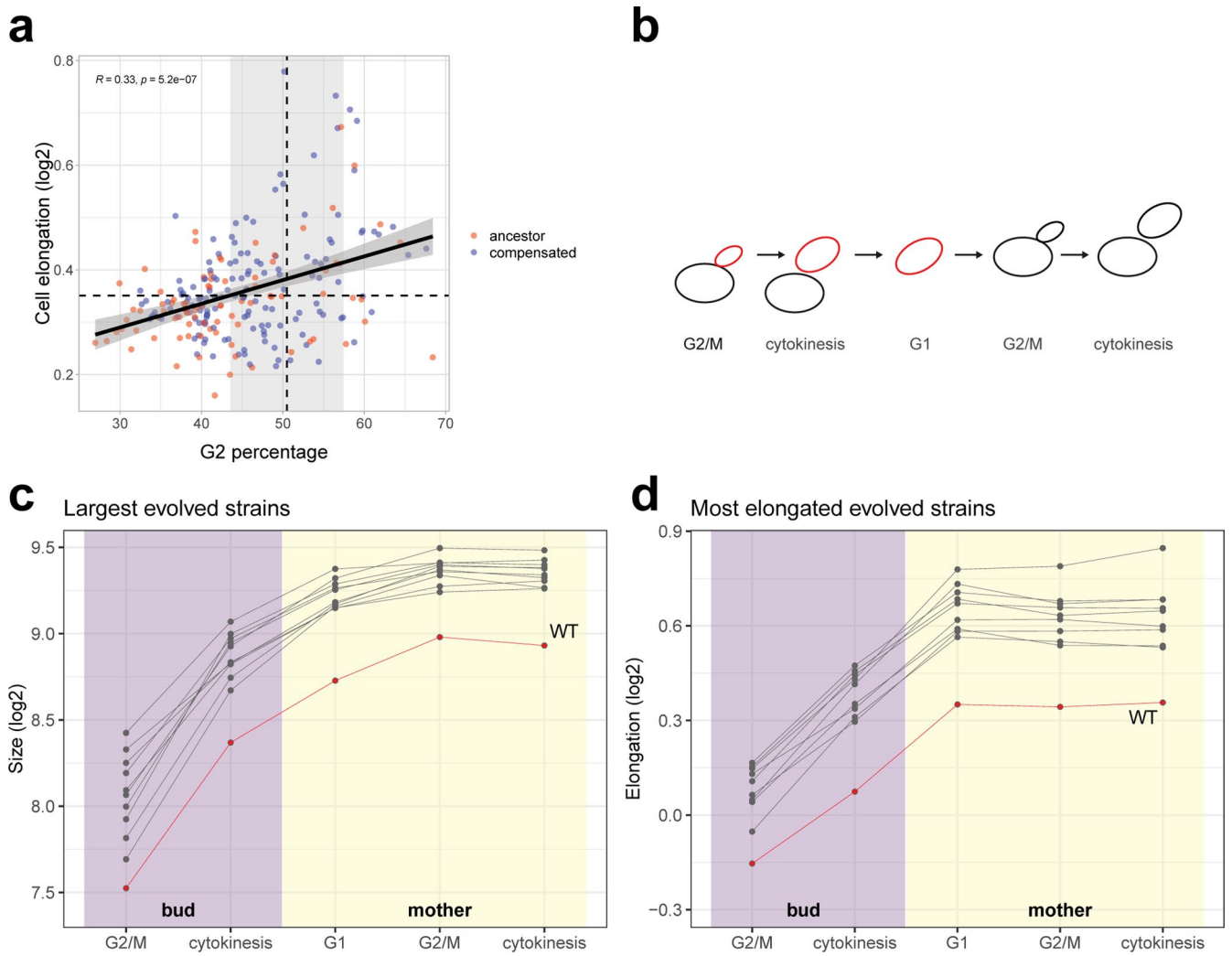
Extended Data Fig. 4 | Clustering of the morphological profiles. (a) Heatmap of morphological profiles. Each row represents the morphological profile of a genotype. Ancestor (red) and compensated (blue) strains are marked along the column next to the dendrogram (labeled as column an-ev). The column left to the heatmap (labeled as column WT) indicates wild-type (red) and control evolved strains (blue). Columns of the heatmap are the first eight principal components with colors representing the principal component scores. The dendrogram is the result of hierarchical clustering with red boxes representing 11 significant clusters (see Methods). Representative images of the wild-type (WT) and strains from the two clusters: i) containing cells with small bud angle (esc2-ev3, Cluster #11) and ii) cells with enlarged bud size relative to mother cell size (mms22-ev1, Cluster #02), are shown. We note that the strains harboring deletion of DNA damage responding genes are 16.5-fold enriched in the latter cluster (GO:0006974, Fisher's exact test, $P < 2 \times 10^{-16}$, Supplementary Data 3). Cell wall and nuclei are colored with green and red, respectively. Scale bar (on image of WT) represents 5 μm distance. **(b) Dendrogram showing hierarchical clustering of genotypes based on single-cell morphology profiles.** The same dendrogram as in panel (a), but also showing the names of the strains and the approximately unbiased probability values (AU p-value) for each cluster. AU p-values were used to define statistically significant clusters (Cluster #01-11) indicated by red boxes (for further details, see Methods). For further information on the clusters, see Supplementary Data 3. High-resolution image of Extended Data Fig. 4 can be found at https://figshare.com/s/a5f1571eb8cc5bada89b.



