

Quantitative analysis of colony morphology in yeast

Pekka Ruusuvuori^{1,2}, Jake Lin^{1,2,5}, Adrian C. Scott², Zhihao Tan^{2,4}, Saija Sorsa¹, Aleksi Kallio³,
Matti Nykter³, Olli Yli-Harja^{1,2}, Ilya Shmulevich^{2,1}, Aimée M. Dudley^{2,4}

¹Department of Signal Processing, Tampere University of Technology, Tampere, Finland

²Institute for Systems Biology, Seattle, WA, USA

³Institute of Biomedical Technology, University of Tampere, Tampere, Finland

⁴Molecular and Cellular Biology Program, University of Washington, Seattle, WA, USA

⁵Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Luxembourg

Correspondence to:

Aimée M. Dudley (aimee.dudley@gmail.com)

Institute for Systems Biology

401 Terry Avenue North

Seattle, WA 98109

Tel: (206) 732-1214

Fax: (206) 732-1299

Key Words: Colony morphology; Image analysis; Software; Yeast; Phenotype; Time-lapse

Abstract

Microorganisms often form multicellular structures, such as biofilms and structured colonies, which can influence the organism's virulence, drug resistance, and adherence to medical devices. Phenotypic classification of these structures has traditionally relied on qualitative scoring systems that limit detailed phenotypic comparisons between strains. Automated imaging and quantitative analysis have the potential to improve the speed and accuracy of experiments designed to study the genetic and molecular networks underlying different morphological traits. We have developed a platform that uses automated image analysis and pattern recognition to quantify phenotypic signatures of yeast colonies. The strategy enables quantitative analysis of individual colonies, measured at a single time point or over a series of time-lapse images, as well as the classification of distinct colony shapes based on image-derived features. Phenotypic changes in colony morphology can be expressed as changes in feature space trajectories over time, thereby enabling the visualization and quantitative analysis of morphological development. To facilitate data exploration, results are plotted dynamically through an interactive web application that integrates the raw and processed images across all time points, allowing exploration of the image-based features and principal components associated with morphological development. The web application YIMAA is available at <http://yimaa.cs.tut.fi>.

Method summary

Our platform enables the automated, quantitative analysis of yeast colony morphology. The method extracts a relatively large number of features from colony images followed by supervised classification in feature space. This computational approach provides an alternative to subjective scoring of colonies, is compatible with high-throughput and time-lapse experimental designs, and provides a web-based application for data exploration.

Introduction

A number of microorganisms, many of them well-known opportunistic pathogens, are able to form highly structured biofilms and multicellular communities (1-4). The formation of these complex and well differentiated structures is thought to increase their resistance to antimicrobial treatments (5), and has been shown to be a key factor in persistent infections (1). Some strains of *Saccharomyces cerevisiae*, a non-pathogenic model organism, also display structured colony morphologies (5) with characteristics of microbial biofilms, including the presence of an extracellular matrix composed largely of complex polysaccharides (6-8), the development of channels in the colony interior (6), and the use of cell-cell communication in colony development (9). The genetic tractability and availability of numerous resources (11) not available for other biofilm forming organisms makes *S. cerevisiae* an attractive organisms in which to study the development of complex morphologies, with the goal of ultimately uncovering the molecular mechanisms underlying biofilm formation (11).

While studies aimed at characterizing the variation in colony morphology in *S. cerevisiae* have been as objective as possible, qualitative classification schemes, such as having a single investigator categorize colonies by eye, are still widely used (12-14). Image analysis tools have been applied to the automated analysis of yeast colonies. The image analysis platform ImageJ (15) offers tools for processing quantifying colony images (16), and the image analysis tool CellProfiler (17) has been used to segment colonies on agar plates and group them based on shape, size and color information. Methods and software for quantifying colony growth combined with statistical analysis have also been presented (18,19). Other model organisms have also been subjected to quantitative, image-based characterization and morphological classification. For example, image analysis has been applied to the automated screening of a variety of phenotypes (including morphology) in *Caenorhabditis elegans* (20), and recently an

application case similar to ours was applied to the study of filamentous fungi using a set of over 30 morphological features (22).

