Article

Genome duplication in a long-term multicellularity evolution experiment

https://doi.org/10.1038/s41586-025-08689-6

Received: 23 April 2024

Accepted: 23 January 2025

Published online: 5 March 2025

Check for updates

Kai Tong[1,2,3,4,10]✉, Sayantan Datta[1,2,10], Vivian Cheng[1,5], Daniella J. Haas[1,6], Saranya Gourisetti[1], Harley L. Yopp[1], Thomas C. Day[7,8], Dung T. Lac[1], Ahmad S. Khalil[3,4,9], Peter L. Conlin[1], G. Ozan Bozdag[1] & William C. Ratcliff[1]✉

Whole-genome duplication (WGD) is widespread across eukaryotes and can promote adaptive evolution[1–4]. However, given the instability of newly formed polyploid genomes[5–7], understanding how WGDs arise in a population, persist, and underpin adaptations remains a challenge. Here, using our ongoing Multicellularity Long Term Evolution Experiment (MuLTEE)[8], we show that diploid snowflake yeast (*Saccharomyces cerevisiae*) under selection for larger multicellular size rapidly evolve to be tetraploid. From their origin within the first 50 days of the experiment, tetraploids persisted for the next 950 days (nearly 5,000 generations, the current leading edge of our experiment) in 10 replicate populations, despite being genomically unstable. Using synthetic reconstruction, biophysical modelling and counter-selection, we found that tetraploidy evolved because it confers immediate fitness benefits under this selection, by producing larger, longer cells that yield larger clusters. The same selective benefit also maintained tetraploidy over long evolutionary timescales, inhibiting the reversion to diploidy that is typically seen in laboratory evolution experiments. Once established, tetraploidy facilitated novel genetic routes for adaptation, having a key role in the evolution of macroscopic multicellular size via the origin of evolutionarily conserved aneuploidy. These results provide unique empirical insights into the evolutionary dynamics and impacts of WGD, showing how it can initially arise due to its immediate adaptive benefits, be maintained by selection and fuel long-term innovations by creating additional dimensions of heritable genetic variation.

Polyploidy, resulting from WGD, is widespread in nature and is an important driver of species adaptation and diversification[1,3]. Most, if not all, living species bear signatures of ancient WGDs[3,9]. However, the establishment of nascent polyploids is rarely successful[10,11], as newly formed polyploids usually face fitness disadvantages when competing in a population of diploids under normal environments[5,12]. Furthermore, nascent polyploids often exhibit genomic instability[5–7] and rapidly revert to diploidy via chromosome losses[13–15] (although certain genetic changes can stabilize polyploidy[16]). This costly and transient nature of nascent polyploidy raises the question of how polyploidy can rise and be maintained over long-term evolution. A central hypothesis is that the immediate phenotypic effects of polyploidy, often stemming from the increased size of polyploid cells, can confer fitness advantages under novel, often stressful, environments[12,17–20]. This has been suggested to contribute to many ancient WGDs that rose during periods of drastic climate change[21]. However, it remains elusive whether selection on the immediate phenotypic effects of nascent polyploidy is sufficient to drive its rise and long-term persistence, especially given the genomic instability of polyploidy that typically erodes recently duplicated genomes.

The instability of nascent polyploid genomes may provide an evolutionary advantage under novel environments by rapidly generating genetic variation, especially via aneuploidy[9,22]. This has been shown to facilitate the rapid evolution of microorganisms[2,23–25] and cancer[4,26,27]. Although this benefit of WGD often arises as the transient polyploids undergo genome reduction towards diploidy, it remains untested whether WGD can also facilitate adaptation via novel aneuploidy when polyploidy is maintained (that is, it still possesses a baseline tetraploid genome content), potentially fuelling longer-term adaptation.

Understanding how WGDs rise, persist and drive both short-term and long-term adaptation is fundamental to our understanding of their evolutionary impacts. Our current knowledge about WGDs is largely based on the comparison of natural polyploids with their diploid relatives, which, although informative, can be confounded by evolution following WGD establishment and concomitant evolutionary processes other than WGD. Experimental evolution provides a novel opportunity to overcome these limitations[14,15]. However, to our knowledge, no previous work has observed how polyploidy arises de novo from diploids, and the rapid losses of polyploidy in laboratory experiments limits our ability to study the long-term maintenance and consequences of

[1]School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA. [2]Interdisciplinary Graduate Program in Quantitative Biosciences, Georgia Institute of Technology, Atlanta, GA, USA. [3]Biological Design Center, Boston University, Boston, MA, USA. [4]Department of Biomedical Engineering, Boston University, Boston, MA, USA. [5]School of Integrative Biology, University of Illinois Urbana-Champaign, Urbana, IL, USA. [6]Mayo Clinic Alix School of Medicine, Rochester, MN, USA. [7]School of Physics, Georgia Institute of Technology, Atlanta, GA, USA. [8]Department of Biological Sciences, University of Southern California, Los Angeles, CA, USA. [9]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA, USA. [10]These authors contributed equally: Kai Tong, Sayantan Datta. ✉e-mail: kaitong@bu.edu; ratcliff@gatech.edu

WGD. Thus, we lack an experimental system for directly examining how WGDs spontaneously arise and subsequently evolve over long evolutionary timescales.

Our ongoing Multicellularity Long Term Evolution Experiment (MuLTEE)[8] provides a unique experimental system to circumvent these long-standing constraints. Initially designed to study the open-ended evolution of a nascent multicellular organism, we subjected mixotrophic and anaerobic populations of snowflake yeast (*S. cerevisiae*), a model of undifferentiated multicellularity, to 1,000 rounds (approximately 5,000 generations) of daily selection for larger size. We found that in all replicate populations, tetraploidy rapidly evolved from diploid ancestors and has been maintained over the rest of the experiment. We found that tetraploidy initially evolved because it produces larger clusters made of larger, longer cells, an immediate advantage under our size-based selection. Continuous selection for larger size also maintained tetraploidy for unprecedented timescales, despite its intrinsic tendency to regress back towards diploidy. Moreover, although polyploidy evolved in both mixotrophic and anaerobic populations, the former remained microscopic and predominantly euploid, whereas the latter evolved extensive aneuploidy, which underpins the subsequent origins of macroscopic multicellular size. Together, we provide direct experimental evidence that selection on the immediate phenotypic effects of polyploidy is sufficient to drive its rise and long-term persistence, and when polyploidy is maintained, its intrinsic genomic instability can fuel long-term adaptation through aneuploidy.

1,000 days of multicellular evolution

We began the MuLTEE (Fig. 1a) with the goal of examining the open-ended evolution of a nascent multicellular organism[8]. The snowflake yeast ancestor that we used to initiate the MuLTEE is a homozygous diploid *S. cerevisiae ACE2* knockout, which causes incomplete cell separation after mitosis[28]. Snowflake yeast grow as clonal multicellular groups and reproduce through branch fragmentation induced by cell packing stress[28,29]. We evolved snowflake yeast with three metabolic treatments: mixotrophic, obligately anaerobic and obligately aerobic, each with five replicate populations (that is, lines). In this study, we focus on the five mixotrophic populations (PM1–5), which were started with a mixotrophic ancestor, and the five anaerobic populations (PA1–5), which descend from a petite mutant with non-functional mitochondria. We subjected these populations to daily cycles of growth and selection for rapid settling through liquid media, favouring both rapid growth and larger multicellular size. Although PAs evolved to be approximately 20,000-fold larger within 600 days (approximately 3,000 generations), forming macroscopic clusters, PMs only increased in size by approximately sixfold[8]. This was because limited oxygen diffusion constrains the evolution of increased size under mixotrophic, but not anaerobic, metabolism[8,30].

We have continued the MuLTEE for 400 more transfers, or 1,000 days (approximately 5,000 generations) in total (Fig. 1b and Extended Data Fig. 1). In our model system, novel multicellular traits (for example, larger, tougher clusters) arise as an emergent property of changes in cell-level traits (for example, increased cell aspect ratio)[8,29–31]. Here we characterized the evolutionary history of key group-level and cell-level traits across the first 1,000 days of the MuLTEE (Fig. 1c–e and Extended Data Fig. 2). We focus particularly on the evolution of cell volume, as this has not been systematically examined in our system and is a common phenotypic effect of WGD[9]. We found that the PMs experienced a 1.9-fold increase in cell volume within the first 200 days of our experiment ($P = 7.53 \times 10^{-5}$, $t_4 = 16.7$, two-tailed one-sample Student's *t*-test), which remained largely unchanged for the rest of the experiment ($P = 0.637$, $F_{3,16} = 0.58$, one-way analysis of variance (ANOVA); Fig. 1d). Similarly, the cell volume in PAs increased by 2.3-fold during the first 200 days ($P = 1.07 \times 10^{-5}$, $t_4 = 27.3$, two-tailed one-sample Student's *t*-test), with little further increase after 400 days ($P = 0.593$, $F_{2,12} = 0.546$, one-way ANOVA; Fig. 1d). Although PMs largely plateaued in cell aspect ratio and cluster size after 200 days ($P = 0.455$ and 0.613, $F_{3,16} = 0.917$ and 0.618, respectively, one-way ANOVA) following an initial increase ($P = 1.61 \times 10^{-4}$ and 8.27×10^{-5}, $t_4 = 13.7$ and 16.3, respectively, two-tailed one-sample Student's *t*-test, comparing t200s and t0), the PAs displayed continuous increases in these two traits over the experiment ($P = 3.17 \times 10^{-6}$ and 5.12×10^{-7}, $r^2 = 0.67$ and 0.73, respectively, linear regression; Fig. 1c,e).

As cell volume increased concomitantly with cell aspect ratio in both PMs and early PAs (Fig. 1d,e), we sought to disentangle their effects on cluster size, using our previously validated biophysical model[8,32]. We found that increased cell volume and aspect ratio both lead to larger clusters (Fig. 1f), but in mechanistically different ways: larger cells give rise to proportionally larger clusters without changing cell packing density or cell number per cluster, whereas longer cells reduce cell packing density and allow growing more cells per cluster (Fig. 1g,h). These results demonstrate how increases in cell volume and aspect ratio, two key cellular traits, contribute to the evolution of larger multicellular size in snowflake yeast through distinct biophysical mechanisms.

Origin and maintenance of polyploidy

In most organisms, WGD results in a twofold increase in cell volume[9]. We thus examined the ploidy of our yeast over the MuLTEE. First, we sequenced the genomes of isolates from t200 and beyond, and we found that most of their point mutations have allele frequencies centred around 0.25, with the others around 0.5, 0.75 or 1 (Fig. 2a), suggesting that they had evolved tetraploid genomes. To further validate this, we developed an imaging-based method for measuring ploidy levels of multicellular yeast strains (Extended Data Fig. 3a). As we used asynchronous, exponential-phase cultures for ploidy measurements, the distribution of cellular DNA contents of a single, clonal strain is expected to contain two peaks, corresponding to G1-phase and DNA-replicated G2-phase cells (Extended Data Fig. 3b). All evolved isolates have roughly double the genome size as their diploid ancestor, suggesting that they have undergone WGD (Fig. 2b). To trace the origin of tetraploidy, we measured ploidy distributions of populations from earlier time points, and we discovered that in all ten populations, tetraploidy had emerged and become dominant by 50 days, fixing by 100 days (Fig. 2c). Together, these results show that in all ten PM and PA lines, tetraploidy emerged very early in the MuLTEE, and it has been maintained for over 950 transfers (and counting). Of note, although PMs were mainly euploid, PAs evolved extensive aneuploidy (Fig. 2d and Extended Data Fig. 4). We examine the temporal dynamics (Fig. 2d) and evolutionary consequences (Figs. 4 and 5) of aneuploidy evolving after WGD in the section 'Aneuploidy underpins macroscopic size', focusing first on the evolution of WGD itself.

To our knowledge, this is the first evolution experiment in which polyploidy convergently evolved from diploid ancestors. Although it has long been known that polyploidy can be mechanistically induced by physical or chemical treatment[33,34], we did not use an experimental treatment that directly induces polyploidy. Moreover, the long-term persistence of polyploidy in our experiment (more than 4,750 generations) is in dramatic contrast to previous evolution experiments, where, under various conditions, tetraploid yeast ancestors are genomically unstable and typically converge to diploidy within a few hundred generations[2,13,16,35,36]. This raises the question regarding what drove the origin and maintenance of polyploidy in the MuLTEE.

Polyploidy brings immediate size benefit

In unicellular *S. cerevisiae*, increased ploidy increases both cell volume[6,37] and cell aspect ratio[37]. In fact, increased cell size is a universal feature of polyploid cells across eukaryotes[9]. As increased cell volume and aspect ratio both contribute to larger snowflake yeast clusters (Fig. 1f), we hypothesize that tetraploidy arose in the MuLTEE because

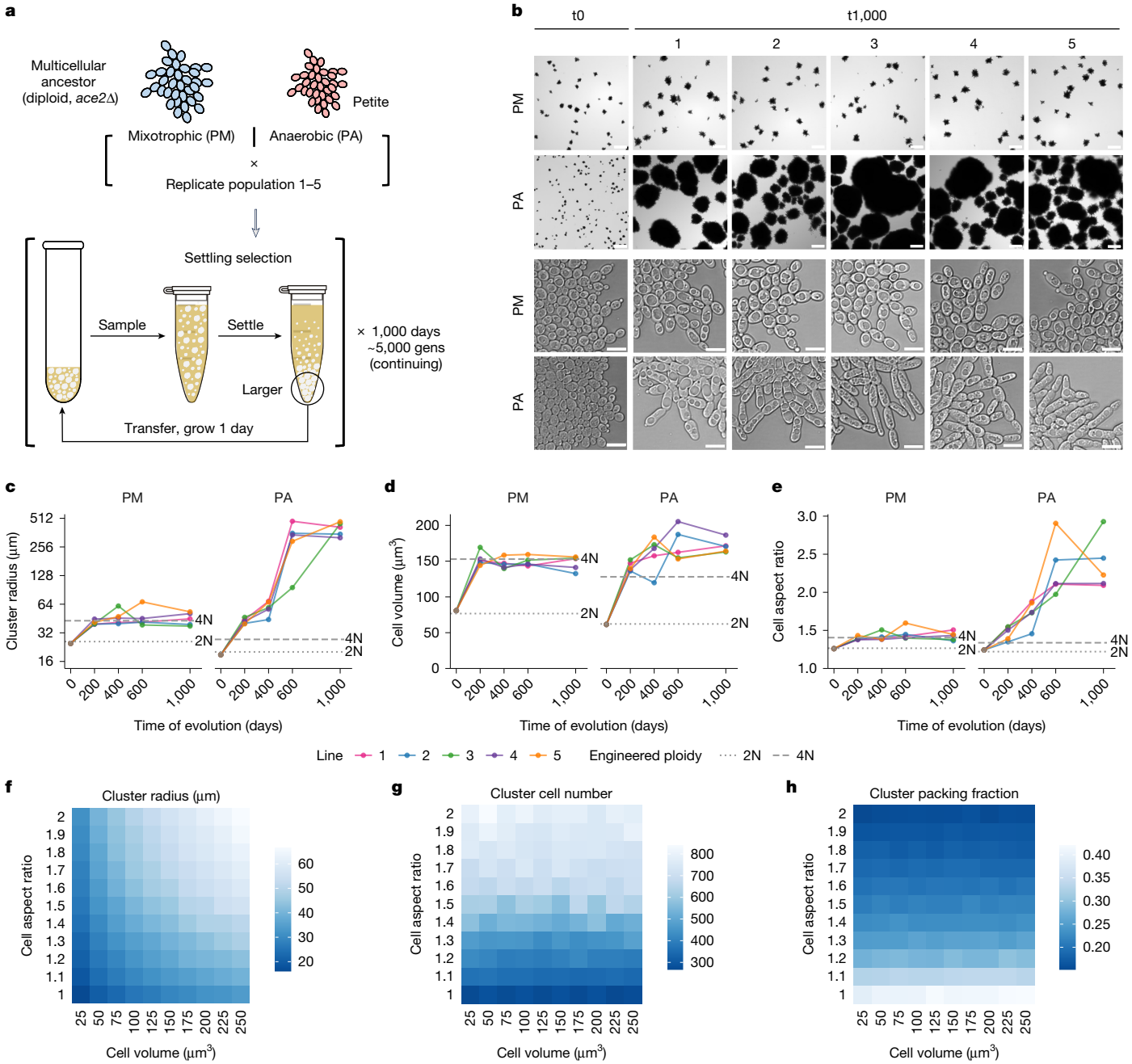


Fig. 1 | Increased cell volume and aspect ratio drive the evolution of larger multicellular groups in the 1,000-day MuLTEE. a, Experimental setup of the MuLTEE. **b**, Representative cluster-level (top two rows) and cell-level (bottom two rows) images of the ancestor (t0) and day 1,000 (t1,000) isolates from the five replicate populations evolved under PM and PA conditions. Scale bars, 200 μm (cluster-level images) and 10 μm (cell-level images). **c**–**e**, Evolutionary dynamics of biomass-weighted mean cluster radius (Methods; **c**), mean cell volume (**d**) and mean cell aspect ratio (**e**), showing the values of PM/PA t0 and PM/PA1–5 t200, t400, t600 and t1,000 isolates (on average, $n = 886$ clusters (**c**) and 1,288 cells (**d**,**e**) measured for each of the 42 strains), with the grey dotted and dashed lines representing artificially constructed diploid and tetraploid strains (mean value of the four biological replicates in Fig. 3e,b,c), respectively. **f**–**h**, Heat maps showing the biophysical simulations of how cell volume and cell aspect ratio affect the mean cluster radius (**f**), cell number per cluster (**g**) and cell packing fraction (fraction of the cluster volume occupied by cells; **h**) of clusters at fragmentation ($n = 50$ simulated clusters per pair of parameter values).

it brings immediate phenotypic effects, generating larger, longer cells that yield larger clusters, which is beneficial under settling selection. To test this, we genetically engineered tetraploidy in a diploid snowflake yeast background (Extended Data Fig. 5a and Supplementary Table 1). Consistent with our hypothesis, under both mixotrophic and anaerobic conditions, tetraploid clusters consist of larger, more elongated cells and are larger than their diploid counterparts (Fig. 3a–f and Extended Data Fig. 5b–e).

