A dynamic network model predicts the phenotypes of multicellular clusters from cellular properties

Highlights

- A dynamic network for the growth of multicellular clusters using cellular features

- Kissing number, the maximum number of physically linked cells, limits cluster size

- The rates at which links break and cells divide set cluster size below this limit

- With two cell types, raising kissing number increases intra-cluster heterogeneity

Authors

Piyush Nanda, Julien Barrere, Thomas LaBar, Andrew W. Murray

Correspondence

awm@mcb.harvard.edu

In brief

Nanda et al. present a framework for the size limits of yeast multicellular clusters. With cells as nodes and cell-cell connections as edges, they simulate cluster growth as a dynamic network of cells. A model with a limit on the maximum number of neighboring cells and age-dependent breakage of cell-cell connections best fits experimental data.
A dynamic network model predicts the phenotypes of multicellular clusters from cellular properties

Piyush Nanda, Julien Barrere, Thomas LaBar, and Andrew W. Murray

INTRODUCTION

The origin of multicellularity was a major transition in Earth’s life history. As differentiated multicellularity evolved, reproductive strategies that preserved key multicellular properties were favored. Unicellular organisms can evolve multicellularity either by colliding and sticking together (aggregative multicellularity) or by failing to separate from each other after cell division (clonal multicellularity). Two simple characteristics of the initial multicellular forms were their size (the number of cells in a cluster) and composition (the proportion of different cell types), but their extinction or evolution makes it difficult to understand how these features evolved. Even in series of phylogenetically related organisms that differ in their degree of multicellularity, like the Volvocales, the original unicellular and multicellular forms no longer exist.

One alternative to comparative approaches is to create and investigate multicellularity in the lab by using genetic engineering or experimental evolution. Recently, the budding yeast, Saccharomyces cerevisiae, has emerged as a model organism for understanding aspects of multicellularity, including cluster size and cluster composition (in clusters with multiple cell types). Selecting for faster settling or production and consumption of public goods leads to the evolution of clonal multicellularity. Cluster size is under selection when faster sedimentation or public goods sharing provides a fitness advantage, while cluster composition affects fitness when division of labor or metabolic interactions provide an advantage. Understanding the parameters that control cluster size and composition will offer insights into the evolution of simple multicellularity.

The growth and division of multicellular clusters results from the balance between cell division, which adds cells, and the fragmentation of clusters due to chemical processes and physical forces that break the linkage between cells in the cluster. For clusters that contain multiple cell types, the transition rates between different cell types and their proliferation rates control cluster composition. Previous work has highlighted genotypes that control maximum cluster sizes and internal physical forces. For example, changes in cell shape or the strength of cell-cell connections affect cluster size. In budding yeast, the failure to promptly dissolve the primary septum, the portion of the cell wall that links daughter cells to their mothers, leads to multicellularity. Experimental evolution of multicellularity has resulted in mutations in genes that affect the linkage between the mothers and daughters (ACE2, CTS1, and GIN4) or control cell
shape (CLB2 and ARP5), but we do not know how these mutations affect cluster growth and fragmentation to produce a steady-state distribution of cluster sizes. Modeling has explored the effect of geometric constraints on individual clusters and suggests that changing cell shape or the geometry of the linkages to a single cell increases cluster size more than increasing strength of the linkages between individual cells. Even more recently, it has been shown that the distribution of cell size to cluster size (geometric size) at a constant number of cells per cluster follows a predictable distribution in snowflake yeast and in Volvox, strongly suggesting the existence of fundamental rules.

While these studies discuss the parameters that set fundamental limits on cluster size, the underlying cell-level variables that control the distribution of cluster sizes remain unclear. An ideal model would enable ab initio prediction of cluster size and its distribution from a few measurable variables that can be directly linked to gene products or pathways in cells. Such predictions can be useful in engineering clusters of desired sizes or understanding the behavior of multicellular clusters in different environments where underlying variables (like rates at which cells grow or dissolve their linkages to other cells) differ. For example, how does a change in budding pattern in yeast alter the cluster-size distribution? How is this distribution affected if the growth rate of cells changes? In clusters where differentiation has produced a division of labor between different cell types, the distribution of cluster compositions will depend not only on the transition rates between different cell types and their rates of proliferation but also on which link breaks to turn one cluster into two. How do the cluster growth and fracture dynamics change the composition of such primitive, differentiated clusters?

We studied the size and composition of multicellular clusters by constructing a dynamic network model that represents cells as nodes and the connections between mothers and daughters as edges. Using simulations and experiments, we analyzed the effect of varying three parameters: the mode of budding (kissing number), link-breaking rate, and doubling time. We show that a model with a steric limit on the kissing number and links that are more likely to break as they age best captures the experimentally observed distribution. The faster cells divide relative to the link-breaking rates, the more closely cluster sizes approach the upper limit set by the kissing number. In differentiating clusters with germ and somatic cells, increasing the kissing number and optimizing the difference in the division rate between the two cell types maximize the fraction of clusters that contain both cell types. We discuss the implications of our work for understanding the evolution of adaptive multicellular phenotypes and extensions of this model to capture complex multicellular properties.
RESULTS

A dynamic-network approach to model the growth of multicellular clusters

We programmed a dynamic network model to simulate the growth and breakage of clusters and predict steady-state distributions of cluster sizes (Figure 1B). In a dynamic network, the edges and nodes can change (appear or disappear) over time. We represent cells as nodes and their connections as edges. Each network represents an individual cluster and grows as cells divide, with division rate (ι), and fragments at a specific edge when a certain set of conditions are met. Each time a cell divides, it creates a new link to a daughter. We subject the dynamic network to several cycles of growth and fracture to reach a steady-state distribution (Figure 1B). Increasing crowding around a mother cell leads to the development of steric forces due to geometric constraints, which we capture as the kissing number, κ, the maximum number of links a cell may have to its daughters. The concept of kissing number is derived from geometric limits on the packing of spheres, i.e., the maximum number of spheres that can surround a given sphere. We define the kissing number as the maximum number of daughter cells that can be physically attached to a single mother cell. For simplicity, we assume that only the interactions between the focal mother cell and her immediate daughter cells apply stress to the links between the mother and her daughter cells. Incorporating budding angles and the exact geometric positions of cells would better estimate the interactions between all the cells in a cluster but would do so at the cost of adding additional parameters. In our simulations, once the number of links to a cell approaches the kissing number, one or more cell-cell connections start to fracture. The fracture can either happen probabilistically, i.e., each connection has a probability of breaking based on an underlying probability distribution, or deterministically, for example, the oldest connection breaks. The fracture partitions the cells in the parental cluster into two new clusters, with the ratio of their sizes determined by which link breaks. We allow cluster growth and fragmentation to continue until the distribution of cluster sizes in a population reaches a steady state. In our model, the cell division rate, the rules for breaking links, and the kissing number can be varied independently, allowing us to explore multiple modalities with different rules for cluster fragmentation.