Here, we describe an automated image analysis pipeline (**Figure 1**) that facilitates the quantitative study of colony morphology dynamics in large, time-lapse datasets. We start with automated image processing and then extract a large, generic set of quantitative descriptors. The combination of high-dimensional feature representation together with a sparse, supervised, logistic regression-based classification model is a powerful platform for the analysis of colony morphology. We have also built a web-based application to facilitate the intuitive exploration of the original raw and segmented time series images, the results of Principal Component Analysis (PCA), and hundreds of individual quantitative features. We test the accuracy of our method by using it to computationally distinguish the complex (“fluffy”) and unstructured (“smooth”) colony phenotypes (6,22) based on image data from both single time points and fine resolution time lapses.

Materials & Methods

Yeast strains and growth conditions

Standard media and methods were used for the growth and genetic manipulation of *S. cerevisiae* (24). All colonies were grown and imaged in a 30°C warm room on YPD (2% glucose) agar plates. Strains used in this study are described in Table 1.

Colony imaging

Colonies used to distinguish the fluffy and smooth phenotype based on a single time point were generated by manually micromanipulating individual cells into a gridded pattern separated by 10 mm in both the x and y directions. Colonies were imaged after 5 days of growth using a PowerShot SX10IS camera outfitted with a DCR-250 macro lens (Raynox).

Colonies used for automated, time-lapse imaging were generated by depositing single cells 12.7 mm apart in a “checkerboard” pattern with a FACSaria II cell sorter (BD Biosciences, Supplemental Materials). These colonies were imaged every 14 minutes for 5 days using a 5d Mark II camera outfitted with a MP-E 65mm 1-5x macro lens (Canon). The camera was attached to a custom built 2-axis gantry that moves the camera over the entire set of plates (Supplemental Materials). Camera settings were held constant at an exposure time of 0.2 seconds and aperture of *f*/16. White balance was set using a grey card. Focus was held constant.

Generating quantitative colony phenotype signatures using image derived features

The first step in our automated pipeline involves segmenting the colony area as the region of interest (Supplemental Materials) and extracting features that describe the colony shape, size, intensity, fractal, and texture. We segment using a straightforward, intensity-based global thresholding operation (25) and then apply an additional size constraint to prevent detecting excessively small or large objects, which can arise from debris on the plate or camera lens flare. We also perform image border clearing to remove false segmentations that occur when colonies located close to plate borders have refraction from the edge of the plate incorrectly assigned to the colony. This first set of segmentation masks (Figure 2A) are used for the first round of feature extraction. The shape and size categories include basic descriptors for object morphology (e.g. area, convex area and roundness). Intensity-based features provide quantitative measures of the intensity distribution (e.g. intensity percentiles and deviation), whereas the texture features (e.g. intensity deviations in local area, texture features from grey-level co-occurrence matrices (26), histogram of oriented gradients (27) and local binary patterns (28)) take the spatial information into account.

The next step involves an additional round of segmentation to detect shapes inside the colonies, visible as intensity changes in 2D projection images, and the extraction of a different set of features from the segmented images. For this segmentation we use a Difference of Gaussians segmentation (29), where the difference of two low-pass filtered versions of the original image (highly blurred and slightly blurred) is thresholded. The two low-pass filters serve as a band-pass filter and the resulting binary image contains areas where intensity changes exist, but in which sharp variation, such as noise, is suppressed (Supplemental Materials). Ideally, the resulting segmentation mask would be empty for a smooth colony and capture the colony shape for a fluffy colony. The features extracted from these second segmentation masks include descriptors containing information about the shapes detected inside the colony (e.g. area of the mask relative to the colony size, mask area in the center and border of the colony, number of objects in the mask, object sizes and deviations).

The combined feature set serves as a quantitative signature of colony phenotype, with colonies derived from the same strain or belonging to the same phenotypic class sharing similar characteristics among many of the features (Fig. 2D). A detailed description of all 427 features is given in the Supplemental Materials. The feature list can be extended or trimmed without changes to the subsequent classification process.

Supervised colony phenotype classification

To transform these quantitative features into biologically meaningful phenotype information, we used a supervised classification strategy. To circumvent the need to specify the features used, we chose a classifier model with built-in feature selection, specifically the l_1 regularized logistic regression (29,30), which produces sparse solutions and thus includes only a subset of the features in the model.