Next, we compared the engineered tetraploids with the evolved isolates to see how much of the phenotypic changes in the MuLTEE can be explained by tetraploidization alone. In PMs, engineering tetraploidy alone recapitulated the increases in cluster size, cell volume and cell aspect ratio over the first 1,000 days of the MuLTEE (Fig. 1c–e). In PAs, tetraploidy contributed to most of the cell volume increase in the first 200 days of the experiment (Fig. 1d), as well as most of the increases in cluster size and cell aspect ratio in the first 50 days (Extended Data

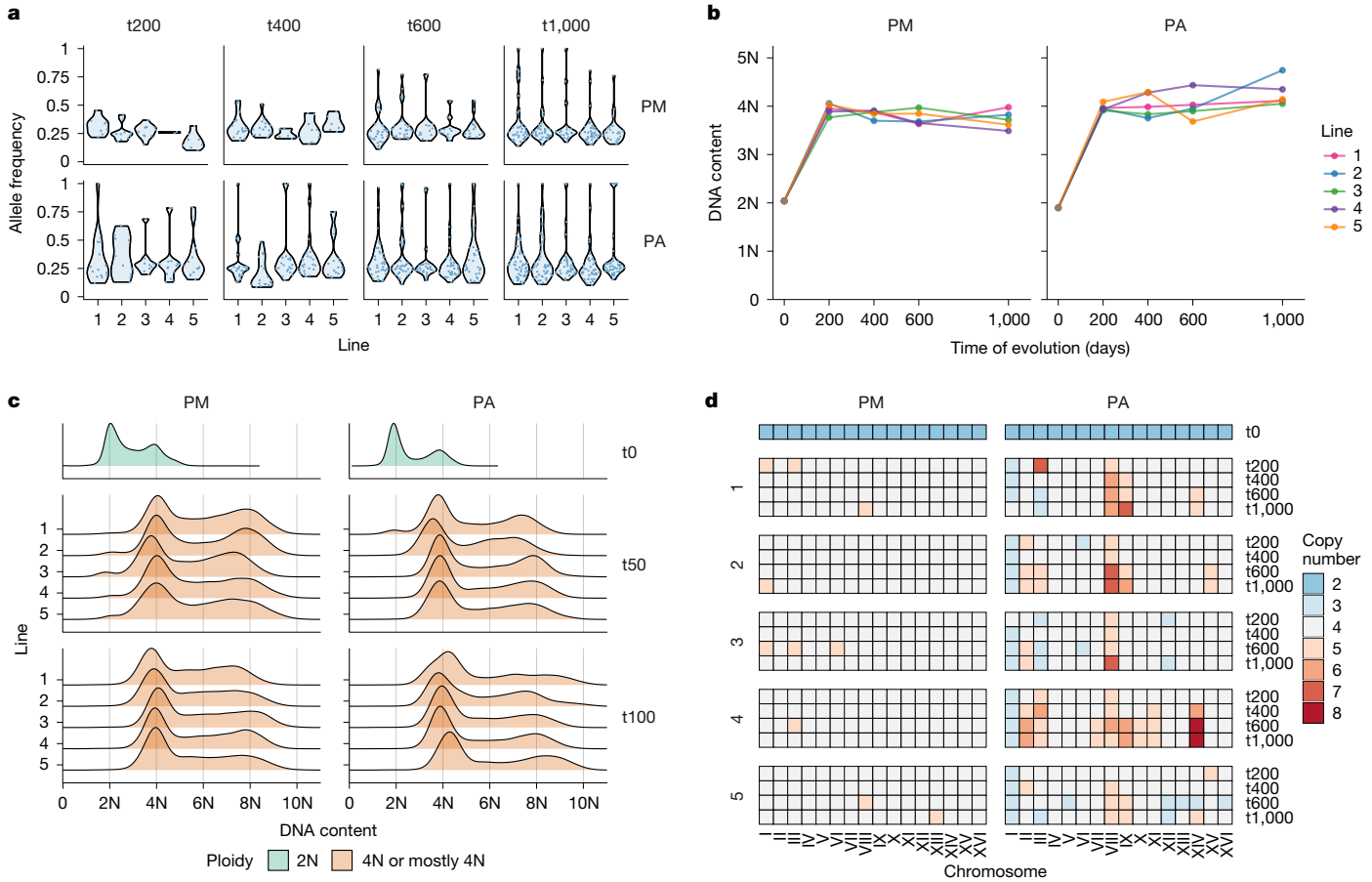


Fig. 2 | Tetraploidy rapidly emerged and was maintained in all PMs and PAs; PAs subsequently evolved extensive aneuploidy. a, Violin plot of the distribution of allele frequencies (each determined as the proportion of reads containing the mutant allele) of novel point mutations identified in PM/PA1–5 t200, t400, t600 and t1,000 isolates, with each dot representing one point mutation. **b**, Evolutionary dynamics of DNA content, showing that of PM/PA t0 and the strains used in panel **a** (on average, $n = 11{,}456$ cells measured for each of the 42 strains). **c**, Ridge plot of the distribution of cellular DNA contents in the PM/PA t0 and PM/PA1–5 t50 and t100 populations (on average, $n = 12{,}922$ cells measured for each of the 22 populations). As we used asynchronous cultures for ploidy measurements, a single genotype in a population would form two distinct DNA-content peaks, corresponding to its G1-phase and G2-phase cells. **d**, Heat maps of the karyotypes of PM/PA t0 and the strains used in panel **a**, determined based on chromosome coverages from the whole-genome sequencing data.

Fig. 5f,g), when tetraploidy became dominant (Fig. 2c). However, tetraploidy alone does not explain the subsequent increases in cluster size and cell aspect ratio in PAs (Fig. 1c,e).

To examine whether tetraploidy indeed confers an immediate fitness benefit under selection for larger size, we competed engineered diploid and tetraploid clusters with and without settling selection, using a label-free fitness assay (Extended Data Fig. 6). Supporting our hypothesis, under both mixotrophic and anaerobic conditions, settling selection strongly favoured tetraploidy (Fig. 3g), increasing its average frequency from 53% to 82% within 3 days.

Selection for larger size drives polyploidy

Our results above suggest that tetraploidy evolved as a mechanism of increasing group size. To test this more directly, we re-evolved the MuLTEE ancestors with selection acting in the opposite direction. This was achieved by growing our yeast on solid media (Extended Data Fig. 7a), a condition known to favour smaller groups[38]. We evolved the PM and PA ancestors, each with four replicate populations, on agar with daily dilutions for 70 days (approximately 500 generations). All populations remained predominantly diploid (Fig. 3h). This is markedly different from the MuLTEE, in which tetraploid strains became dominant within the first 50 days (approximately 250 generations) of settling selection (Fig. 2c).

Although tetraploidy initially evolved as a mechanism to increase group size in our experiment, nascent tetraploidy, especially in *S. cerevisiae*, is notoriously unstable[2,13,16,35,36]. What explains its maintenance over nearly 5,000 generations in the MuLTEE is still unknown. We have two central hypotheses: tetraploidy may be actively maintained by selection for larger size, or molecular mechanisms could have evolved that stabilize the duplicated genomes, like in Lu et al.[16].

To disentangle these hypotheses, we performed a similar reverse selection experiment, evolving the tetraploid t1,000 isolates with selection against larger size. We evolved the ten PM and PA t1,000 isolates, each with four replicate populations, on solid media with daily transfers for 70 days (approximately 500 generations; Extended Data Fig. 7a), and most populations indeed evolved reduced cluster size (Extended Data Fig. 7b). In total, 22 of 40 evolved populations exhibited ploidy reduction, including nine populations that reverted to a genome size similar to the diploid ancestor (Fig. 3i). Furthermore, ploidy reductions were observed in 8 of 10 genetically distinct t1,000 isolates (Fig. 3i). These results suggest that the long-term maintenance of tetraploidy over the MuLTEE was due to sustained selection for larger size. Of note, the presence of intermediate ploidy levels (approximately 3N) in multiple evolved populations (Fig. 3i) suggests that the ploidy reductions were probably mediated by non-meiotic chromosome losses (like in Gerstein et al.[13,35]). Although we cannot rule out the contribution of meiosis to ploidy reduction, we do not expect it to have a major role: PAs are

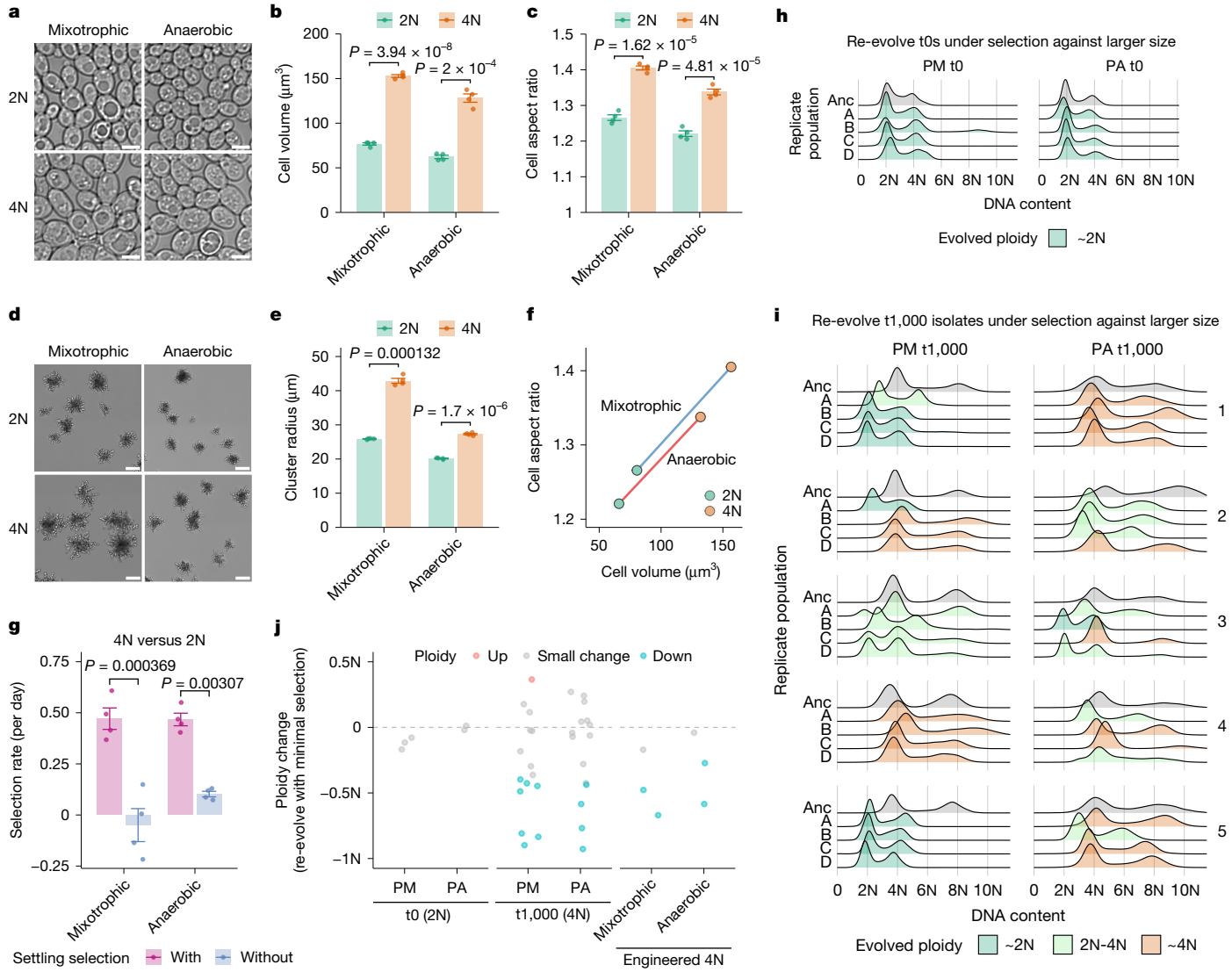


Fig. 3 | Tetraploidy confers immediate phenotypic effects that are beneficial under selection for larger size, driving the origin and maintenance of tetraploidy despite its genomic instability. a,**d**, Cell-level (**a**) and cluster-level (**d**) images of the engineered diploid and tetraploid clusters under mixotrophic and anaerobic conditions. Scale bars, 5 µm (**a**) and 50 µm (**d**). **b**,**c**,**e**, Mean cell volume (**b**), mean cell aspect ratio (**c**) and biomass-weighted mean cluster radius (**e**) of the strains in panels **a**,**d**. **f**, Relationship of the changes in cell volume and cell aspect ratio due to tetraploidization, under mixotrophic and anaerobic conditions. Values are means of the four replicates or petite mutants in panels **b**,**c**. **g**, Per day selection rate of competing engineered tetraploid clusters against diploid counterparts with and without settling selection for 3 days, under mixotrophic and anaerobic conditions. For **b**,**c**,**e**,**g**, values are mean ± s.e.m. (n = 4 biological replicates for mixotrophic condition or 4 independent petite mutants for anaerobic condition; on average, 2,458 cells (**b**,**c**) and 1,077 clusters (**e**) were measured per replicate or petite mutant, and 381 clusters (**g**) were measured per sample). P values were calculated by two-tailed Welch's t-test. **h**,**i**, Distribution of cellular DNA contents in the ancestral (Anc) and evolved populations, initiated with PM/PA t0 (**h**) and PM/PA1–5 t1,000 isolates (**i**) with four replicate populations (A–D), after selecting against larger size by evolving on agar for 70 days (on average, n = 14,231 cells measured per population). **j**, Ploidy (G1 peak) changes in the evolved populations, initiated with PM/PA t0, PM/PA1–5 t1,000 isolates, and engineered mixotrophic and anaerobic tetraploids, each with three replicate populations, after evolving with minimal selection for 56 days or 28 bottlenecks (on average, n = 15,848 cells measured per population). Cohen's d = 0.8 is used as the cut-off for large and small ploidy changes. Grey dashed line indicates no ploidy change.

mitochondrial petites and thus cannot sporulate[39], and sporulation in PMs is rare when growing on rich media.

To further test whether polyploidy arises and persists in the MuLTEE as the result of selection, rather than intrinsic genomic properties of our yeast that favour genome duplication and maintenance, we performed a mutation accumulation experiment. By repeatedly passaging populations through single-cell bottlenecks, mutation accumulation experiments allow evolution to proceed in the near-absence of selection, revealing the inherent tendencies of genomic changes. We passaged three replicate populations for each of the diploid PM and PA ancestors, all five of their t1,000 isolates, and engineered tetraploid strains through 28 consecutive single-colony bottlenecks over 56 days. The genome size of the ancestors remained generally constant, whereas at least one replicate population of engineered tetraploids and 8 of 10 of tetraploid t1,000 isolates underwent considerable ploidy reduction (Cohen's d > 0.8) by approximately 0.5–1N (Fig. 3j and Extended Data Fig. 8), indicative of non-meiotic chromosome losses. These results demonstrate that the diploid ancestors have no intrinsic tendency for tetraploidization, whereas the evolved tetraploids remained genomically unstable and prone to ploidy reduction, further underscoring that selection for larger size (Fig. 3a–i) is the main driver of the rise and maintenance of polyploidy in the MuLTEE.

