

RESEARCH ARTICLE

Open Access



# Data-driven multiscale modeling reveals the role of metabolic coupling for the spatio-temporal growth dynamics of yeast colonies

Jukka Intosalmi<sup>1\*</sup>, Adrian C. Scott<sup>2</sup>, Michelle Hays<sup>3</sup>, Nicholas Flann<sup>4</sup>, Olli Yli-Harja<sup>5,6</sup>, Harri Lähdesmäki<sup>1</sup>, Aimée M. Dudley<sup>2,3</sup> and Alexander Skupin<sup>7,8\*</sup> 

## Abstract

**Background:** Multicellular entities like mammalian tissues or microbial biofilms typically exhibit complex spatial arrangements that are adapted to their specific functions or environments. These structures result from intercellular signaling as well as from the interaction with the environment that allow cells of the same genotype to differentiate into well-organized communities of diversified cells. Despite its importance, our understanding how this cell–cell and metabolic coupling lead to functionally optimized structures is still limited.

**Results:** Here, we present a data-driven spatial framework to computationally investigate the development of yeast colonies as such a multicellular structure in dependence on metabolic capacity. For this purpose, we first developed and parameterized a dynamic cell state and growth model for yeast based on on experimental data from homogeneous liquid media conditions. The inferred model is subsequently used in a spatially coarse-grained model for colony development to investigate the effect of metabolic coupling by calibrating spatial parameters from experimental time-course data of colony growth using state-of-the-art statistical techniques for model uncertainty and parameter estimations. The model is finally validated by independent experimental data of an alternative yeast strain with distinct metabolic characteristics and illustrates the impact of metabolic coupling for structure formation.

**Conclusions:** We introduce a novel model for yeast colony formation, present a statistical methodology for model calibration in a data-driven manner, and demonstrate how the established model can be used to generate predictions across scales by validation against independent measurements of genetically distinct yeast strains.

**Keywords:** Multicellular systems, Multiscale modeling, Yeast colony, Metabolic coupling, Diauxic shift, Markov chain Monte Carlo, Bayesian optimization

## Background

Multicellular organisms and colonies of unicellular microbes are able to form characteristic structures by specialized cell types [1, 2]. While it is generally accepted that the structure and functions of tissue and organs are genetically encoded, more recently it has been demonstrated

that the morphologies of biofilms like *Saccharomyces cerevisiae* yeast colonies have a strong genetic component [3–5]. Together with the frequently observed growth medium dependency of yeast cultures, these findings further underpin the importance of genome–environment interactions for phenotype development [6]. To understand the underlying mechanisms, it is key to investigate how metabolic coupling is influencing individual cell states and instructing structure formation.

Yeast colonies and biofilms represent an efficient experimental model system to investigate how metabolic

\*Correspondence: [jukka.intosalmi@gmail.com](mailto:jukka.intosalmi@gmail.com); [alexander.skupin@uni.lu](mailto:alexander.skupin@uni.lu)

<sup>1</sup>Department of Computer Science, Aalto University, P.O.Box 15400, FI-00076, Aalto, Finland

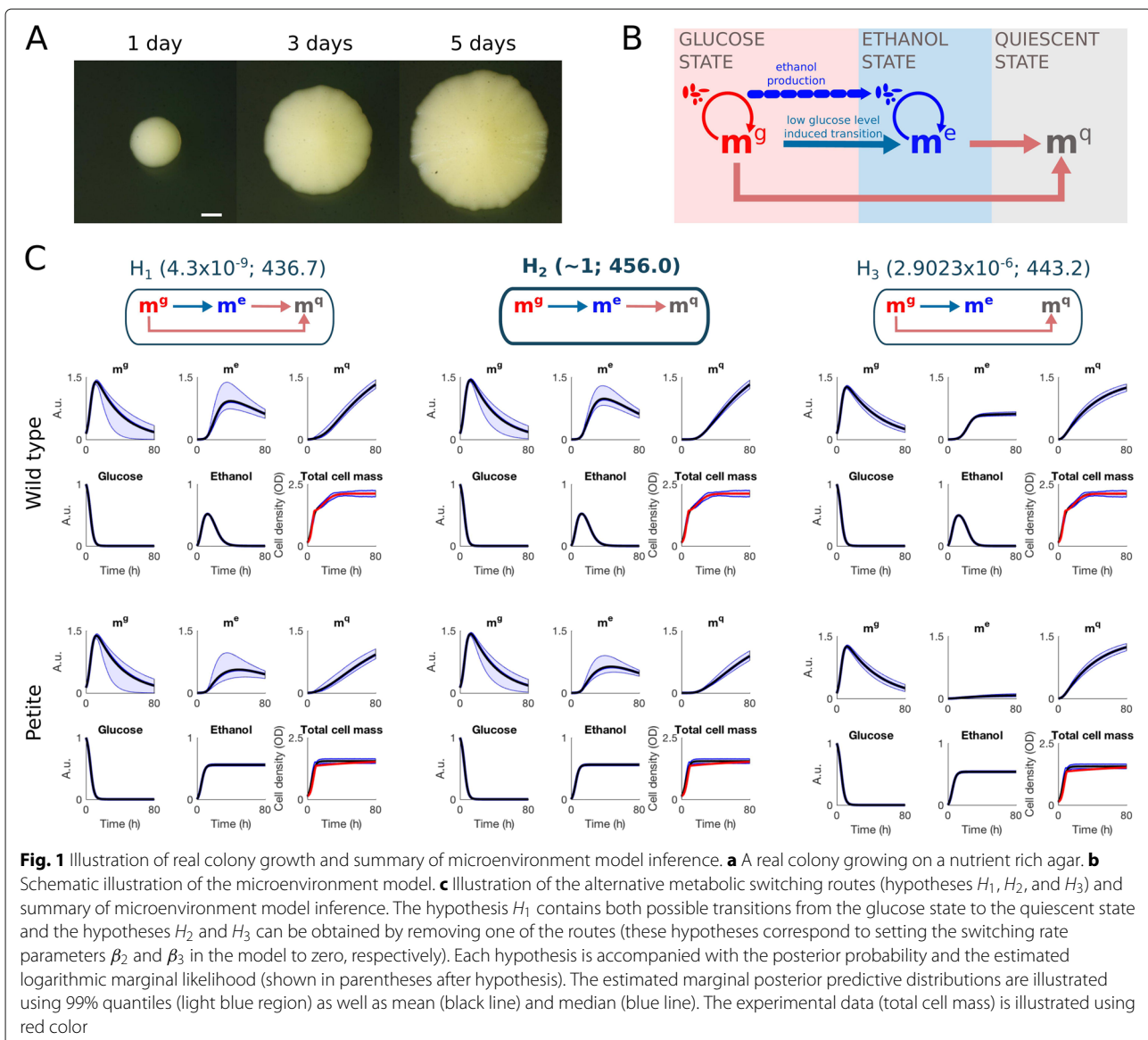
<sup>7</sup>Luxembourg Centre for Systems Biomedicine, University of Luxembourg, 2, avenue de l'Université, L-4365 Esch-sur-Alzette, Luxembourg

Full list of author information is available at the end of the article



dynamics and spatial coupling determine morphogenesis because yeast exhibits (i) cell state transition in dependence on the environment by switching from glucose to ethanol metabolism and quiescence [7], (ii) fast growth, and (iii) can be easily genetically modified (Fig. 1A, B). Although yeast is a unicellular organism, it can form rather complex colony structures including heterogeneous cell states in a strain specific manner [8, 9]. Recently, we have shown that predominant changes in morphology from smooth to wrinkled “fluffy” structures can be induced by aneuploidy as a multicellular phenotype switch [10]. Despite the systematic genetical characterization of this switch, the question how the gain or loss of a chromosome copy leads to a significant change in morphology is not understood.

Mathematical modeling can provide essential insights into the underlying processes as it allows quantitative investigation of the coupling between metabolic and spatial growth dynamics. A general challenge is thereby to cover and parameterize the relevant scales ranging from intra- and intercellular interactions to population and environment dynamics. Existing multiscale modeling approaches for complex multicellular systems typically rely on large sets of physiological parameters that are often not easily accessible in experiments [11, 12]. Other spatiotemporal modeling approaches are based on homogeneity assumption and simulate partial differential equations neglecting the discrete properties of cells. While being useful in building a general understanding of different mechanisms across the scales, most of these



approaches do not allow for direct experimentally-based model construction and validation. Such experimental data driven model constructions have been successfully applied in the context of mechanistic modeling of molecular mechanisms [13–15] and extending these approaches to more complex multiscale models will be essential for methodological advancement in systems biology [16].