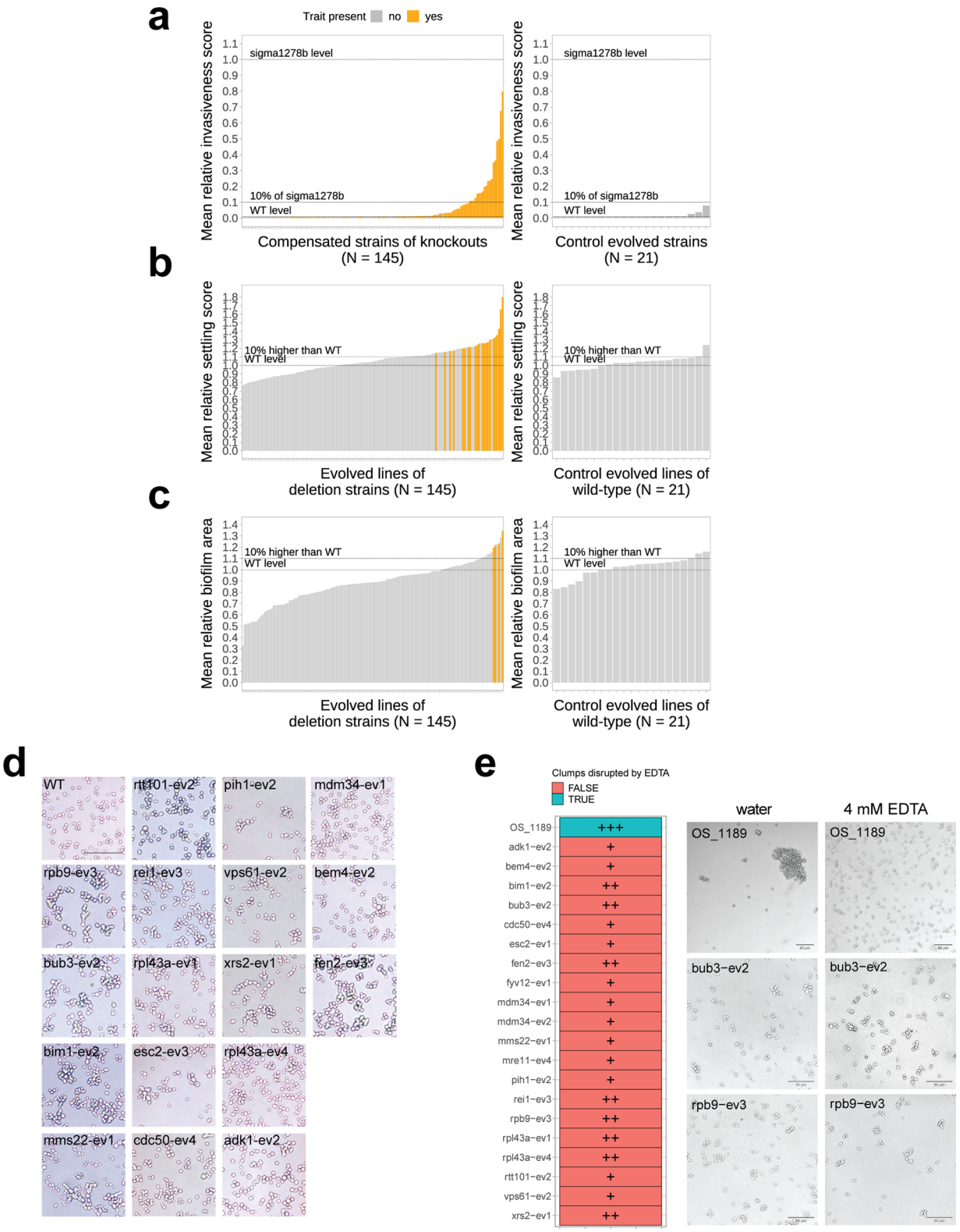
Extended Data Fig. 5 | Morphological changes are specific to compensatory evolution. (a-b-c-d) Evolved control strains show limited change in cellular morphology. Distribution of cell size (**a**), cell elongation (**b**) and neck position (**c**) for the evolved controls (initiated from the wild-type background, WTevo) and compensated strains (KOevo). Each dot represents average trait value for an individual strain. Changes of the above parameters in the evolved controls are negligible in comparison to a large number of compensated strains. Horizontal line and grey area denote the average value and average ± 2 standard deviation of the wild-type replicates, respectively. Morphological traits correspond to the same CalMorph parameters as in Fig. 1c. **(d)** Distribution of Euclidean distance (from the wild-type) of the control evolved (WTevo) and compensated strains (KOevo). Degree of morphological changes between the wild type (WT) and evolved controls is smaller than between the WT and most of the compensated strains (Brunner-Munzel test, $P = 3 \times 10^{-14}$). Degree of morphological change is measured by Euclidean distance between morphological profiles (see Methods). Red dots and error bars show the average and 95% confidence interval for the two strain sets. **(e) Morphological divergence during compensatory evolution is independent of the number of accumulated mutations**. The figure shows the Euclidean distance of the 18 compensated strains from their corresponding ancestors as a function of the number of mutations accumulated during the course of compensatory evolution[8]. The left and right panel shows the number of mutations including and excluding the synonymous ones, respectively. We found a lack of significant correlation between the number of accumulated mutations and the overall morphological distance, indicating that large morphological changes are often accessible in a few mutational steps.