We began by asking if all cluster sizes in a yeast population could grow and proliferate to produce the full distribution of cluster sizes. We tested whether clusters of different sizes, selected from the same distribution, can give rise to the parental distribution. Using a strain with reduced Cts1 expression, which makes clusters of 3–10 cells because it reduces the expression of the chitinase (Cts1p) that helps to degrade the primary septum, we used fluorescent-activated cell sorting (FACS) to separate the population into four quartiles (Q1, Q2, Q3, and Q4) by cluster size, allowed them to proliferate independently for 48 h, and measured the distribution of cluster sizes (Figures S1A–S1C). All four populations recapitulate the original distribution, strongly suggesting that it is generated from an underlying process whose parameters are similar across the population.

We considered five possible modalities of cluster fracture (Figure 1C) and modeled their effect on the cluster-size distribution. There are two types of constraints that could limit cluster size: (1) steric limited (S), the clusters grow and reach a packing limit at which steric forces fracture the cluster to relieve the strain. (2) Non-steric limited (NS), the links between cells decay over time, and the clusters fracture because links are chemically degraded rather than because any of its cells exceed the kissing number. The models are then subdivided by specifying the rules that determine which link breaks to fracture a cluster. One choice is deterministic: in the SO (steric, oldest), the oldest link in the entire cluster breaks, separating the two cells in the cluster with the highest number of links as the older of these cells has reached the maximum number of cells it can be surrounded with. Alternatively, a stochastic process assigns a fracture probability to each link. In one pair of models, once the maximum number of connections per cell (SR) or breaking time (NSR) is reached, one or more randomly chosen links fracture. Finally, if the chronological age (A) of the link matters, the longer it has been present in the cluster, the more likely it is to have fractured. In these models (SA or NSA), the probability that a link breaks per unit time is a constant given by the link-breaking rate, δ, and the probability that a link survives declines exponentially with its age: \( p(\text{survival to time } t) = \exp(-t \delta) \). In our simulations, each link is assigned a randomly chosen survival time from this exponential distribution. Note that when a cluster breaks into two new clusters, many of the links are already several generations old and are thus more likely to break in the future.

We set out to test which of the models best explains experimental cluster-size distributions. We engineered strains where we can control the biological equivalents of the three model parameters: the cell division rate (ι), the kissing number (κ), and the link-breaking rate (δ). We altered the kissing number by controlling the expression of BUD421 (Figures 1D, S1D, and S1E) in a strain deleted for ACE2 to switch the budding pattern from axial (κ = 5, BUD4 ON) to bipolar (κ = 8, BUD4 OFF). In principle, kissing number (κ) in our model can take values from κ = 2 to the maximum number of division cycles before a cell stops proliferating. In practice, the kissing number reflects steric constraints imposed by multiple buds attached to the same mother cell. We controlled the link-breaking rate (δ) by tuning the expression of the chitinase gene (CTS1) and controlled the cell division rate (ι) by altering the concentration of the carbon source in the media. Finally, we compared the simulated cluster-size distribution obtained from all the models with the experimental distribution using an imaging pipeline that determines the cluster-size distribution (Figures S2A–S2E).

Steric-limited age-dependent fracture (SA)-based dynamic network model predicts cluster-size distributions

To test model predictions, we first varied the kissing number parameter (κ) while keeping the other two parameters (ι and δ) constant. We changed the levels of Bud4p protein in an ace2Δ strain by using a β-estradiol-inducible promoter.23 Bipolar budding (BUD4-OFF) allows a greater number of daughter cells to remain attached to mother cells before the onset of steric forces as buds emerge from opposite poles. We confirmed the axial and bipolar budding patterns by imaging calcofluor white-stained clusters (Figures 2A and 2D). Using our imaging pipeline (Figures S2A–S2E), we obtained the cluster-size distributions corresponding to BUD4-ON and BUD4-OFF. For BUD4-OFF,
The maximum number of cells in a cluster ranged between 128 and 256. This implies a kissing number of 8, as exceeding this would produce clusters with more than 256 cells. Meanwhile, the maximum number of cells per cluster for BUD4-ON clusters was between 16 and 32, implying a kissing number of 5. The cell division rate ($\lambda$) of the clusters was estimated from the bulk growth curves for specific mutants. The link-breaking rate ($\delta$) for the SA model was set to a very low value ($1/\delta = 300$ h for a kissing number of 8 and $1/\delta = 100$ for a kissing number of 5) such that link breaking does not limit the size of the clusters. A sensitivity analysis was performed with respect to values of $\delta$ for SA/SR models to make sure the distribution of cluster sizes is insensitive to these parameter changes.

Figure 2. Sterically limited, age-dependent (SA) model recapitulates experimental cluster size statistics

(A) Three-dimensional CLSM (confocal laser scanning microscope) rendering of a multicellular cluster with uninduced BUD4 expression showing the location of bud scars (the triangles indicate bud scar location).

(B) Cluster-size distribution statistics for different models and experimental data (Exp.) for high kissing number ($k = 8$).

(C) Distribution obtained from models (SA and SR) and experimental measurements for uninduced BUD4 expression (n = 109 clusters). The faint lines show simulation repeats with independent seed values; the darker lines show their average.

(D) Three-dimensional rendering of multicellular clusters with fully induced BUD4 expression (the triangles indicate bud location).

(E) Cluster-size distribution statistics for different models and experimental data for low kissing number ($k = 5$).

(F) Distribution obtained from models (SA and SR) and experimental measurements for full induction (n = 114 clusters). The faint lines show simulation repeats with independent seed values; the darker lines show their average.

(G) Cluster-size distribution for different levels of Cts1p controlled through β-estradiol induction.

(H) Relation between mean and standard deviation of cluster-size distribution obtained in (G) for two independent biological replicates (n = 400 clusters per condition). The linear regression fit is to an Erlang’s distribution with shape parameter $k = 2.5$. r is the correlation coefficient between the mean and the standard deviation.