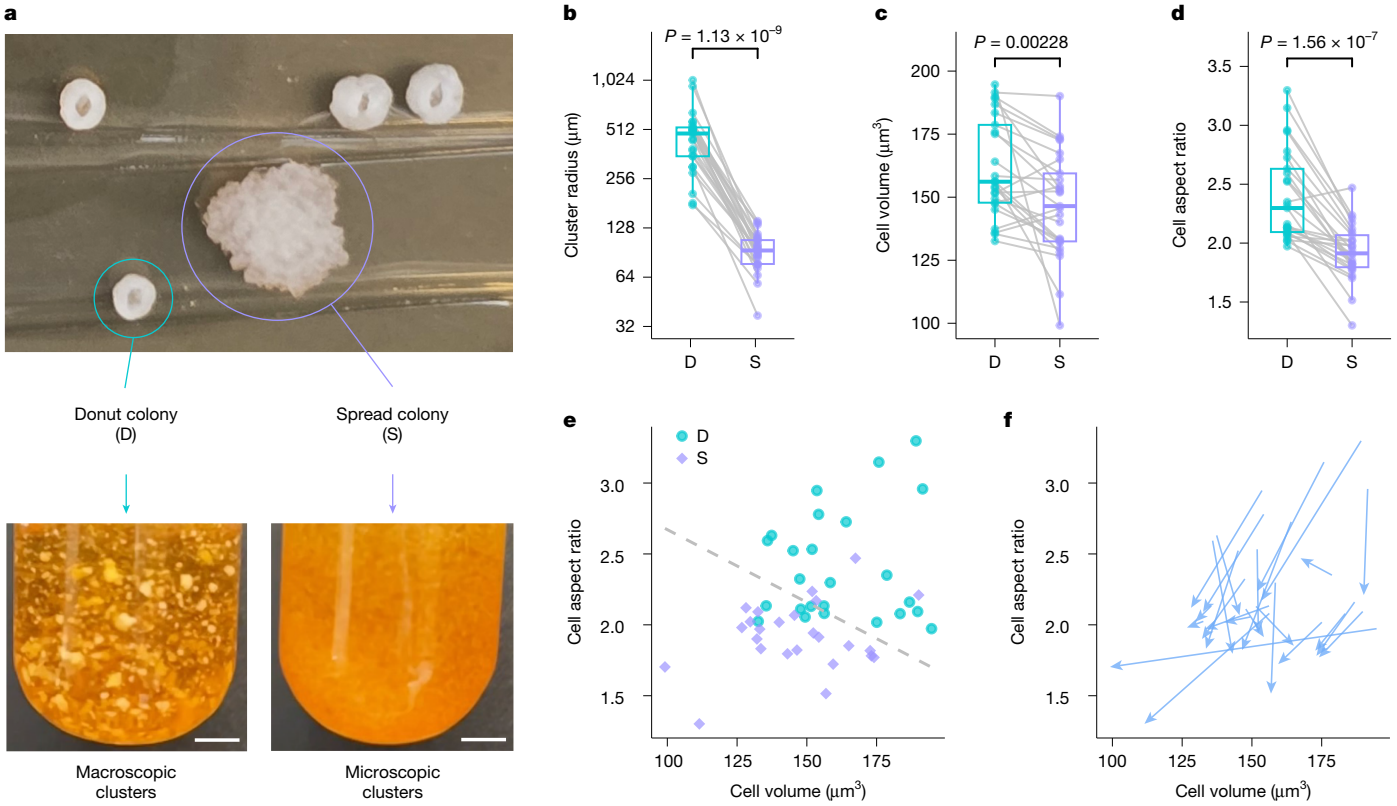


Fig. 4 | Changes in colony morphology allow screening for spontaneous losses of macroscopic multicellularity in PAs. a, Evolved macroscopic isolates (PA3 t1,000 is shown) form 'donut' colonies and occasionally 'spread' colonies that cannot form macroscopic clusters. Scale bars, 5 mm. **b**–**d**, Biomass-weighted mean cluster radius (**b**), mean cell volume (**c**) and mean cell aspect ratio (**d**) of the 25 D–S pairs (each connected by a line) from nine macroscopic isolate backgrounds. The boxes denote the interquartile range (IQR), the centre lines indicate the median, and the whiskers denote the values within 1.5× the IQR of the first and third quartiles. *P* values were calculated by two-tailed paired Student's *t*-test (on average, $n = 605$ clusters (**b**) and 1,294 cells (**c**,**d**) measured per strain). **e**,**f**, Relationship of how cell volume and aspect ratio changed in D–S transitions, with donut and spread strains separated by a linear boundary generated by support vector machine (**e**), and with each D–S transition connected by an arrow pointing from donut to spread (**f**).

Aneuploidy underpins macroscopic size

Polyploidy is associated with much higher rates of aneuploidy than diploidy[40–42]. Nascent polyploids tend to rapidly undergo genome reduction, and the resulting aneuploidy can offer a novel genetic route for rapid adaptation to novel environments[2,25]. However, it remains untested whether polyploids, when maintained, can also adapt by generating aneuploidy. The MuLTEE, with its long-term maintenance of polyploidy, presents a unique system to directly test this hypothesis. Consistent with the instability of tetraploid genomes, we observed cases of aneuploidy (Fig. 2d) and segmental aneuploidy (Extended Data Fig. 4) in the evolved isolates in the MuLTEE. Of note, although PMs were euploid or near-euploid, PAs displayed extensive aneuploidy (Fig. 2d). This distinction is correlated with their divergent multicellular evolution over 1,000 days: whereas PMs remained microscopic, PAs evolved macroscopic size (Fig. 1c). We thus hypothesize that the evolved aneuploidy in PAs might have contributed to their origin of macroscopic multicellular size, arguably the most striking phenotypic innovation that has evolved in the MuLTEE so far.

Parallel changes in aneuploidy and multicellular size in PAs over the MuLTEE suggest a potential link between the two. The level of aneuploidy in PAs generally increased over evolution (linear regression against time for number of aneuploid chromosomes, total number of chromosome copies deviating from euploid tetraploidy, and coefficient of variation of chromosome copy numbers, $P = 0.023$, 0.017 and and 0.020, $r^2 = 0.10$, 0.11 and 0.11, respectively; Fig. 2d), mirroring the trends in cluster size and cell aspect ratio (Fig. 1c,e). However, after the evolution of macroscopic size (which occurred by t600 for all lines except PA3), these complex karyotypes stayed remarkably conserved between t600 and t1,000 in each line (Fig. 2d). The exception, PA5, in which considerably different karyotypes were observed between its t600 and t1,000 isolates (Fig. 2d), also displayed substantially different cluster sizes and cell aspect ratios between them (Fig. 1c,e). These observations suggest that aneuploidy might have had a critical role in the evolution of macroscopic size in PAs.

To test this hypothesis more directly, we used a forward-genetics approach, selecting for losses of macroscopic size in evolved isolates and examining whether karyotype changes are involved. This is enabled by our accidental finding that changes in colony morphology offer a convenient marker for screening spontaneous losses of macroscopic size: all macroscopic PA isolates typically form ring-shaped colonies ('donut' colonies), interspersed with rare, larger, flattened colonies ('spread' colonies) that can no longer form macroscopic clusters when cultured in liquid media (Fig. 4a). We randomly selected three donut colonies from each of the nine macroscopic PA t600 and t1,000 isolates (excluding the still microscopic PA3 t600 isolate) and then isolated one spread colony derived from each donut colony. All spread strains had dramatically reduced cluster size, no longer forming macroscopic clusters (Fig. 4b). Consistent with the previously demonstrated effects of cell-level traits on cluster size, the donut-to-spread transitions were associated with overall decreases in cell volume and, more significantly and consistently, in cell aspect ratio (Fig. 4c–f).

To identify the genetic changes underlying the spontaneous losses of macroscopic size, we sequenced and compared the genomes of each donut–spread pair. We found no strong statistical evidence that point mutation changes had a systematic role in this process (Extended Data

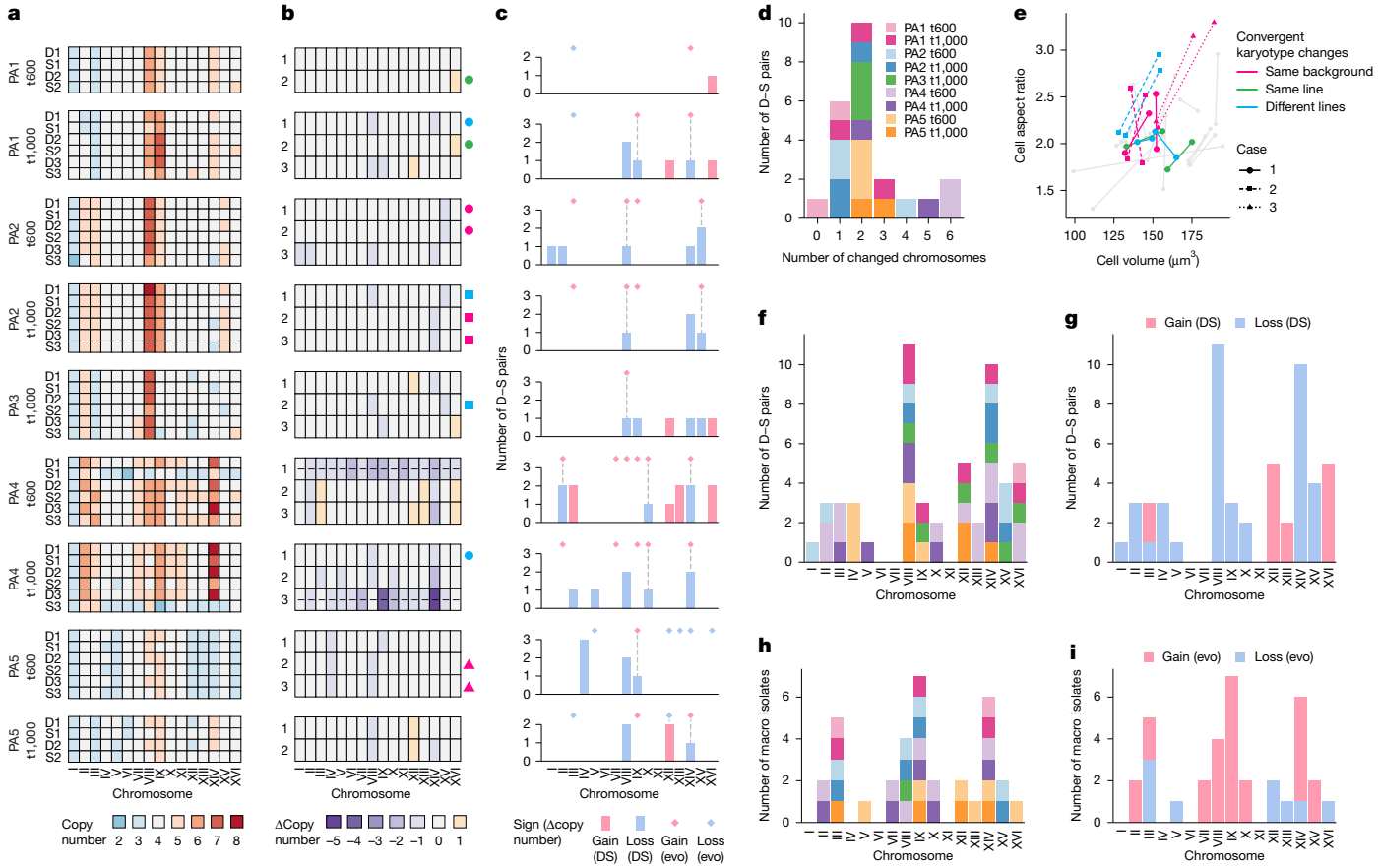


Fig. 5 | Aneuploidy promotes the origin of macroscopic multicellularity in PAs. a,**b**, Karyotypes of D–S pairs (**a**) and karyotype changes in D–S transitions (**b**). Inside the heatmap (**b**), dashed lines indicate two D–S transitions with near-triploidization that were excluded from subsequent analyses. Symbols beside the heatmap (**b**) indicate cases of convergent karyotype changes where two D–S pairs came from the same background (magenta), different backgrounds but the same PA line (green) or different lines (blue), with each case in a different symbol shape. **c**, For each background, chromosomes with copy number changes in the evolution (evo; diamonds) and loss (bars, showing number of occurrences) of macroscopic size, with the diamond and the bar connected by a dashed line when the copy number of a chromosome changes in both the evolution and loss of macroscopic size. **d**, Histogram of the number of chromosomes with copy number changes in each D–S pair, coloured by background. **e**, Changes in cell volume and aspect ratio compared between D–S pairs with convergent karyotype changes, with symbol colour–shape code consistent with panel **b** and line colour–shape code similarly assigned. **f**–**i**, Aggregating all backgrounds in panel **c**, the number of times each chromosome changes its copy number in the loss (**f**,**g**) and evolution (**h**,**i**) of macroscopic size, coloured by background (**f**,**h**; share a colour legend with panel **d**) or direction of change (**g**,**i**).

Fig. 9 and Supplementary Note 1). By contrast, 24 of 25 donut-to-spread transitions were associated with karyotype changes (Fig. 5a,b). Two spread strains in PA4 even lost almost an entire set of chromosomes and partially triploidized (Fig. 5a,b). We excluded these two extreme cases from subsequent analyses, due to their divergence from our otherwise tetraploid background and confounding effect of dramatically decreased DNA content. Karyotype changes in spread strains were mostly limited to copy number changes in one to two chromosomes (Fig. 5d), and among the 53 chromosome copy number changes in total, all but two changed by only one copy (Fig. 5b).

Surprisingly, despite the numerous ways that karyotypes could potentially change, we observed six cases of convergent karyotype changes among just 23 donut–spread pairs ($P = 8.24 \times 10^{-6}$, probability of observing six or more cases with identical karyotype changes, each with changes in no more than two chromosomes, assuming equal probability of gains and losses of each of the 16 chromosomes; Fig. 5b). Three of these six cases even involved donut–spread pairs derived from different genetic backgrounds (either different evolutionary time points in the same line or different lines; Fig. 5b). Most of these convergent karyotype changes were also associated with similar changes in cell volume and aspect ratio (Fig. 5e), suggesting a link between karyotype changes and phenotypic changes. Moreover, we also observed convergent changes in the copy numbers of certain chromosomes across donut-to-spread transitions (Fig. 5f), and strikingly, each chromosome (except for chromosome III) that underwent copy number changes always changed in the same direction across all spread strains (that is, a chromosome was either always gained or always lost; Fig. 5g). This is most clearly reflected in chromosomes VIII and XIV, which were convergently lost in almost half of all spread strains involving all five PA lines ($P = 6.28 \times 10^{-7}$ and 5.05×10^{-6}, respectively, binomial probability of observing 11 or more, or 10 or more losses of any given chromosome among 53 chromosome copy number changes, assuming equal probability of gains and losses of each of the 16 chromosomes; Fig. 5f,g). Together, the prevalent and convergent changes at the karyotype and chromosome level during donut-to-spread transitions, as well as their correlations with phenotypic changes, strongly suggest that changes in aneuploidy underpin the spontaneous losses of macroscopic size.

To investigate the role of aneuploidy in the evolution of macroscopic size in the MuLTEE, we compared the chromosome copy number changes associated with the emergence of macroscopic size during the MuLTEE to those involved in the spontaneous loss of macroscopic size in the donut-to-spread transitions. We first identified the candidate chromosomes that are directly involved in the evolution of macroscopic size in each line, by identifying the chromosomes with copy number

changes that are absent in any of the microscopic isolates (t200 and t400, and for PA3, also t600) but present in at least one of the macroscopic isolates (t600 and t1,000). This yields one set of candidate chromosomes for each of the nine macroscopic isolates (Fig. 5c, shown as diamonds). Of note, for 13 of 18 candidate chromosomes (exceptions are mostly within PA5), the evolutionary changes of their copy numbers observed in the t600 isolate were maintained in the t1,000 isolate of the same line (Fig. 5c). Among the 53 chromosome copy number changes in all donut-to-spread transitions (Fig. 5c, shown as bars), 20 occurred in the candidate chromosomes of the corresponding isolate background (Fig. 5c, indicated by lines connecting diamonds and bars). Of these, all changed in the opposite direction to their evolutionary changes in the MuLTEE (Fig. 5c, contrasting colours between diamonds and bars connected by lines), suggesting that copy number changes of these candidate chromosomes contributed to the evolution of macroscopic size. If these changes were due to intrinsic mutational biases, we would expect that they would show consistent directionality. Instead, their opposing patterns during the evolution of macroscopic size in the MuLTEE versus the donut-to-spread transitions suggest that they were the result of selection.

