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
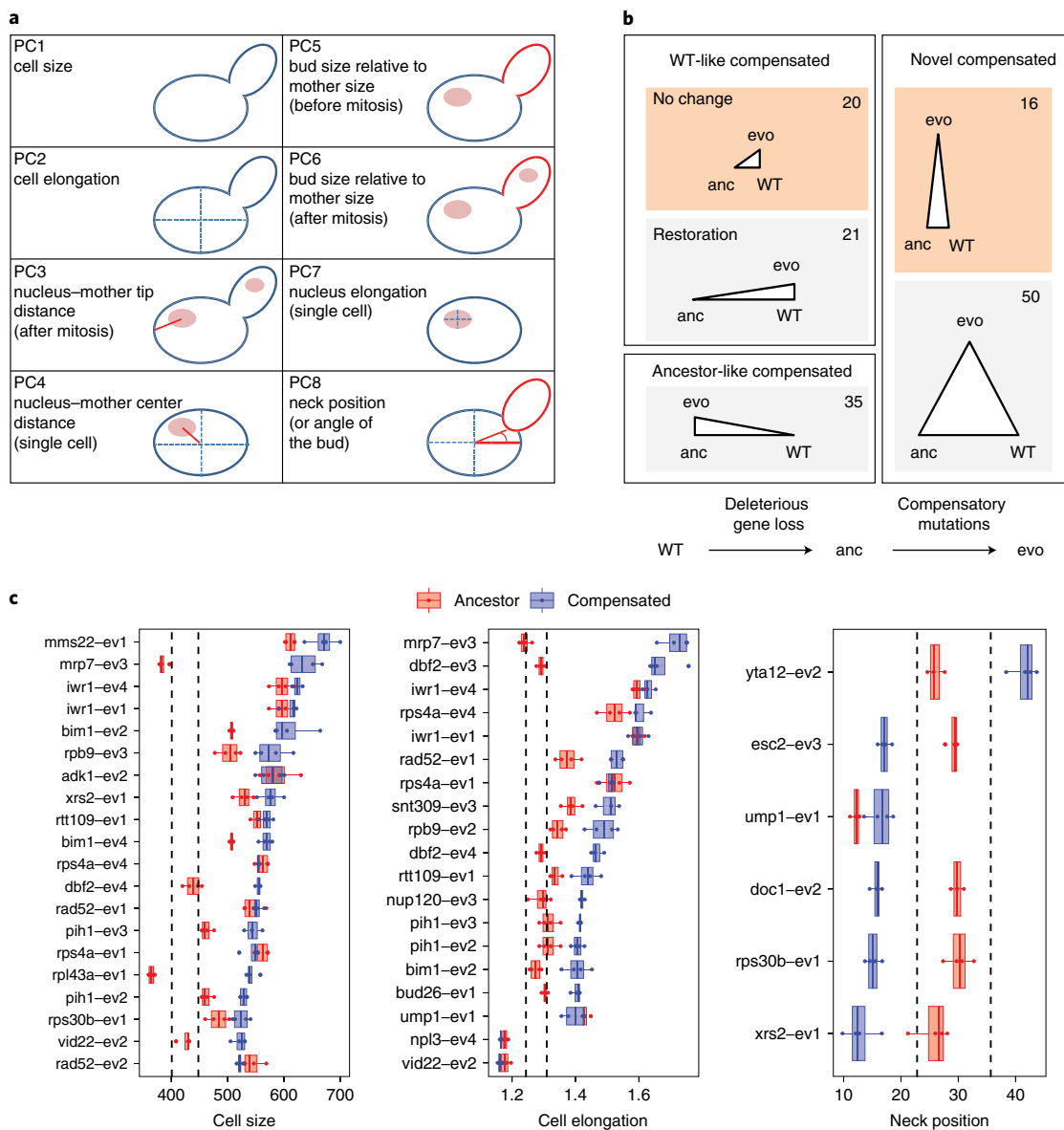
Fig. 1 | Compensatory evolution alters cellular morphology. a, Schematic representation of the principal components (PC1–PC8). The figure shows the intuitive meaning of the first eight principal components based on the top contributing morphological traits (Extended Data Fig. 2b). Separate morphological 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
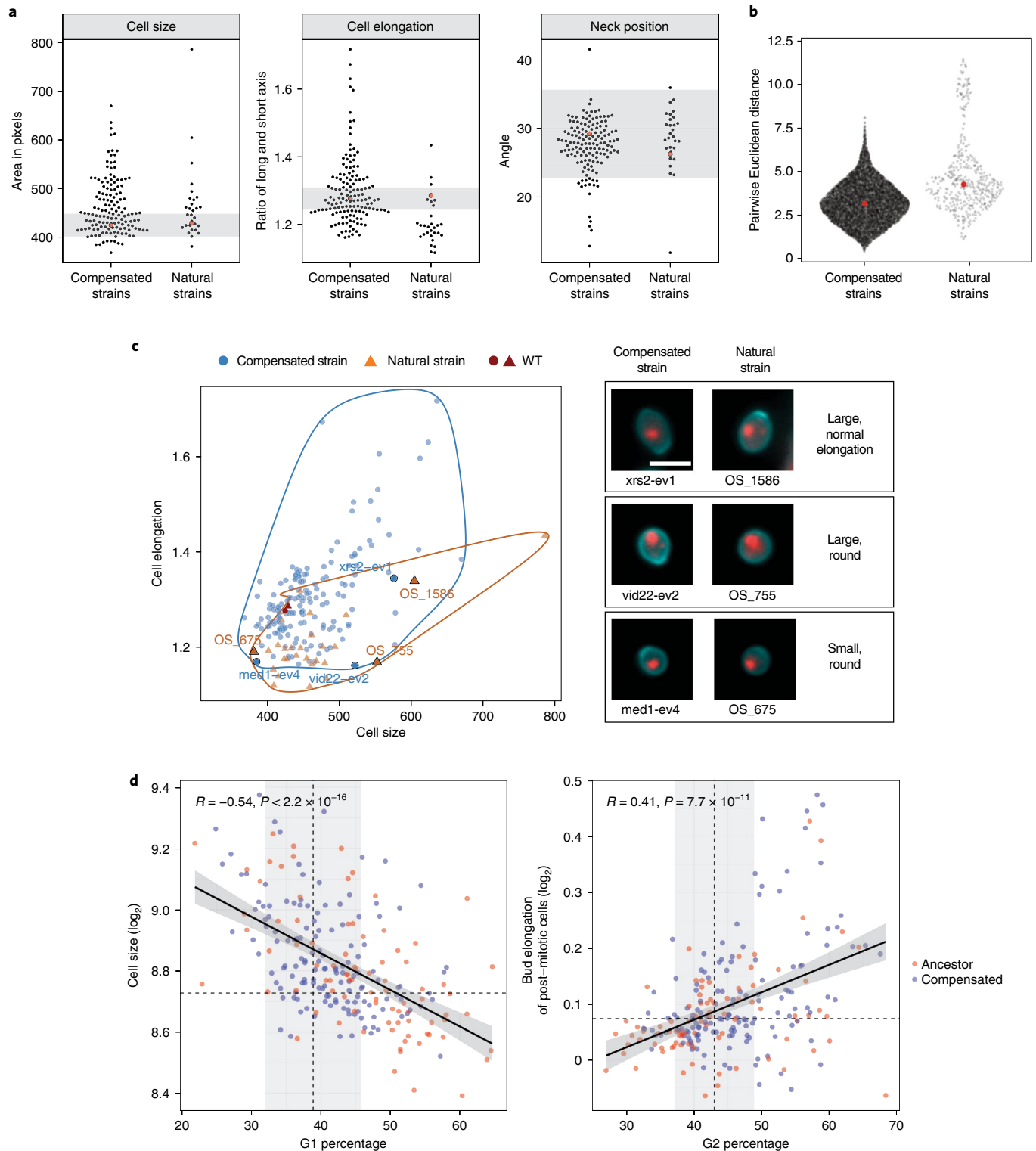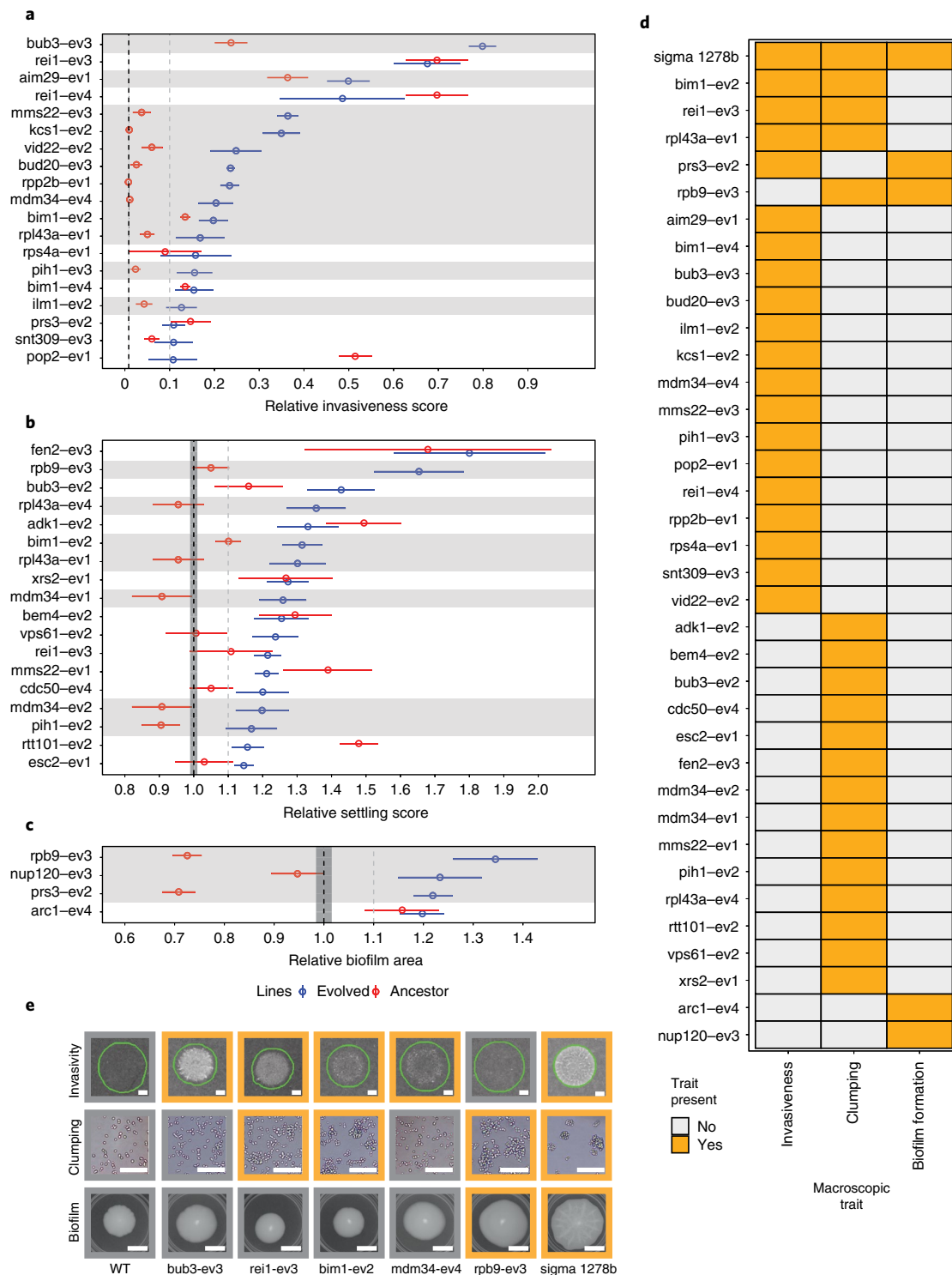
Fig. 3 | Compensatory evolution promotes the emergence of multicellular phenotypes. a,b,c, Plots show trait values for three types of multicellular form: invasiveness (**a**), multicellular aggregates as measured by settling (**b**) and biofilm area (**c**). Along with the corresponding ancestor strains (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.
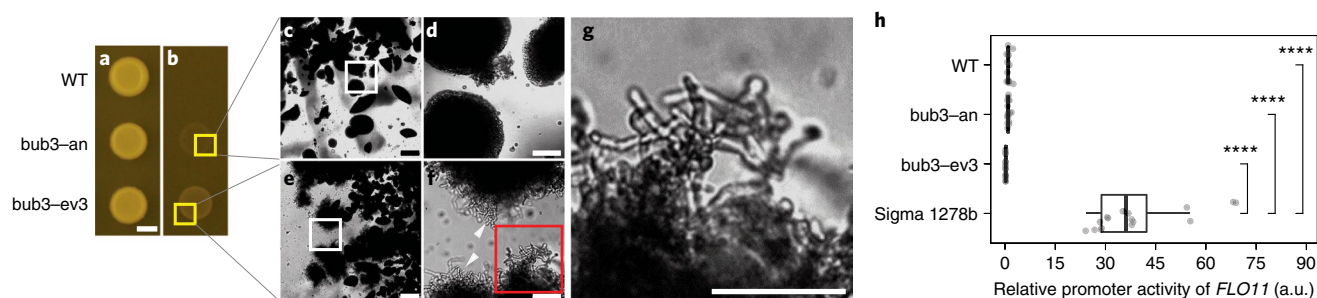
Fig. 4 | High invasiveness of a compensated strain. a–f, Filament formation appeared in a compensated strain of Δ*bub3*. **a,b**, The figure shows yeast spots (illuminated from the