In logistic regression based classification, a feature vector x can be classified based on the conditional probability of belonging to the fluffy class given by the logistic regression algorithm as follows:

$$p(x) = \frac{1}{1 + \exp(\beta_0 + \boldsymbol{\beta}^T x)}$$

where $p(x)$ is the probability for the positive class given the feature vector x (i.e. $p(x) = P(\text{"fluffy"}|X = x)$), and the parameters β_0 and $\boldsymbol{\beta}$ are estimated by maximizing the l_1 penalized log-likelihood

$$\sum_{x \in F} \log p(x) + \sum_{x \in S} \log(1 - p(x)) - \lambda \|\boldsymbol{\beta}\|_1$$

where F denotes the fluffy class training samples, S is the smooth (non-fluffy) class training set, and λ is the parameter regularizing the sparsity of the solution. In practice, the solution is typically very sparse leading to computationally efficient models (32), with only a small subset of features receiving a non-zero weight in vector $\boldsymbol{\beta}$. Further, the use of logistic regression enables the extension to multi-class cases with more than two different strains or phenotypes.

Quantitative analysis of colony spatiotemporal dynamics

Time-lapse image sequences are processed frame by frame as individual colony images once the colonies are large enough to be visible in the image (approximately one day of growth). The most obvious effect of colony growth is colony size, which also affects the quantification process. All features are extracted in the same manner from both small and large colonies. Feature trajectories are visualized by reducing the dimensionality with principal component analysis. Finally, a spatiotemporal profile of the yeast colony's development is built in which the spatial locations of the colony shapes are visualized over time by taking a cumulative sum of the colony shape segmentation masks. Details can be found in the Supplemental Materials.

Web application for data and result browsing

We have developed the Yeast Image Analysis (YIMAA) web application that serves as an interface for the original and binary segmentation images together with the time-lapsed plotting of quantitative phenotypic results. YIMAA is built using the open source components Highcharts.js, jQuery, and jQuery plugins. The design of YIMAA focuses on interactivity and integration of images with dynamic time series plotting. Quantitative results are retrieved using AJAX. Image data are stored as assets organized by experiment and fetched on demand. The YIMAA web application is available at <http://yimaa.cs.tut.fi>. The source code for the project, including the implementation of the image analysis pipeline can be found at <http://code.google.com/p/yimaa/>.

Results and Discussion

Our study aimed to develop a generalized method for quantitatively representing the properties of microbial colonies. For this purpose, we selected a general feature set that is not tailored to a single strain or classification task. Extracting a large set of image-derived features that measure different characteristics of the colony also helps ensure that changes in the experiment or objects being studied, e.g. different magnifications, illumination settings, or strains, do not require significant alterations to the computational framework. Such generalization will facilitate its use in a variety of applications.

Our own research on yeast colony morphology has two experimental designs in which this general framework could be applied. First, the classification of colonies into “smooth” and “fluffy” classes at a single time point, which was performed manually in our previous work (23), could be performed more objectively and in higher throughput using image-derived features. Second, an automated image analysis pipeline could be used to extract quantitative features for many individual colonies as they grow and change shape over a series of time-lapsed images. In this

framework, features extracted from the images form a vector of numerical values for each colony, where an element of the vector represents a feature value at the time point sampled. Both descriptions of colony morphology could be used to inform the genetic analysis of a relatively large number of yeast strains under a variety of environmental conditions.

To assess the discriminating power of our morphological signatures, we first tested whether the method could distinguish the smooth and fluffy morphologies using static images acquired at a single time point (**Figure 2**). Smooth (YPG339, YPG 344, YPG348, YPG352, YPG356 and YPG360) and fluffy (F7, F11, F18, F25, F29, F31, F45, F47 and F49) yeast strains (Table 1) were grown on solid YPD medium. Twenty replicates (colonies) of each strain were photographed daily, and day five was selected as the static time point. Colonies that failed to grow were removed from subsequent analysis, yielding a dataset of 251 colony images. This dataset was analyzed and uploaded to the YIMAA web application. Representative images are shown in Fig. 2A, with a fluffy colony in the upper left and a smooth colony in the upper right. The ternary-valued segmentation images (below the colony images) illustrate the region-of-interest identified by two rounds of segmentation, with the grey area corresponding to the intra-colony shapes (Methods). Quantitative features were then extracted from the images and normalized to zero-mean and unit variance.