Here, we develop such a new multiscale modeling framework for multicellular yeast structure formation that allows for experimentally-based model construction and validation. In contrast to previous approaches that simulate individual cells [17], our framework is based on an approximation that discretizes the spatial domain into elementary cubes and allows us to model the heterogeneous microenvironment dynamics under the local homogeneity assumption. Furthermore, the elementary cube approximation enables us to model the information flows (like nutrient transport or the flow of signaling molecules) and mass transfer (movement of the growing cell mass) by means of computationally efficient flux mechanisms. The presented model represents a first approach to simulate colony growth in a data driven manner but does not address aneuploidy particularly as the underlying mechanism at this stage.

To construct a growth and cell state model for the homogeneous microenvironment dynamics, we combine ordinary differential equation (ODE) modeling with experimental data using advanced statistical techniques and, by means of this objective approach, infer the metabolic switching mechanisms as well as the corresponding model parameterization directly from the data. The calibrated microenvironment model is subsequently embedded into the spatial framework which allows for predictions of cell mass, cell state, nutrient, and metabolic distributions throughout the colony formation process after model calibration by colony growth data.

Our model construction process utilizes measurements from two different yeast strains. First, we calibrate the model using time-course data from wild-type yeast cells (YAD145) and subsequently the calibrated model is validated against independent measurements from a respiratory deficient (petite) yeast strain (YAD479). These genotypically different training and validation strains are known to result in distinct colony morphologies and therefore the validation approves that our multiscale model captures essential mechanisms across the scales spanning from microenvironment dynamics to the spatiotemporal colony formation dynamics.

## Results

### Dynamic model construction for cell growth and metabolic switching in homogeneous medium

Depending on external conditions and their intracellular state, yeast cells can either metabolize glucose or

ethanol for growth or remain in the so-called quiescent state. The diauxic shift between the different metabolic states is determined by nutrient sensing pathways and if the extracellular glucose level becomes low, cells change their metabolic wiring towards a state that allows growth on ethanol produced during growth on glucose [7, 18]. Cells can also switch to a quiescent state in which they act as passive by-standers that do not grow nor produce any aromatic alcohols. The metabolically distinct glucose, ethanol, and quiescent cell states are the starting point in our model construction and a schematic illustration of the dynamic interactions between these states is shown in Fig. 1B.

The dynamics of the different cellular metabolic states cannot be easily observed directly but it is rather straightforward to monitor cell growth by optic growth curve measurements [19] (see “Methods” section). With the help of mathematical modeling, we are able to infer the switching behavior between the metabolic states and the related nutrient dynamics from time-course data. This is done by constructing alternative quantitative growth models with different metabolic switching mechanisms between the states and testing these hypothetical models against time-course data by means of statistical techniques. In the following, we construct a mathematical model that describes yeast cell growth on glucose and ethanol and couples the growth dynamics with transient switching between three distinct metabolic states: (i) glucose, (ii) ethanol, and (iii) quiescent state (Fig. 1B).

We model the cell growth and switching between different metabolic states by ODEs. We start by considering the glucose state in which the cells grow on glucose and denote the cell mass in this state by  $m^g$ . Given that glucose intake is sufficiently fast, the cell mass dynamics in the glucose state can be modeled as

$$\frac{dm^g}{dt} = \mu_1 m^g g - \beta_1 \frac{1}{g + K} m^g - \beta_2 m^g, \quad (1)$$

where  $g$  denotes the level of available glucose and the first term,  $\mu_1 m^g g$ , describes the actual growth kinetics with the rate parameter  $\mu_1$ . If the glucose signal drops to a low level, cells start to switch gradually to the ethanol state. This switching is reflected by the second term in Eq. 1 with the switching rates  $\beta_1$  and  $K$ . Analogously, the third term in Eq. 1 describes the potential switching to the quiescent state with the rate parameter  $\beta_2$ . In a typical experimental setting, a fixed amount of glucose is provided to cells in the beginning and the glucose level decreases when it is used for growth. Subsequently, the glucose concentration is governed by

$$\frac{dg}{dt} = -\frac{\mu_1}{\gamma_1} m^g g, \quad (2)$$

where  $\gamma_1$  is a parameter that determines the yield of glucose to the produced biomass. Growth in the ethanol state occurs in an analogous manner as in the glucose state. We denote the cell mass in the ethanol state by  $m^e$  and the cell mass dynamics in this state is modeled as

$$\frac{dm^e}{dt} = \mu_2 m^e e + \beta_1 \frac{1}{g + K} m^g - \beta_3 m^e. \quad (3)$$

Here, the first term describes the actual growth kinetics with the rate parameter  $\mu_2$ , the second term corresponds to the cell mass entering the ethanol state from the glucose state, and the third term describes the possible switching from the ethanol state to the quiescent state with the rate parameter  $\beta_3$ . Ethanol is typically not added to a cell culture, but it is produced as a by-product of growth on glucose. Thus, the ethanol dynamics is given by

$$\frac{de}{dt} = \frac{\mu_1}{\gamma_2} m^g g - \frac{\mu_2}{\gamma_3} m^e e, \quad (4)$$

where the first term represents ethanol production during the growth on glucose and the second term considers the decrease due to biomass production. The parameters  $\gamma_2$  and  $\gamma_3$  determine the production and decrease, respectively. The above expressions for  $m^g$  and  $m^e$  dynamics include switching to a quiescent state. We denote the cell mass in the quiescent state by  $m^q$  and describe the cell mass dynamics in this state by

$$\frac{dm^q}{dt} = \beta_2 m^g + \beta_3 m^e, \quad (5)$$

with the terms introduced in Eqs. 1 and 3. Given the three distinct metabolic states, the total cell mass reflecting directly the experimental time-course measurements is given by  $m = m^g + m^e + m^q$ . In experiments, cells are initially put in glucose rich medium and we therefore assume that all cells are initially in the glucose state and the initial glucose level is high. Consequently, we assume that only the model variables  $m^g$  and  $g$  have non-vanishing initial values. These properties are also used in the reparameterization of the mathematical model which is presented in detail in Additional file 1. The model output, i.e. the total cell mass as a function of time, is denoted by  $m(t, \theta)$  where  $\theta$  is a parameter vector containing the parameters that result from the reparameterization.

### Statistical inference for model parameters and metabolic transitions in homogeneous medium

The mechanisms that are included in the mathematical model are illustrated in Fig. 1B. The full model contains the essential transition from the glucose state to the ethanol state and allows the cells also to switch to the quiescent state directly from the glucose and ethanol states. However, detailed information about the switching mechanisms to the quiescent state is not available and, consequently, there remains notable uncertainty about the

routes that cells may use to enter the quiescent state. To treat this uncertainty accurately, we consider three alternative hypotheses ( $H_1$ ,  $H_2$ , and  $H_3$ ) regarding the switching routes between the metabolic states (schematic illustrations of corresponding switching models are shown in Fig. 1C) and investigate the feasibility of these hypotheses by quantitative statistical testing. In the following, we outline the experimental data used for model calibration and explain how we infer the structure and parameterization of the microenvironment model.

To obtain dynamic data on total cell mass that can be used in the microenvironment model inference, we measured growth curves for wild-type and petite yeast strains (see “Methods” section). The petite yeast strain differs genetically from the wild-type strain and is not capable to grow on ethanol [10, 20]. In the context of our microenvironment model, this means that the growth rate parameter  $\mu_2$  should tend to zero when the petite strain is considered but all other parameters can be expected to be shared between these two strains. Given this direct connection between the wild-type and petite strains, we can carry out the statistical inference using the wild-type data and subsequently test the predictive performance of our models against the petite strain which is not included in the model calibration.