top) of wild type, the ancestor (bub3-an) and a compensated strain (bub3-ev3) of Δ*bub3* on solid medium after seven days 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 Δ*bub3* 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 Δ*rpb9* (ref. [8]) via relief of growth inhibition. Intriguingly, deletion of *WHI2* not only improves fitness but also yields elongated cells in Δ*rpb9*, 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 (Δ*rpb9*) and the compensatory mutation (Δ*whi2*) 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 Δ*bub3* 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 Δ*bub3* genetic backgrounds. The mutation significantly improved fitness in Δ*bub3* 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 Δ*bub3 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
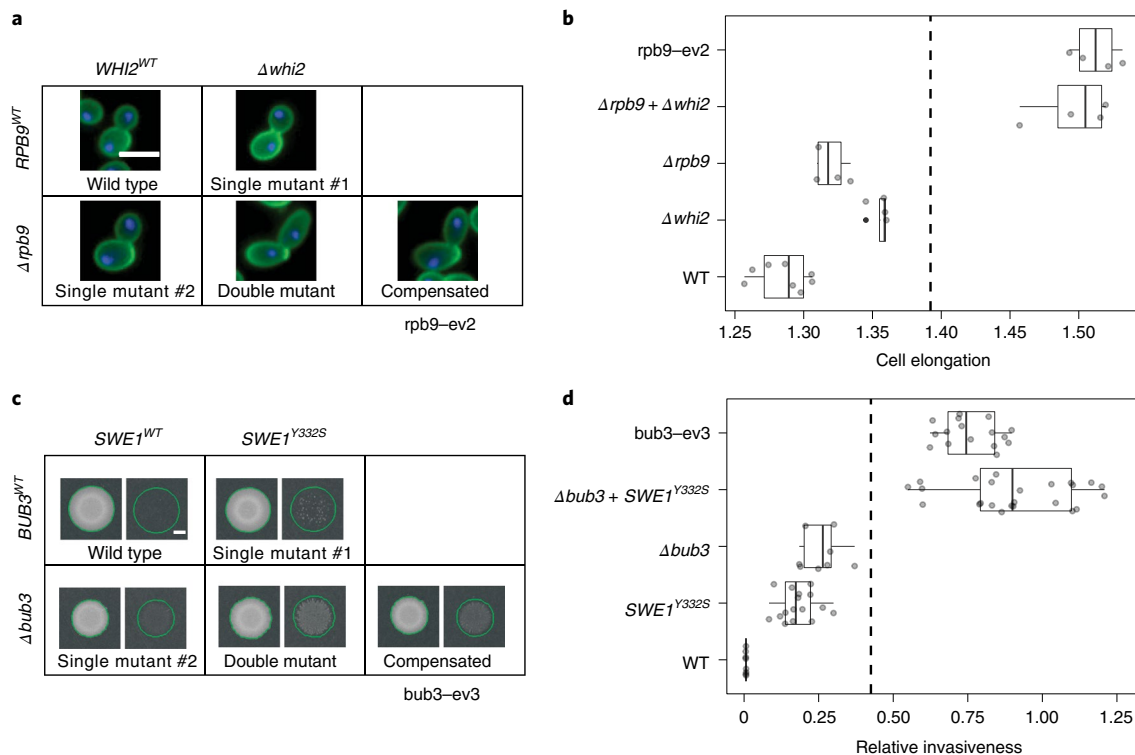
Fig. 5 | Synergistic epistasis underlying morphological changes in two compensated strains. a, Representative images of yeast cells of five selected genotypes, including the wild type, two single mutants (Δrpb9 and Δwhi2) and a reconstructed double mutant (Δrpb9 + Δwhi2). The fifth genotype is the compensated 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$_2$HPO$_4$, 1.8 mM KH$_2$PO$_4$, 1 mM CaCl$_2$, 0.5 mM MgCl$_2$) and P buffer (10 mM Na$_2$HPO$_4$, 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 $\Delta bub3$ and $SWE1^{Y332S}$ at the level of invasiveness, we performed an agar invasion assay using the wild type, the two single mutants ($\Delta bub3$ and $SWE1^{Y332S}$) and the reconstructed double mutant ($\Delta bub3 + SWE1^{Y332S}$) 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.

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

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.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
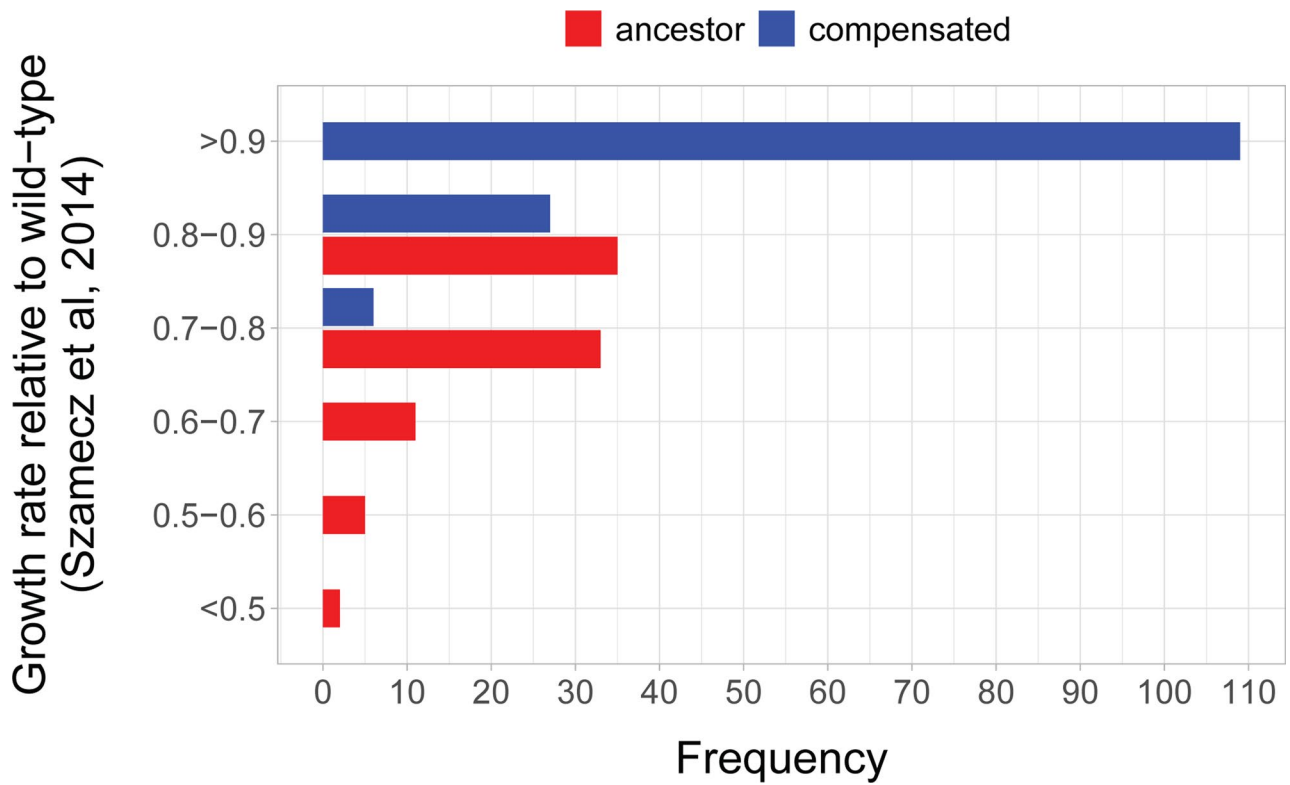
Extended Data Fig. 1 | Fitness distribution of the investigated strains. The barplot shows the distribution of relative fitness of initial knock-out mutant strains (i.e. ancestors, red) and the compensated strains (blue). Data from our previous study[8] is re-plotted here. Relative fitness was estimated by growth rate in liquid medium relative to the wild-type.