Furthermore, when comparing across the five PA lines, we found that some candidate chromosomes and their directions of copy number changes over evolution are shared among multiple lines (for example, chromosomes VIII, IX and XIV), whereas the other candidate chromosomes are more line specific (Fig. 5h,i). In the donut-to-spread transitions, all the chromosomes (candidate or not) that changed copy numbers in more than one spread strain changed consistently in the opposite direction to their evolutionary changes (comparing Fig. 5g and Fig. 5i). As the copy numbers of some candidate chromosomes can rapidly change in the donut-to-spread transitions, their maintenance from t600 to t1,000 isolates in the MuLTEE suggest that, just like polyploidy, at least some part of the evolved aneuploidy was maintained by sustained selection for larger multicellular size despite its intrinsic instability.

Discussion

The MuLTEE was initially designed to study the open-ended evolution of increasingly complex multicellular life. Our results reveal that the MuLTEE also serves as a unique polyploidy long-term evolution experiment, providing direct empirical insights into the short-term and long-term evolutionary dynamics of nascent WGDs. We found that tetraploidy evolved in our system due to its immediate phenotypic effects, producing larger, longer cells that give rise to larger multicellular clusters, which is adaptive under our selection regime. Despite its intrinsic genomic instability, tetraploidy was maintained over thousands of generations by sustained selection for larger organismal size. Furthermore, we discovered that extensive, convergently evolving and evolutionarily conserved aneuploidy had a key role in the subsequent evolution of macroscopic multicellularity in our model system.

Our results have several important implications for understanding the evolution of WGD. First, we show that nascent polyploidy is not necessarily transient, even before the evolution of genome-stabilizing mechanisms. WGD has long been known to bring immediate phenotypic changes, many stemming from increased cell size (a universal feature of polyploid cells and central to evolved polyploids in the MuLTEE), which can confer fitness benefits under novel environments[12,17,18,20,21]. Here we have experimentally demonstrated that selection favouring the immediate phenotypic effects of WGD is sufficient to drive both its origin and maintenance, despite the intrinsic instability of polyploid genomes. As a result, the duration of the environment favouring polyploidy may be a key determinant of whether a nascent WGD will be transient or persistent. This may also explain the increased abundance of polyploids in certain environments[14,43] (for example, tetraploid yeast in bakeries[44]). Unlike cases in which genome-stabilizing mechanisms evolve, the polyploid genomes actively maintained under environmental selection can still be intrinsically unstable, rapidly reverting towards diploidy in the absence of the polyploidy-favouring selection.

Building on previous works showing that WGD can profoundly influence evolutionary dynamics by altering the rate, spectrum and fitness effect of mutations[2,14,15,25], our experiment is the first to directly examine how recently duplicated genomes evolve when polyploidy is maintained over long evolutionary timescales. In the MuLTEE, we found that the point mutations arising in the polyploid background are highly heterozygous. Some mutations can reach high allele frequencies (more than 0.75) within individual genomes, which may be the result of positive selection given evidence of extensive gene conversion in our system (Supplementary Note 1). We also found that polyploidy, even when maintained, can still exhibit genomic instability and generate aneuploidy, exploring novel karyotypic space and facilitating long-term adaptation. Moreover, our findings challenge a previous notion that aneuploidy necessarily serves as a temporary solution to novel environments[40,45]. Instead, we demonstrate that even complex aneuploid karyotypes can be maintained for thousands of generations, despite the inherent instability of aneuploid genomes. The polyploid backgrounds in our system may also facilitate the maintenance of aneuploidy by better tolerating its potential costs through reduced gene dosage effects[2,3]. Our observation also offers an alternative explanation for the aneuploidy frequently found in natural polyploids[40]: rather than being a transient, non-adaptive consequence of genomic instability, in some cases they may be the result of selection, preserved over long timescales when the selective environment remains stable.

Although previous work has revealed the adaptive roles of aneuploidy in promoting stress resistance and virulence[40,41], our results show that aneuploidy can also contribute to the evolution of morphological innovations (that is, macroscopic multicellularity), expanding its functional repertoire. Further work will be required to explore the mechanistic basis of how aneuploidy drives this phenotypic change. Possible mechanisms may include gene dosage effects and epistatic interactions between structural variations and point mutations. From an evolutionary perspective, it will be interesting to examine whether the divergent karyotypes evolved and maintained in different PA lines result from historical contingency and represent local fitness optima in the karyotypic space. In addition, it is worth investigating why the aneuploidy in PMs is limited. PMs do not appear to be constrained, however, in their intrinsic capacity of generating aneuploidy, as most PM t1,000 isolates underwent rapid ploidy reduction in our counter-selection and mutation accumulation experiments (Fig. 3i,j). The near-euploidy in PMs may instead be because the evolution of increased multicellular size is constrained under this condition[8,30], and aneuploidy is selected against if its benefits do not outweigh its costs[46].

The MuLTEE is now the longest-running evolution experiment not just in the evolution of multicellularity, but also in the evolution of polyploidy, offering direct empirical insights into how WGDs can arise, persist and drive short-term and long-term adaptation. We anticipate that this open-ended, multi-condition, long-term evolution experiment will continue to serve as a source of inspiration and a testbed for various hypotheses regarding WGD, a fundamental process in eukaryotic evolution.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-025-08689-6.

1. Otto, S. P. The evolutionary consequences of polyploidy. *Cell* **131**, 452–462 (2007).
2. Selmecki, A. M. et al. Polyploidy can drive rapid adaptation in yeast. *Nature* **519**, 349–352 (2015).

3. Van de Peer, Y., Mizrachi, E. & Marchal, K. The evolutionary significance of polyploidy. *Nat. Rev. Genet.* **18**, 411–424 (2017).
4. Vittoria, M. A., Quinton, R. J. & Ganem, N. J. Whole-genome doubling in tissues and tumors. *Trends Genet.* **39**, 954–967 (2023).
5. Comai, L. The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* **6**, 836–846 (2005).
6. Storchova, Z. et al. Genome-wide genetic analysis of polyploidy in yeast. *Nature* **443**, 541–547 (2006).
7. Buggs, R. J. et al. Rapid, repeated, and clustered loss of duplicate genes in allopolyploid plant populations of independent origin. *Curr. Biol.* **22**, 248–252 (2012).
8. Bozdag, G. O. et al. De novo evolution of macroscopic multicellularity. *Nature* **617**, 747–754 (2023).
9. Fox, D. T., Soltis, D. E., Soltis, P. S., Ashman, T. L. & Van de Peer, Y. Polyploidy: a biological force from cells to ecosystems. *Trends Cell Biol.* **30**, 688–694 (2020).
10. Mayrose, I. et al. Recently formed polyploid plants diversify at lower rates. *Science* **333**, 1257 (2011).
11. Levin, D. A. Why polyploid exceptionalism is not accompanied by reduced extinction rates. *Plant Syst. Evol.* **305**, 1–11 (2018).
12. Clo, J. & Kolar, F. Short- and long-term consequences of genome doubling: a meta-analysis. *Am. J. Bot.* **108**, 2315–2322 (2021).
13. Gerstein, A. C., Chun, H. J., Grant, A. & Otto, S. P. Genomic convergence toward diploidy in *Saccharomyces cerevisiae*. *PLoS Genet.* **2**, e145 (2006).
14. Gerstein, A. C. & Sharp, N. P. The population genetics of ploidy change in unicellular fungi. *FEMS Microbiol. Rev.* **45**, fuab006 (2021).
15. Todd, R. T., Forche, A. & Selmecki, A. Ploidy variation in fungi: polyploidy, aneuploidy, and genome evolution. *Microbiol. Spectr.* **5**, 5–4 (2017).
16. Lu, Y. J., Swamy, K. B. & Leu, J. Y. Experimental evolution reveals interplay between Sch9 and polyploid stability in yeast. *PLoS Genet.* **12**, e1006409 (2016).
17. Bomblies, K. When everything changes at once: finding a new normal after genome duplication. *Proc. Biol. Sci.* **287**, 20202154 (2020).
18. Doyle, J. J. & Coate, J. E. Polyploidy, the nucleotype, and novelty: the impact of genome doubling on the biology of the cell. *Int. J. Plant Sci.* **180**, 1–52 (2019).
19. Levin, D. A. Polyploidy and novelty in flowering plants. *Am. Nat.* **122**, 1–25 (1983).
20. Mortier, F. et al. Understanding polyploid establishment: temporary persistence or stable coexistence? *Oikos* https://doi.org/10.1111/oik.09929 (2024).
21. Van de Peer, Y., Ashman, T. L., Soltis, P. S. & Soltis, D. E. Polyploidy: an evolutionary and ecological force in stressful times. *Plant Cell* **33**, 11–26 (2021).
22. Storchova, Z. Ploidy changes and genome stability in yeast. *Yeast* **31**, 421–430 (2014).
23. Gerstein, A. C. et al. Polyploid titan cells produce haploid and aneuploid progeny to promote stress adaptation. *mBio* **6**, e01340-15 (2015).
24. Hirakawa, M. P., Chyou, D. E., Huang, D., Slan, A. R. & Bennett, R. J. Parasex generates phenotypic diversity de novo and impacts drug resistance and virulence in *Candida albicans*. *Genetics* **207**, 1195–1211 (2017).
25. Scott, A. L., Richmond, P. A., Dowell, R. D. & Selmecki, A. M. The influence of polyploidy on the evolution of yeast grown in a sub-optimal carbon source. *Mol. Biol. Evol.* **34**, 2690–2703 (2017).
26. Bielski, C. M. et al. Genome doubling shapes the evolution and prognosis of advanced cancers. *Nat. Genet.* **50**, 1189–1195 (2018).
27. Fujiwara, T. et al. Cytokinesis failure generating tetraploids promotes tumorigenesis in p53-null cells. *Nature* **437**, 1043–1047 (2005).
28. Ratcliff, W. C., Fankhauser, J. D., Rogers, D. W., Greig, D. & Travisano, M. Origins of multicellular evolvability in snowflake yeast. *Nat. Commun.* **6**, 6102 (2015).
29. Jacobeen, S. et al. Cellular packing, mechanical stress and the evolution of multicellularity. *Nat. Phys.* **14**, 286–290 (2018).
30. Bozdag, G. O., Libby, E., Pineau, R., Reinhard, C. T. & Ratcliff, W. C. Oxygen suppression of macroscopic multicellularity. *Nat. Commun.* **12**, 2838 (2021).
31. Day, T. C. et al. Morphological entanglement in living systems. *Phys. Rev.* **14**, 011008 (2024).
32. Day, T. C. et al. Cellular organization in lab-evolved and extant multicellular species obeys a maximum entropy law. *eLife* **11**, e72707 (2022).
33. Harrison, B. D. et al. A tetraploid intermediate precedes aneuploid formation in yeasts exposed to fluconazole. *PLoS Biol.* **12**, e1001815 (2014).
34. Ramsey, J. S. & Schemske, D. W. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Ann. Rev. Ecol. Syst.* **29**, 467–501 (1998).
35. Gerstein, A. C., McBride, R. M. & Otto, S. P. Ploidy reduction in *Saccharomyces cerevisiae*. *Biol. Lett.* **4**, 91–94 (2008).
36. Voordeckers, K. et al. Adaptation to high ethanol reveals complex evolutionary pathways. *PLoS Genet.* **11**, e1005635 (2015).
37. Galitski, T., Saldanha, A. J., Styles, C. A., Lander, E. S. & Fink, G. R. Ploidy regulation of gene expression. *Science* **285**, 251–254 (1999).
38. Rebolleda-Gomez, M. & Travisano, M. The cost of being big: local competition, importance of dispersal, and experimental evolution of reversal to unicellularity. *Am. Nat.* **192**, 731–744 (2018).
39. Treinin, M. S. G. Mitochondrial activity is required for the expression of IME1, a regulator of meiosis in yeast. *Curr. Genet.* **23**, 223–227 (1993).
40. Gilchrist, C. & Stelkens, R. Aneuploidy in yeast: segregation error or adaptation mechanism? *Yeast* **36**, 525–539 (2019).
41. Vande Zande, P., Zhou, X. & Selmecki, A. The dynamic fungal genome: polyploidy, aneuploidy and copy number variation in response to stress. *Annu. Rev. Microbiol.* **77**, 341–361 (2023).
42. O'Donnell, S. et al. Telomere-to-telomere assemblies of 142 strains characterize the genome structural landscape in *Saccharomyces cerevisiae*. *Nat. Genet.* **55**, 1390–1399 (2023).
43. De Chiara, M. et al. Domestication reprogrammed the budding yeast life cycle. *Nat. Ecol. Evol.* **6**, 448–460 (2022).
44. Albertin, W. et al. Evidence for autotetraploidy associated with reproductive isolation in *Saccharomyces cerevisiae*: towards a new domesticated species. *J. Evol. Biol.* **22**, 2157–2170 (2009).
45. Yona, A. H. et al. Chromosomal duplication is a transient evolutionary solution to stress. *Proc. Natl Acad. Sci. USA* **109**, 21010–21015 (2012).
46. Li, R. & Zhu, J. Effects of aneuploidy on cell behaviour and function. *Nat. Rev. Mol. Cell Biol.* **23**, 250–265 (2022).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2025

Methods

MuLTEE

The details of how we constructed the ancestor strains and conducted experimental evolution with settling selection up to 600 days have been previously described[8]. In brief, we constructed a multicellular diploid *S. cerevisiae* strain from a unicellular diploid Y55 strain by *ACE2* deletion. From this grande strain, which grows both aerobically and anaerobically in YPD media (1% yeast extract, 2% peptone and 2% dextrose), we isolated a spontaneous petite mutant that grows anaerobically in YPD media even in the presence of oxygen. We confirmed the anaerobic phenotype of the petite mutant by showing that it cannot grow on glycerol and does not consume oxygen during growth using direct oxygen measurements. We evolved five replicate populations of mixotrophic and anaerobic snowflake yeast (referred to as PM1–5 and PA1–5, respectively) with daily selection for increased size over 1,000 days. Every day, we grew each population in 10 ml of YPD media at 30 °C with shaking at 250 rpm for 24 h, and then we transferred 1.5 ml of the culture into a 1.5-ml Eppendorf tube, let clusters settle on the bench for 3 min and transferred the bottom 50 µl into 10 ml of fresh YPD media for the next day of growth. Once PA1–5 evolved macroscopic clusters that settle rapidly, we used wide-bore pipette tips to minimize breaking clusters during pipetting, and we also decreased the settling time to 30 s to allow sustained selection for increased size, with this change occurring on approximately 350 days for PA2 and PA5 and approximately 500 days for PA1, PA3 and PA4. For simplicity, starting from day 850, we sampled 1 ml, instead of 1.5 ml, from each culture for daily settling selection. We archived a glycerol stock for each population every 10–15 days. We extracted one representative clonal isolate from each of the ten populations archived on days 200, 400, 600 and 1,000 for subsequent analyses.

Measuring cluster size

We revived strains from glycerol stocks by growing them on YPD plates at 30 °C for 2 days. Then, we inoculated each strain into 10 ml of YPD media and grew them at 30 °C with 250 rpm shaking for 3 days with daily settling selection before transferring to fresh media, recapitulating how the strains grow during the evolution experiment. On the last day, after settling selection and transfer, we grew the cultures for 24 h and sampled them at 4 h (exponential phase) and 24 h (stationary phase) for measuring cluster size. Unless otherwise noted, the 24-h measurements are used throughout the paper, as they represent the states of the cultures right before settling selection.

Before imaging, we gently shook each culture by hand (without vortexing, which may break clusters) and added an appropriate volume of the culture (1–250 µl, depending on cluster density) into H_2O containing 10 µl of 16% (w/v) formaldehyde (28906, Thermo Scientific) in a 24-well plate, making up 510 µl per well. For macroscopic strains, we sampled the culture using manually cut wide-bore 1-ml tips to avoid breaking macroscopic clusters during pipetting, and we typically sampled twice from each culture and transferred them into two wells to reduce the randomness in sampling macroscopic clusters. We gently shook the 24-well plate to evenly spread out the clusters and allowed 5–10 min for clusters to settle down.