We determined the average classification accuracy (98.79%) by performing a 4-fold cross validation for 5,000 repetitions with Monte Carlo random sampling on the 251 colony images described above. The upper panel of Fig. 2B illustrates the distribution of classification accuracies for the validation partitions in the 5,000 loop trials. The lower panel of Fig. 2B shows the distribution of probability values (also obtained from the 5,000 cross validation repetitions), where the probability of a sample x belonging to the “fluffy” class, $p(x)$, is given by the logistic regression classifier. Classification is performed by dividing the probability space into two

classes. In practice, $p(\mathbf{x}) < 0.5$ corresponds to a smooth classification. Since the classifier is learned using $\frac{3}{4}$ of the samples chosen randomly at each repetition, the actual classification model varies between the trials and the values of model weight vector β change within the validation loop. To analyze the model behavior and learn which features are most informative, we collected the model parameter values in all 5000 trials. As expected, only a small number of features were used in the classifier model during the cross validation, with six features receiving a non-zero weight value in the model weight vector β (Supplemental Materials).

Next, we hierarchical clustered (in feature space) the colony image samples using the subset of six features shown to contribute to the classifier model during cross validation. The clustering (Fig. 2C) showed a clear separation between the fluffy and smooth strains, and the heat map reveals that colonies with the same phenotype share similar feature values. The selection counts confirm that, as expected based on the applied regularization, the logistic regression classifier produced a sparse model using only a small subset of the features. Thus, the classification results obtained with the regularized logistic regression classifier show that the features comprising phenotypic signatures can be used as a basis of classifying complex phenotypes in an automated manner when training samples are available.

Interestingly, the histogram of probability values in Fig. 2B appeared to consist of two main distributions (large peaks on both the smooth and fluffy side) with additional, smaller peaks on each side. Such behavior suggested the existence of phenotypic subclasses or outlier samples. To explore this possibility, we analyzed the images that comprised these small peaks manually and discovered that they corresponded to cases of respiratory deficient mutants (RDM) that had arisen spontaneously from the corresponding “parental” strain. Since the ability to respire drastically affects colony size as well as the ability to form fluffy colonies (23), we removed all images from RDM samples. Repeating the classification procedure described above on the

remaining 238 images resulted in a near perfect average classification accuracy (Fig. 2D), with only 5 false predictions out of 300,000 classifications during cross validation. These probability distributions included only two modes, and together with the improved classification accuracy, suggested that the respiratory deficient mutants were indeed not covered by the two-class model. Finally, we tested whether the logistic regression classification framework could be used to define a third class consisting of respiratory deficient mutants (13 samples). With a limited sample size, we chose a simple leave-one-out cross validation, yielding 96.41% overall accuracy, with all fluffy and smooth samples classified correctly but only 4/13 RDM samples classified correctly. Thus, in this dataset considering the RDM samples separately gives improved classification accuracy for the fluffy and smooth phenotypes, but evaluating the applicability of the proposed framework for automated classification of RDM samples would require a larger dataset.

To test the ability of the method to analyze the spatiotemporal dynamics of colonies as they grow and change shape, we acquired a set of 18 time-lapse image sequences of four different strains (FY4, F29, F45 and YO779), where each sequence contained between one and three colonies. Features were then extracted over the course of the time-lapse, providing a quantitative representation (in feature space) of the morphological dynamics of colonies over time (Fig. 3A). Examples of fluffy and smooth colonies at different times during development are shown in Fig. 3B. We also generated “strain summaries” for each strain at each time point by taking the median value for each feature across all replicates. Both the feature profiles of each individual replicate (colony) and these strain summaries were then analyzed by principal component analysis allowing the trajectories in feature space as the colony develops to be visualized in reduced dimensions (Fig. 3C). The time-lapse results (**Figure 3**) demonstrate that the feature dynamics quantified for fluffy and smooth colonies differ in the two example features, and that the PCA plots reveals different feature trajectories for different phenotypes.