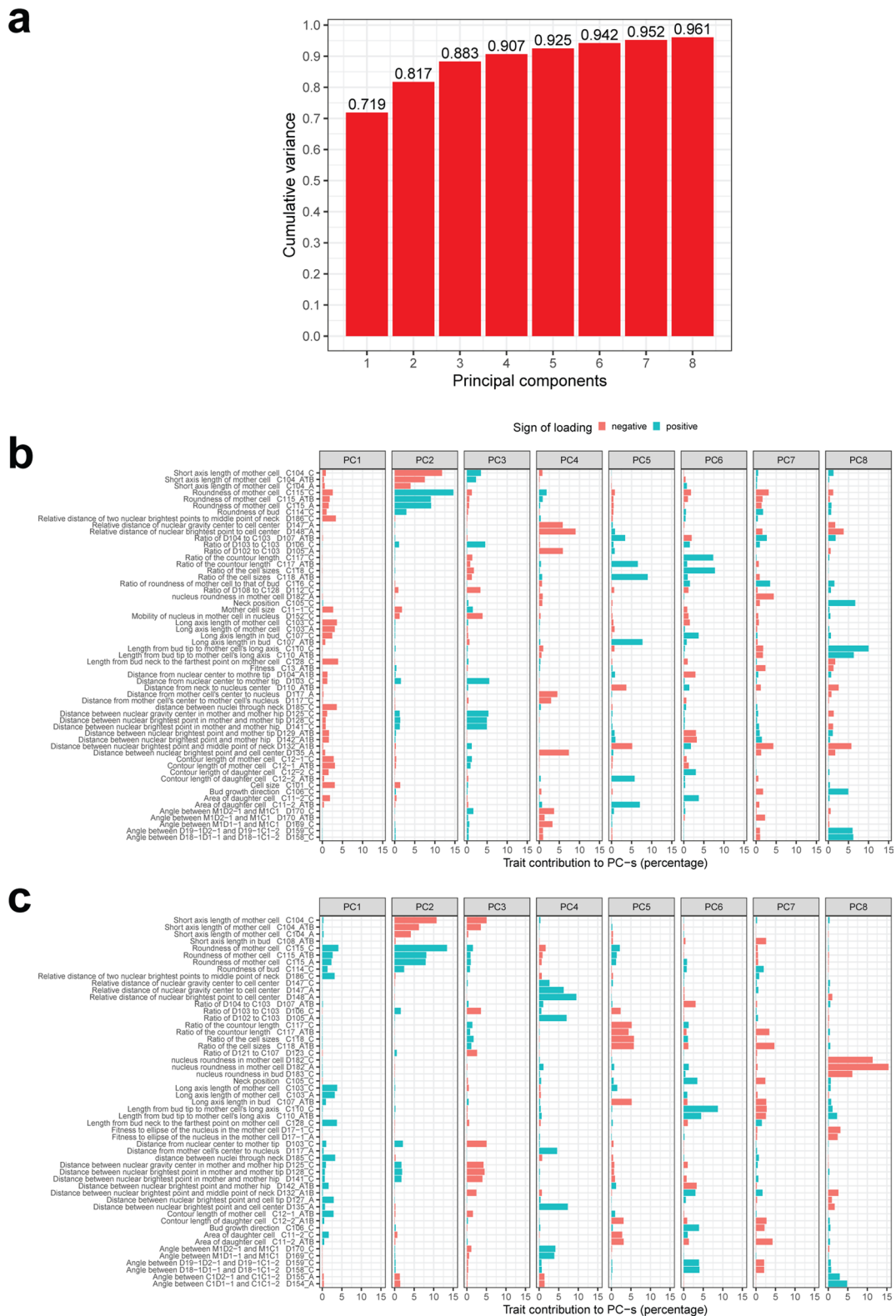
Extended Data Fig. 2 | Results of principal component analysis on single-cell morphology. (a) Cumulative variance of all single-cell morphological traits explained by the first 8 principal components in a principal component analysis (PCA). Note that PCA was performed on all genotypes, including wild-type 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.
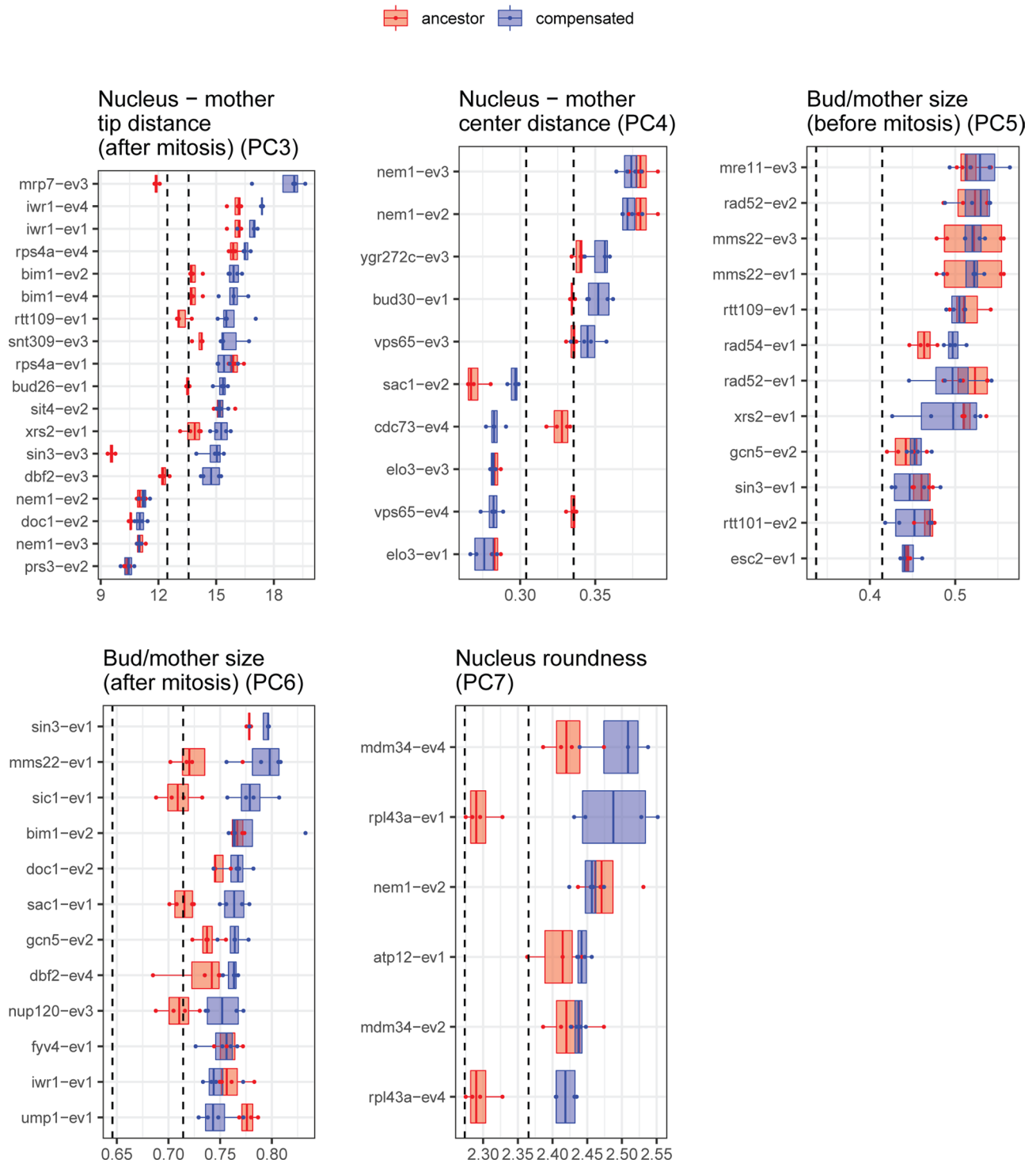
Extended Data Fig. 3 | Evolution of five representative morphological traits in compensated strains with the most extreme trait values. The boxplots display the trait values in the compensated strains and corresponding ancestors compared to that of the wild-type (based on n = 3 or n = 4 biological replicates 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.