We used a Nikon Eclipse Ti inverted microscope (NIS-Elements v4.30.01) to scan the whole wells by taking and stitching 5 × 8 brightfield, shading-corrected images with 10% overlap at ×4 magnification, scanning at 1 mm s^{-1}. We developed a semi-automated image analysis pipeline using ImageJ (v1.54f) to (1) segment clusters using auto-local thresholding (which allows detecting clusters ranging from tiny branches to macroscopic clusters) and split touching clusters using seed-based watershed, (2) perform manual correction using a custom, user-friendly toolkit for improving the speed and reproducibility of manual correction, and (3) measure the cross-sectional area of each cluster. We removed segmented objects with an area below 40 µm^2. Using R (v4.1.2), we converted cluster areas to cluster volumes and radiuses, treating clusters as perfect spheres. For each strain, we calculated its biomass-weighted mean cluster radius[8] by first calculating its mean cluster volume weighted by cluster volume and then converting it to radius.

Measuring cell volume and aspect ratio

We revived strains from glycerol stocks by growing them on YPD plates at 30 °C for 2 days. Then, we inoculated each strain into 10 ml of YPD media and grew them at 30 °C with 250 rpm shaking for 3 days with daily settling selection before transferring to fresh media. On the last day, we transferred 100 µl of culture (without settling selection) to 10 ml of fresh media and grew it for 12 h, following Bozdag et al.[8,30].

To prepare samples for imaging, we transferred 25 µl of each culture to a 1.5-ml microcentrifuge tube. For macroscopic strains, we sampled the culture using 100-µl wide-bore tips to allow pipetting macroscopic clusters, and we broke clusters into tiny pieces by pipetting with 100-µl regular-bore tips to facilitate crushing them into a single layer of cells later. We pelleted the clusters by spinning at 5,000*g* for 1 min, washed them with 1 ml of 1× PBS, and incubated them in 100 µl of 5 µM Calcofluor White (a blue fluorescent cell wall stain) in 1× PBS at room temperature. We gently crushed 5 µl of stained clusters (without shearing, which can lyse cells) into a single layer of cells between a microscope slide and a coverslip.

We used a Nikon Eclipse Ti inverted microscope (NIS-Elements v4.30.01) to take brightfield and fluorescent (UV channel) images of at least five fields of view (FOVs) per strain at ×40 magnification. For each FOV, we set the focus based on the brightfield channel by manually moving to a *z*-plane where cells look grey (rather than brighter or darker) and then moving down by 0.7 µm to get sharp fluorescent signals of the cell wall. We developed a semi-automated image analysis pipeline using Cellpose (v2.2.2)[47] and ImageJ (v1.54f) to perform (1) automated cell segmentation using Cellpose (specify cell diameter as 80 pixels for images with 0.073 µm per pixel), run on GPUs provided by Georgia Tech's Phoenix Cluster, (2) manual correction using Cellpose's Graphical User Interface (GUI), and (3) measurement of the area, lengths of major and minor axis, aspect ratio, and solidity of each cell in ImageJ. We removed segmented objects with an area below 1 µm^2 or a solidity below 0.8. Using R (v4.1.2), we calculated cell volumes using the formula $V = 4/3 \times \pi ab^2$ (treating yeast cells as perfect ellipsoids), where a and b are the lengths of the major and minor axis of each cell, respectively.

Biophysical simulation

We adapted our previously published biophysical model[8,32] to disentangle and elucidate the effects of cell volume and cell aspect ratio on cluster size. In our model, we added new cells (modelled as prolate ellipsoids) in the characteristic snowflake yeast budding pattern until a certain total amount of overlap between cells is reached (representing the threshold for cluster fragmentation), whereupon the simulation is terminated. Using this model, we swept through a range of overlap thresholds to find a value that recapitulated our empirical data. Then, using that value, we swept through a range of cell volumes (from 25 to 250 µm^3) and aspect ratios (from 1 to 2). We simulated 50 clusters at each pair of parameter values, and measured cluster volume (estimated by the volume of the convex hull bounding the cluster), cell number and packing fraction of each simulated cluster upon fragmentation.

Whole-genome sequencing and analysis

To sequence yeast genomes, we first extracted yeast genomic DNA using VWR's DNA purification kit. Genomic library preparation using the Illumina DNA Prep kit and sequencing using the Illumina NovaSeq 6000 and NovaSeq X Plus sequencer with 150-bp paired-end reads were performed by the Microbial Sequencing Center at the University of Pittsburgh. We obtained FASTQ files of genome reads at an average coverage of 160X.

To analyse the genome reads, we first filtered and trimmed low-quality reads using Trimmomatic (v0.39). Next, we aligned the reads to the S288C reference genome[48] (R64-3-1 build) using BWA-MEM (v0.7.17)[49]. Following GATK's Best Practices pipeline[50], we generated binary alignment files (BAM) and marked duplicates using Picard Toolkit (v2.27.5). Next, we used GATK Haplotypecaller (v4.2.4.1) to call for variants. To filter out variants with low genotype quality, we used VCFTOOLS (v0.1.16)[51]. Next, we extracted novel mutations that are only present in the evolved genomes by comparing them against the ancestral genome via bcftools-isec (v1.10)[52]. Finally, we annotated novel mutations using snpEff (v5.0)[53] (including assigning mutation impacts as high, moderate, low or modifier) and calculated their allele frequencies by dividing the read depth of the mutant allele with the total read depth in each particular position.

To determine the karyotype of each strain, we first computed the read depth of each base position (base coverage) in the genome using Bamtools Stats (v2.5.1) and then estimated chromosome copy numbers in R (v4.1.2). Specifically, we calculated the mean base coverage in each 1-kb non-overlapping bin along each chromosome (bin coverage), and then normalized the bin coverages in each genome by dividing it with the median bin coverage in the whole genome. Next, we calculated bin copy number by multiplying the normalized bin coverage with the baseline ploidy level of the strain (determined by imaging-based ploidy measurement below). Finally, we estimated the copy number of each chromosome as its rounded median bin copy number.

For donut and spread strains, we also estimated mutation allele copy number in R (v4.1.2), using mutation allele frequency and chromosome copy number calculated above. For a mutation with its corresponding chromosome having copy number N, its allele copy number would be 0 or N if its allele frequency is 0 or 1, respectively. For other values of allele frequency, we estimated the mutation allele copy number (C) as the integer between 1 and $N-1$ where C/N is the closest to the allele frequency. In cases of ties (observed in 7 of 2,003 mutations in 46 donut and spread strains), we chose the allele copy number that has the least change from the other strain in the corresponding donut–spread pair. We also used C/N as the corrected allele frequency.

To calculate the distribution of mutation impacts of a point mutation randomly introduced into the yeast genome (S288C reference genome, R64-3-1 build), we only considered single-nucleotide substitutions, as multi-nucleotide substitutions and indels are much rarer and were not observed in the mutations that were gained in donut-to-spread transitions. We calculated the probability of the random mutation having high, moderate, low or modifier impact as the probability of it being a nonsense, missense, synonymous or intronic, or non-coding/intronic mutation.

Yeast strain construction

To construct isogenic grande diploid and tetraploid snowflake yeast, we started with two *S. cerevisiae* grande haploid unicellular strains (*MATa hoΔ::hphNT1* and *MATα hoΔ::hphNT1*), derived from the Y55 background (from which the MuLTEE ancestors were also derived). We first deleted the *ACE2* open reading frame from these two haploid strains using a *kanMX* cassette, which made them multicellular, and then mated them to form a grande diploid strain (a/α). Next, we transformed this diploid strain with a pGAL1-HO-tADH1-natNT2 plasmid containing *CEN/ARS* elements, and then transiently induced *HO* expression by galactose to switch the mating type from a/α to a/a and α/α (determined by a halo assay), followed by selection for plasmid loss. We then mated the a/a and α/α diploid strains to form a grande tetraploid strain (a/a/α/α).

To generate petite diploid and tetraploid snowflake yeast, we isolated spontaneous petite mutants from the grande diploid and tetraploid strains, respectively, by selecting colonies that are smaller than normal colonies on a YPD plate (1% yeast extract, 2% peptone, 2% dextrose and 1.5% agar) and cannot grow on a YP-glycerol plate (the same as the YPD plate but with the dextrose replaced by 2.5% glycerol). As petite mutation is not isogenic, we isolated four independent petite mutants per strain, each of which is derived from a different single colony of the parental grande strain.

Imaging-based ploidy measurement of snowflake yeast

Common protocols for ploidy measurement of unicellular yeast are based on flow cytometry of DNA-stained single cells, in which fluorescent intensity of the DNA stain such as propidium iodide scales linearly with DNA content[15,54]. However, these protocols cannot be readily applied to multicellular yeast without an efficient method for dissociating clusters into single cells, so we developed an imaging-based protocol for measuring the ploidy level of snowflake yeast strains. Basically, we used fluorescent images of propidium iodide-stained, flattened clusters to quantify the propidium iodide intensity of G1-phase nuclei, which was compared with strains with known ploidy to estimate the DNA content of the focal strain.

We adapted the sample preparation procedure before fluorescent imaging from Todd et al.[54]. In every experiment, two control strains with known ploidy, namely, the engineered grande diploid and tetraploid snowflake yeast, also went through the same procedure of sample preparation and imaging. We first grew each strain of interest in YPD media to mid-log phase, and then transferred 250 µl of the culture into a 1.5-ml microcentrifuge tube. We pelleted the clusters by spinning at 5,000*g* for 1 min and washed them with 1 ml of H_2O. For macroscopic strains, we broke the clusters into small pieces by pipetting vigorously using 100-µl regular-bore tips before washing. We fixed and permeabilized the cells in 1 ml of 70% ethanol at room temperature for 2 h with end-to-end rotation on a mini-rotator (Bio RS-24, BioSan) at the maximum speed, followed by washing with 1 ml of 50 mM sodium citrate twice. We resuspended the clusters in 200 µl of 50 mM sodium citrate containing 0.5 mg ml^{-1} RNase A (101076, MP Biomedicals) and incubated them in a 37 °C heat block for 2 h with gentle inversion every 30 min (as clusters settle over time). After RNA digestion, we added 5 µl of 1 mg ml^{-1} propidium iodide (P1304MP, Thermo Fisher) and incubated the mixture in a 30 °C incubator in dark overnight with rotation on a mini-rotator (Bio RS-24, BioSan) at the minimum speed to keep clusters from settling. Propidium iodide-stained clusters can be stored at 4 °C for no longer than 1 week before imaging.

For fluorescent imaging, we crushed 5 µl of propidium iodide-stained clusters into a single layer of cells between a microscope slide and a coverslip. For each sample, we took 14-bit images of approximately 10 FOVs at ×20 magnification using a Nikon Eclipse Ti inverted microscope (NIS-Elements v4.30.01). For each FOV, we set the focus by manually moving to a *z*-plane in which cells look grey (rather than brighter or darker) in the brightfield channel and then moving down by 1 µm to get sharp fluorescent signals of the nuclei. We imaged the flattened clusters in the red fluorescent channel (exposure of 600 ms and gain of 2.2×) at the focal plane as well as one *z*-plane 0.3 µm above and below the focal plane (three *z*-planes in total) to detect the propidium iodide-stained nuclei, and then imaged in the brightfield channel (exposure of 100 ms and gain of 4.1×) at the focal plane. We set the exposure and gain of the fluorescent channel such that the brightest pixels in the tetraploid control strain are approximately 80% of the maximal allowed pixel value while minimizing photobleaching and noise.

We performed quantitative image analysis using ImageJ (v1.54f). We first performed maximum intensity projection of the three *z*-planes taken in the fluorescent channel and used the resulting image for segmentation and fluorescence quantification. We segmented nuclei and filtered them to include only single round nuclei. We measured the total propidium iodide fluorescence intensity of each nucleus with background subtraction, in which the background fluorescence is the median propidium iodide fluorescence intensity of the cytoplasm in the cluster that the focal nucleus is in. The resulting nuclear propidium iodide intensity scales linearly with the DNA content.

We analysed the image analysis results in R (v4.1.2). For each sample, we removed tiny segmented objects with areas two median absolute deviations below median (which are nucleus segmentation artefacts), and we manually removed FOVs with outlier distributions of nuclear propidium iodide intensity. The distribution of nuclear propidium iodide intensity in a clonal strain contains two peaks that correspond to G1-phase and G2-phase cells. We estimated the DNA content of each nucleus by dividing its propidium iodide intensity with the propidium iodide intensity of a haploid genome, which was estimated by averaging across the two ploidy control strains, that is, mean(G1 peak intensity of diploid control strain/2, G1 peak intensity of tetraploid control strain/4). We estimated the DNA content of a clonal strain as the DNA content of its G1 peak.

Competition assay

To perform competition assays between the engineered diploid and tetraploid snowflake yeast without fluorescent labelling (which may incur fitness cost), we developed an imaging-based method for distinguishing clusters with different ploidy levels, utilizing the fact that tetraploid clusters contain larger cells than diploid counterparts. Specifically, before imaging, we transferred an appropriate volume (12–18 µl, depending on cluster density) of the culture containing diploid and tetraploid clusters into a 24-well plate with 500 µl of H_2O per well. We gently shook the 24-well plate to evenly spread out the clusters and allowed 5–10 min for clusters to settle down. To image the centre of each well, we used a Nikon Eclipse Ti inverted microscope (NIS-Elements v4.30.01) to take and stitch 6 × 6 brightfield, shading-corrected images with 5% overlap at ×20 magnification, scanning at 1 mm s^{-1}. We imaged at three *z*-planes, including the *z*-plane where the cells touching the well bottom look slightly bright, as well as 5 µm and 10 µm above it. Using ImageJ (v1.53q), we first segmented the clusters using the middle *z*-plane (with manual correction), and then we segmented the bright cells in the cluster edges in all three *z*-planes, thus capturing cells in different focal planes. Next, for each cluster, we calculated the mean area of the five largest cells detected across all three *z*-planes. This value is sufficient to distinguish the engineered diploid and tetraploid clusters in either grande or petite form, with the cut-off determined in a preliminary test.

To compete the engineered diploid and tetraploid snowflake yeast under mixotrophic and anaerobic conditions, we competed grande diploid versus grande tetraploid with four biological replicates, and we paired the four diploid and tetraploid petite mutants to form four competing pairs as our four replicates. We first revived the strains from glycerol stocks by growing them on YPD plates at 30 °C for 2 days. Then, we inoculated four replicate tubes for each grande strain (diploid and tetraploid) as well as one replicate tube for each petite mutant (diploid and tetraploid), and we grew them in 10 ml of YPD media at 30 °C with 250 rpm shaking for 2 days with daily 1:100 dilution. To start the competition of each of the eight competing pairs, we mixed 500 µl of the 24-h culture of each competing strain (1 ml in total), from which we inoculated 100 µl into 10 ml of YPD media in each of two culture tubes and grew them for 3 days, transferring daily with and without settling selection, respectively. For competitions with settling selection, we recapitulated the settling selection scheme in the MuLTEE: every day, we transferred 1.5 ml of each 24-h culture into a 1.5-ml microcentrifuge tube, let clusters settle for 3 min and transferred the bottom 50 µl into 10 ml of YPD media to grow for another day. For competitions without settling selection, we grew each culture with daily 1:100 dilution. We measured the ratio of the total cluster volume in each competing strain at the start and the end of the 3-day competition, using the above imaging-based method to distinguish diploid and tetraploid clusters. Finally, we calculated the per-day selection rate of tetraploid versus diploid strain, using the formula r = [log(% of tetraploids on day 3/% of tetraploids on day 0) − log(% of diploids on day 3/% of diploids on day 0)]/3 days.