See also Figures S1–S3 and Table S1.
to the choice of $\lambda$ (Figures S3G and S3H) as the link-breaking rate is not the limiting factor. In simulations, the number of growth cycles was sufficient to allow the simulated cluster-size distributions to reach a steady state (Figures S3E and S3F). Based on these parameter values (Table S1), the models were simulated, and the resulting cluster-size distributions were compared to the experimental distribution. The Kolmogorov-Smirnov test was used to compare the experimental and simulated distributions as it is insensitive to the binning strategy used and therefore a robust measure of similarity between two distributions.

When comparing experimental data and simulations for the two different kissing numbers ($BUD4$-OFF and $BUD4$-ON), the SA model best predicted the cluster-size distributions in both cases (Figures 2B, 2C, 2E, and 2F). The correlated increase between cluster size and kissing number strongly suggests that cluster size is steric limited.

None of the other models could accurately fit the experimental data (Figures S3A–S3C). The SR model, which assigns an equal probability of breaking to all the edges, produces a cluster-size distribution that has a higher maximum value than experimentally seen: breaking the oldest edge produces daughter clusters of equal sizes, whereas breaking the newest link produces highly asymmetric daughter clusters leading to a coefficient of variation (CV) in cluster size larger than the experimental data (Figures 2C and 2F). All the NS models, which lack a steric component, fail to explain the difference in cluster size between the $BUD4$-ON and $BUD4$-OFF data (Figures S3A and S3C). For example, model NSA produces some clusters that are larger than those seen in any experimental condition (>500 cells per cluster). Model NSR produces a cluster-size distribution restricted to 2–6 cells per cluster. Model SO produces a bimodal distribution of cluster sizes corresponding to 128 and 64 cells per cluster. Breaking at the oldest node will always produce clusters of equal sizes as the cluster is symmetric about the oldest edge. Therefore, the clusters would have 128 cells per cluster when they are about to break or 64 cells per cluster after they break.

One potential drawback of our modeling strategy is that all the cells in a cluster divide synchronously. In real clusters, asynchronous cell divisions in a cluster can influence the cluster-size distribution. To test whether asynchronous cell division significantly affects the predicted cluster-size distribution, we simulated asynchronous cell division in the dynamic network. During a given division event, only a fraction of cells, chosen stochastically (based on the probability of cell division), were allowed to divide (Figure S3I). The probability of cell division per generation varied from 0.1 to 1. This parameter only significantly affected cluster-size distribution when the probability was less than 0.7. A probability value of 0.7 or less implies that a significant fraction (>30%) of the cells are lagging by unit division time, which is more than the variation in doubling time for cells grown under the conditions of our experiments.

Cts1p performs the majority of the chitin hydrolysis activity in yeast and is controlled by the Ace2p transcription factor. Controlling Cts1p expression independently of Ace2p allows for specific control of the chitin hydrolysis activity, which influences the link-breaking rate ($\lambda$). We tested the effect of varying the link-breaking rate on cluster-size distribution by using a strain whose Cts1p expression can be tuned by altering the $\beta$-estradiol concentration in the media (Figures 2G and S3D). If links are hydrolyzed progressively, their fracture probability will increase with age, making older links more likely to break than younger ones. To confirm that an exponential distribution captures the decay statistics of the links, we titrated the rate of decay (Cts1p expression levels) and tested if the cluster size follows a predictable distribution. If an exponential distribution governs the probability of link breakage, the clusters will register a fracture event at intervals that follow a gamma distribution. Therefore, increasing the expression of Cts1p, which increases the rate constant of the exponential distribution, would lead to a reduced time interval between two fracture events. The length of this time interval determines the size of the cluster during constant cellular growth and follows a gamma distribution. As addition of cells leads to discrete increases in cluster size, we used a special version of gamma distribution: Erlang’s distribution. For random variables that follow Erlang’s distribution, the mean and standard deviation are linearly correlated, and for our model, the slope should be $1/k$. We observed that the means and the standard deviations of the cluster-size distributions, produced at different levels of CTS1 expression, followed Erlang’s distribution (Figure 2H), supporting the choice of exponential distribution for modeling the time for which a link survives.

Because the SA model predicts that clusters preferentially fragment at their older links, it implies that the two daughter clusters should be of similar size. The ratio between the size of the smaller of two daughter clusters ($S_{\text{daughter}}$) and their mother cluster ($S_{\text{mother}}$) gives an indication of whether the breaking mechanism is biased toward the oldest link. If the cluster always breaks at its oldest link the ratio is very close to 0.5. The SA model predicted that the distribution would be biased toward higher ratios. We tested this prediction by time-lapse imaging of $BUD4$-OFF clusters labeled with CFP (cyan fluorescent protein) during their growth and fragmentation cycles (Figure 3A). The clusters were imaged at a low magnification to increase the depth of field, allowing us to use the total signal from a cluster as a proxy for the number of cells it contains (Figures 3A, S4A, and S4B). Consistent with our prediction, we found that the ratio ($S_{\text{daughter}}/S_{\text{mother}}$) distribution is biased toward higher ratios (Figure 3B). We tested whether the SA model accurately captures the details of cluster fracture: the distribution of $S_{\text{daughter}}/S_{\text{mother}}$ measured experimentally (Figure 3C). To accomplish this, we simulated the dynamic network models for SA and SR (for comparison) and obtained the distribution of $S_{\text{daughter}}/S_{\text{mother}}$. We compared this distribution to the experimental quantile values using root-mean-squared error (RMSE) as a figure of merit. This analysis was performed using a link-breaking rate in accordance with the range obtained from Figures S3G and S3H. The SA model showed good correspondence with the experimental distribution (Figure 3C) while the SR model did not (Figure 3D). This result is consistent with the proposition that the SA model captures both the overall statistics of the cluster-size distribution (Figure 2C) and the fracture-level details. Sensitivity analysis of the results with respect to link-breaking rate showed comparatively high correspondence between the SA model fracture statistics and experimental estimates at the higher limit of the link-breaking rate (Figures S4C–S4F). The ability of the SA model to accurately capture the distribution of the number of cells per cluster resembles the weak-link theory in predicting cluster...
volumes. In principle, the cluster volumes predicted by this theory should be proportional to the number of cells we predict using our model.

Our results show that the SA model captures the overall distribution of cluster sizes and the details of fracture. However, it does not provide a biophysical mechanism that sensitizes older links toward breakage. Oldest links are more likely both to be densely packed and to have undergone increased chitin degradation (due to remnant chitinase activity). It remains to be determined which of these forces are primarily responsible for this age-dependent behavior of links.