In addition to the image analysis software, we also developed a web application (YIMAA, Supplemental Materials) that allows investigators to easily explore the results of the quantitative analysis alongside the raw input images from their experiment. The default page plots the PCA analysis results for an example from this study (strain F29). Users can also select multiple strains from the drop down list and their PCA results are plotted instantly. The plot can be animated to display points in order across the time series allowing the user to explore the PCA values over time. This animation has “pause” and “play” functions. As the plotting advances, the gallery container shows the raw and segmented image of the most recently plotted point. YIMAA can also plot a time series of any of the several hundred individual features captured by the image analysis pipeline, and clicking on any time point brings up the associated images. Within the gallery panels, choosing a second strain permits side-by-side image comparison. A user guide and screen shots of the YIMAA web application are included in the Supplemental Materials.

Thus, we have developed a platform for the quantitative analysis of yeast colony morphology and demonstrated its use for visualizing changes in colony morphology in feature space. We have also shown that these quantitative colony morphology signatures can be used for supervised classification of colony phenotypes. These methods add statistical rigor to the analysis of colony morphology and will enable the use of a variety of computational tools, such as the classification and visualization tools described here, for the automated analysis of colony shapes. The automated aspect of the software can also enable studies at scales not possible for manual scoring, i.e. extremely large numbers of images. Finally, a web application has been built for easy and rapid sharing of results. This integrative environment for data exploration can be extended to other large-scale image analysis projects and to other colony-forming microorganisms.

Acknowledgments

The authors thank Drs. Cecilia Garmendia-Torres and Alexander Skupin for helpful discussions, and Tapiro Manninen for advice with data analysis. This work was funded by a National Institutes of Health Award (P50 GM076547/Center for Systems Biology) and a strategic partnership between the Institute for Systems Biology and the University of Luxembourg; P.R. is funded by Academy of Finland (project #140052); Z.T. is funded by the Agency for Science, Technology, and Research (Singapore).

Competing interest statement

The authors declare no competing interests.

References

1. **Costerton, J.W., P.S. Stewart, and E.P. Greenberg.** 1999. Bacterial biofilms: a common cause of persistent infections. *Science* 284:1318-1322.
2. **Parsek, M.R. and E.P. Greenberg.** 2005. Sociomicrobiology: the connections between quorum sensing and biofilms. *Trends Microbiol* 13:27-33.
3. **Lachke, S.A., S. Joly, K. Daniels, and D.R. Soll.** 2002. Phenotypic switching and filamentation in *Candida glabrata*. *Microbiology* 148:2661-2674.
4. **Fries, B.C., D.L. Goldman, R. Cherniak, R. Ju, and A. Casadevall.** 1999. Phenotypic switching in *Cryptococcus neoformans* results in changes in cellular morphology and glucuronoxylomannan structure. *Infect Immun* 67:6076-6083.
5. **Cavalieri, D., J.P. Townsend, and D.L. Hartl.** 2000. Manifold anomalies in gene expression in a vineyard isolate of *Saccharomyces cerevisiae* revealed by DNA microarray analysis. *Proc Natl Acad Sci U S A* 97:12369-12374.
6. **Kuthan, M., F. Devaux, B. Janderova, I. Slaninova, C. Jacq, and Z. Palkova.** 2003. Domestication of wild *Saccharomyces cerevisiae* is accompanied by changes in gene expression and colony morphology. *Mol Microbiol* 47:745-754.
7. **Karunanithi, S., N. Vadaie, C.A. Chavel, B. Birkaya, J. Joshi, L. Grell, and P.J. Cullen.** 2010. Shedding of the mucin-like flocculin Flo11p reveals a new aspect of fungal adhesion regulation. *Curr Biol* 20:1389-1395.
8. **Vachova, L., V. Stovicek, O. Hlavacek, O. Chernyavskiy, L. Stepanek, L. Kubinova, and Z. Palkova.** 2011. Flo11p, drug efflux pumps, and the extracellular matrix cooperate to form biofilm yeast colonies. *J Cell Biol* 194:679-687.
9. **Vopalenska, I., V. St'ovicek, B. Janderova, L. Vachova, and Z. Palkova.** 2010. Role of distinct dimorphic transitions in territory colonizing and formation of yeast colony architecture. *Environ Microbiol* 12:264-277.
10. **Botstein, D. and G.R. Fink.** 2011. Yeast: an experimental organism for 21st Century biology. *Genetics* 189:695-704.