Evolution experiment with selection against larger size

We revived PM/PA t0 and PM/PA1–5 t1,000 isolates from glycerol stocks by growing them on YPD plates at 30 °C for 2 days. Then, we inoculated each of the 12 strains into 10 ml of YPD media and grew them at 30 °C with 250 rpm shaking overnight, from which we transferred 20 µl to inoculate each of the four replicate populations for the evolution experiment. We grew the 48 populations in a 30 °C static incubator for 70 days (approximately 500 generations) in a spatially structured environment using 24-well plates containing 2 ml of YPD agar per well. Every 24 h, we resuspended each population in 1 ml of saline solution (0.85% NaCl), pre-diluted it in a microcentrifuge tube, and then transferred 20 µl of the diluted resuspension onto fresh YPD agar and spread it out by gently rocking the 24-well plate. We initially used a pre-dilution factor of 1:2 (for a total dilution of 1:100 after accounting for the volume plated), but we increased it to 1:4 (total 1:200) on day 36 after the populations had adapted to this selection regime. We prepared glycerol stocks of each population every 7 days by mixing 800 µl of the undiluted saline resuspension with 400 µl of 70% glycerol and stored them at −80 °C. Before routine protocols of measuring ploidy level and cluster size, we revived the populations by scraping a big chunk from the glycerol stocks and growing them in 10 ml of YPD to saturation for 2 days (without transferring after the first 24 h).

Mutation accumulation experiment

We first streaked the strains from glycerol stocks onto YPD plates. Every 2 days of growth on YPD plates at 30 °C, we picked the colony farthest from the initial streak point and streaked it onto a new YPD plate. This method ensures that we were unbiased in colony selection and that we were sampling colonies founded by small snowflake yeast branches that each descended from a single cell[28], providing a unicellular genetic bottleneck at each transfer. We ran the mutation accumulation experiment for 28 transfers (56 days) and prepared glycerol stocks for the evolved populations every seven transfers. In this experiment, replicate population B of PA t0 displayed greatly reduced cell size that made it difficult to resolve the nuclei during ploidy measurements, and this population was thus removed from our results.

Isolating donut and spread colonies

To isolate one pair of donut and spread colonies from a macroscopic strain, we first streaked the strain from glycerol stocks onto a YPD plate and grew it at 30 °C for 2 days. As streaking or plating a multicellular strain does not ensure that a colony is founded by a single cell, we used chitinase (C8241, Sigma) to break clusters down to single cells (although some tiny branches exist after digestion, they should have gone through a recent single-cell bottleneck) before plating cells and isolating donut or spread colonies below.

To isolate a donut colony, we picked a single donut colony on the YPD plate and transferred it into a 1.5-ml microcentrifuge tube containing 800 µl of chitinase solution (0.25 mg ml^{-1} chitinase in 50 mM potassium phosphate buffer, pH 6.0). We broke apart the colony by pipetting with 100-µl regular-bore tips and vortexing, and then incubated it at 30 °C for 8 h with gentle rotation on a mini-rotator (Bio RS-24, BioSan) at level 5. After chitinase digestion, we washed the cells twice with 1 ml of H_2O and spread them on YPD plates with three dilution factors (1:400, 1:1,000 and 1:2,500). We incubated the plates at 30 °C for 3 days. Then, we picked a single donut colony from the plates and grew it in 10 ml of YPD media at 30 °C with 250 rpm shaking for 24 h, from which we prepared glycerol stocks by mixing 500 µl of the culture and 500 µl of 70% glycerol. This served as our donut isolate.

To isolate a spread colony, we transferred 100 µl of the above 24-h culture into a 1.5-ml microcentrifuge tube, washed it twice with 1 ml of H_2O and resuspended it in 800 µl of chitinase solution. We performed chitinase digestion, washing, plating and plate incubation in the same way as above for the donut isolate, except for plating at

dilution factors of 1:800, 1:2,000 and 1:5,000 with three plates each. From these nine plates that contained approximately 1,000 colonies, we could usually find more than one spread colony. However, as chitinase does not perfectly digest clusters into single cells, it is challenging to accurately estimate the frequency of donut-to-spread transitions. We picked one single spread colony and streaked it on a YPD plate to confirm the spread colony morphology and to apply an additional round of clonal isolation. We incubated the plate at 30 °C for 2 days. Then, we picked a single spread colony, and we grew it in YPD media and prepared glycerol stocks in the same way as above for the donut isolate. This served as our spread isolate for the corresponding donut isolate. As each spread colony was isolated from plating the 24-h liquid culture inoculated from a donut colony, we limited the number of cell generations between donut and spread phenotypes, which allowed us to better pinpoint the relevant genetic changes and also showed how rapidly macroscopic sizes can be lost.

In total, we prepared 25 donut–spread pairs from the nine macroscopic PA isolates, including two pairs from the PA1 t600 isolate (because the donut strain in the third pair did not grow macroscopic in later experiments) and the PA5 t1,000 isolate (because we were only able to obtain two pairs after many attempts) as well as three pairs from each of the other seven isolates.

Statistics

We performed statistical tests using R (v4.1.2), with the details described in the main text and figure legends. We implemented support vector machine in Fig. 4e using scikit-learn (v1.3.2) in Python (v3.11.6). We ran the biophysical simulations in MATLAB (R2019a).

Reporting summary

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

Data availability

Underlying data used to make the figures as well as raw experimental data are available on GitHub (https://github.com/ktong25/WGD_in_MuLTEE). Raw Illumina sequencing reads are available at the NIH Sequence Read Archive under the accession numbers PRJNA943273 (for MuLTEE evolved isolates) and PRJNA1093477 (for donut and spread strains). All raw microscopy images are publicly available at the BioImage Archive under the accession number S-BIAD1559. The yeast genome (S288C reference genome, R64-3-1 build) used for genome sequence analysis was downloaded from the *Saccharomyces* Genome Database.

Code availability

All code used for image analysis, genome sequence analysis, biophysical modelling, data analysis and figure generation are available on GitHub (https://github.com/ktong25/WGD_in_MuLTEE).

47. Pachitariu, M. & Stringer, C. Cellpose 2.0: how to train your own model. *Nat. Methods* **19**, 1634–1641 (2022).
48. Cherry, J. M. et al. *Saccharomyces* Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res.* **40**, D700–D705 (2012).
49. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at https://arxiv.org/abs/1303.3997 (2013).
50. Van der Auwera, G. A. & O'Connor, B. D. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra* (O'Reilly Media, 2020).
51. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
52. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
53. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).
54. Todd, R. T., Braverman, A. L. & Selmecki, A. Flow cytometry analysis of fungal ploidy. *Curr. Protoc. Microbiol.* **50**, e58 (2018).

Acknowledgements We thank the Microbial Sequencing Center at University of Pittsburgh for sequencing the genomes; the Partnership for an Advanced Computing Environment (PACE) at Georgia Institute of Technology for providing the research cyberinfrastructure resources and services; and all members of the Ratcliff laboratory, and H. Chen and F. Mortier of Yves Van de Peer's laboratory for feedback on the study. This work was supported by US National Institutes of Health (NIH) grant R35-GM138030 to W.C.R., Human Frontiers Science Program grant RGY0080/2020 to W.C.R., Packard Fellowship for Science and Engineering to W.C.R., NIH grant R35-GM138354 to P. J. Yunker, NIH grant R01AI171100 to A.S.K., and Schmidt Sciences Polymath award G-22-63292 to A.S.K.

Author contributions K.T. and W.C.R. conceived the project. K.T., W.C.R., P.L.C., S.D. and A.S.K. designed the experiments. G.O.B. performed the MuLTEE, analysed the point mutations in the evolved isolates and identified the genomic signature of tetraploidy. V.C., D.J.H. and K.T. genetically engineered the tetraploid snowflake yeast. S.G., H.L.Y. and P.L.C. performed the evolution experiments with selection against larger size and measured the evolved populations. S.D. and K.T. isolated donut and spread strains, characterized their phenotypes and analysed their point mutations. T.C.D. performed the biophysical simulations. D.T.L. performed the genomic DNA extraction and helped with genome analysis. S.D. and K.T. performed the mutation accumulation experiment. K.T. performed the rest of the experiments, analysed data and made figures. K.T. prepared the first draft of the manuscript. All authors contributed to the revision of the manuscript.

Competing interests The authors declare no competing interests.

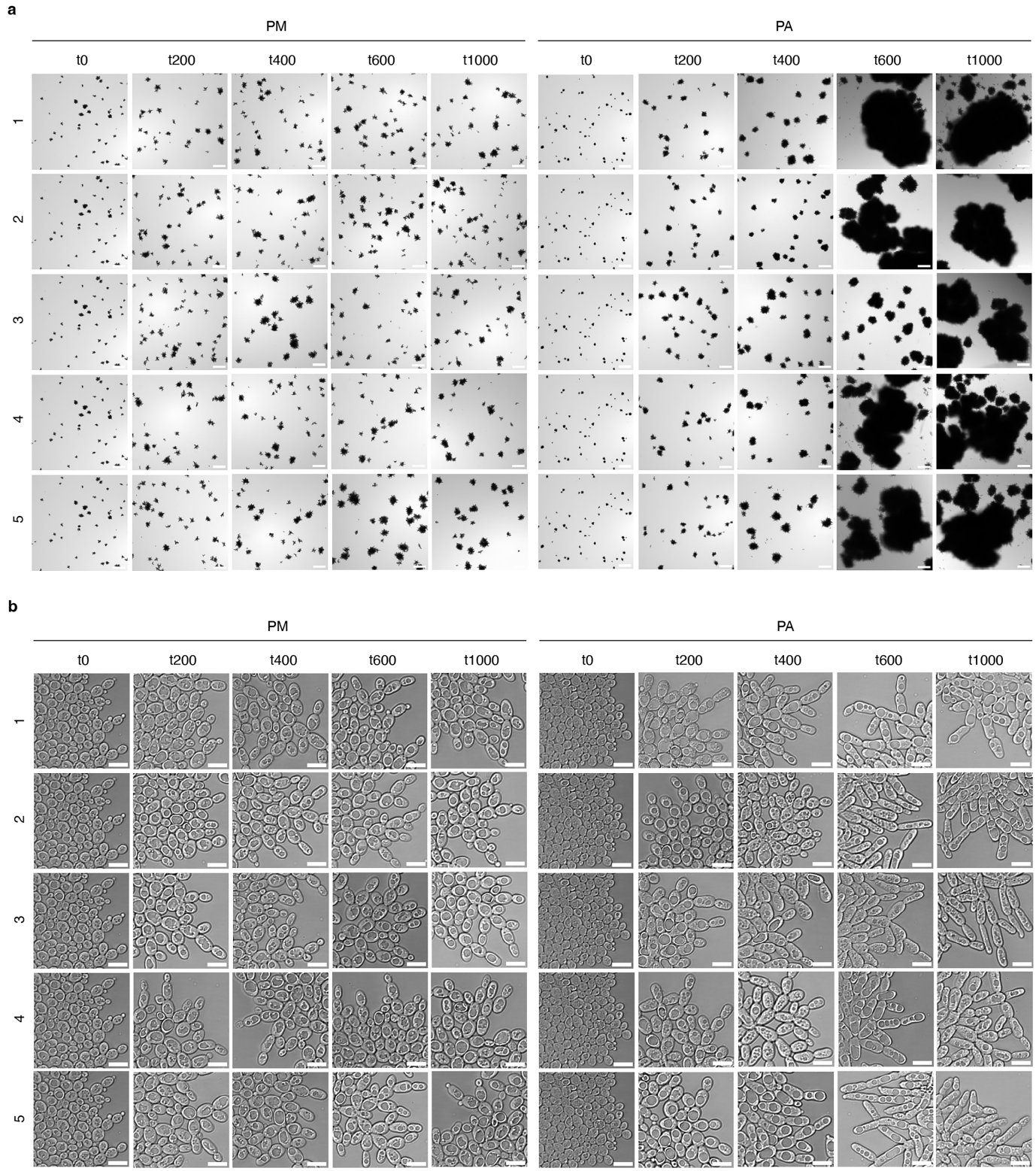
Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-025-08689-6.

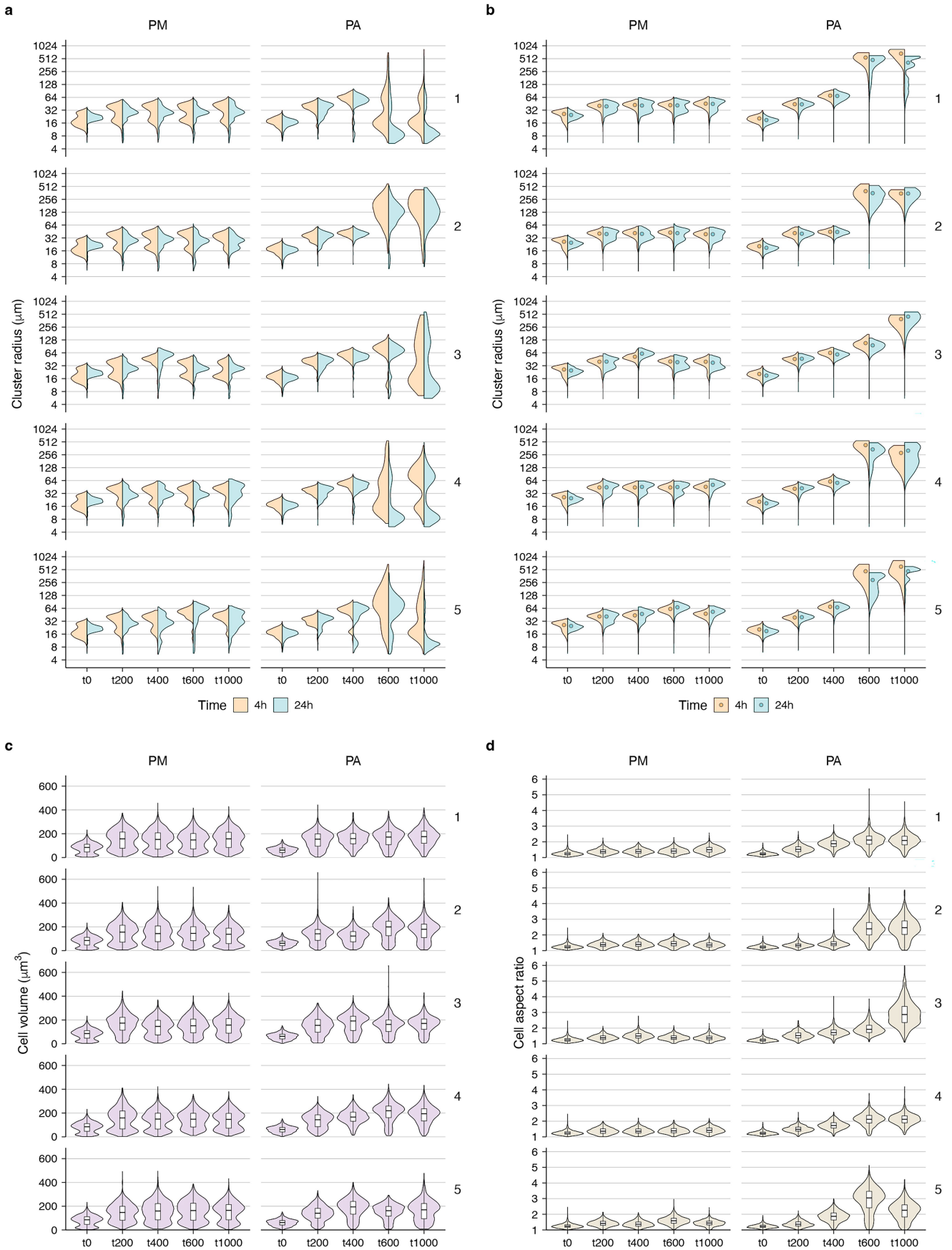
Correspondence and requests for materials should be addressed to Kai Tong or William C. Ratcliff.

Peer review information *Nature* thanks Omaya Dudin, Gianni Liti and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

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

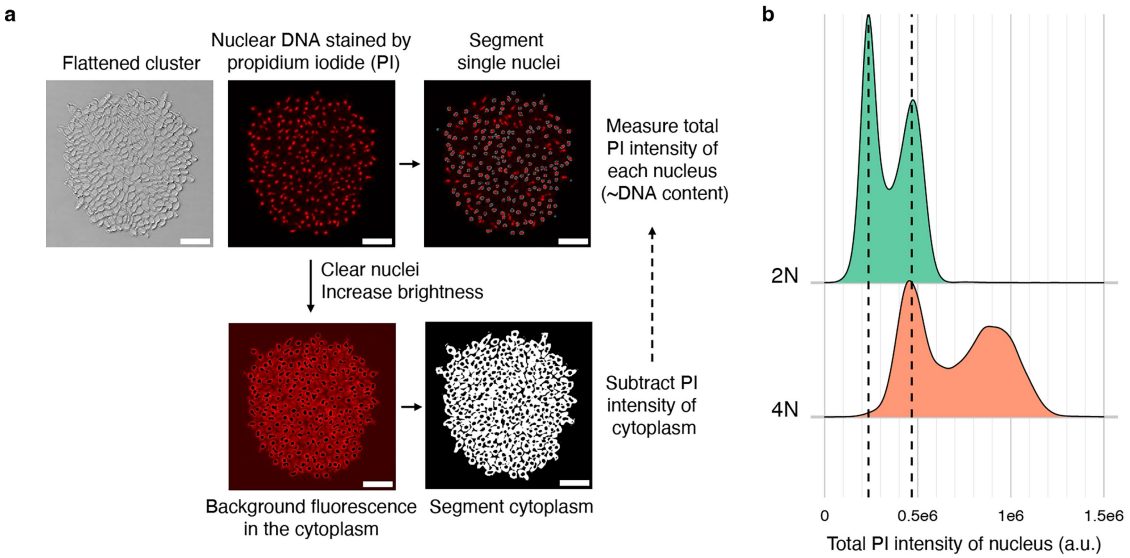


Extended Data Fig. 1 | Images of the ancestors and evolved isolates in the MuLTEE. a,**b**, Representative cluster-level (**a**) and cell-level (**b**) images of PM/PA t0 and PM/PA1-5 t200, t400, t600, and t1000 isolates. The images of PM/PA t0 are reused for five replicate populations. Scale bars, 200 µm (**a**) and 10 µm (**b**).