For model inference, we first collect the wild-type growth curve data into the data vector  $D_k$ . The elements of this data vector contain the average total cell mass at time points  $t_k, k = 1, \dots, N$ . The average cell mass as well as the corresponding sample variances  $v_k$  are computed over 6 replicates (see Additional file 1: Figure S1 for details about data pre-processing). From previous studies [5, 18, 21] the relative fractions of cells in ethanol and quiescent states at steady state (reached in our setting at  $t_N = 80$  hours) can be taken to be approximately  $29 \pm 6\%$  and  $62 \pm 6\%$ , respectively. We denote these relative fractions by  $\alpha^e = 0.29$  and  $\alpha^q = 0.62$  and the corresponding standard deviations representing uncertainty about the exact values by  $\sigma_{\alpha^e} = 0.02$  and  $\sigma_{\alpha^q} = 0.02$ . These wild-type data, which are used in the model calibration and hypothesis testing, can be combined with the model output under alternative metabolic switching hypothesis  $H_1$ ,  $H_2$ , and  $H_3$  by assuming independent normally distributed measurement errors and defining the likelihood function

$$\begin{aligned} \pi(D|\theta_{H_i}, H_i) &= \prod_{k=1}^N \mathcal{N}(D_k | m_{H_i}(t_k, \theta_{H_i}), v_k) \\ &\times \mathcal{N}\left(\alpha^e \left| \frac{m_{H_i}^e(t_N, \theta_{H_i})}{m_{H_i}(t_N, \theta_{H_i})}, \sigma_{\alpha^e}^2\right.\right) \\ &\times \mathcal{N}\left(\alpha^q \left| \frac{m_{H_i}^q(t_N, \theta_{H_i})}{m_{H_i}(t_N, \theta_{H_i})}, \sigma_{\alpha^q}^2\right.\right), \end{aligned} \quad (6)$$



where  $D = \{D_k, v_k, \alpha^e, \sigma_{\alpha^e}, \alpha^q, \sigma_{\alpha^q}\}$  is the data,  $\theta_{H_i}$  is the parameter vector under the hypothesis  $H_i$ , and  $\mathcal{N}(\cdot | \mu, \sigma^2)$  is the normal probability density function with the mean  $\mu$  and variance  $\sigma^2$ . We next construct a Bayesian statistical model by combining the likelihood function with uninformative but proper prior distributions where we do not assume any prior dependencies between the parameters and use standard normal prior distributions in logarithmic parameter space. The selected prior distribution introduces a soft lower bound for the parameters. Thus, if a certain rate parameter is present in the model, its value cannot be infinitely close to zero. We estimate the parameter posterior distributions and posterior probabilities of alternative hypotheses by means of population-based Markov chain Monte Carlo (MCMC) sampling and thermodynamic integration (see “Methods” section for details).

### Quantitative hypothesis testing reveals the most likely metabolic switching mechanisms

The posterior analysis is first carried out independently for each alternative metabolic switching mechanism (hypotheses  $H_1$ ,  $H_2$ , and  $H_3$ ). The resulting approximations for the parameter posterior distributions show that the models are identifiable under all three metabolic wiring scenarios (Additional file 1: Figures S2-S4 and a summary about convergence diagnostics in Figure S5). In general, the predictions in all three scenarios are in a good agreement with the experimental wild-type data (see predicted total cell mass in Fig. 1C, wild type). The posterior predictive distributions (PPDs) are very similar under the hypotheses  $H_1$  and  $H_2$  and the only notable difference is a larger dynamical variability under  $H_1$  (Fig. 1C, Wild type). This finding is consistent since the models are nested and the additional switching route under hypothesis  $H_1$  increases the model flexibility. The PPD under hypothesis  $H_3$  exhibits less variability and additionally a distinct dynamic behavior of  $m^e$  compared to the other two scenarios. Furthermore, Fig. 1C shows the PPDs also for the petite strain and we can conclude that under all three hypotheses we are capable of predicting the total cell mass dynamics of the petite strain even though the dynamics of the non-observed model components may differ significantly. Consequently, we can conclude that the predictive performance of our models is good for both the training and the validation data sets. However, based on visual inspection, it is impossible to judge which hypothesis is most likely and, therefore, we perform statistically rigorous quantitative hypothesis testing over the hypotheses  $H_1$ ,  $H_2$ , and  $H_3$ .

Despite the non-distinguishable model predictions in the data space, the posterior analysis over different metabolic switching hypotheses shows significantly more evidence for  $H_2$  (Fig. 1C) with a posterior probability

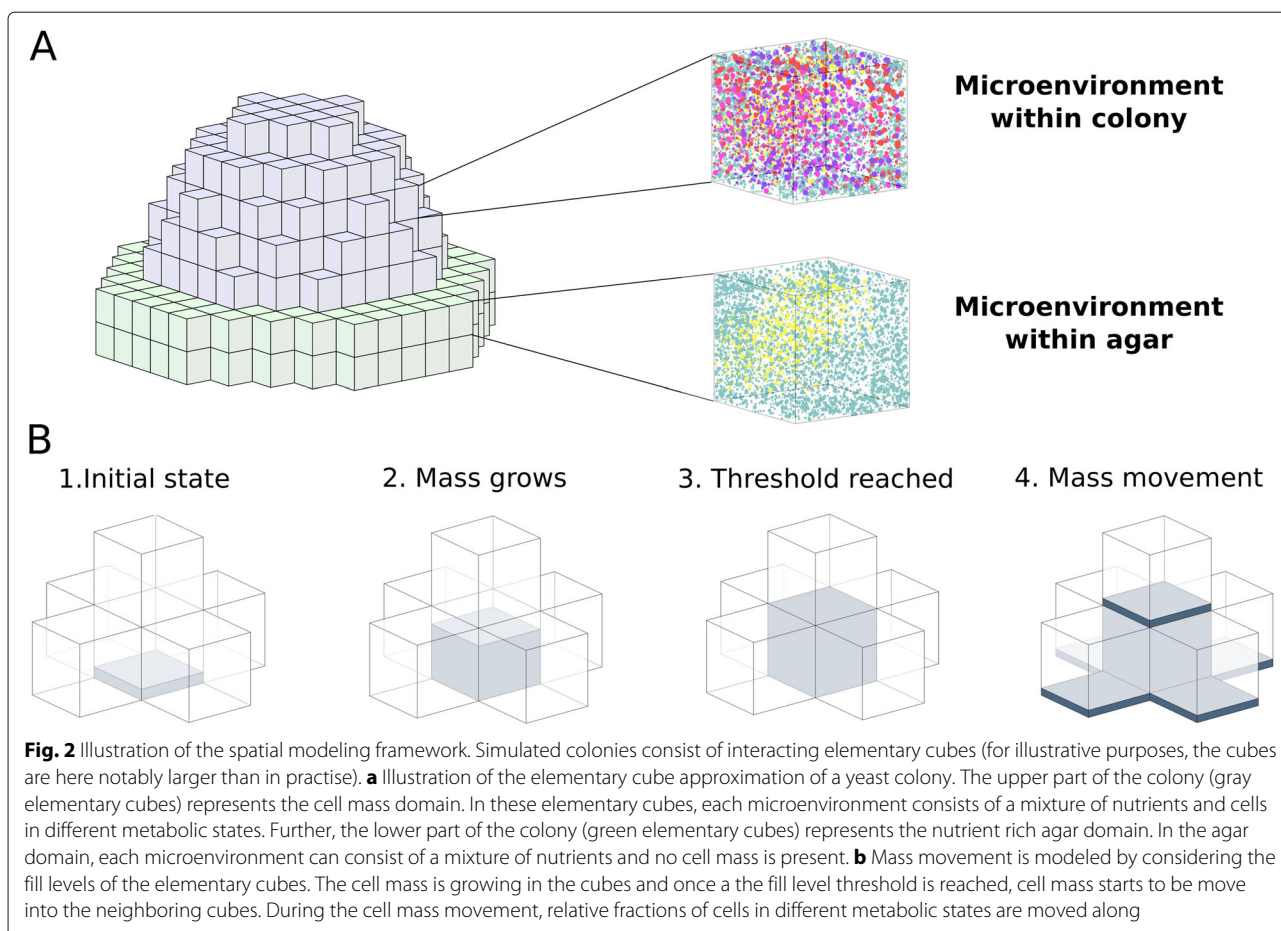
of  $H_2$  very close to 1 (the posterior probabilities as well as the estimated logarithmic marginal likelihoods are shown in parentheses after the hypothesis labels in Fig. 1C). This strong statistical evidence for  $H_2$  suggests that the metabolic switching to the quiescent state in wild-type yeast cells occurs always through the ethanol state in agreement with the current biological interpretations [7, 18, 22].

### Spatial modeling framework to study colony formation

In our experimental setup, yeast cells grow on a glucose rich agar plate and form 3d colonies (Fig. 1A) but the underlying growth mechanisms in terms of metabolic activity and cell state transitions are not understood. To address this challenge, we construct a spatial modeling framework which allows us to predict three dimensional cell state and nutrient distributions during the colony formation process based on our inferred microenvironment model. In addition to cell mass and nutrient dynamics within the colony, we also model the nutrient dynamics within the agar.

To setup the spatial model, we discretize the space into elementary cubes (Fig. 2A). Since the size of the elementary cubes is chosen appropriately, the growth dynamics within each cube (microenvironment) can be modeled under the homogeneity assumption. In other words, each elementary cube consists of a homogeneous mixture of nutrients and cells in distinct metabolic states (Fig. 2A) and the time-evolution of these local components can be described using the microenvironment model developed above. The spatial colony formation is subsequently determined by the dynamics of interacting neighboring cubes with information exchange by the flow of nutrient signals and movement of the growing cell mass.