11. **Reynolds, T.B. and G.R. Fink.** 2001. Bakers' yeast, a model for fungal biofilm formation. *Science* 291:878-881.
12. **Granek, J.A. and P.M. Magwene.** 2010. Environmental and genetic determinants of colony morphology in yeast. *PLoS Genet* 6:e1000823.
13. **St'ovicek, V., L. Vachova, M. Kuthan, and Z. Palkova.** 2010. General factors important for the formation of structured biofilm-like yeast colonies. *Fungal Genet Biol* 47:1012-1022.
14. **Voordeckers, K., D. De Maeyer, E. van der Zande, M.D. Vinces, W. Meert, L. Cloots, O. Ryan, K. Marchal, and K.J. Verstrepen.** 2012. Identification of a complex genetic network underlying *Saccharomyces cerevisiae* colony morphology. *Mol Microbiol* 86:225-239.
15. **Schneider, C.A., W.S. Rasband, and K.W. Eliceiri.** 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 9:671-675.
16. **Dymond, J.S., S.M. Richardson, C.E. Coombes, T. Babatz, H. Muller, N. Annaluru, W.J. Blake, J.W. Schwerzmann, et al.** 2011. Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* 477:471-476.
17. **Lamprecht, M.R., D.M. Sabatini, and A.E. Carpenter.** 2007. CellProfiler: free, versatile software for automated biological image analysis. *Biotechniques* 42:71-75.
18. **Memarian, N., M. Jessulat, J. Alirezaie, N. Mir-Rashed, J. Xu, M. Zareie, M. Smith, and A. Golshani.** 2007. Colony size measurement of the yeast gene deletion strains for functional genomics. *BMC Bioinformatics* 8:117.
19. **Dittmar, J.C., R.J. Reid, and R. Rothstein.** 2010. ScreenMill: a freely available software suite for growth measurement, analysis and visualization of high-throughput screen data. *BMC Bioinformatics* 11:353.
20. **Wahlby, C., L. Kamentsky, Z.H. Liu, T. Riklin-Raviv, A.L. Conery, E.J. O'Rourke, K.L. Sokolnicki, O. Visvikis, et al.** 2012. An image analysis toolbox for high-throughput *C. elegans* assays. *Nat Methods* 9:714-716.
21. **Posch, A.E., O. Spadiut, and C. Herwig.** 2012. A novel method for fast and statistically verified morphological characterization of filamentous fungi. *Fungal Genet Biol* 49:499-510.
22. **Tan, Z., M. Hays, G.A. Cromie, E.W. Jeffery, A.C. Scott, V. Ahyong, A. Sirr, A. Skupin, and A.M. Dudley.** 2013. Aneuploidy underlies a multicellular phenotypic switch. *Proc Natl Acad Sci U S A* 110:12367-12372.
23. **Rose, M.D., F.M. Winston, P. Hieter, and Cold Spring Harbor Laboratory.** 1990. Methods in yeast genetics : a laboratory course manual. Cold Spring Harbor Laboratory Press, [Cold Spring Harbor, N.Y.].
24. **Otsu, N.** 1979. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* 9:62-66.
25. **Haralick, R.M., K. Shanmugam, and I. Dinstein.** 1973. Textural Features for Image Classification *IEEE Trans. Syst. Man Cybern. SMC-3:610-621.*
26. **Ludwig, O., D. Delgado, V. Gonçalves, and U. Nunes.** 2009. Trainable classifier-fusion schemes: An application to pedestrian detection. *12th International IEEE Conference on Intelligent Transportation Systems, ITSC '09.* p1-6
27. **Ojala, T., M. Pietikainen, and T. Maenpaa.** 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* 24:971-987.
28. **Russ, J.C.** 2011. The image processing handbook. CRC Press, Boca Raton.
29. **Tibshirani, R.** 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* 58:267-288.
30. **Friedman, J., T. Hastie, and R. Tibshirani.** 2010. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* 33:1-22.
31. **Manninen, T., H. Huttunen, P. Ruusuvuori, and M. Nykter.** 2013. Leukemia prediction using sparse logistic regression. *PLoS One* 8:e72932.
32. **Winston, F., C. Dollard, and S.L. Ricupero-Hovasse.** 1995. Construction of a set of convenient *Saccharomyces cerevisiae* strains that are isogenic to S288C. *Yeast* 11:53-55.
33. **Fay, J.C. and J.A. Benavides.** 2005. Evidence for domesticated and wild populations of *Saccharomyces cerevisiae*. *PLoS Genet* 1:66-71.
34. **Liti, G., D.B. Barton, and E.J. Louis.** 2006. Sequence diversity, reproductive isolation and species concepts in *Saccharomyces*. *Genetics* 174:839-850.