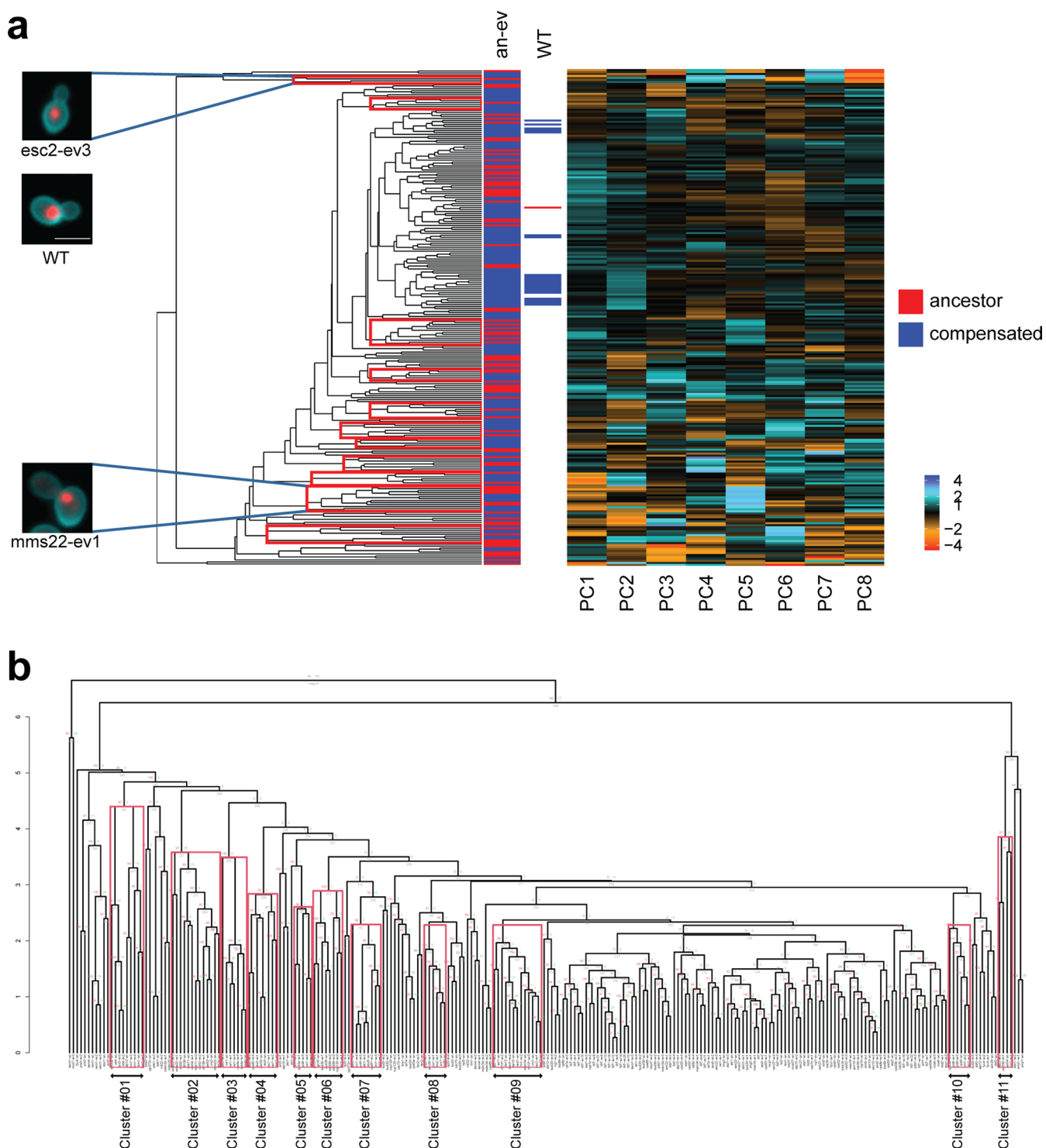
Extended Data Fig. 4 | Clustering of the morphological profiles. (a) Heatmap of morphological profiles. Each row represents the morphological profile of a genotype. Ancestor (red) and compensated (blue) strains are marked along the column next to the dendrogram (labeled as column an-ev). The column left 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 ×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.
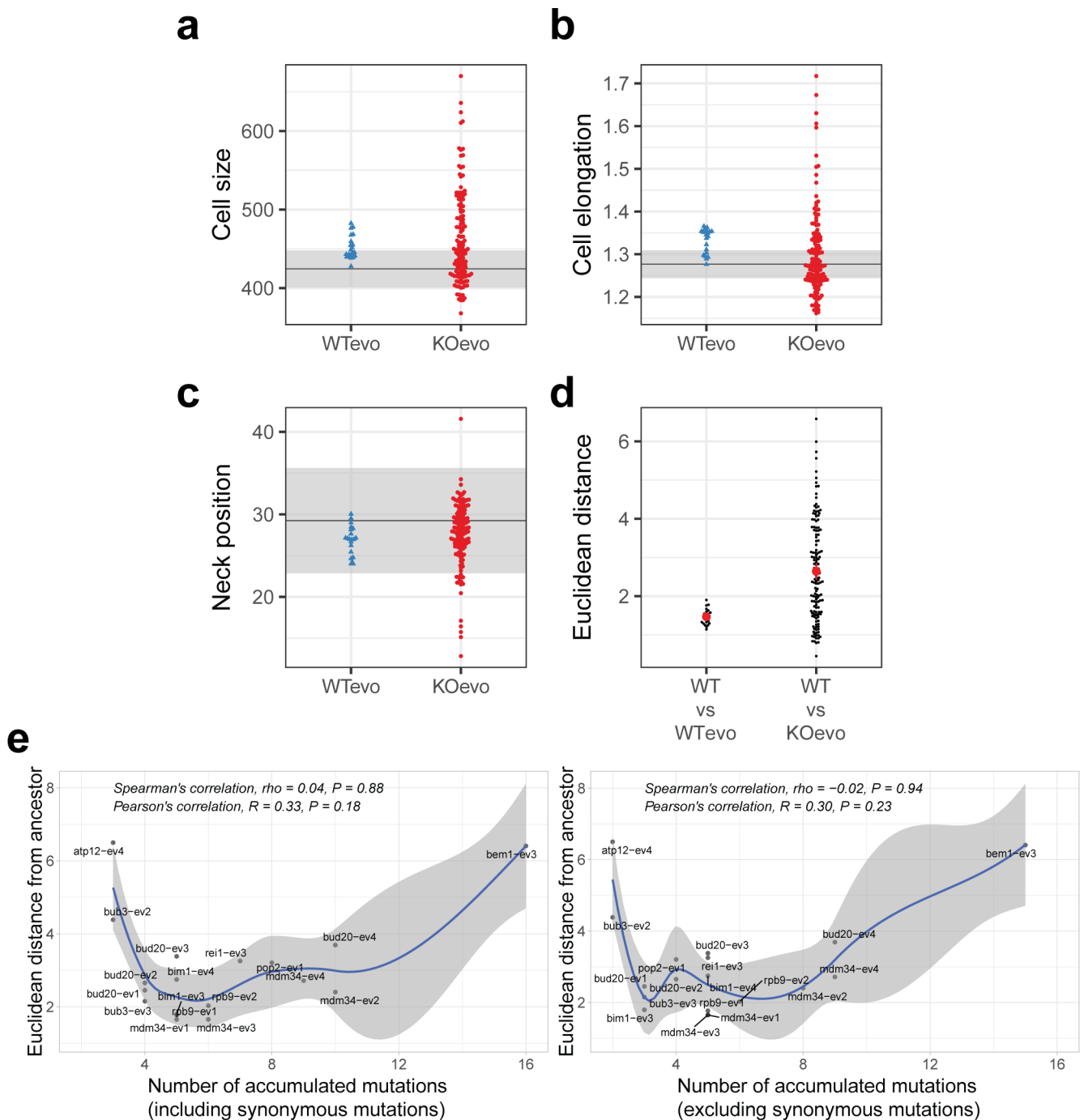
Extended Data Fig. 5 | Morphological changes are specific to compensatory evolution. (a-b-c-d) Evolved control strains show limited change in cellular morphology. Distribution of cell size (**a**), cell elongation (**b**) and neck position (**c**) for the evolved controls (initiated from the wild-type 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 ×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.