Extended Data Fig. 6 | Larger field of views for microscopy images. (a) Similar cellular morphology of compensated strains and natural isolates. The figure shows wider field of views for Fig. 2c. Images show pairs of compensated and natural strains that display similar morphological trait combinations (cell wall and nuclei are colored with green and red, respectively): (i) large cells with normal, wild-type-like elongation: xrs2-ev1, OS_1586 isolate from tree leaves, (ii) large round cells: vid22-ev2, OS_755 wine yeast isolate, (iii) small round cells: med1-ev4, OS_675 isolate from human blood. Scale bar represents 10 µm. **(b) Synergistic epistasis underlying morphological changes in a compensated strain of Δrpb9**. The figure shows wider field of view images for Fig. 5a. Images show 5 selected genotypes, including the wild-type (WT), two single mutants (Δrpb9 and Δwhi2) and a reconstructed double mutant (*Δrpb9*+*Δwhi2*). The fifth genotype is the compensated strain of *Δrpb9* (rpb9-ev2) that harbors the *whi2*^{S133*} loss-of-function allele. Cell wall and nuclei are colored with green and blue, respectively.

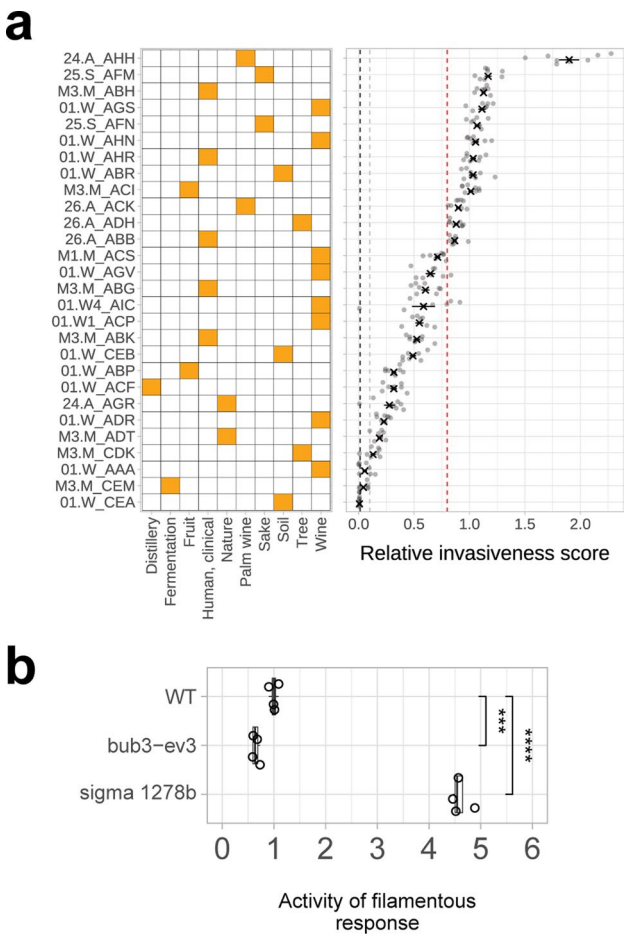


Extended Data Fig. 7 | Cell morphology progression through the cell cycle. (a) Pearson's correlation between cell elongation and G2 percentage (as measured by flow-cytometry). Cell elongation corresponds to CalMorph trait C115-A. WT denotes the wild-type strain. Ancestors and compensated strains are colored by red and blue, respectively. Dashed line represents the average of the WT. Grey area represents the WT average ± 2 standard deviations. We estimated standard deviation using the pool of strainwise centered replicate measurements of all investigated strains. **(b)** Scheme of bud growth stages through the cell cycle. **(c)** Plot shows cell size in different cell cycle stages of the 10 largest compensated strains. Importantly, genotypes with large mother cells also have larger buds than that of the wild-type (red line). Note that the extent of cell size increase throughout the cell cycle stages varies somewhat across the compensated strains. **(d)** Compensated strains displaying the most elongated mother cells reach their elongated shape during the G2/M and cytokinesis phase of the bud growth. Note that several strains show more intense bud elongation than the wild type. Size of the mother cell and bud corresponds to CalMorph traits C11-1 and C11-2, respectively. Elongation of the mother cell and bud corresponds to CalMorph traits C114 and C115, respectively. Cell cycle stages G1, G2/M and cytokinesis indicated on the plots correspond to stages A, A1B and C of the CalMorph software, respectively.