Table 1. *S. cerevisiae* strains used in this study.

Name	Genotype	Source
FY4	MAT α , Prototroph	F. Winston (33)
F7 (YPG385)	MAT α <i>hoΔ::HphMX6, SPS2:EGFP:KanMX4</i>	This study ¹
F11 (YPG407)	MAT α <i>hoΔ::HphMX6, SPS2:EGFP:NatMX4</i>	This study ¹
F18 (YPG490)	MAT α <i>hoΔ::HphMX6, SPS2:EGFP:NatMX4</i>	This study ¹
F25 (YPG542)	MAT α <i>hoΔ::HphMX6, SPS2:EGFP:NatMX4</i>	This study ¹
F29 (YPG586)	MAT α <i>hoΔ::HphMX6, SPS2:EGFP:NatMX4</i>	This study ¹
F31 (YPG583)	MAT α <i>hoΔ::HphMX6, SPS2:EGFP:NatMX4</i>	This study ¹
F45 (YPG725)	MAT α <i>hoΔ::HphMX6, SPS2:EGFP:NatMX4, unmapped serine auxotrophy</i>	A. Dudley (23) ¹
F47 (YPG746)	MAT α <i>hoΔ::HphMX6, SPS2:EGFP:KanMX4</i>	This study ¹
F49 (YPG755)	MAT α <i>hoΔ::HphMX6, SPS2:EGFP:KanMX4</i>	This study ¹
YPG339	MAT α <i>hoΔ::HphMX6, SPS2:EGFP:NatMX4</i>	This study ¹
YPG344	MAT α <i>hoΔ::HphMX6, SPS2:EGFP:NatMX4</i>	This study ¹
YPG348	MAT α <i>hoΔ::HphMX6, SPS2:EGFP:KanMX4</i>	This study ¹
YPG352	MAT α <i>hoΔ::HphMX6, SPS2:EGFP:NatMX4</i>	This study ¹
YPG356	MAT α <i>hoΔ::HphMX6, SPS2:EGFP:KanMX4</i>	This study ¹
YPG360	MAT α <i>hoΔ::HphMX6, SPS2:EGFP:NatMX4</i>	This study ¹
YO779	MAT α <i>hoΔ::HphMX6, SPS2:EGFP:NatMX4, unmapped serine auxotrophy, ρ- or ρ⁰</i>	This study; a respiratory deficient isolate of F45

¹ Haploid segregant of a cross between UC5 (34) and DBVPG1853 (35)

Figure Legends:

Figure 1. The components of the platform for automated, quantitative analysis of yeast colonies.

Figure 2. Phenotype analysis of colonies from static images. **A)** Example images of fluffy and smooth phenotypes and the corresponding segmentation results. **B)** Classification accuracies (top) and probability values for class representing the complex phenotypes (bottom) during the 5000 repetitions. **C)** Hierarchical clustering of the selected feature subspace shows how the features chosen by the logistic regression classifier separate the phenotypes and how the colonies within phenotype show similar feature patterns. **D)** Classification accuracies (top) and probability values for class representing the complex phenotypes (bottom) during 5000 repetitions of hold-out error estimation after excluding respiratory deficient mutants.

Figure 3. Analysis of spatiotemporal dynamics of yeast colonies extracted from time course data. **A)** Examples of features 'mean intensity' and 'energy' during a time-lapse measurement; green lines are for three replicates of the complex F29 strain and blue for the smooth strain YO779. **B)** Snapshots of colonies of F29 (top) and YO779 (bottom) at three times during development (indicated by bold vertical lines in **A**). **C)** Dimensionality reduction of time-lapse feature trajectories using principal component analysis. The trajectories shown are for strain summaries, which are obtained by taking median across all individual colonies at each time point. Strains F29 and F45 are fluffy, while strains S1 (FY4, Table 1) and S2 (YO779, Table 1) are smooth.