The ratio between growth rate and link-breakage rate determines cluster size

We examined the effect of growth and link-breaking rates on cluster-size distribution. There are two regimes: when the link-breaking rate is low, steric limitations govern cluster size and altering either the growth or link-breaking rates has little effect, and higher link-breaking rates, where slowing the growth rate produces substantial decreases in cluster size (Figure 4A). We lowered the glucose concentration (to 0.025% or 0.05%) to increase the doubling time (Figure S5C). In addition, we uncoupled Cts1p production from its normal transcription factor, Ace2p, in two ways: deleting ACE2 or expressing CTS1 from a β-estradiol-regulated promoter. In both manipulations, the rate of chitin hydrolysis should be independent of the growth rate. We measured the cluster-size distribution in a BUD4-OFF ace2Δ strain growing exponentially in different concentrations of glucose. Consistent with our hypothesis, the cluster-size distribution shifted to smaller clusters (reflected in the mean of the distribution) in the lower glucose concentrations (0.025% or 0.05%) (Figures 4B and S5A).

To study the effect of growth rate on the mean cluster size, we explored parameter space and asked how the kissing number and link-breaking rates influence the relation between growth rate and mean cluster size. Simulations suggested that, for a fixed cell division rate, the mean number of cells per cluster increases monotonically as the link-breaking rate falls (Figure 4C). Thus, the relative values of the link-breaking rate (β) and cell division rate (λ) alter the mean size of the clusters for a fixed kissing number. The kissing number (κ) sets the maximum value consistent with the idea that clusters cannot accommodate cells beyond κ.

We simultaneously varied the link-breaking rate (β, by altering the estradiol concentration) and the cell division rate (λ, by varying glucose concentrations) to test the validity of this model. The cell division rate (λ) was calculated from the growth curves for bulk cultures in different glucose concentrations (Figure S5B). We examined all the combinations of three different β-estradiol concentrations, 0, 4, and 8 nM, with three different cell division rates. We found that the cluster-size distribution shifted to smaller clusters (Figure 4D).
rates ($\lambda$), 0.36 h$^{-1}$, 0.3 h$^{-1}$, and 0.25 h$^{-1}$. The mean number of cells per cluster was used as a proxy for the effect of cell division rates and link-breaking rate on cluster-size distribution. At a given link-breaking rate, increasing the growth rate increased cluster size, and at a given growth rate, increasing link-breaking rate decreased cluster size (Figures 4D and 4E). We predicted that the effect of increasing the link-breaking rate should be stronger at lower growth rates, but this is not what we see. Our interpretation is that there are two components to the link-breaking rate, one due to the expression of CTS1 and the other due to the expression of other hydrolytic enzymes. At the lowest growth rate, 0.25 h$^{-1}$, we argue that the CTS1-independent link-breaking rate is already high enough to substantially diminish both the size of the clusters and the response to graded CTS1 expression (Figures 4D and 4E). Leaky expression of CTS1, even in the absence of $\beta$-estradiol, is likely to explain why cluster sizes do not converge for various growth rates as predicted by the simulations (Figure 4C). Changing the growth rate by growing cells in either 2% galactose or 0.02% galactose also led to a decrease in the mean cluster size with growth rates (Figure S5D).

Higher kissing number minimizes compositional heterogeneity in differentiating multicellular clusters

A three-step model for the evolution of clonal multicellularity posits that the first step is the failure of cell separation, the second is the division of labor between cells, and the third is the separation of germ and somatic lineages, with irreversible differentiation from the faster dividing germ cells to more slowly dividing somatic cells. When the presence of both cell types maximizes the rate of reproduction, clusters that are all somatic...
Cells are disadvantaged because they can never produce germ cells.

We studied how the parameters governing cluster size affect the distribution of germ and somatic cells using an engineered yeast strain that mimics irreversible differentiation. In this strain, the germ cells express a cycloheximide-resistance gene that can be used to set the relative division rates of germ and somatic cells, and Cre-mediated recombination converts germ cells into somatic cells by excising this gene. In its original form, the somatic cells produced a public good by converting sucrose into glucose and fructose, but we performed our experiments in glucose so that the germ and somatic cells had the same metabolic capabilities. The two cell states are distinguished by the fluorescent protein they express: germ cells express RFP and somatic cells express YFP. We define the switching rate from the germ state to the somatic state as \( p \) and the relative growth advantage of germ cells with respect to somatic cells as \( m \). \( p \) can be experimentally modulated by changing the expression of Cre-recombinase using \( \beta \)-estradiol and \( m \) can be controlled by changing the cycloheximide concentration.

Sterile clusters, which only contain somatic cells, are formed when the link between a cell that has differentiated and its undifferentiated mother cell breaks (Figure 5A). When the kissing number is low, the number of cells connected to a given cell will be low. We therefore expect the probability of the differentiating and fracturing branch coinciding to be high even if differentiation and breakage are independent processes. As the kissing number rises, this probability falls as there will be more branches to distribute the events among. We modified the SA model to incorporate the dynamics of differentiation. Briefly, when a new node is added to a growing network, it is assigned a phenotypic state (germ or somatic) based on the value of the kissing number and the expression of Cre-recombinase.

Figure 5. Kissing number (\( k \)) affects compositional heterogeneity in differentiating clusters
(A) Two scenarios for the fragmentation of clusters where cells can differentiate irreversibly: in scenario 1, one daughter cluster contains only germ cells and the other only somatic cells; in scenario 2, both daughter clusters contain both cell types.
(B) Simulations with the SA-based model at two different values of the ratio of the growth rates between germ and somatic cells, \( \mu_0 \). The effect of increasing the kissing number and relative growth advantage on the proportion of cluster types: only germ, only somatic, or mixed.
(C) Experimental composition of clusters (germ only, somatic only, and mixed) with varied kissing number and relative growth rates determined experimentally (\( n = 3 \)).
(D) Images for clusters with two cell types for various conditions shown in (C); dotted circles show mixed clusters. See also Figure S6.
of $\pi_{g\rightarrow s}$. Because differentiation is irreversible, $\pi_{s\rightarrow g} = 0$. When germ cells have no growth advantage over somatic cells ($\mu_s = 1$), increasing the kissing number only marginally reduced the fraction of sterile clusters that solely contained germ cells (Figure 5B). When germ cells divide faster ($\mu_g > 1$), the effect of kissing number on the fraction of sterile clusters was more pronounced (Figure 5B). Simulations also predicted that increased kissing number would increase the fraction of clusters having both cell types (mixed).