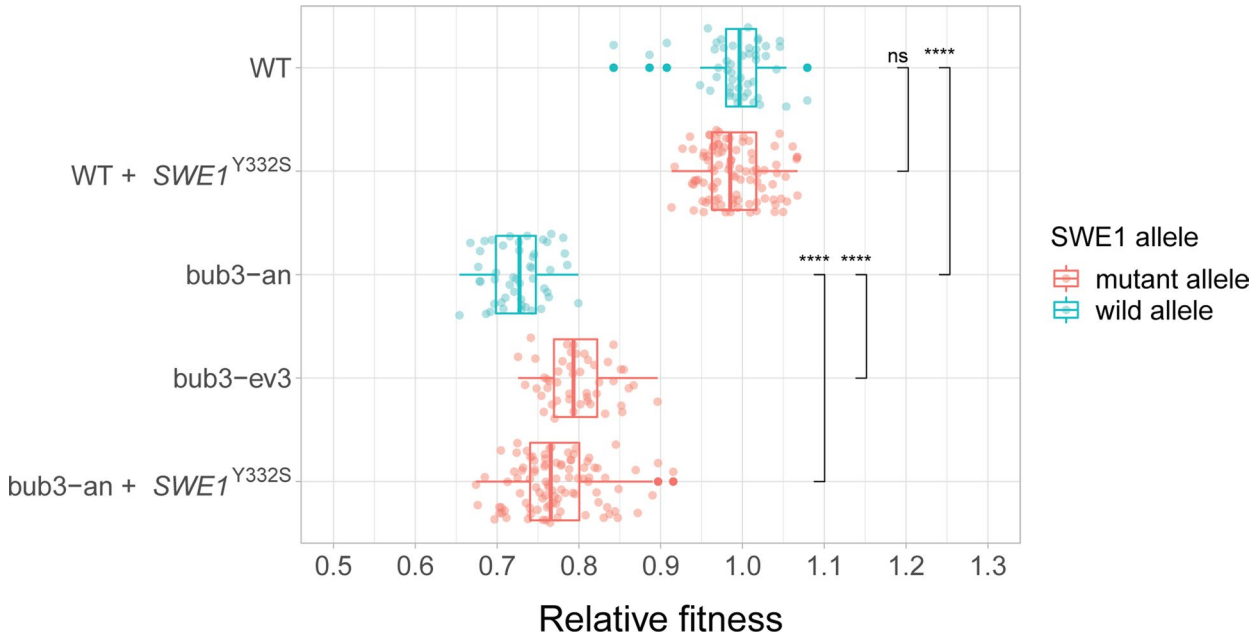


Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Multicellular morphologies of compensated strains. (a-b-c) Systematic screening of multicellular morphology. Barplots show the relative invasiveness (**a**), the relative settling score (**b**) and the relative biofilm area (**c**) of the compensated strains (initiated from knockout backgrounds, left panel) and control evolved strains (initiated from WT, right panel), respectively. Relative invasiveness score was calculated by normalizing the invasiveness score of the strains to that of the positive control strain (sigma1278b). Relative settling score (a proxy of cell aggregation) was calculated by normalizing the settling of the strains to that of the wild type strain. Relative biofilm area was calculated by normalizing the biofilm area of the strains to that of the WT. Orange color marks those compensated strains that display the corresponding trait (see Methods). **(d) Imaging multicellular aggregation**. The label-free microscopy images shows wider field of views for Fig. 3e, involving clump-forming compensated strains and the non-clumping WT. **(e) Flocculation assay**. Heatmap on the left summarizes the response of multicellular clumps to a deflocculation agent (4 mM EDTA) that can disrupt clumps formed via Ca^{2+}-dependent flocculation (see Methods). Deflocculation resulted in clear separation of the multicellular flocs into single / few cells (green) in a well-flocculating positive control strain (OS_1189 soil isolate, described in a previous study[28]). In contrast, there was no obvious change in the phenotypes of the compensated strains forming multicellular aggregates (red). Compensated strains were grouped into 3 different classes: +++/++/+ show the largest/medium-sized/smallest multicellular clumps, respectively. Microscopic images on the right show the deflocculation assay of two representative compensated strains that displayed significant settling (bub3-ev2 and rpb9-ev3), along with a flocculation positive strain (OS_1189). For microscopy analysis of the flocculation positive control strain and the compensated strains, a 10x and a 20x objective was used, respectively. Scale bar represents 50 μm distance.