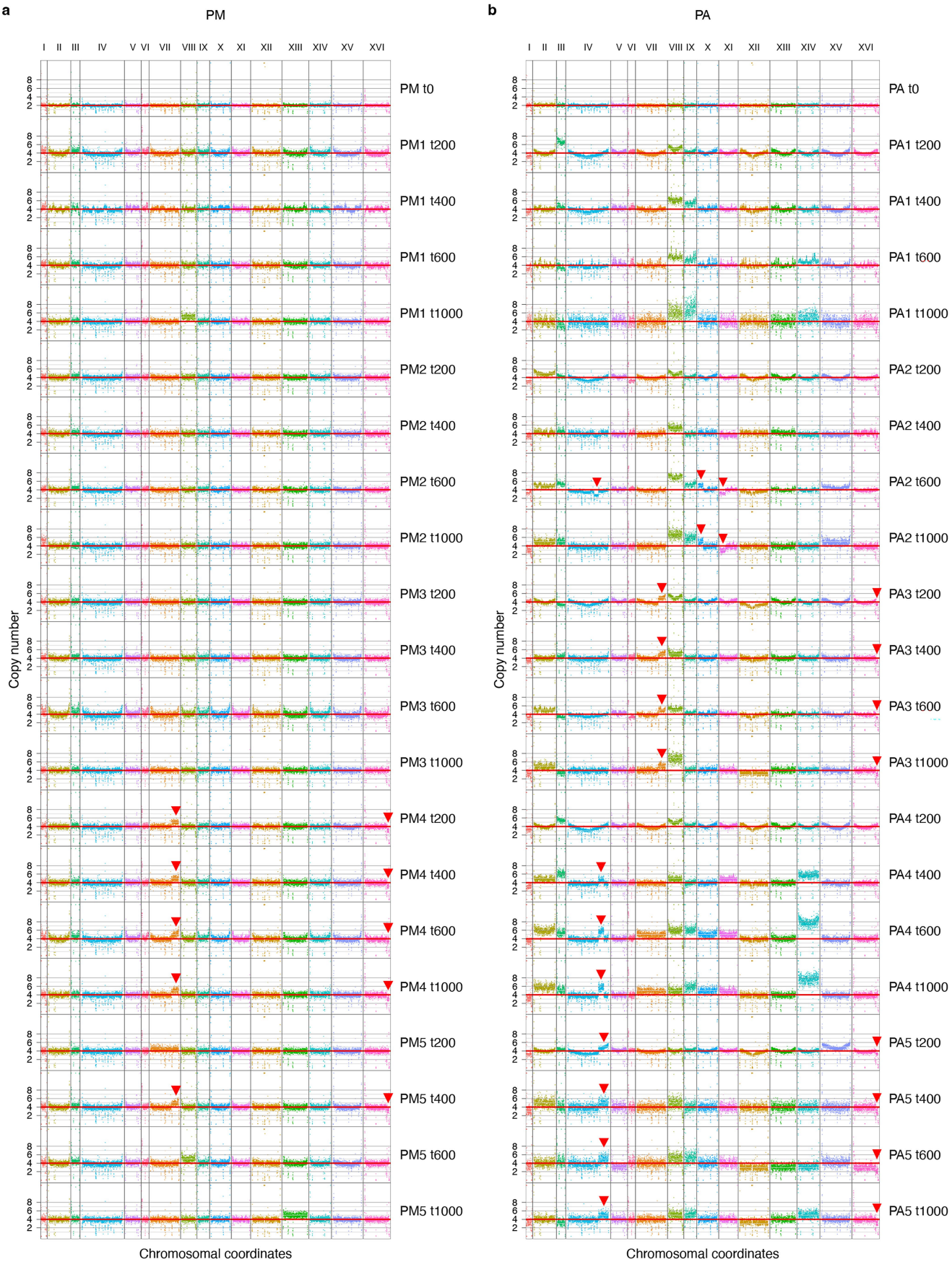


Extended Data Fig. 2 | See next page for caption.

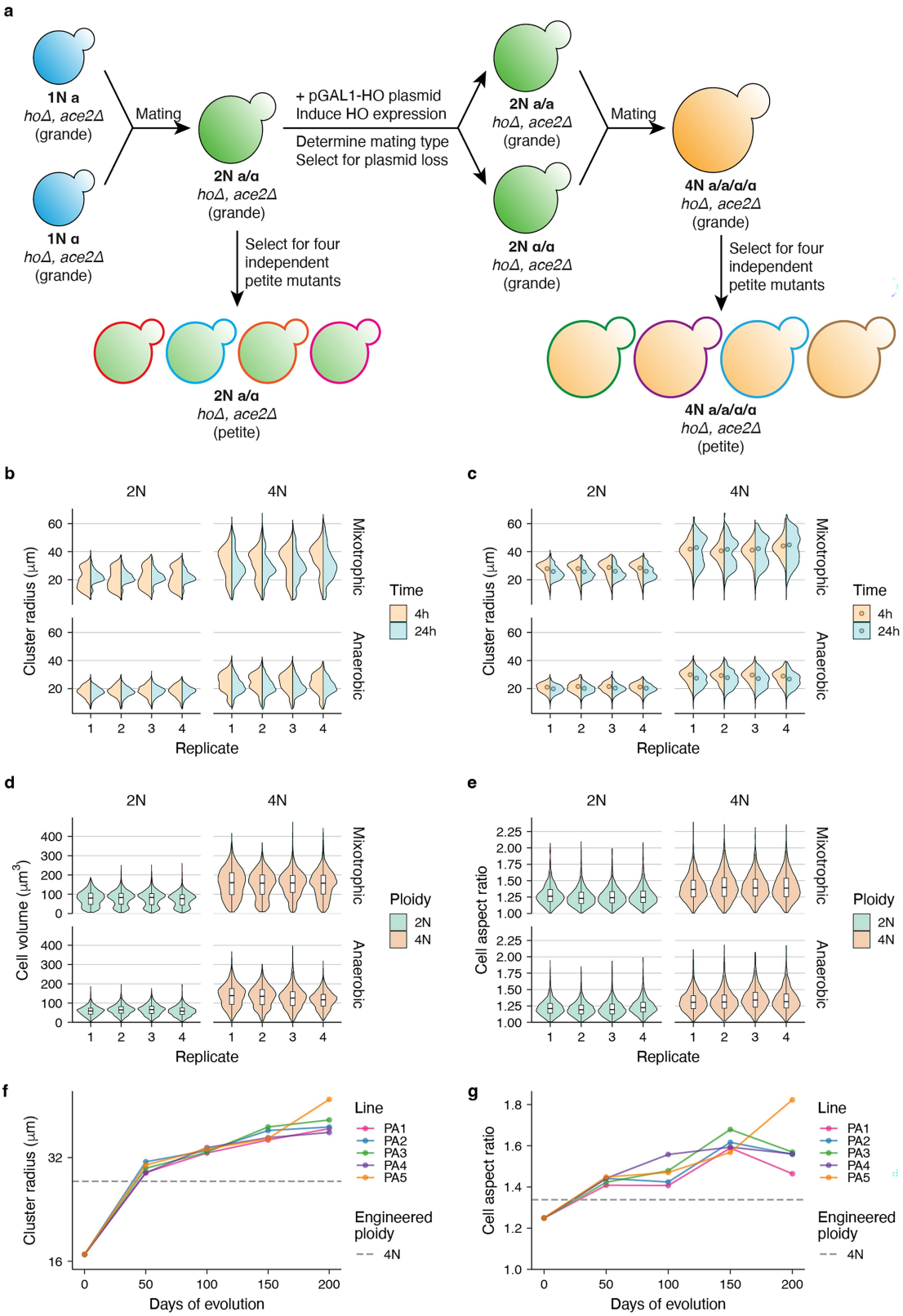
Extended Data Fig. 2 | Phenotypic characterization of the ancestors and evolved isolates in the MuLTEE. a-d, Violin plots showing the distributions of cluster radius (**a**,**b**, where **b** is weighted by cluster volume), cell volume (**c**), and cell aspect ratio (**d**) in PM/PA t0 and PM/PA1-5 t200, t400, t600, and t1000 isolates (on average, $n = 789$ clusters (**a**,**b**) and 1,288 cells (**c**,**d**) measured per sample). The distributions of PM/PA t0 are reused for five replicate populations. For **a**,**b**, we measured cluster radius at 4 h (exponential phase) and 24 h (stationary phase) after transferring the culture to fresh media, and the 24-hour measurements (corresponding to the states of the cultures right before settling selection) are used throughout the paper unless otherwise noted. For **b**, filled circles show biomass-weighted mean cluster radius (the 24-hour values are the same as the values in Fig. 1c). For **c**,**d**, boxes, IQR; center lines, median; whiskers, values within 1.5 × IQR of the first and third quartiles.



Extended Data Fig. 3 | Imaging-based method for measuring ploidy level of snowflake yeast. a, Overview of imaging and image analysis workflow. Snowflake yeast clusters are crushed into a single layer of cells and imaged at the brightfield channel and fluorescent channel, with the latter showing the nuclear DNA stained by propidium iodide (PI). The nuclei in the fluorescent image are segmented and filtered to get single round nuclei, outlined in cyan. The fluorescent image is also nuclei-cleared and brightness/contrast-enhanced to show the background fluorescence in the cytoplasm, and the cytoplasm is segmented, shown in white, for background subtraction. The total fluorescence intensity of PI in each nucleus is quantified and background-subtracted. Scale bar, 20 µm. **b**, Distribution of the nuclear PI intensity (arbitrary unit) of the engineered diploid and tetraploid mixotrophic clusters (n = 14,276 and 10,031 nuclei, respectively), as a validation for this ploidy measurement method. Since asynchronous, exponential-phase cultures are used for ploidy measurements, each strain shows two peaks that correspond to G1- and G2-phase nuclei of the actively-dividing cells, and the G2 peak has double the fluorescent intensity of the G1 peak. Also, the G2 peak of diploid clusters aligns nicely with the G1 peak of tetraploid clusters.

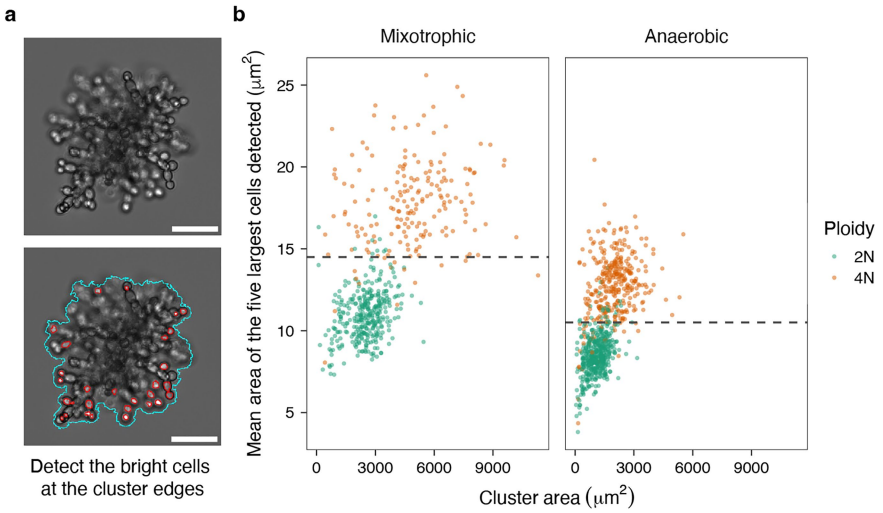


Extended Data Fig. 4 | Copy number variation of the ancestors and evolved isolates in the MuLTEE. a,b, Estimated copy number of each 1-kb non-overlapping bin in each chromosome in PM/PA t0 and PM/PA1-5 t200, t400, t600, and t1000 isolates (**a**, PM; **b**, PA). Estimated bin copy numbers above 12 are shown as 12, indicated by little triangles. Red horizontal line, baseline ploidy of each strain (i.e., 2 for PM/PA t0 and 4 for all evolved isolates). Red arrowhead, incidence of segmental aneuploidy.

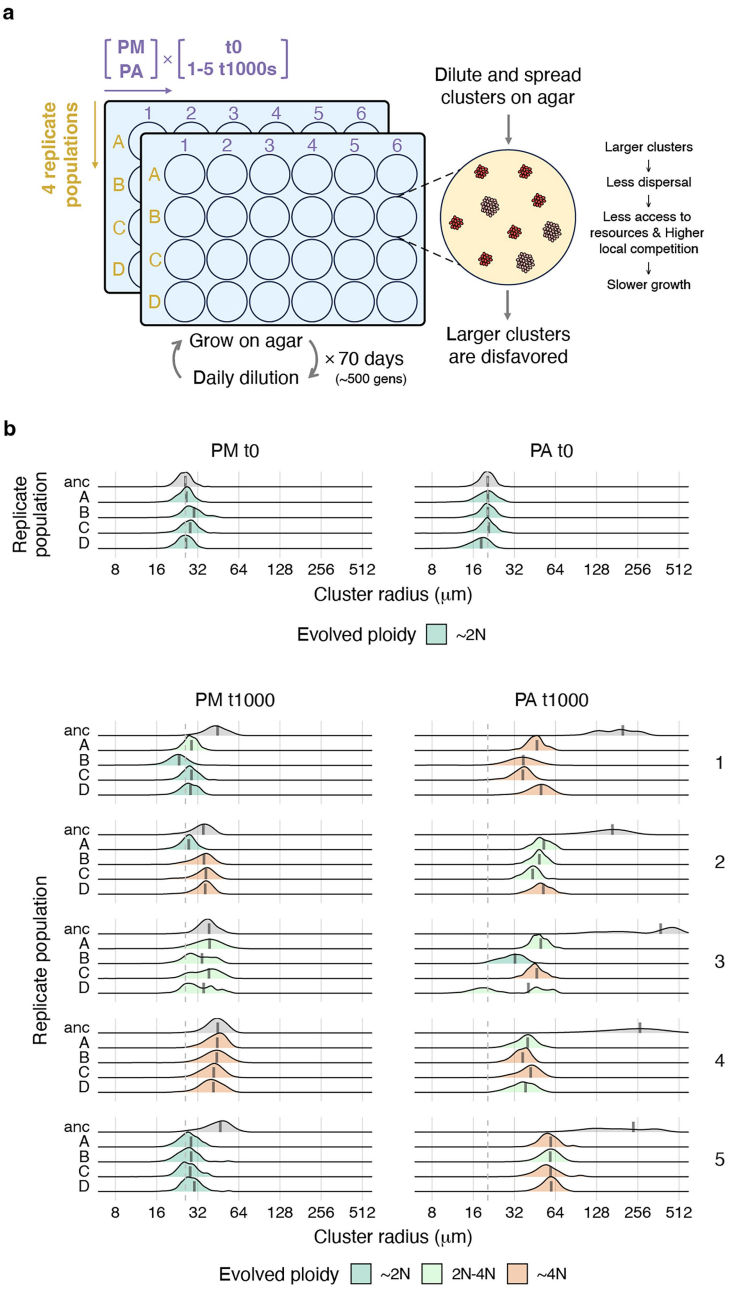


Extended Data Fig. 5 | See next page for caption.

