

遺伝子破壊株イメージ・マイニング

東京大学大学院新領域創成科学研究科

○森下 真一 大矢 禎一

Image Mining for *Saccharomyces cerevisiae* Mutants with Abnormal Morphology

Shinichi Morishita and Yoshikazu Ohya

Graduate School of Frontier Sciences, University of Tokyo

To study the global regulation of cell morphology, a number of groups have recently reported genome-wide screening data for yeast mutants with abnormal morphology. Despite the relatively simple ellipsoidal shape of yeast cells, in the past, cell morphology researchers have processed information on cells manually. These time-consuming, entirely subjective tasks motivated us to develop image-processing software that automatically extracts yeast cells from micrographs and processes them to measure key morphological characteristics such as cell size, roundness, bud neck position angle, nuclear DNA localization and actin localization. We have completed to retrieve 1,907,842 cells from 91,653 micrographs of 4,780 non-lethal mutants using our software, and the results are available in the *Saccharomyces cerevisiae* Morphological Database (SCMD, <http://yeast.gi.k.u-tokyo.ac.jp/>). Our striking observation is that cell morphology of 1,924 among 4,780 mutants are significantly altered with respect to at least one of such 380 parameters that no abnormality of any parameter is detected in wild type after 126 independent trials of experiments. In addition, data mining analysis of these parameter values uncovered a number of genes relevant with the control of cell morphology.

1. はじめに

生物個体の最小単位である細胞の増殖は、非常にたくさんの遺伝子産物が、複雑に相互作用し、巧みに連携することによって制御されていることが明らかにされつつある。幾つかのモデル生物種においては、ゲノム配列の解読が完了し、ゲノムにコードされる遺伝子数が明らかにされている。ポストゲノムの更なる展開としては、様々な生命現象の各事象に着目し、個々の事象に関与する新規因子の同定を目指すことによってその分子メカニズムを明らかにしようとする方向性と、細胞を一つのシステムとして捉え、その枠組みの中で個々の遺伝子産物がどのように機能しているのかを体系的に理解しようとする方向性の二つの局面を迎えている。

後者の流れでは、個々の遺伝子発現や遺伝子産物間の相互作用、遺伝子産物の細胞内局在に関する網羅的な情報収集が行われ、近年では、これらの情報をもとにした遺伝子機能予測の試みがさかんに行われている。従来、個々の遺伝子機能を明らかにする手法としては、ある遺伝子を欠失させた細胞が、正常に働いている細胞と比べて、どのような変化を示すのかという表現型を詳細に解析し、その因果関係を追及するという方法が取られてきた。しかし、この手法を網羅的に行うためには、一つ一つの遺伝子を欠損

させた細胞が必要であり、また、いかに各遺伝子の機能欠損を反映した表現型を効率よく見つけ出すことができるかという点で非常に困難である。従って、網羅的な遺伝子欠損株の表現型解析は遅れているのが現状であり、表現型情報の不足は遺伝子機能の体系的な理解を遅らせている原因と考えられた。

そこで、我々は遺伝学的解析手法が強力に発揮できる出芽酵母の系を用い、既に構築されている網羅的な遺伝子破壊株を実験材料として、個々の遺伝子破壊株が示す細胞形態の異常性に着目した解析を行うことにした。細胞形態の異常性を表現型として解析する理由は、その制御機構が生育環境や増殖過程と密接に連携しており、非常に多くの遺伝子産物の協調的な働きによって維持されていると考えられることから、これを解析することで、より多くの遺伝子機能に関する知見が得られると考えられたためである。各組織や諸器官を形成する細胞種特有の多様な形態形成メカニズムを解明することは生物学の大きなテーマの一つであるが、そのメカニズムの全貌を明らかにするためにも、単一細胞である出芽酵母の形態形成メカニズムを理解することが必須であると考えられる。従って、網羅的に作製された遺伝子破壊株の形態的特徴を表現型の一つとして客観的かつ定量的な情報で表し、迅速に抽出して比較分類することにより、形態制御に関わる新たな因子を同定し、各遺伝子産物の機能予測を行うための情報基盤を確立することが重要である。