The cell mass movement is modeled by considering fluxes between neighboring cubes determined by thresholded fill levels of the neighboring cubes where cell mass is moving from a high to low concentration (for illustration see Fig. 2B with parameters given in Table 1). The thresholding is essential because the size of elementary cubes is fixed and it is reasonable to assume that the mass movement does not occur until a certain amount of cell mass has accumulated locally and the resulting pressure starts to push cells forward. In our implementation, the fluxes are computed between six neighboring cubes in each spatial direction and the time-evolution of the full mass distribution is modeled using an ODE system which is determined by the net impact of the individual fluxes. The fluxes are always computed based on the thresholded total mass distribution and the proportions of metabolic states moving along the cell mass are proportional to the proportions of cell states in the cube from which the cell mass is moving. On top of the agar, cell mass can move



only to five directions because mass movement into the agar is excluded.

The nutrient transfer is modeled using the same flux-based model as the cell mass movement. However, the thresholding is not needed for the nutrient transfer because it can be assumed that nutrients can diffuse freely over the domain. The domain for glucose diffusion is the union of the agar domain and the elementary cubes with positive cell mass. In addition, it is assumed that the ethanol which is produced as a by-product during growth on glucose can diffuse freely over the positive cell mass.

**Table 1** Parameters of the spatial framework. Bounds are given for parameters that are estimated

	Parameter	Value	Bounds
mass movement threshold	$th$	1	–
mass movement rate	$\lambda_{\text{mass}}$	$20 \text{ h}^{-1}$	–
nutrient transfer rate within agar	$\lambda_{\text{agar}}$	$25.42 \text{ h}^{-1}$	[5, 75]
nutrient transfer rate within colony	$\lambda_{\text{col}}$	$0.05 \text{ h}^{-1}$	[0.005, 1]
elementary cube edge length	$h$	0.1 mm	–
initial glucose conc. in the agar	$g_0^{\text{agar}}$	1	–

A formal derivation of the mass movement and nutrient transfer models can be found in the “[Methods](#)” section.

#### Data-driven calibration of the spatial model

As explained in detail above, the spatial model consists of interacting elementary cubes and within each cube we consider an approximately homogeneous mixture of cells in different metabolic states and nutrients. Local dynamics in each elementary cube are modeled using the microenvironment model whose structure and parameterization is calibrated using growth curve data and population composition information at 80 hours. More specifically, we use the microenvironment model under metabolic switching hypothesis  $H_2$  which was ranked the highest in the statistical testing. The parameterization of this model is fixed to the maximum a posteriori values that were obtained as a by-product of the posterior analysis. Once the microenvironment model is parameterized, we are left with several unknown parameters that are needed for the spatial framework. These parameters are the mass movement rate, the nutrient transfer rates in the agar and within the cell mass, and the initial glucose level in the agar (Table 1). Because practically no pressure is accumulating inside the colony, we set a high value for the

mass movement rate ( $20 \text{ h}^{-1}$ ). This means that the cell mass is distributed at the same rate as the cells are growing and local crowding does not occur. Furthermore, we assume that the glucose reserve in the agar can be modeled by means of a disc with thickness of 0.2 mm and a diameter of 1 cm. Then the local initial glucose level in the elementary cubes in the agar domain can be normalized to equal one and, consequently, we are left with two free parameters: the nutrient transfer rate in the agar and the nutrient transfer rate within the cell mass.

To estimate the free parameters of the spatial framework, we measure the colony footprint as the area under the growing wild type colony over time (see “Methods” section for details) and optimize the free parameters by minimizing the difference of the experimental measured footprint and the area under the simulated colony. Hence, we minimize the cost function

$$\xi(\lambda_{\text{agar}}, \lambda_{\text{col}}) = \log \left( \sum_{i=1}^n (A_{t_i}^{\text{sim}}(\lambda_{\text{agar}}, \lambda_{\text{col}}) - A_{t_i}^{\text{meas}})^2 \right), \quad (7)$$

where  $\lambda_{\text{agar}}$  and  $\lambda_{\text{col}}$  are the transfer rates within the agar and the colony, and  $A_{t_i}^{\text{sim}}(\lambda_{\text{agar}}, \lambda_{\text{col}})$  and  $A_{t_i}^{\text{meas}}$  are the simulated and measured areas at time  $t_i$ , respectively. Because objective initialization of the cell state and nutrient distribution above the agar is practically impossible, we initialize one elementary cube with cell mass in the glucose state up to the cell mass movement threshold and set the initial glucose level in this cube to one.

We minimize the cost function using Bayesian optimization [23]. The optimization is initialized by evaluating the cost function at 20 points which are sampled within the bounds (Table 1) using Latin hypercube sampling. After initialization, the optimal parameter values (Table 1) are obtained after 9 iterations of the algorithm. Figure 3A exhibits the fitted footprint area against the experimental data. The model fit is in a good agreement with the data even though at the late time points the model shows saturating behavior that is not present in the real data. This slight disagreement suggests that there is some fraction of cells in a metabolic state which is not included in the model. However, the calibrated model does not only fit well to the wild type data but is also in an excellent agreement with two replicates of our petite strain validation data (see red curves in Fig. 3A). The third replicate can clearly be seen as an outlier and may indicate a low efficiency of biomass production [20] described in the model by the yield parameter  $\gamma_1$ . Based on the good fits, we conclude here that our model successfully captures essential dynamics also with respect to the colony size over time.

### Predicting nutrient and metabolic state distributions

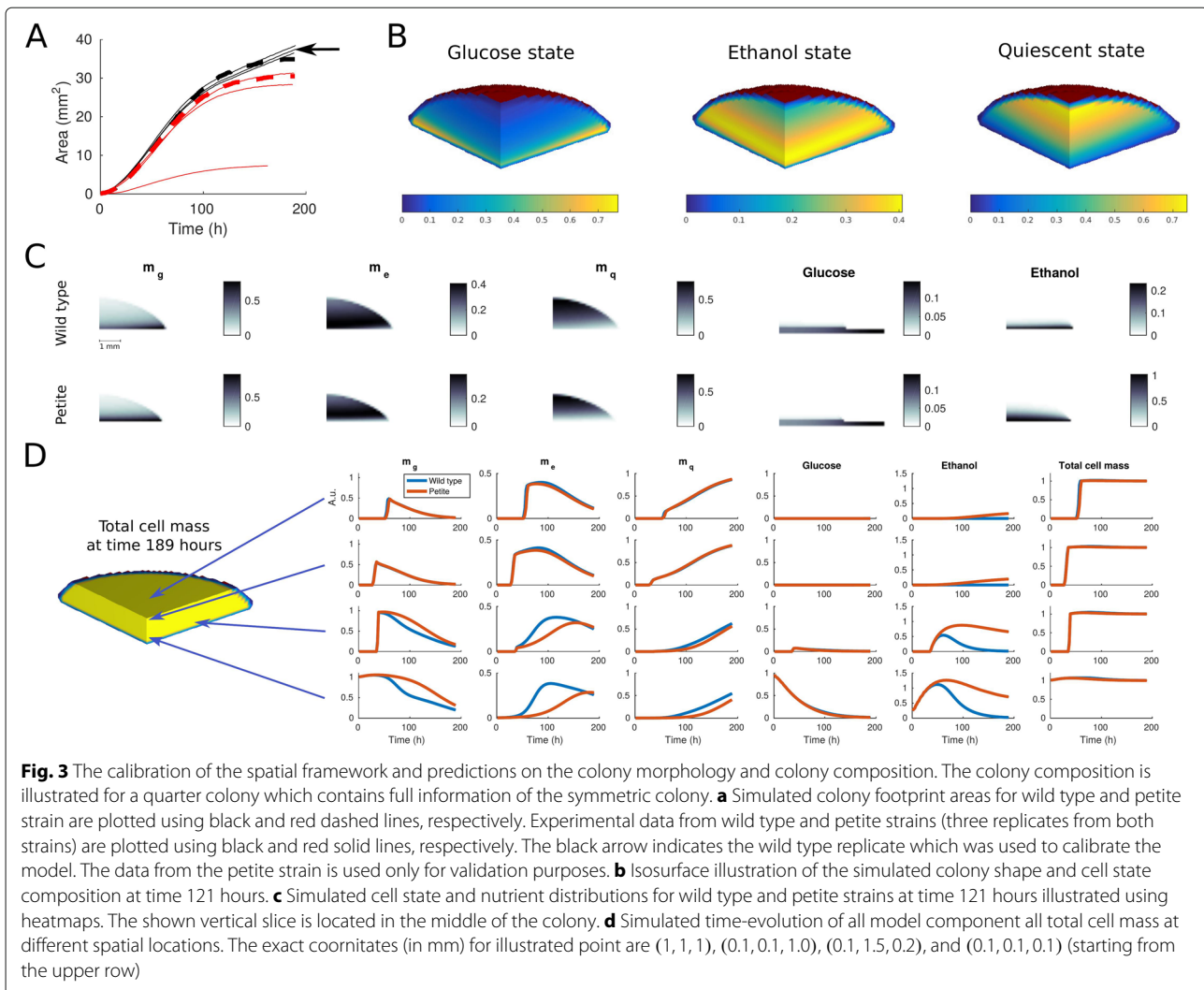
The calibrated model provides us with rich information about the spatial organization within the colony as well as the colony morphology over time. Figure 3B illustrates the colony shape and cell state composition at 121 hours. In our data-driven simulation, we observe three distinct regions in which the different cell states are concentrated. The cells in glucose state are present mainly close to the agar, the cells growing on ethanol are located in the middle of the colony, and, in the upper part the colony, we see a high concentration of cells in quiescent state.