To experimentally test these hypotheses, we modified the kissing number of differentiating clusters. We could not delete BUD4 because its effect on kissing number is dependent on the absence of ACE2 and ACE2 is required for the differentiation system used here. Instead, we deleted either BUD2 (bud2Δ cells bud at random sites on cell surface) or BUD3 (bud3Δ cells bud at bipolar sites) in a cts1Δ background. These clusters have increased kissing numbers, and thus increased maximum attainable size, due to a switch from axial budding to bipolar or random budding. The fraction of clusters that have exclusively germ or somatic cells or a mixture of both can be identified by flow cytometry based on the total RFP or YFP signal emitted for each cluster (Figure S6). We measured the fraction of each cluster type in glucose-containing medium at increasing concentrations of cycloheximide (which elevates $\mu_s$) for both cts1Δ and bud2Δ cts1Δ clusters. Even though the system was designed to exploit sucrose hydrolysis as a public good,12 we used glucose as the sole carbon source to ensure that both cell types had equal metabolic capacities. This was to eliminate any selection of cluster types due to growth on sucrose.12 Consistent with the prediction from simulations (Figure 5B), increasing the kissing number (comparing cts1Δ and bud2Δ cts1Δ clusters) reduced the fraction of sterile clusters (Figure 5C). We confirmed these results using microscopy (Figure 5D). The fraction of clusters that are exclusively germ cells disagreed between the simulations and the experiments: while the model predicted that this fraction would fall at high kissing numbers, the experiment revealed an increase. This could be due to missing interactions between the variables in the differentiating version of the dynamic networks or second-order effects of the perturbations or differences between model parameters (switching rates and growth rates) and the experimental set-up.

We asked whether the model parameters influence the ratio of the cell types in individual clusters and the variability in this ratio across clusters in the population, and whether they can be modulated in a predictable manner. In conditions where metabolic division of labor or cross-feeding occurs, the ratio of two cell types in a cluster is particularly relevant.31,32 An unbalanced ratio would underfeed one of the two cell types and reduce the performance of the cluster. The cell-type ratio is relevant in any scenario where cell types perform specific tasks that influence overall fitness of the multicellular organism.

We modified our imaging pipeline (Figures S2A–S2E) to measure the ratio of germ and somatic cells in each cluster. This change allowed us to determine the fraction of a given cell type in individual clusters based on YFP and RFP signals across hundreds of clusters in a population and for several different perturbations (Figure S7A). First, we tested how parameters specific to differentiating clusters, i.e., $\pi_{g\rightarrow s}$ and $\mu_g$, alter this mean fraction of somatic cells per cluster. We express the net effect of these two parameters as $\pi_{g\rightarrow s} = \pi_{g\rightarrow s}/\mu_g$. Therefore, $\pi_{g\rightarrow s}$, the net switching rate, describes the net effect of the formation ($\pi_{g\rightarrow s}$) and slower proliferation ($\mu_g$) of the somatic cell type. The SA dynamic network model predicted that the mean fraction of somatic cell type in clusters would monotonically increase with the net switching rate (Figure 6A). As the model predicts, any desired ratio could be attained by fixing the system at a given net switching rate ($\pi_{g\rightarrow s}$). This relation seemed to be largely independent of both the kissing number ($\kappa$) and the link-breaking rate ($\beta$). To test the validity of this relation, we set up an experimental system to alter the net switching rate by varying the concentrations of β-estradiol and cycloheximide. We used the switching rates and relative growth rates previously reported for the differentiation system.12 We measured the fraction of somatic cells for hundreds...
of clusters per sample, for all combination of strains, switching rate, and relative growth advantage, a total of 9 conditions. Consistent with model predictions, the mean somatic fraction across clusters scaled monotonically with net switching rate independent of the strain background (Figure 6B). We note a difference in the curvature of mean somatic fraction versus the net switching rate (Figure 6B) relative to simulations (Figure 6A), which could be attributed to missing parameters in the model or second-order effects of cycloheximide or β-estradiol.

The fraction of somatic cells changed from 0.25 for the lowest $\sigma_{g-\lambda}$ to 0.8 for the highest $\sigma_{g-\lambda}$. We also explored the cluster-to-cluster variability for each condition. As this fraction is bounded between 0 and 1, we used the Gini coefficient to measure cluster-to-cluster variability in the fraction of somatic cells. The Gini coefficient is adopted from econometrics where it is used to quantify income variability within populations. We found a strong negative correlation between mean somatic fraction (Figure 6C) and the Gini coefficient for all the conditions measured ($r^2 = 0.87$). This implies a scaling relation between mean and variability like that observed for the mean and noise in gene expression. A higher somatic fraction automatically leads to a reduction in cluster-to-cluster variability. We obtained a similar relationship using conventional measures like CV, which is typically used for unbounded measurements (Figure S7B). Changing kissing number or link-breaking time in the model did not alter this correlation (Figures S7C–S7F).

**DISCUSSION**

Multicellularity has evolved multiple times in nature and has been studied in the lab by experimental evolution and engineering. These studies prompted us to examine the variables that determine the size and composition of multicellular clusters in budding yeast using simulation and experiments. The model that best predicted the experimental data relied on a combination of steric constraints and an age-dependent fall in the probability of a link surviving. We found that clusters exist in two regimes: in one the link-breaking rate is lower than the cell division rate and the kissing number sets the cluster size, whereas in the other, the link-breaking rate is greater than the division rate, leading to smaller clusters. Finally, the compositional heterogeneity of differentiating clusters was minimized by increasing the kissing number and the growth advantage of germ over somatic cells.

Experiments that have evolved multicellularity show remarkable differences in cluster-size distributions across independent lines. While several populations show convergent evolution, like the inactivation of ACE2, cluster-size distribution varies across evolution lines. Our simulations and experiments show that cluster-size distribution and composition can be predicted from a small number of parameters. Multicellular growth cycles involve the growth of cells as an outcome of cellular metabolism and fractures of the linkage between cells as an outcome of chemical decay of the links and physical forces. While the details of the location and magnitude of physical forces are an interesting question, they are dispensable for predicting cluster-size distribution. We argue that the few phenomenological parameters used in our model, which can be directly linked to specific proteins or sets of proteins, are sufficient to recapitulate the experimentally observed distribution across a set of conditions. The model, with additional features, could potentially help in predicting which specific parameter (kissing number, link-breaking rate, or cell division rate) would be most likely to change to produce a cluster-size distribution that maximizes fitness.