2. 研究開発の成果概要

2.1 酵母細胞の自動認識と細胞形態パラメータの定量化

出芽酵母遺伝子破壊株の細胞形態及び細胞内構造の状態を細胞の写真から自動認識するには、まず、細胞の外郭、アクチン細胞骨格、および核をそれぞれ FITC-ConA、Rh-ph、及び DAPI 蛍光試薬を用いて三重染色し、各々の染色画像を撮影することが第1のステップである。つづいて FITC-ConA 像から、細胞の大きさ、細長さ、芽の位置などの形態パラメータに基づく個々の細胞形態情報を取得する[1]。3次元の構造をしている細胞を2次元画像として取得するため、小さな芽細胞に焦点を合わせることが困難な場合が多く、芽細胞の認識率を上げる特殊なアルゴリズムの研究に力を注ぎ、エラー率を1%未満に抑えることに成功した。次に、Rh-ph 像からアクチン細胞骨格の細胞内局在に関する情報を、DAPI 像からは核の位置に関する情報を抽出することができ(図1)、細胞外郭の情報と組み合わせて分類することにより、細胞極性や細胞周期の進行についての情報も得ることが可能になった(図2)。4784株の非必須遺伝子破壊株のうち細胞の凝集が激しいため画像解析が困難な4株を除いた4780株を対象に、形態パラメータを定量化し、WWW から公開中である[2]。

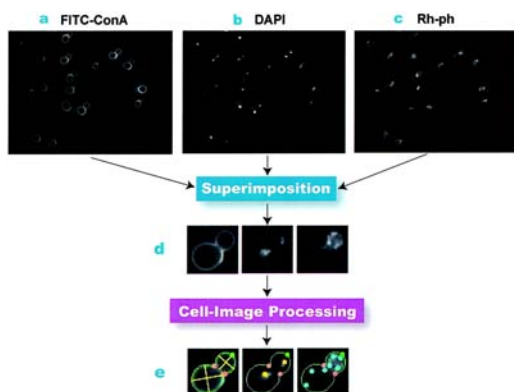


図1 細胞壁(a) 核(b) アクチン(c) を染色し、個々の細胞を認識し、画像を処理する流れ

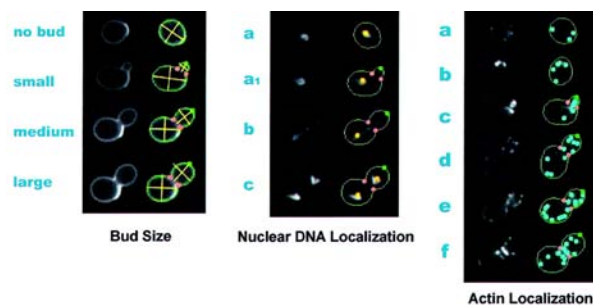


図2 芽細胞の大きさ、核の位置、アクチン分布の情報から各細胞を細胞周期進行に分類

2.2 異常な形態をもつ非必須遺伝子破壊株

野生株に比べて、形態が異常に変化する破壊株はいくつぐらい存在するか？ この疑問に答えるためには、対象とする形態パラメータを選択する必要がある、我々は生物学的に意味があると考えられる形態パラメータを501個選択した。つづいて、野生株を基準に破壊株の異常性を検出する目的に叶うように、野生株における形態パラメータの分布を調査した。野生株から200個以上の細胞を抽出する作業を独立に126回繰り返えし、各形態パラメータについて126個の値を収集し解析したところ様々な分布をもつことがわかった。理由として、画像解析プログラムの出力に誤差が含まれること、野生株の実験条件の微妙な差が生みだす揺らぎなどが考えられる。一方、4780の破壊株ごとに200個以上の細胞を抽出する作業を1回だけおこない、各形態パラメータの異常性を、野生株から得られた126個の値の分布に照らし合わせて判定した。この過程で、以下の2つの問題を解決した。

- 1) 野生株から得られた126個の値の分布の分散が、破壊株から得られた4780個の値の分散を上回る形態パラメータが501個中82個も存在した。これらの形態パラメータに対して、野生株は形態異常と誤判断しないような有意水準を設定し、破壊株の形態異常をみつけることは困難である。したがって除外し、残りの419個を候補とした。
- 2) 野生株から得られた126個の分布は様々な形をしているため同じ基準で異常性の判定を行うことが困難である。そこで各形態パラメータの分布を **power transformation** により正規分布に近似し、標準化することを試みた。そして、近似後の分布の正規分布との適合度を **Shapiro-Wilk** 検定で調べ、確率が閾値 α より小さい形態パラメータを正規分布への近似が難しいため異常性の判定基準の設定が困難なパラメータと判定し除外した。 α が0.1, 0.3, 0.5に応じて、正規分布へ変換できた形態パラメータの数は各々380個, 340個, 285個であった。