A more detailed view on the spatial organization within the colony is given in Fig. 3C which shows the simulated cell state and nutrient distributions for wild type and petite strains in the middle of the colony at time 121 hours. The nutrient distributions show that glucose is mainly present close to the agar and this indicates that most of the glucose growth and consumption occur in this region. Further, the ethanol distributions show that the ethanol level is much higher in case of the petite strain due to lacking consumption. The snapshot distributions for wild type and petite strains look quite similar but essential differences become visible when we observe the time-evolution of model components at different spatial locations (Fig. 3D). Besides the differing ethanol dynamics for wild type and petite strains, also the cell state dynamics differ notably at many spatial locations. The driver for these differences is the growth in the ethanol state which on its behalf affect the switching between the different metabolic states.

### Discussion

We introduced here a novel coarse-grained multiscale model for yeast colony formation demonstrated how computational modeling can be used to integrate experimental information across scales. In particular, we used here growth measurements in homogenous fluid medium to first infer statistically the growth and cell state transition dynamics. For this purpose, we constructed 3 alternative microenvironment models with respect to cell state transitions (Fig. 1C) and implemented rather simple mass action like behavior allowing for efficient model identification. Interestingly, our unbiased approach supports the current biological perspective of the diauxic shift which assumes that cells reach quiescence through the ethanol state [7, 18] by the identification of model  $H_2$ .

We subsequently used the microenvironment model in our spatially coarse-grained framework to investigate the impact of metabolic coupling on colony formation. Our coarse-grain approximation can be considered as a compromise between agent based modeling of individual cells typically based on many unknown parameters and continuum strategies by computational expensive partial differential equations. In particular, the coarse-



grained cubes allowed the direct incorporation of the inferred microenvironment model and information flux based coupling between neighboring cubes enabled efficient parameter calibration by Bayesian optimization in a data-driven manner. While data-driven model parameterization would be also possible for the other spatial modeling strategies by means of exhaustive parameters sweeps, our approach can save a notable amount of computational resources and increase accuracy by Bayesian optimization. The role of this efficient optimization technique will become even more important when rich data across the scales becomes available and a larger fraction of model parameters can be calibrated together with the spatial parameters.

Here, we applied our modeling strategy to yeast structure formation and could predict how metabolic coupling is shaping the resulting multicellular entity in terms of cell state and nutrient distributions in a dynamical manner (Fig. 3D). Given the available data, our model

cannot explicitly discriminate between dead and quiescence cells and therefore does not include growth on dead cell material. As a consequence our predictions do not directly match reported colony organization [24] but correspond to experimental results on yeast biofilm structures where quiescent cells are located at the periphery [9]. This discrepancy indicates the potential consequence of metabolic coupling on multicellular development and distinct regulatory mechanisms. In the current version, our model does also not include the effect of the extracellular matrix (ECM) that can induce intra-colony chambers and channels for nutrient transport to different spatial locations and may support growth under nutrient limited conditions [10]. In this context, our finding on pronounced ethanol and quiescent states within the middle of the colony suggests a preferred interaction of these cells with the ECM and will be investigated in future work. In a similar way, the current version of the model does not include agar invasion of the yeast cells which would induce spatial



nutrient gradients leading to structural inhomogeneities. These aspects will be investigated in an extended version of the model by incorporating corresponding data on cell death and agar invasion.

Even though our computational framework is presented in the context of yeast colony modeling, our approach is fully general and can be applied to model any multicellular system. For instance, an interesting future application for our method could be to study the role of metabolic coupling and cellular heterogeneity during human glioblastoma tumor growth [25]. Our ultimate goal is to develop a spatial framework that would allow simultaneous calibration of local and global parameters. Careful formulation of the related statistical inference problem would also enable at least semi-automatic experimental design planning. In other words, the model calibration could be carried out iteratively so that every iteration would not only provide information about the parameters but also probabilistic predictions on the most beneficial future measurements in terms of quantities and time points.

## Conclusions

In this study, we present a data-driven spatial framework to computationally investigate the development of multicellular yeast structure. Throughout the model development process, we used state-of-the-art statistical techniques to handle the uncertainty of model structure and parameterization. Using our unbiased approach, we could validate the underlying mechanism of the diauxic shift and validated the model predictions against independent experimental data illustrating the importance of metabolic coupling in colony formation.

## Methods

### Growth curve data

The experimental procedures are detailed elsewhere [10]. In brief, growth curves in suspension of the FY4 (YAD145) strain and its petite version YAD479 (that is unable to metabolise non-fermentable carbon sources like ethanol) were measured on a TECAN Sunrise (Tecan). Initially, 200  $\mu$ L were seeded from running cultures at  $5 \times 10^5$  cells/mL and cell number was monitored by optical density (OD) every 15 min for 88 hours in YPD medium containing 2% glucose at 30°C. The growth data are provided in a machine readable format as a part of the computational implementation and details about preprocessing can be found in Additional file 1: Figure S1.

### Colony footprint area data

Colony formation was measured by our custom built colony imaging system (Scott and Dudley, unpublished results). Colonies from single cells were grown on YPD-agar plates with 5% glucose for 7.5 days in an incubator at 30 °C, and photographed every 20 minutes. Typical

distances of colonies at the end point measurements were around 1 cm. Colony areas were extracted from each image by a script for NIH ImageJ [26]. (See Additional file 1 for details on image capture and analysis.)

### Bayesian techniques for ODE model calibration

The parameters and structure of ODE models are calibrated within the Bayesian framework (see e.g. [27]). In brief, we link the model output with time-course data  $D$  via the likelihood function  $\pi(D|\theta_{H_i}, H_i)$  where  $\theta_{H_i}$  is the parameter vector under the hypothesis  $H_i$  about the model structure ( $i = 1, \dots, n$ ). A Bayesian statistical model can be constructed by combining the likelihood function with a prior distribution over the parameters,  $\pi(\theta_{H_i}|H_i)$  [28]. Bayes' theorem yields the parameter posterior distribution  $\pi(\theta_{H_i}|D, H_i) = \pi(D|\theta_{H_i}, H_i)\pi(\theta_{H_i}|H_i)/\pi(D|H_i)$ , where  $\pi(D|H_i)$  is the marginal likelihood. The marginal likelihoods  $\pi(D|H_i)$  can be used to compute the posterior distribution over the hypotheses, i.e.  $\pi(H_i|D) = \pi(D|H_i)\pi(H_i)/\sum_i^n \pi(D|H_i)\pi(H_i)$ , where  $\pi(H_i)$  is the prior distribution over the alternative models. In this study, the prior distribution over the alternative hypotheses is assumed to be uniform.

### Population-based Markov chain Monte Carlo sampling

Neither the posterior distributions nor the marginal likelihoods can be analytically solved for our models and, consequently, the posterior analysis needs to be carried out using numerical techniques. For this purpose, we use the population-based Markov chain Monte Carlo (MCMC) sampling and thermodynamic integration [29, 30].