a

b
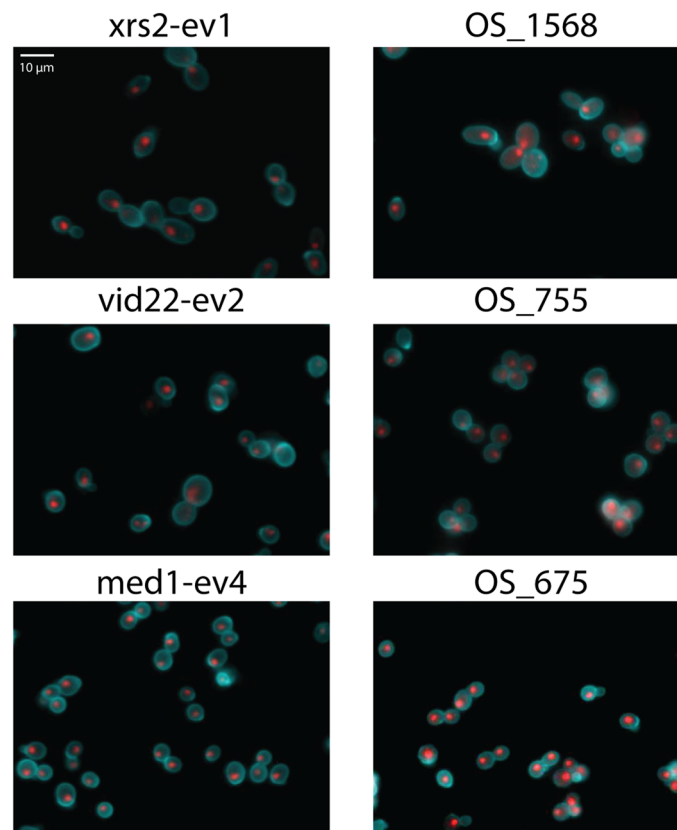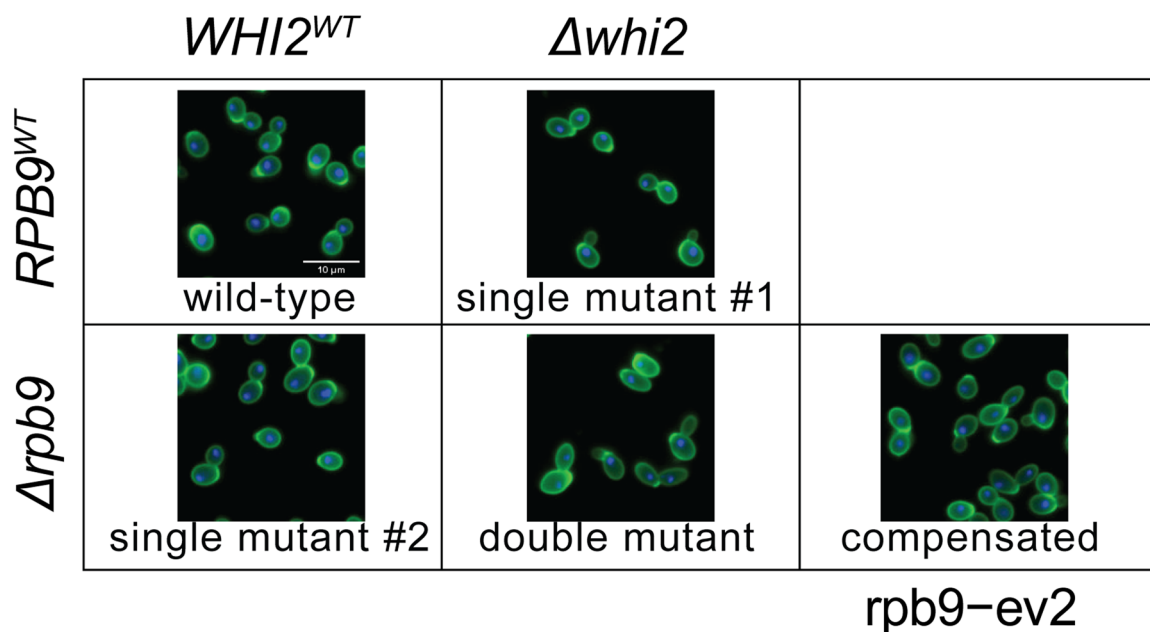
Extended Data Fig. 6 | Larger field of views for microscopy images. (a) Similar cellular morphology of compensated strains and natural isolates. The figure shows wider field of views for Fig. 2c. Images show pairs of compensated and natural strains that display similar morphological trait combinations (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.