In this study, the dynamic network model was programmed such that all three of its parameters, kissing number, link-breaking rate, and cellular growth rate, are independent of each other. One possible extension of this model is a case where the model parameters are interdependent. For example, using this model, one can simulate what would happen if the cell division ceased at the core of the cluster where high cell wall stress is present. In this scenario, the cell division rate ($\lambda$) is set to zero when the number of connections a cell makes reaches the kissing number ($\kappa$). Simulations suggest this relationship allows the clusters in the population to accommodate a greater number of cells (Figure 7) than would be possible if all the parameters were independent (SA model). We speculate that such correlations might be present in evolved multicellular clusters with sizes much greater than those observed in this study. The effect of increasing kissing number due to changes in cell aspect ratio must also be considered for testing this prediction. We did not model the effect of changes in cell aspect ratio on cluster-size distribution due to its non-linear relation to kissing number and the necessity to account for the physical location of cells within a growing cluster, two parameters that are difficult to experimentally measure across a population of clusters. A fully mechanistic model that incorporates details of cell location, the actual mechanics of fracture, and cell shape is likely to be more accurate than the models presented in this paper.

A limitation of our study is the assumption of tree-like growth of the clusters of cells, which is idiosyncratic to multicellular yeast. In more complex life forms, the cells in the clusters are cemented together by an extracellular matrix comprising collagen, cellulose, and other polymers. This allows cells to form connections with multiple other surrounding cells, unlike yeast clusters where a cell is only linked to its mother and its daughters. Moreover, cells can sense mechanical forces around them and use this information to control cell division. This regulation and external cues that control the rate and planes of cell divisions allow them to form non-symmetric 3D structures like organs. Given the ease
of adding more features to the dynamic network model, i.e., as attributes to nodes and edges, it would be interesting to ask what other features need to be incorporated to explain more sophisticated multicellular phenotypes. For example, would incorporating morphogen sensing \textsuperscript{38} as a node attribute based on geometric location allow us to predict cellular patterns?

\textbf{STAR+METHODS}

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead contact
  - Materials availability
  - Data and code availability
- **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**
  - Yeast strains and media
- **METHOD DETAILS**
  - Estimating number of cells per clusters through high-throughput imaging
  - Timelapse imaging
  - Confocal laser scanning microscopy of 3D clusters
  - Flow cytometry and sorting
  - \textit{j}-estradiol titration
  - \textit{j}-estradiol Cre induction and differentiation
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
  - Dynamic network simulations and analysis
  - Segmentation and estimation of cluster attributes
  - Measuring mother-daughter cluster size correlation
  - Estimating fraction of somatic, germ and mixed clusters
  - Statistical calculations

\textbf{SUPPLEMENTAL INFORMATION}

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2024.05.014.

\textbf{ACKNOWLEDGMENTS}

The authors thank all members of the Murray lab for their feedback during lab meetings. We would also like to thank Boris I. Shraiman, Michael T. Laub, Jane Kondiev, Alon Klein, and Michael M. Desai for critically reading the manuscript and for their feedback. We want to thank the Harvard F.A.S. staff, especially Zach Niziolek from the Bauer flow cytometry core facility, for assistance in FACS and flow cytometry assistance. We thank Christian Heinegel from Harvard Biological Imaging Center (HCBI)/Zeiss. We sincerely thank Dr. and Mrs. Daniel Simmons for awarding the Simmons Award to P.N. for performing advanced microscopy in HCBI. P.N., J.B., and A.W.M. were supported by NIH/NIGMS R01 GM043987. A.W.M. is also supported by the NSF-Simons Center for the Mathematical and Statistical Analysis of Biology (NSF #1764269, Simons #594598). T.L. was the Candy and William Raveis Fellow of the Damon Runyon Cancer Research Foundation while this research was performed. Figure 1D was created with BioRender.com.

\textbf{AUTHOR CONTRIBUTIONS}

Conceptualization, P.N. and A.W.M.; methodology, P.N., J.B., and A.W.M.; investigation, P.N., J.B., and T.L.; writing – original draft, P.N. and A.W.M.; writing – review & editing, P.N., J.B., T.L., and A.W.M.; funding acquisition, A.W.M.; supervision, A.W.M.

\textbf{DECLARATION OF INTERESTS}

A.W.M. is on the advisory board of \textit{Current Biology}.

Received: November 1, 2023
Revised: March 20, 2024
Accepted: May 8, 2024
Published: May 31, 2024

\textbf{REFERENCES}


Please cite this article in press as: Nanda et al., A dynamic network model predicts the phenotypes of multicellular clusters from cellular properties, \textit{Current Biology} (2024), https://doi.org/10.1016/j.cub.2024.05.014


STAR METHODS

KEY RESOURCES TABLE

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemicals, peptides, and recombinant proteins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>β-estradiol</td>
<td>SIGMA ALDRICH</td>
<td>Cat#E8875</td>
</tr>
<tr>
<td>Concanavalin A</td>
<td>MP BIO</td>
<td>Cat#02195283</td>
</tr>
<tr>
<td>Cycloheximide</td>
<td>SIGMA ALDRICH</td>
<td>Cat# 239763-M</td>
</tr>
<tr>
<td>Yeast Nitrogen Base</td>
<td>BD (Becton Dickinson)</td>
<td>Cat# 233520</td>
</tr>
<tr>
<td>Calcofluor White</td>
<td>SIGMA ALDRICH</td>
<td>Cat#18909</td>
</tr>
</tbody>
</table>

Deposited data

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Python, MATLAB and SLURM scripts</td>
<td>This study</td>
<td><a href="https://doi.org/10.5281/zenodo.11007363">https://doi.org/10.5281/zenodo.11007363</a></td>
</tr>
<tr>
<td>Raw Image Files</td>
<td>This study</td>
<td><a href="https://doi.org/10.5281/zenodo.11007610">https://doi.org/10.5281/zenodo.11007610</a></td>
</tr>
</tbody>
</table>

Experimental models: Organisms/strains

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yeast Strains</td>
<td>This study</td>
<td>Table S2</td>
</tr>
</tbody>
</table>

Recombinant DNA

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERBD plasmid</td>
<td>Addgene</td>
<td>Addgene ID #58431</td>
</tr>
<tr>
<td>β-estradiol plasmid</td>
<td>Addgene</td>
<td>Addgene ID #58434</td>
</tr>
</tbody>
</table>

Software and algorithms

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yeast segmentation Code</td>
<td>GitHub</td>
<td><a href="https://github.com/alexxijielu/yeast_segmentation.git">https://github.com/alexxijielu/yeast_segmentation.git</a></td>
</tr>
<tr>
<td>MATLAB 2022a</td>
<td>Mathworks</td>
<td>N/A</td>
</tr>
<tr>
<td>Fiji</td>
<td>ImageJ</td>
<td>N/A</td>
</tr>
</tbody>
</table>

RESOURCE AVAILABILITY

Lead contact
Further information or request for resources should be forwarded to the lead contact, Andrew Murray (awm@mcb.harvard.edu).