To implement a population-based Markov chain Monte Carlo sampler, we consider a product form of the target density

$$\pi^* \left( \theta_{\beta_1}, \theta_{\beta_2}, \dots, \theta_{\beta_{N_\beta}} | D, H \right) = \prod_{i=1}^{N_\beta} \pi_{\beta_i} \left( \theta_{\beta_i} | D, H \right), \quad (8)$$

where  $\pi_{\beta_i}(\theta|D, H) \propto \pi(D|\theta H)^{\beta_i} \pi(\theta|H)$  is the power posterior for fixed temperatures  $0 = \beta_1 < \dots < \beta_{N_\beta} = 1$  [29, 30]. The distributions  $\pi_{\beta_i}$ , including the posterior distribution  $\pi(D|\theta, H)\pi(\theta|H)$ , are marginal distributions of the product form of the target density. By means of population-based MCMC sampling, we draw samples from the individual marginal distributions as well as allow global moves between neighboring temperatures (for details, see [29, 30]).

In this study, we select the temperatures according to the formula

$$\beta_i = \left( \frac{i-1}{N_\beta-1} \right)^5, \quad i = 1, \dots, N_\beta, \quad (9)$$

and use altogether 30 temperatures ( $N_\beta = 30$ ). Before running the sampler, we use local gradient-based deterministic multistart optimization to determine the highest peak in each temperature and the corresponding points are then used as an initial state for the sampler. For the multistart optimization, we use our own optimization routine which is implemented in Matlab according to the guidelines given in references [31, 32]. The actual sampling is run in two parts. First,  $10^5$  samples are drawn so that the normal proposal distributions are adaptively tuned based on the estimated covariance of the previous 7500 samples. After this burn-in and adaption period, the proposal distributions are fixed and every 1000th sample is collected until 2500 samples are obtained. We run four independent samplers under each alternative hypothesis and the convergence of the chains is monitored by means of the potential scale reduction factors [33] and visual inspection over all temperatures. After checking the convergence, the samples from four independent runs are combined and the posterior analysis is carried out using all  $10^4$  samples.

### Bayesian optimization

The parameters of the spatial model are optimized by using the Bayesian optimization technique which is tailored for global optimization of cost functions [23, 34].

To calibrate the spatial model, we need to minimize a target function  $y(\mathbf{x}) : \mathbb{R}^d \rightarrow \mathbb{R}$  with respect to the parameters  $\mathbf{x}$  (we note here that this notation applies only to this subsection). The evaluation of the target function is computationally costly and, to be able to find the minimum using as few as possible function evaluations, we approximate  $y(\mathbf{x})$  by means of a Gaussian process  $f(\mathbf{x})$ . Formally, we can write

$$f(\mathbf{x}) \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}', \boldsymbol{\theta})), \quad (10)$$

where

$$k(\mathbf{x}, \mathbf{x}', \boldsymbol{\theta}) = \theta_{d+1} \exp\left(-\sum_{k=1}^d \frac{(x_k - x'_k)^2}{2\theta_k^2}\right) \quad (11)$$

is the squared exponential kernel function and  $\boldsymbol{\theta} \in \mathbb{R}^{d+1}$  is a parameter vector (for details about Gaussian processes, see e.g. [35]). We assume that the approximation error is normally distributed i.e.

$$y(\mathbf{x}) = f(\mathbf{x}) + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma_{\text{error}}^2). \quad (12)$$

Based on the above definitions, the prior distribution for the approximated function values  $f_n = f(\mathbf{x}_n)$ ,  $n = 1, \dots, N$  is the zero-mean multivariate normal distribution, i.e.

$$\mathbf{f}|\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \Sigma_{\mathbf{X}, \mathbf{X}}), \quad (13)$$

where  $\mathbf{f} = [f(\mathbf{x}_1), f(\mathbf{x}_2), \dots, f(\mathbf{x}_N)]'$ ,  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$ , and  $\{\Sigma_{\mathbf{X}, \mathbf{X}}\}_{ij} = k(\mathbf{x}_i, \mathbf{x}_j, \boldsymbol{\theta})$ ,  $i, j = 1, \dots, N$ . It follows also that

$$\mathbf{y}|\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \Sigma_{\mathbf{X}, \mathbf{X}} + \sigma_{\text{error}}^2 \mathbf{I}), \quad (14)$$

where we have used the above notation,  $\mathbf{y} = [y(\mathbf{x}_1), y(\mathbf{x}_2), \dots, y(\mathbf{x}_N)]'$ , and  $\mathbf{I}$  is the identity matrix. The marginal likelihood is  $p(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}, \sigma_{\text{error}}^2)$  where we have explicitly added the kernel parameters  $\boldsymbol{\theta}$  and error variance  $\sigma_{\text{error}}^2$  to emphasize that the distribution and the marginal likelihood depend on this parameterization.

Given a set of evaluated function values at certain points given by  $\mathbf{y} = [y(\mathbf{x}_1), y(\mathbf{x}_2), \dots, y(\mathbf{x}_N)]'$ , we can generate a probabilistic prediction on the function value  $y(\mathbf{x}^*)$  at an arbitrary point  $\mathbf{x}^*$  in the domain. The prediction about the function value  $y(\mathbf{x}^*)$  can be generated in form of a random variable  $y^*$  which follows the joint distribution in Eq. 14. By conditioning  $y^*$  on the evaluated values, we obtain

$$y^*|\mathbf{X}, \mathbf{y}, \mathbf{x}^* \sim \mathcal{N}(\Sigma_{\mathbf{x}^*, \mathbf{X}} (\Sigma_{\mathbf{X}, \mathbf{X}} + \sigma_{\text{error}}^2 \mathbf{I})^{-1} \mathbf{y}, (\Sigma_{\mathbf{x}^*, \mathbf{x}^*} + \sigma_{\text{error}}^2) - \Sigma_{\mathbf{x}^*, \mathbf{X}} (\Sigma_{\mathbf{X}, \mathbf{X}} + \sigma_{\text{error}}^2 \mathbf{I})^{-1} \Sigma_{\mathbf{X}, \mathbf{x}^*}), \quad (15)$$

where  $\Sigma_{\mathbf{x}^*, \mathbf{X}} = [k(\mathbf{x}^*, \mathbf{x}_1, \boldsymbol{\theta}), k(\mathbf{x}^*, \mathbf{x}_2, \boldsymbol{\theta}), \dots, k(\mathbf{x}^*, \mathbf{x}_N, \boldsymbol{\theta})]$ ,  $\Sigma_{\mathbf{X}, \mathbf{x}^*} = \Sigma_{\mathbf{x}^*, \mathbf{X}}'$ , and  $\Sigma_{\mathbf{x}^*, \mathbf{x}^*} = k(\mathbf{x}^*, \mathbf{x}^*, \boldsymbol{\theta})$ . The probabilistic nature of the prediction makes it also possible to predict the next point at which it is most beneficial to evaluate the function value in the context of a minimization problem [23]. The optimal evaluation point can be chosen by finding the point  $\mathbf{x}^*$  which maximizes the expected improvement function

$$E[I(\mathbf{x}^*)] = E[\max(y_{\min} - Y, 0)], \quad (16)$$

where  $y_{\min}$  is the minimum of the so far evaluated function values and  $Y = y^*|\mathbf{X}, \mathbf{y}, \mathbf{x}^*$  (see e.g. [23] for details and illustrative examples). The expected improvement (Eq. 16) can be expressed in the closed form

$$E[I(\mathbf{x}^*)] = (y_{\min} - \hat{y}) \Phi\left(\frac{y_{\min} - \hat{y}}{s}\right) + s \phi\left(\frac{y_{\min} - \hat{y}}{s}\right), \quad (17)$$

where  $\phi$  and  $\Phi$  are the standard normal density and distribution function, respectively, and  $\hat{y}$  and  $s$  are the mean and standard deviation of the normal distribution in Eq. 15, respectively [23].

The actual optimization routine consists of two steps: (i) fitting the response surface by maximizing  $p(\mathbf{y}|\mathbf{X})$  (Eq. 14) with respect to the hyperparameters  $(\boldsymbol{\theta}, \sigma_{\text{error}}^2)$  and (ii) finding the optimal point for next function evaluation by maximizing the expected improvement (Eq. 16). The steps are carried out sequentially and the response surface is always fitted using a set of evaluated function values which are standardized to have a zero mean and standard deviation of one. In our implementation, the hyperparameters of the Gaussian process model and the next evaluation point with respect to the expected improvement are optimized using `fminunc` and `fmincon` optimization routines in Matlab, respectively. The hyperparameter optimization is initialized using parameter values  $\theta_1 = \theta_2 = \theta_3 =$

1,  $\sigma_{\text{error}} = 0.1$  which correspond to a smooth Gaussian process response surface. In the context of expected improvement optimization, we utilize a multistart optimization strategy for which the initial points are obtained by means of Latin hypercube sampling (lhsdesign function in Matlab). The sequential procedure is repeated until the expected improvement goes under a threshold ( $10^{-46}$  in this study) or the maximum number of iteration steps (i) and (ii) is reached.