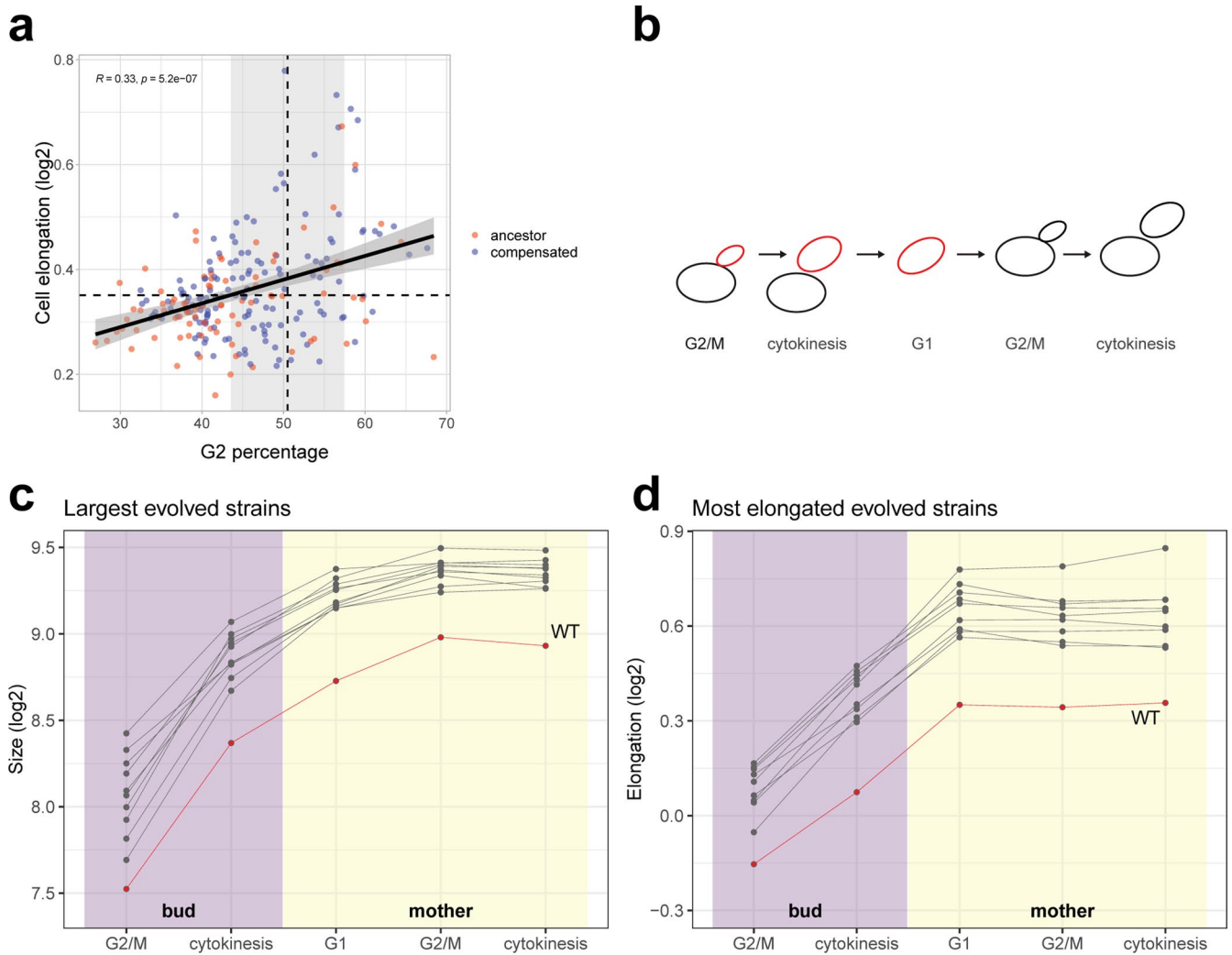
Extended Data Fig. 7 | Cell morphology progression through the cell cycle. (a) Pearson's correlation between cell elongation and G2 percentage (as measured by flow-cytometry). Cell elongation corresponds to CalMorph trait C115-A. WT denotes the wild-type strain. Ancestors and compensated strains are 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.
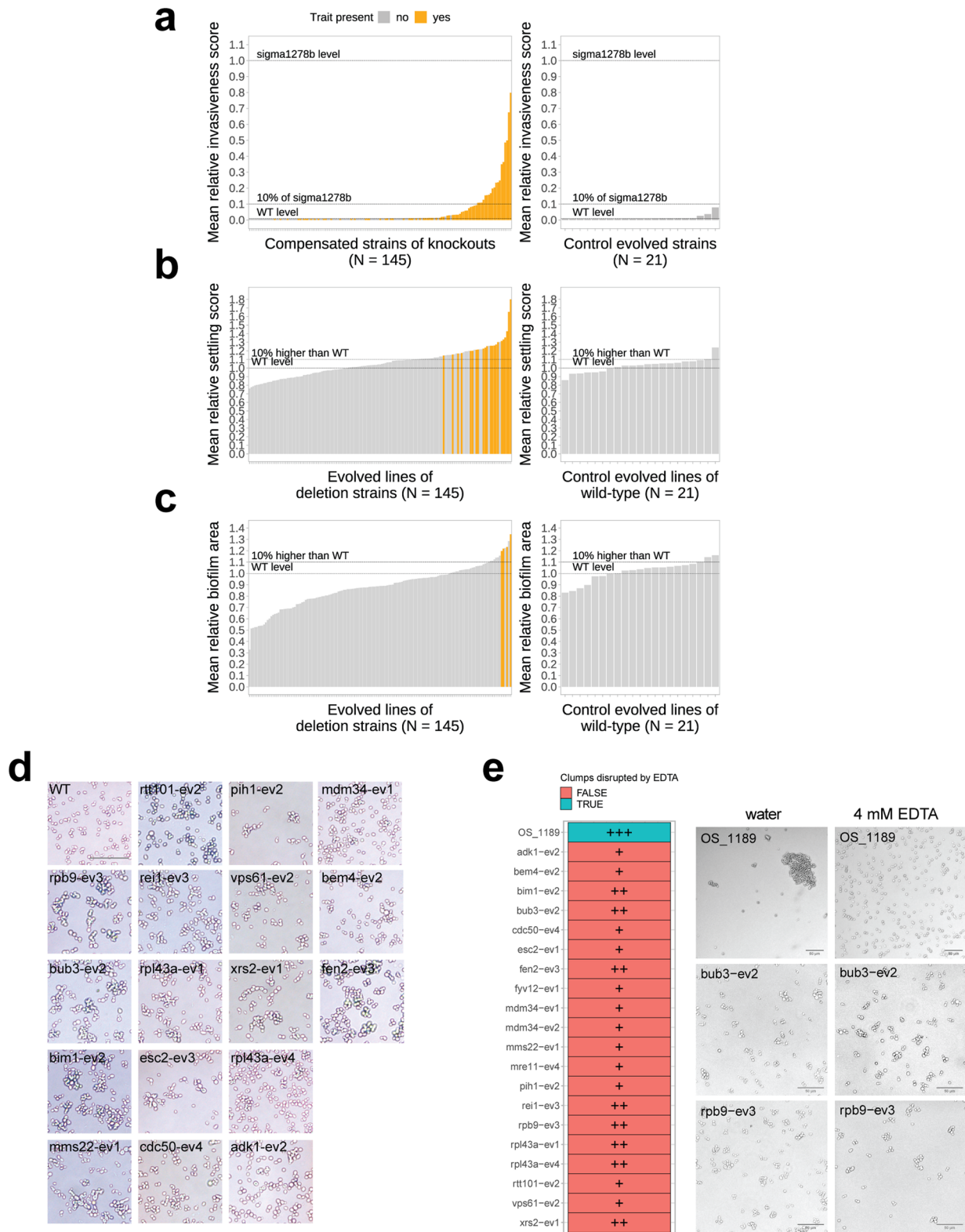
Extended Data Fig. 8 | See next page for caption.