Materials availability
Yeast strains and plasmid built for this project are available on request. Please contact the lead contact for further details.

Data and code availability

- Original microscopy data with the associated metadata have been uploaded to Zenodo. The DOI is listed in the key resources table.
- All the scripts used for processing data or generating figures are uploaded to GitHub. A version of record of the GitHub repository has been uploaded to Zenodo. The DOI is listed in the key resources table.
- Any additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Yeast strains and media
All experiments were performed using strains constructed from a modified W303 background. Briefly, the functional version of BUD4 allele from S288C was engineered into the endogenous locus in W303. The β-estradiol-responsive transcription factor was introduced at the HIS3 locus through an integrative plasmid. Genes of interest were put under the control of β-estradiol promoter by replacing their endogenous promoter by homologous recombination. Constitutively expressed fluorescent markers were introduced at the HO locus using homologous recombination through the KanMX marker. Gene deletions were performed by amplifying the KanMX resistance marker with 500 bp upstream and downstream sequence from yeast deletion collection and introduced in the background of interest by homologous recombination. Confirmation PCR was performed to both check the integration of the resistance cassette and removal of target gene. Replica plating was performed...
for each successive transformation to confirm retention of previous markers. Table S2 contains genotype information of the strains used in this study.

Strains were grown in synthetic minimal media supplemented with the specified carbon source at various concentrations. Media were prepared freshly by diluting 10X stocks of the refrigerated constituents and prewarmed before inoculation. For every experiment, strains were freshly streaked out from \(-70^\circ C\) glycerol stocks (15\%) and allowed to recover for at least 2 days at 30°C. Strains were cultured in desired media for at least 10 generations before any measurements were performed. For some specific experiments, strains were grown for longer than 10 generations by maintaining them in exponential phase to allow the cluster size distribution to reach steady state. For experiments in the different concentrations of glucose, cells were periodically diluted back to fresh media to keep them proliferating exponentially. All the experiments were performed with two biological replicates for high-throughput imaging of size distributions. Replicate correlations have been described in supplementary figures.

**METHOD DETAILS**

**Estimating number of cells per clusters through high-throughput imaging**

Simulations using the models described in the main text produce a cluster size distribution, which we compared with experimental distributions. It is challenging to image individual cells in an entire three-dimensional cluster due to excessive light scattering and interference by the yeast cell wall. To overcome this challenge, the clusters were stained with calcofluor white (CW) and squeezed between a coverslip and a slide. The fluid volume between the coverslip and slide was optimized so that the cells from a single cluster formed a monolayer without lysing (Figure S2). We performed a timelapse of clusters disintegrating to a monolayer to check for any lysis event or mixing of cells from different clusters (Video S1). By optimizing the density of clusters in a field of view, we spatially resolved cells belonging to individual clusters in a monolayer (Figures S2B–S2E). We used a previously established Mask-RCNN pipeline (Region-based Convolutional Neural Network) images to segment yeast cells using images obtained from the CW channel (Figure S2). The codes and associated files were acquired from https://github.com/alexxijielu/yeast_segmentation. We collected multiple fields of view (>60 FOVs, i.e., \(\approx 200\) clusters) to obtain an accurate distribution of cluster sizes. The minimum number of clusters to be obtained to claim a detectable difference in two conditions or genetic backgrounds was calculated based on Cohen’s d (Figure S2F). For all the experiments, the expected effect size was <0.4 and therefore, 200 clusters were enough to derive a representative distribution. The centroid of cells in segmented images was estimated and used to cluster cells based on their spatial location using DBSCAN (Density-Based Spatial Clustering of Applications with Noise). The number of cells corresponding to each cluster was used to generate a discrete distribution of cluster sizes. We obtained reproducible cluster distributions for various genetic backgrounds (corresponding to size distributions) (Figures S2G and S2H).

Strains were grown in the indicated media for at least 10 generations to reach their steady state distribution of cluster sizes. For imaging, exponentially grown cells were incubated with 1 volume of calcofluor white for 1 min at room temperature. Approximately, 2.8 \(\mu\)L of the mixture was squeezed between a microscope slide and a 20 mm \(\times\) 20 mm coverslip. The coverslip and microscope slide were cleaned with an air jet before experiments to remove dust particles. A Nikon Eclipse TiE inverted fluorescence microscope with a Hamamatsu EMCCD camera was used for all imaging. The microscope was controlled through MetaMorph software. An automated script written in MetaMorph scanned 49 positions on the slides in an arrayed fashion. The density of cultures was optimized to yield at least 4–5 clusters per field of view. Monolayer formation was manually verified for each sample by imaging z-stacks and confirming the absence of more than one layer of cells.

**Timelapse imaging**

All timelapse imaging was performed using the microscope described above. Concanavalin-A coated wells in a 384 well plate were used to perform lineage tracing of cluster growth and division using a 10X objective with 10 mm working distance. Fluorescence from a constitutively expressed CFP was used as a proxy for total number of cells per cluster. The depth of field was calculated using excitation wavelength and magnification values. The plates were filled with at least 40 \(\mu\)L of media to reduce evaporative loss. Strains were grown in specified media for 10 generations and were spun down on a 384 well plate at 300 g for 2 min. Images were taken every 10 min for a 12 h period.

**Confocal laser scanning microscopy of 3D clusters**

A Zeiss LSM900 microscope was used for 3D imaging of multicellular clusters. Briefly, cells were stained with calcofluor white and loaded onto a concavity slide in 0.2% agarose to keep clusters stationary during imaging. The clusters were imaged using the DAPI channel with a 2 stack width of 0.2 \(\mu\)m. The 3D images were rendered using Fiji (ImageJ) and custom scripts.