### Formal definition of the spatial framework

We discretize the space by dividing it into finite size elementary cubes each having a constant volume (see Fig. 2 for illustration). The cubes are indexed by their location in a 3D array i.e. mass in different metabolic states at different spatial locations can be expressed by writing

$$m_{i,j,k}^{\{n\}}, \quad i = 1, \dots, N_i, \quad j = 1, \dots, N_j, \quad k = 1, \dots, N_k,$$

where  $\{n\} \in \{g, e, q\}$  denotes the metabolic state. The total mass at each location can be computed by summing the cell masses in distinct metabolic states, i.e.

$$m_{i,j,k} = m_{i,j,k}^g + m_{i,j,k}^e + m_{i,j,k}^q.$$

The cubes interact through their fill levels and the cell mass is flowing from a high concentration to a low concentration once a certain threshold is exceeded. The amount of mass exceeding the threshold can be interpreted as pressure that pushes the cell mass onwards. This pressure is computed based on a thresholded total mass distribution over the space. The thresholded total mass at a certain spatial location is defined by

$$m_{i,j,k}^{\text{th}} = \max(m_{i,j,k} - th, 0),$$

where  $th$  is the threshold parameter.

### Mass movement

For mass movement modeling, the moving cell mass has to reflect the fractions of cells in different metabolic states. The fractions carried along can be taken to be proportional to the cell state fractions in the source cubes (the cubes from which the mass is moved). Consequently, the mass movement is modded by

$$\begin{aligned} \frac{dm_{i,j,k}^{\{n\}}}{dt} = & \lambda_m [F(m_{i,j,k}, m_{i-1,j,k}, m_{i-1,j,k}^{\{n\}}, m_{i,j,k}^{\{n\}}) \\ & + F(m_{i,j,k}, m_{i+1,j,k}, m_{i+1,j,k}^{\{n\}}, m_{i,j,k}^{\{n\}}) \\ & + F(m_{i,j,k}, m_{i,j-1,k}, m_{i,j-1,k}^{\{n\}}, m_{i,j,k}^{\{n\}}) \\ & + F(m_{i,j,k}, m_{i,j+1,k}, m_{i,j+1,k}^{\{n\}}, m_{i,j,k}^{\{n\}}) \\ & + F(m_{i,j,k}, m_{i,j,k-1}, m_{i,j,k-1}^{\{n\}}, m_{i,j,k}^{\{n\}}) \\ & + F(m_{i,j,k}, m_{i,j,k+1}, m_{i,j,k+1}^{\{n\}}, m_{i,j,k}^{\{n\}})], \quad (18) \end{aligned}$$

where  $\lambda_m$  is the mass movement rate parameter,

$$F(m, m', m^{\{n\}}, m'^{\{n\}}) = \begin{cases} 0, & g(m) = g(m') \\ (g(m') - g(m)) \frac{m^{\{n\}}}{m}, & g(m) > g(m') \\ (g(m') - g(m)) \frac{m'^{\{n\}}}{m'}, & g(m) < g(m') \end{cases} \quad (19)$$

and  $g(m) = \max(m - th, 0)$  is a function which takes care of the thresholding with the parameter  $th$ . At the agar-cell mass interface, the mass movement into the agar is prevented by mapping the corresponding values of the function  $F$  to zero.

To show that the mass is conserved through the movement, we can consider mass movement between two elementary cubes  $m$  to  $m'$ . Based on our model structure, we have

$$m = m^g + m^e + m^q \quad (20)$$

$$m' = m'^g + m'^e + m'^q \quad (21)$$

and the thresholded total cell masses in these two cubes are

$$m^{\text{th}} = \max(m - th, 0) \quad (22)$$

$$m'^{\text{th}} = \max(m' - th, 0). \quad (23)$$

Without losing any generality, we can assume  $m^{\text{th}} > m'^{\text{th}}$ . Now

$$\frac{dm^{\{n\}}}{dt} = \lambda_m F(m, m', m^{\{n\}}, m'^{\{n\}}) = \lambda_m \frac{m^{\{n\}}}{m} (m'^{\text{th}} - m^{\text{th}}) \quad (24)$$

and

$$\begin{aligned} \frac{dm'^{\{n\}}}{dt} &= \lambda_m F(m', m, m'^{\{n\}}, m^{\{n\}}) \\ &= \lambda_m \frac{m'^{\{n\}}}{m'} (m^{\text{th}} - m'^{\text{th}}). \end{aligned} \quad (25)$$

From Eqs. 24 and 25, we can deduce

$$\frac{dm^{\{n\}}}{dt} = - \frac{dm'^{\{n\}}}{dt}, \quad (26)$$

which proofs mass conservation during the movement. Since the net mass movement defined in Eq. 18 is a sum of six pairwise movements, the mass is conserved also for the net movement.

### Nutrient transfer

The nutrient transfer can be described in a similar manner as the mass movement but, in this context, we do not need to threshold the distribution because nutrient diffusion occurs freely in the media. Furthermore, nutrient transfer can be simply defined by fluxes between the neighboring cubes whereas in the context of mass movement we had to take the fractions of different cell types into account. If we consider the nutrient concentrations

$n_{i,j,k}$ ,  $i = 1, \dots, N_i$ ,  $j = 1, \dots, N_j$ ,  $k = 1, \dots, N_k$ , the nutrient transfer can be described by

$$\begin{aligned} \frac{dn_{i,j,k}}{dt} = & f(n_{i,j,k}, n_{i-1,j,k}, \lambda_{\text{agar}}, \lambda_{\text{col}})I(m_{i-1,j,k}) \\ & + f(n_{i,j,k}, n_{i+1,j,k}, \lambda_{\text{agar}}, \lambda_{\text{col}})I(m_{i+1,j,k}) \\ & + f(n_{i,j,k}, n_{i,j-1,k}, \lambda_{\text{agar}}, \lambda_{\text{col}})I(m_{i,j-1,k}) \\ & + f(n_{i,j,k}, n_{i,j+1,k}, \lambda_{\text{agar}}, \lambda_{\text{col}})I(m_{i,j+1,k}) \\ & + f(n_{i,j,k}, n_{i,j,k-1}, \lambda_{\text{agar}}, \lambda_{\text{col}})I(m_{i,j,k-1}) \\ & + f(n_{i,j,k}, n_{i,j,k+1}, \lambda_{\text{agar}}, \lambda_{\text{col}})I(m_{i,j,k+1}). \end{aligned} \quad (27)$$

Here,

$$\begin{aligned} f(n_{i,j,k}, n'_{i',j',k'}, \lambda_{\text{agar}}, \lambda_{\text{col}}) = \\ \begin{cases} \lambda_{\text{col}}(n' - n), & \text{if } k > h + 1 \text{ or } k = h + 1; k' = k + 1, \\ \lambda_{\text{agar}}(n' - n), & \text{otherwise} \end{cases} \end{aligned} \quad (28)$$

where  $\lambda_{\text{col}}$  and  $\lambda_{\text{agar}}$  are the nutrient transfer rate parameters within the colony and agar, respectively, and  $h$  is the height of the agar given as the number of elementary cube layers. The domain in which the nutrient transfer occurs is determined by the indicator function

$$I(m) = \begin{cases} 1, & \text{if } m > 0 \\ 0, & \text{otherwise.} \end{cases} \quad (29)$$

In other words, the mass distribution dependent domain for the nutrient transfer consists of the cubes which have a positive cell mass concentration.

### Computational implementation

Mathematical models, population-based MCMC sampler, and Bayesian optimization were implemented in Matlab (The MathWorks Inc., Natick, MA, USA). ODE systems were solved using the ode15s solver and the full multiscale model was simulated using the Euler method with a time-step of 0.0025 h.

### Supplementary information

Supplementary information accompanies this paper at <https://doi.org/s12860-019-0234-z>.

**Additional file 1:** The supplementary information consists of three sections: 1. Colony footprint area data, 2. Reparameterization of the microenvironment model, and 3. Supplementary Figures.

### Abbreviations

MCMC: Markov chain Monte Carlo; OD: Optical density; ODE: Ordinary differential equation; PPD: Posterior predictive distribution; YPD: Yeast extract peptone dextrose

### Acknowledgements

The authors acknowledge the computational resources provided by the Aalto Science-IT project.