Extended Data Fig. 9 | Analyzing invasive growth phenotype of bub3-ev3 and natural yeast isolates. (a) Invasive growth assay of 29 haploid natural yeast isolates. Natural baker's yeast isolates were selected from a previous study[28] and represent several phylogenetic clades (N = 8) and ecological origins (N = 10), indicated on the left panel. Relative invasiveness score (right panel) was calculated by normalizing the invasiveness of the strains to the mean of the positive control strain (sigma 1278b). The black cross and the point-range represent the mean and the standard error of the invasiveness score of at least four biological replicates (separate grey points). The red dashed line mark the mean invasiveness score of the compensated strain (bub3-ev3) that displays the strongest invasive growth phenotype. For further details, see Methods. For strain abbreviations, see Supplementary Data 5. **(b) Measuring the activity of the filamentous growth pathway**. Boxplot shows the activity of the FRE-lacZ reporter across several genotypes including WT, bub3-ev3 line, and a positive control strain (sigma 1278b). Activity of the FRE (Tec1p-dependent filamentous response element) gives information about activity of the filamentous growth pathway. The level of the filamentous response was estimated by measuring the β-galactosidase activity on protein extracts of yeast colonies after 3 days of incubation. To assess β-galactosidase activity, an established ONPG assay was used. Relative FRE-lacZ activity was calculated by normalizing the Miller Units of the investigated genotypes to that of the WT. Boxplots show the median, first and third quartiles, with whiskers showing the 5th and 95th percentiles of at least four biological replicates for each of the genotypes. Significant differences were assessed by two-sided Student's t-tests (***/**** indicates $P < 0.001/0.0001$). The P values are 5.3×10^{-4} and 4.2×10^{-6} for comparing WT with bub3-ev3 and sigma 1278b, respectively.



Extended Data Fig. 10 | Mutation in *SWE1* partially compensates the fitness defect of the *Δbub3* ancestor strain. Boxplot shows the relative fitness across several genotypes, including wild-type (WT), the ancestor (bub3-an) and a compensated strain of *Δbub3* (bub3-ev3), and strains harboring the reconstructed *SWE1*^{Y332S} mutant allele. As a proxy for fitness, colony size after 72 h of incubation on solid medium was measured as previously[8]. Briefly, ordered arrays of strains at 768-density were spotted onto YPD solid medium with medium-density (2%) agar. After 48 h of acclimatization to the medium at 30 °C, plates were replicated again onto the same medium. Digital images of the plates were taken with a camera after 72 h of incubation at 30 °C. The images were then processed to calculate colony sizes, after correcting for potential systematic biases[8]. Genotype fitness was estimated by the mean colony size of six biological replicates (i.e. six independent colonies). Relative fitness was calculated by normalizing the absolute colony sizes (see Methods) to that of the wild type strain. Significant differences were assessed by two-sided Wilcoxon rank-sum tests (**** indicates $P < 0.0001$, ns = non-significant). The P values are 0.15 and 3.11×10^{-28} for comparing WT with WT + SWE1^{Y332S} and bub3-an, respectively, while the P values are 6.92×10^{-13} and 4.16×10^{-8} for comparing bub3-an with bub3-ev3 and bub3-an + SWE1^{Y332S}, respectively. Boxplots show the median, first and third quartiles, with whiskers showing the 5th and 95th percentiles.