**Flow cytometry and sorting**

All flow cytometry experiments were performed using a BD LSR II. The photomultiplier tube (PMT) voltages for forward and side scatter were optimized for each experiment to record maximum signal from the biggest clusters. The area signal from forward scatter (FSC-A) was used as a proxy for cluster size for preliminary measurements. The flow cytometry tubes were vigorously vortexed to obtain a uniform suspension of clusters before loading the samples.
Cell sorting was performed on multicellular clusters using a BD FACS Aria II using signal from the FSC-A channel. The FSC-A distribution was segmented into 4 quartiles and sorted into 4 different collection tubes at RT with minimal media. The cluster fractions were grown in flasks at 30°C for 48 h. The evolution of the size distribution over time was recorded every 12 h using flow cytometry.

β-estradiol titration
Strains were grown overnight in minimal media without β-estradiol. They were sub-cultured next day into minimal media with desired beta-estradiol concentrations. β-estradiol stocks were prepared in 100% ethanol and stored at −20°C. The stock was diluted to desired concentrations in minimal media. All measurements were performed after overnight growth (at least 10 generations) in presence of inducer. The dynamic range of inducer concentrations was estimated by performing a log-scale dose-response curve with the cluster size distribution as the output. The relative link breaking time for Cts1p titration experiment (Figures 4D and 4E) was calculated as
$$\frac{1}{\beta} \ln \left( \frac{C_{0}}{C_{1}} \right)$$
This assumes that beta-estradiol concentration is directly proportional the Cts1p enzyme activity which sets the link breaking time.

β-estradiol Cre induction and differentiation
Strains were freshly streaked out from −70°C glycerol stock. Single colonies were picked and checked for spontaneous Cre recombination events. Colonies with the pure germ state were grown in complete synthetic media with 2% Dextrose (prepared from 10X stocks of constituents) overnight and diluted 100-fold into fresh media supplemented with desired concentrations of β-estradiol (inducing Cre recombinase expression) and cycloheximide. Cycloheximide stocks were made in 100% ethanol and stored at −20°C. The strains were grown for at least 14 generations (two transfers) to allow clusters to reach steady state fraction and size. Differentiation was confirmed by microscopy and flow cytometry.

The ratio of cell types in individual clusters was determined using a modified version of method described in Figure 2. Instead of calcofluor white, a constitutively expressed CFP present in both germ and somatic was used as segmentation marker and RFP (germ) and YFP (somatic) were used to identify the cell types. The number of cells of a given cell-type was estimated using the pipeline described in the quantification and analysis section.

QUANTIFICATION AND STATISTICAL ANALYSIS

Dynamic network simulations and analysis
Custom scripts were written in Python to perform dynamic network simulations using the NetworkX library (https://github.com/networkx/networkx). Briefly, a recursive program was run in which nodes (cells) were added every division time to extant nodes. Cell properties were encoded in node attributes and link properties were encoded in edge attributes. A function was used to update the attribute based on a trigger event: i) whether a kissing number was reached ii) the links reached a point of breaking. The network growth is initiated from a single cell but continues for at least 10 generations, after which at least 150 clusters have formed. Disconnected modules in the network were determined by finding set of nodes which don’t have connections with outside nodes. A disconnected module was identified as a cluster and the number of cells it contained was determined. Every simulation was performed for multiple random number seeds. The same random number seed was used for comparing effects of two different parameter sets on network properties. The growth rates of some of the mutant strains vary due to pleotropic effects of deleting ACE2. Therefore, the growth rates of ACE2 and ace2a strains were measured, and the appropriate values were used in creating simulations to compare with experiments. The Kolmogorov-Smirnov (KS) test was used to test the agreement between simulated and measured distributions. The purpose of this test was to ascertain the relative goodness of fit amongst different models (Figure 1C).

Segmentation and estimation of cluster attributes
Segmentation of cells in clusters were performed using a yeast optimized pipeline based on DeepRetina. The pipeline was executed in a supercomputing cluster (FASRC) using custom SLURM scripts and images acquired were directly transferred for segmentation without any preprocessing or selection. The aspect ratio of images was preserved, and images acquired through the DAPI channel or CFP was used for segmentation.

MATLAB 2022B was used for image processing and analysis. Briefly, centroids corresponding to cells in a cluster was determined and used for performing DBSCAN (Density Based Spatial Clustering And Noise). DBSCAN assigns the cells to specific cluster based on its geometric location. Clusters with cells close to the image border (in the 50-pixel proximity) were rejected from downstream analysis pipeline.

For estimating the fraction of somatic cells per cluster, the signal from YFP channel was used. Briefly, the cutoff for YFP ON was determined by thresholding. Cells having YFP intensity greater than the threshold were assigned ‘somatic state’. DBSCAN was performed as described earlier and fraction of cells belonging to somatic category was determined. The signal from RFP channel (corresponding to germ state) was not used due to remnant signal from degraded protein that interfered with accurate determination of germ cell status.

The MATLAB script was executed from SLURM in a supercomputing cluster and all images from the same experiment were processed in the same batch. No manual filtering of images were performed at any stage of this pipeline.
Measuring mother-daughter cluster size correlation
The net fluorescence from individual clusters was determined using Fiji and lineages were manually assigned. A custom script written in Python was used to calculate the mother cluster-daughter cluster size correlations. The total size of mother cluster was estimated by summing up the sizes of daughter cluster right after the breaking event was recorded.

Estimating fraction of somatic, germ and mixed clusters
Flow cytometry files were analyzed using FlowCytometryTools library in Python3.7. The threshold for YFP and RFP channel was determined from the histogram of size (FSC-A) normalized fluorescence values. A cluster was assigned ‘Somatic only’ type if the YFP intensity was greater than the threshold and RFP intensity was less than the threshold. Similarly, a cluster was assigned ‘Mix’ type if both the intensity values were greater than the corresponding thresholds. Separate thresholds were determined for different samples and strains.

Statistical calculations

\[
\text{Coefficient of Variation (CV)} = \frac{\text{Standard Deviation}}{\text{Mean}}
\]

\[
\text{Cohen's } d = \frac{\text{Mean}_2 - \text{Mean}_1}{\text{Pooled Standard Deviation}}
\]

\[
\text{Exponential Distribution : } P(X = t) = \lambda e^{-\lambda t}
\]