### Authors' contributions

Jl, ACS, NF, AMD, and AS conceived and designed the study. Jl, ACS, NF, and AS designed the mathematical models. Jl and HL designed statistical techniques for model calibration. Jl designed and implemented the numerical

methods as well as performed the computational experiments. ACS and MH carried out the wet-lab experiments. Jl, ACS, MH, and AS preprocessed and analyzed the data. Jl, ACS, AMD, and AS wrote the paper. NF, OY, HL, AMD, and AS supervised the study. All authors read and approved the final manuscript.

### Funding

This work has been supported by the Academy of Finland [Centre of Excellence in Molecular Systems Immunology and Physiology Research (2012-2017) and the project 275537], a grant from the National Institutes of Health (P50 GM076547), and a strategic partnership between the Institute for Systems Biology and the University of Luxembourg.

### Availability of data and materials

The datasets generated and analyzed during the current study as well as the computational implementation to reproduce the results are available at <https://research.cs.aalto.fi/csb/software/>.

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Department of Computer Science, Aalto University, P.O.Box 15400, FI-00076, Aalto, Finland. <sup>2</sup>Pacific Northwest Research Institute, 720 Broadway, 98122 Seattle, WA, USA. <sup>3</sup>Molecular and Cellular Biology Program, University of Washington, 98195 Seattle, WA, USA. <sup>4</sup>Department of Computer Science, Utah State University, 4205 Old Main Hill, 84322 Logan, UT, USA. <sup>5</sup>BioMediTech and Faculty of Biomedical Sciences and Engineering, Tampere University of Technology, P.O.Box 553, 33101 Tampere, Finland. <sup>6</sup>Institute for Systems Biology, 1441N 34th Street, 98103-8904 Seattle, WA, USA. <sup>7</sup>Luxembourg Centre for Systems Biomedicine, University of Luxembourg, 2, avenue de l'Université, L-4365 Esch-sur-Alzette, Luxembourg. <sup>8</sup>University of California San Diego, 9500 Gilman Dr, 92093 La Jolla, CA, USA.

Received: 17 June 2019 Accepted: 24 October 2019

Published online: 19 December 2019

### References

- Ratcliff WC, Denison RF, Borrello M, Travisano M. Experimental evolution of multicellularity. *Proc Natl Acad Sci*. 2012;109(5):1595–600.
- Komin N, Skupin A. How to address cellular heterogeneity by distribution biology. *Curr Opin Syst Biol*. 2017;3:154–60.
- Stoviccek V, et al. General factors important for the formation of structured biofilm-like yeast colonies. *Fungal Genet Biol*. 2010;47(12):1012–22. <https://doi.org/10.1016/j.fgb.2010.08.005>.
- Graneck JA, et al. The genetic architecture of biofilm formation in a clinical isolate of *Saccharomyces cerevisiae*. *Genetics*. 2013;193(2):578–600.
- Taylor MB, Ehrenreich IM. Genetic interactions involving five or more genes contribute to a complex trait in yeast. *PLOS Genetics*. 2014;10(5):1–8.
- Zahn LM, Purnell BA. Genes under pressure. *Science*. 2016;354:52. <https://doi.org/10.1126/science.354.6308.52>.
- DeRisi JL, et al. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*. 1997;278:680–6.
- Vachova L, et al. Architecture of developing multicellular yeast colony: spatio-temporal expression of *ato1p* ammonium exporter. *Environ Microbiol*. 2009;11:1866–1877. <https://doi.org/10.1111/j.1462-2920.2009.01911.x>.
- Vachova L, et al. Flo11p, drug efflux pumps, and the extracellular matrix cooperate to form biofilm yeast colonies. *J Cell Biol*. 2011;194:679–87. <https://doi.org/10.1083/jcb.201103129>.
- Tan Z, et al. Aneuploidy underlies a multicellular phenotypic switch. *Proc Natl Acad Sci USA*. 2013;110(30):12367–72.
- Kang S, et al. Biocellion : accelerating computer simulation of multicellular biological system models. *Bioinformatics*. 2014;30(21):3101. <https://doi.org/10.1093/bioinformatics/btu498>.



12. Doloman A, et al. Modeling de novo granulation of anaerobic sludge. *BMC Syst Biol.* 2017;11:69. <https://doi.org/10.1186/s12918-017-0443-z>.
13. Schulz EG, et al. Sequential polarization and imprinting of type 1 T helper lymphocytes by interferon-gamma and interleukin-12. *Immunity.* 2009;30(5):673–83. <https://doi.org/10.1016/j.immuni.2009.03.013>.
14. Intosalmi J, et al. Analyzing Th17 cell differentiation dynamics using a novel integrative modeling framework for time-course RNA sequencing data. *BMC Syst Biol.* 2015;9(81):.
15. Chan YH, et al. A subpopulation model to analyze heterogeneous cell differentiation dynamics. *Bioinformatics.* 2016;32(21):3306. <https://doi.org/10.1093/bioinformatics/btw395>.
16. Skupin A, et al. Calcium signals driven by single channel noise. *PLoS Comput Biol.* 2010;6(8):1000870. <https://doi.org/10.1371/journal.pcbi.1000870>.
17. Walther T, et al. Mathematical modeling of regulatory mechanisms in yeast colony development. *J Theor Biol.* 2004;229:327–38. <https://doi.org/10.1016/j.jtbi.2004.04.004>.
18. Galdieri L, et al. Transcriptional regulation in yeast during diauxic shift and stationary phase. *OMICS.* 2010;14:629–38. <https://doi.org/10.1089/omi.2010.0069>.
19. Jung PP, Christian N, Kay DP, Skupin A, Linster CL. Protocols and programs for high-throughput growth and aging phenotyping in yeast. *PLoS One.* 2015;10(3):0119807.
20. Day M. Yeast petites and small colony variants: for everything there is a season. *Adv Appl Microbiol.* 2013;85:1–41. <https://doi.org/10.1016/B978-0-12-407672-3.00001-0>.
21. Alvarez-Vasquez F, et al. Coordination of the dynamics of yeast sphingolipid metabolism during the diauxic shift. *Theor Biol Med Model.* 2007;4:42. <https://doi.org/10.1186/1742-4682-4-42>.
22. Aragon AD, et al. Characterization of differentiated quiescent and nonquiescent cells in yeast stationary-phase cultures. *Mol Biol Cell.* 2008;19:1271–80. <https://doi.org/10.1091/mbc.E07-07-0666>.
23. Jones DR, et al. Efficient Global Optimization of Expensive Black-Box Functions. *J Global Optim.* 1998;13(4):455–92.
24. Váchová L, Cáp M, Palková Z. Yeast colonies: a model for studies of aging, environmental adaptation, and longevity. *Oxidative Med Cell Longevity.* 2012;2012:601836. <https://doi.org/10.1155/2012/601836>.
25. Dirkse A, Golebiewska A, Buder T, Nazarov PV, Muller A, Poovathingal S, Brons NH, Leite S, Sauvageot N, Sarkisjan D, et al. Stem cell-associated heterogeneity in glioblastoma results from intrinsic tumor plasticity shaped by the microenvironment. *Nature Commun.* 2019;10(1):1787.
26. Schneider CA, et al. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods.* 2012;9:671–5.
27. Girolami M. Bayesian inference for differential equations. *Theor Comput Sci.* 2008;408(1):4–16. <https://doi.org/10.1016/j.tcs.2008.07.005>.
28. Robert CP. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*, 2nd: Berlin and Heidelberg; 2007. <https://www.springer.com/gp/book/9780387952314>.
29. Jasra A, et al. On population-based simulation for static inference. *Stat Comput.* 2007;17(3):263–79. <https://doi.org/10.1007/s11222-007-9028-9>.
30. Calderhead B, Girolami M. Estimating Bayes factors via thermodynamic integration and population MCMC. *Comput Stat Data An.* 2009;53(12):4028–45. <https://doi.org/10.1016/j.csda.2009.07.025>.
31. Raue A, et al. Lessons learned from quantitative dynamical modeling in systems biology. *PLoS ONE.* 2013;8(9):74335. <https://doi.org/10.1371/journal.pone.0074335>.
32. Raue A, et al. Data2dynamics: a modeling environment tailored to parameter estimation in dynamical systems. *Bioinformatics.* 2015;31(21):3558–60.
33. Gelman A, et al. *Bayesian Data Analysis*, 3rd: Boca Raton. Texts in Statistical Science; 2013.
34. Ghahramani Z. Probabilistic machine learning and artificial intelligence. *Nature.* 2015;521(7553):452–9. <https://doi.org/10.1038/nature14541>.
35. Rasmussen CE, Williams C. *Gaussian Processes for Machine Learning*: MIT Press; 2006.

## Publisher's Note

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

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