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.
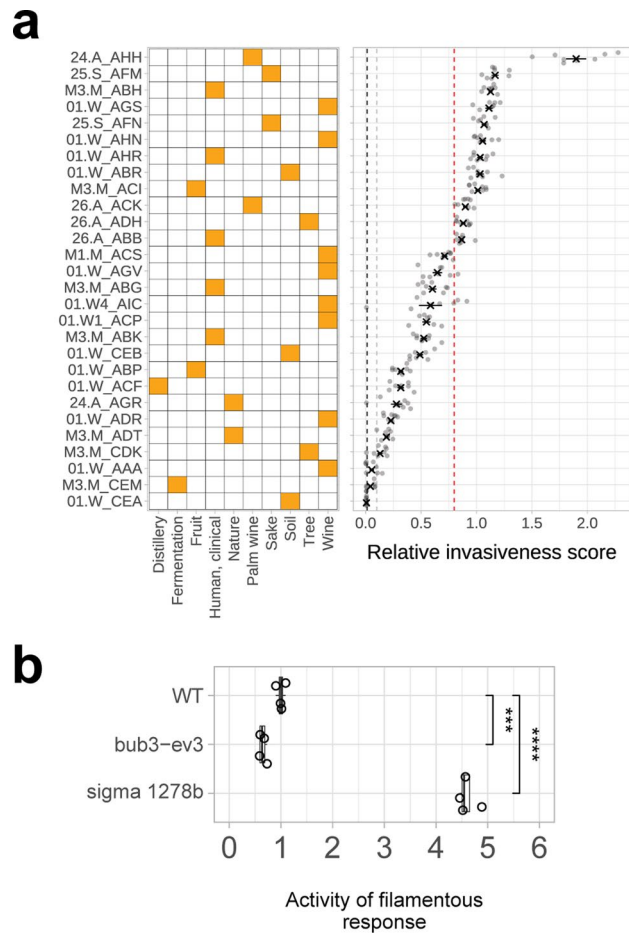
Extended Data Fig. 9 | Analyzing invasive growth phenotype of bub3-ev3 and natural yeast isolates. (a) Invasive growth assay of 29 haploid natural yeast isolates. Natural baker's yeast isolates were selected from a previous study[28] and represent several phylogenetic clades (N = 8) and ecological origins (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.
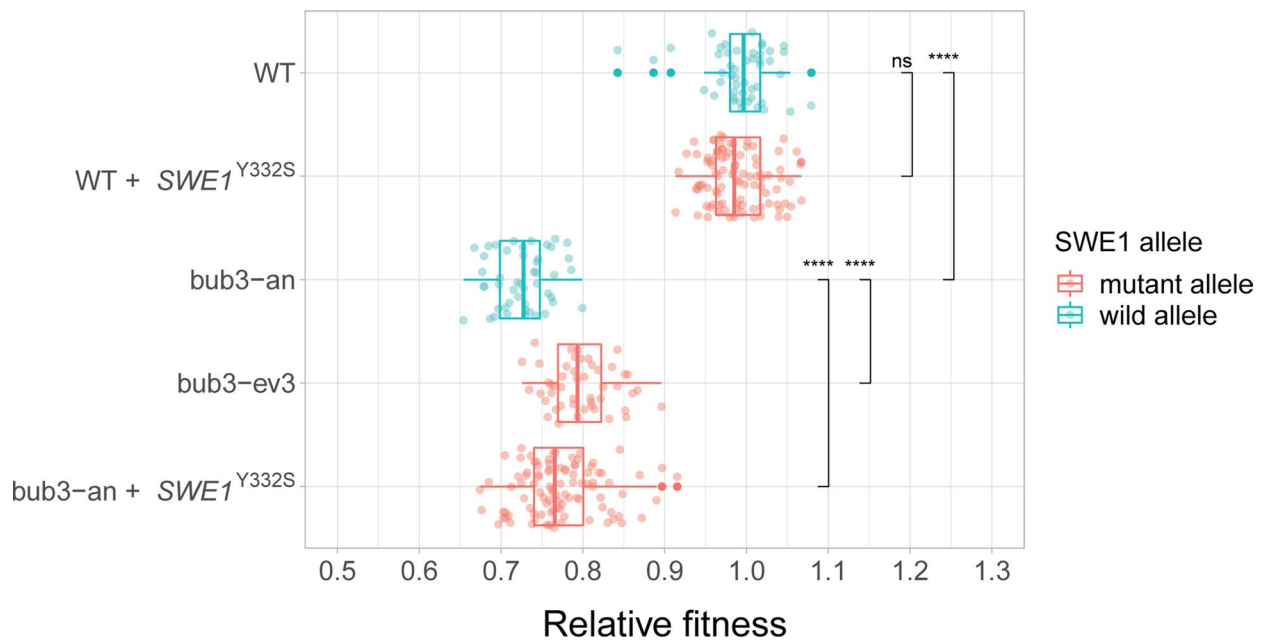
Extended Data Fig. 10 | Mutation in *SWE1* partially compensates the fitness defect of the *Δbub3* ancestor strain. Boxplot shows the relative fitness across several genotypes, including wild-type (WT), the ancestor (bub3-an) and a compensated strain of *Δbub3* (bub3-ev3), and strains harboring the reconstructed *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 $\times10^{-28}$ for comparing WT with WT + SWE1^{Y332S} and bub3-an, respectively, while the P values are 6.92 $\times10^{-13}$ and 4.16 $\times10^{-8}$ 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.

nature portfolio

Corresponding author(s): Balázs Papp, Csaba Pál

Last updated by author(s): 26/02/22

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Raw tiff images of yeast cells were processed using a custom Matlab script. CalMorph 1.2 software was used to extract single-cell morphological data. Pixel intensities for invasive growth assay were estimated with a custom-developed image analysis pipeline in Matlab.
Data analysis	Custom codes were written in R programming language (R version 4.1.0) and 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.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All data are available in the main text or the supplementary materials. Scripts used in the analysis of microscopic images and cell level morphological data 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.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For microsopic morphology we used at least 3 biological replicates following previous studies (e.g. PMID: 23822767), which were found to be sensitive enough to find morphological differences. For macrosopic mesurements the number of biological replicates (3 or 4) was high enough to typically provide significant p-values with the effect size we used.
Data exclusions	All data exclusions are described in the following Methods section: Quality control of raw morphological data
Replication	Biological replicates were used to handle random errors, randomization was used to preclude systematic errors/batch effects.
Randomization	Random layout was used to assign biological replicates of strains to well positions.
Blinding	Not applicable, objective automated image processing was used to quantify morphological differences.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Exponentially growing yeast cells were fixed and stained with Sytox-Green nucleic acid stain to investigate ploidy level.
Instrument	Beckman Coulter's CytoFLEX S
Software	CytExpert for CytoFLEX Acquisition and Analysis Software (DNA content analysis), Kaluza Analysis 2.1 software from Beckman Coulter (Beckman Coulter Diagnostic, CA, USA; cell-cycle analysis)
Cell population abundance	Not applicable

Gating strategy | Not applicable

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.