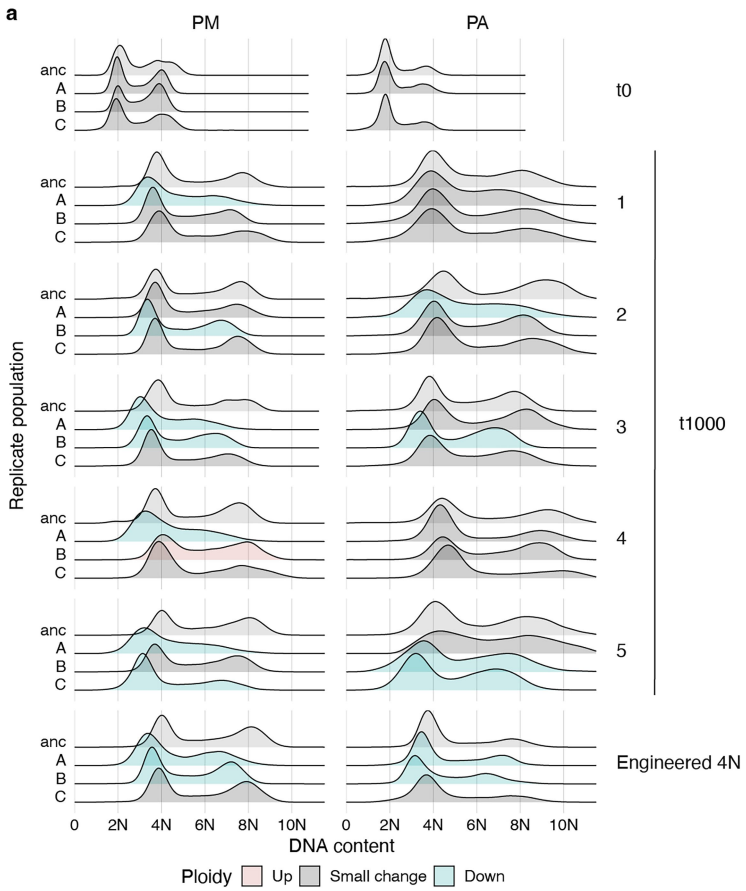
Extended Data Fig. 5 | Genetic construction and phenotypic characterization of diploid and tetraploid clusters. a, Procedure for engineering isogenic grande diploid and tetraploid clusters, from each of which four independent petite mutants were isolated. Isolating multiple petite mutants is important because petite mutations are not isogenic and may confound ploidy-phenotype map. Grande and petite clusters correspond to mixotrophic and anaerobic conditions, respectively. **b**-**e**, Violin plots showing the distributions of cluster radius (**b**,**c**, where **c** is weighted by cluster volume), cell volume (**d**), and cell aspect ratio (**e**) in engineered diploid and tetraploid clusters under mixotrophic and anaerobic conditions (on average, $n = 922$ clusters (**b**,**c**) and 2,458 cells (**d**,**e**) measured per sample). Four biological replicates were measured for the mixotrophic condition, and the four independent petite mutants (each with one biological replicate) were measured for the anaerobic condition. For **b**,**c**, we measured cluster radius at 4 h (exponential phase) and 24 h (stationary phase) after transferring the culture to fresh media, and the 24-hour measurements are used throughout the paper unless otherwise noted. For **c**, filled circles show biomass-weighted mean cluster radius (the 24-hour values are the same as the values in Fig. 3e). For **d**,**e**, boxes, IQR; center lines, median; whiskers, values within 1.5 × IQR of the first and third quartiles. **f**,**g**, Comparison of the biomass-weighted mean cluster radius (**f**) and mean cell aspect ratio (**g**) of the engineered petite tetraploid clusters (mean of the four independent petite mutants, the same values as those in Fig. 1c,e) to the PA t0 and PA1-5 t50, t100, t150, and t200 populations (data from Bozdag et al.[8]).



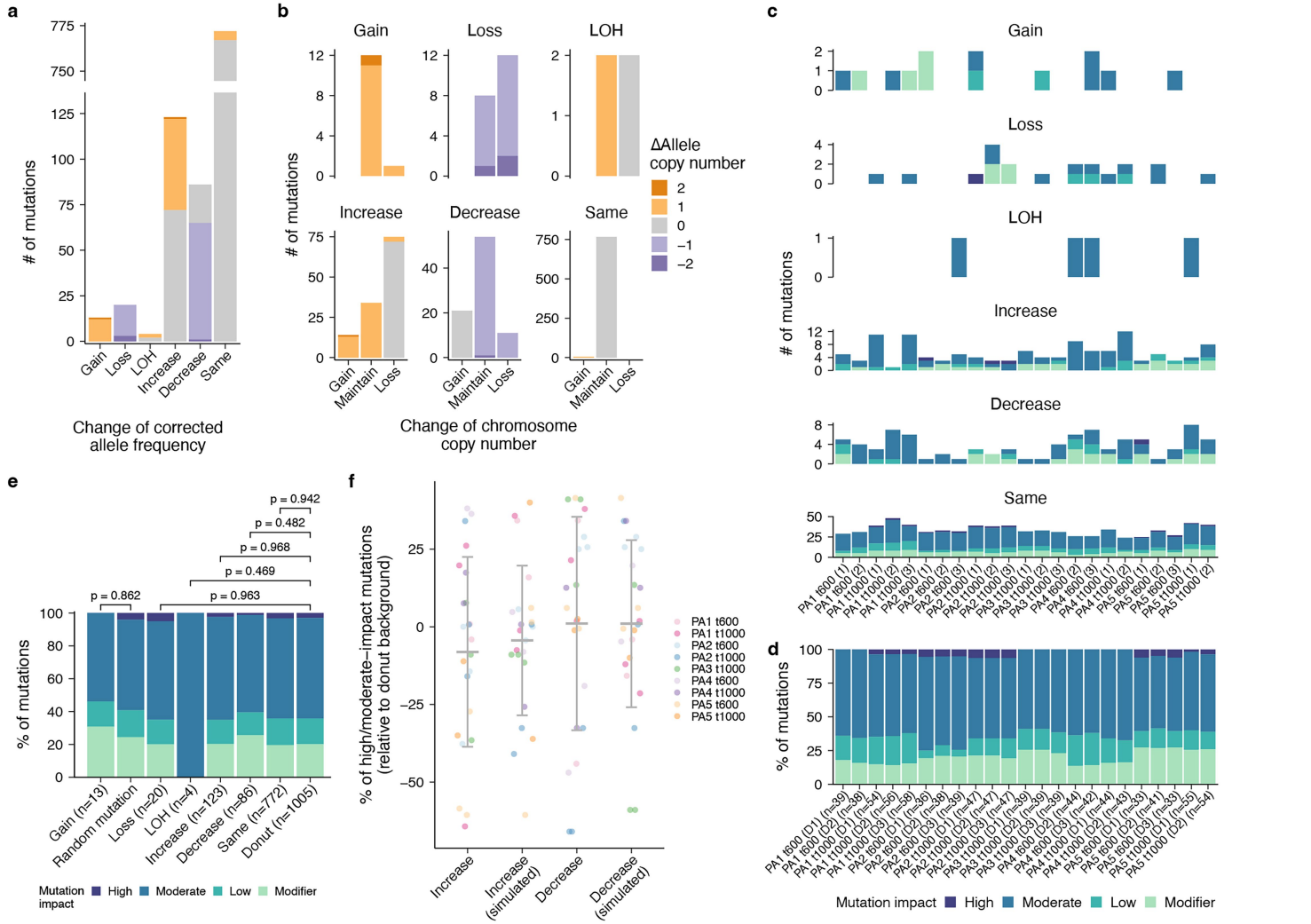
Extended Data Fig. 6 | Label-free method for distinguishing engineered diploid and tetraploid clusters in competition assays. a, Brightfield image of a snowflake yeast cluster (an engineered tetraploid mixotrophic cluster is shown) (top), whose bright cells in the cluster edges are detected (bottom). Scale bar, 30 μm. On average, 381 clusters were segmented per sample using this approach. **b**, Mean area of the five largest cells detected in a cluster can be used to distinguish between the engineered tetraploid and diploid clusters under both mixotrophic and anaerobic conditions, with the dashed line indicating the manually-chosen decision boundary.



Extended Data Fig. 7 | Experimental evolution of the MuLTEE ancestors and t1000 isolates with selection against larger size. a, Experimental setup. We evolved PM/PA t0 and PM/PA1-5 t1000 isolates, each with four replicate populations (A, B, C, D), under selection against larger size for 70 days (~500 generations) by growing them on agar in 24-well plates with daily dilution. **b**, Distributions of cluster radius (weighted by cluster volume) in the ancestral ("anc") and evolved populations (on average, $n = 406$ clusters measured per population). Vertical thick solid line, biomass-weighted mean cluster radius of each population. Vertical dashed line, biomass-weighted mean cluster radius of PM/PA t0. Color codes for the evolved ploidy in each t0 population and in each t1000 population are the same as those in Fig. 3h,i, respectively, and the ancestral populations are colored in gray.



Extended Data Fig. 8 | Experimental evolution of the MuLTEE ancestors, t1000 isolates, and engineered tetraploids with minimal selection. We evolved PM/PA t0, PM/PA1-5 t1000 isolates, and engineered mixotrophic and anaerobic tetraploids, each with three replicate populations (A, B, C), with minimal selection for 56 days or 28 bottlenecks, by growing them on agar with picking and streaking single colonies every two days. **a**, Distributions of cellular DNA contents in the ancestral ("anc") and evolved populations (on average, $n = 15{,}848$ cells measured per population). Color code for ploidy change categories is the same as that in Fig. 3j, and the ancestral populations are colored in light gray.



Extended Data Fig. 9 | Point mutation changes in donut-to-spread transitions. Two donut-to-spread transitions with near-triploidization were excluded, and mutation allele frequency refers to the corrected allele frequency, calculated by dividing estimated allele copy number with copy number of the chromosome that carries the mutation. LOH, loss of heterozygosity. **a**,**b**, Number of mutations in each category of change in allele frequency in all donut-to-spread transitions combined, colored by change in allele copy number (**a**), and how change in allele frequency is associated with change in chromosome copy number (**b**). **a**,**b** share the color code. **c**, Number of mutations in each category of change in allele frequency in each donut-to-spread transition, colored by mutation impact. **d**, Percentage of mutations in each mutation impact category in each donut background, whose total number of mutations is indicated in the brackets. **c**,**d** share the color code. **e**, Comparison of the distribution of mutation impacts, between the mutations that were gained in all donut-to-spread transitions combined and the mutations randomly introduced into yeast genome, as well as between the mutations that underwent loss, LOH, increase, decrease, or maintenance in terms of allele frequency in all donut-to-spread transitions combined and the mutations in all donut backgrounds combined. Number of mutations is indicated in the brackets. *P* values were calculated by two-sided chi-squared test. **f**, For each donut-to-spread transition, the percentage of high/moderate-impact mutations in the mutations that increased or decreased in allele frequency is on average not significantly larger than the percentage of high/moderate-impact mutations in the donut background (for increase and decrease, respectively, $P = 0.891$ and 0.442, $t_{22} = -1.27$ and 0.147, one-tailed one-sample Student's *t*-test), and is largely explained by random sampling of mutations in the donut background (simulation with random seed = 1). Values are mean ± s.d. ($n = 23$ donut-to-spread transitions).

nature portfolio

Corresponding author(s): Kai Tong, William C. Ratcliff

Last updated by author(s): Jan 3, 2025

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	Multicellular (or cluster) size, cell volume and aspect ratio, and ploidy data were collected using a Nikon Eclipse Ti inverted microscope (NIS-Elements v4.30.01). Whole genomes were sequenced using Illumina NovaSeq 6000 and NovaSeq X Plus platforms.
Data analysis	ImageJ v1.54f and Cellpose v2.2.2 were used for image analysis, except that ImageJ v1.53q was used for image analysis in competition assay. Trimmomatic v0.39, BWA-MEM v0.7.17, Picard Toolkit v2.27.5, GATK Haplotypecaller v4.2.4.1, VCFTOOLS v0.1.16, bcftools-isec v1.10, snpEff v5.0, Bamtools Stats v2.5.1, and Linux bash shell were used in genome sequence analysis. R v4.1.2 was used for statistical tests and plotting. scikit-learn v1.3.2 and Python v3.11.6 were used for support vector machine. MATLAB R2019a was used for biophysical simulation. All custom code used for image analysis, genome sequence analysis, biophysical modeling, data analysis, and figure generation are publicly available at https://github.com/ktong25/WGD_in_MuLTEE.

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

Underlying data used to make figures as well as raw experimental data are available at https://github.com/ktong25/WGD_in_MuLTEE. Raw Illumina sequencing reads are available at the NIH Sequence Read Archive under accession number PRJNA943273 (for MuLTEE evolved isolates) and PRJNA1093477 (for donut and spread strains). All raw microscopy images are publicly available at the BioImage Archive under accession number S-BIAD1559. Yeast genome (S288C reference genome, R64-3-1 build) used for genome sequence analysis was downloaded from Saccharomyces Genome Database (SGD).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	n/a
Reporting on race, ethnicity, or other socially relevant groupings	n/a
Population characteristics	n/a
Recruitment	n/a
Ethics oversight	n/a

Note that full information on the approval of the study protocol must also be provided in the manuscript.

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

Ecological, evolutionary & environmental sciences study design

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

Study description	We conducted an evolution experiment over 1000 days to select microscopic snowflake yeast clusters for large multicellular size. To study the effects of oxygen availability, we evolved these yeast under mixotrophic (PM1-5) or anaerobic (PA1-5) condition, making a total of 2 x 5 = 10 replicate populations. Tetraploidy rapidly emerged and was maintained in all ten populations. PM1-5 remained microscopic and are mostly euploid, while PA1-5 evolved macroscopic size with extensive aneuploidy. This study examines how polyploidy arises, becomes maintained and drives long-term adaptation through phenotypic characterization, genome sequencing, synthetic reconstruction, biophysical modeling and evolution experiments.
Research sample	We used the budding yeast Saccharomyces cerevisiae as our model organism. This organism is a well-established eukaryotic model system, making it genetically tractable and phenotypically accessible. To initiate our experiment with a multicellular phenotype, we deleted ACE2 gene from the yeast.
Sampling strategy	We began our evolution experiment with five replicate populations and two metabolic treatment groups. For phenotypic characterization and genome sequencing, we isolated a single clonal snowflake yeast cluster from these populations at day 200, 400, 600, and 1000 and sequenced those isolates. To measure multicellular size, we collected data from an average of 886 multicellular clusters per isolate. To measure cell volume and aspect ratio, we collected data from an average of 1288 cells per isolate. To measure ploidy, we collected data from an average of 12922 cells per isolate.
Data collection	GOB performed the long-term evolution experiment and stored a -80C archive of each population every 10-15 days. The data were collected by KT, SD, SG, HLY, TCD, DL, and PLC without any manipulation by the experimenter. Cluster size, cell volume and aspect ratio, and ploidy data were collected using a Nikon Eclipse Ti inverted microscope at the Ratcliff Lab. Genome sequencing data was generated at the Microbial Sequencing Center in University of Pittsburgh.

Timing and spatial scale	Long-term evolution experiment (the first 1000 days) was performed between 2017-2021 (paused during COVID-19). Multicellular size, cell volume and aspect ratio, ploidy and genome data collection and analysis were performed in 2022-2024.
Data exclusions	No data were excluded from the analysis.
Reproducibility	All ten populations (PM1-5, PA1-5) evolved and maintained polyploidy. All PM1-5 remained microscopic and mostly euploid, while all PA1-5 evolved macroscopic clusters with extensive aneuploidy. This parallel trend not only demonstrates the reproducibility of our data but also provides confidence that, if these populations were to be re-evolved, the same outcome (polyploidy and aneuploidy) would arise. We have kept a frozen record of all those strain isolates and populations in our -80C freezer, which are available for further analysis, such as replicating our results.
Randomization	To ensure masked sample collection, we assigned random labels to samples at all stages of data collection.
Blinding	Blinding was not necessary for our study as we did not need to make any decisions based on phenotypic analysis. We collected multicellular size, cell volume and aspect ratio, and ploidy data by sampling a significant portion of the clusters from the cultures.

Did the study involve field work? ☐ Yes ☒ No

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
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

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

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	Other organisms are used: Saccharomyces cerevisiae - a non-pathogenic microbial eukaryote. Stain: Y55.
Wild animals	n/a
Reporting on sex	n/a
Field-collected samples	n/a
Ethics oversight	n/a

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Plants

Seed stocks	n/a
Novel plant genotypes	n/a
Authentication	n/a