正規分布への変換が可能であった野生株での形態パラメータの分布を使って、片側確率 β で棄却される破壊株を形態異常と判定する。 α と β の値から少なくとも1つのパラメータで形態異常と判定された破壊株と野生株の個数の実数、野生株の分布から推定される形態異常破壊株の個数の期待値を表1に示す。

表 1 形態異常と判定される破壊株および野生株の個数

α	0.1			0.3			0.5		
パラメータ数	380 個			340 個			285 個		
	破壊株		野生株	破壊株		野生株	破壊株		野生株
β	実値	期待値	実値	実値	期待値	実値	実値	期待値	実値
1.E-02	4,703	4,778	110	4,691	4,775	110	4,663	4,765	106
1.E-03	3,788	2,546	35	3,727	2,360	33	3,570	2,078	25
1.E-04	2,657	350	2	2,573	314	0	2,446	265	0
1.E-05	1,924	36	0	1,849	32	0	1,735	27	0
1.E-06	1,460	4	0	1,383	3	0	1,296	3	0

β が1.0E-05 以下であれば、独立に観測された126個の野生株で形態異常と判定される株は存在せず、画像処理ソフトウェアの誤り率と野生株間での揺らぎの影響を受けずに、野生株が形態異常であると誤認定する可能性を排除できたことを意味する。一方、ある破壊株が本当は形態異常を起こしていない

にもかかわらず、形態異常であると誤って判定してしまうことが確率的には起こりうる。単純化した仮定として、全パラメータが独立で、破壊株も野生株と同じ分布に従うとすると、形態異常であると判定される破壊株の個数の期待値($(1-(1-2\beta)^{\text{パラメータ数}}) \times 4780$) は0ではないため、本当は形態異常を起こしていないにもかかわらず、形態異常と誤って認定してしまう可能性があり、その期待値を表1は示している。ただ、 β が1.0E-05の場合、誤認定の期待値は実値の2%未満に過ぎない。1つの遺伝子を破壊しただけで形態異常を示す破壊株が1,924株も存在するという結果は新しく、従来過小評価されがちであった非必須遺伝子の破壊が、これほどまでに形態に多くの変化をもたらすことを示す驚くべき知見である。

2.3 遺伝子機能の予測

多くの遺伝子破壊株に形態変化が見られたことから、遺伝子産物の機能と細胞形態に密接な関係を見いだすことができるかどうかを網羅的に調べることにした。そのために、全非必須遺伝子破壊株の定量的な形態情報をもとに、どのような遺伝子機能の欠損が形態変化を引き起こしやすいのかを精査した。具体的には、ある機能に関連する遺伝子群の変異株に共通する形態的特徴に着目し、その形態的特徴が類似している変異株を、その機能欠損の候補株として予測し、各候補遺伝子に関する既知の情報を精査した。その結果、特に DNA 修復に関わる遺伝子群、細胞壁合成に関与する遺伝子群、アクチン機能に関連する遺伝子群の三カテゴリーについては、既に明らかにされている遺伝子の機能から判断して、予測の正答率が非常に高いと考えられたが、予測された遺伝子の中には既知の情報がない遺伝子もあった。この予測結果については、さらに実験的な検証を行うことによって、いくつかの遺伝子機能に関する裏付けが得られつつある。また画像解析ソフトウェアを用いることで、我々は細胞壁合成をモニターする新規の細胞周期チェックポイントを発見し、これまでチェックポイントと全く関係がないと思われていた因子群をつなぐ新たなシグナル伝達経路の存在を提示することができた[3]。

3. まとめ

過去3年間に、遺伝子破壊株の画像解析ソフトウェアを完成し、4780の非必須遺伝子破壊株中で約2,000株が野生株に比べ形態が異常に変異している事実をつきとめた。また新規の細胞周期チェックポイントを発見した。今後は、形態パラメータによる遺伝子のグループ化を行い、細胞内局在および Gene Ontology との対応を調べ、機能未知の遺伝子の機能を推定するイメージマイニング法を確立することを目指し、またヘテロ二倍体遺伝子破壊株を用いて必須遺伝子破壊株からも形態パラメータを取得する計画である。

4. 研究開発実施体制

代表研究者 森下 真一(東京大学 大学院新領域創成科学研究科 情報生命科学専攻)

研究開発題目

(1) 画像処理/データマイニングソフトウェア、ウェブサーバーの設計・開発 (情報グループ)

グループリーダー 森下 真一(東京大学 大学院新領域創成科学研究科 情報生命科学専攻)

(2) 出芽酵母遺伝子破壊株の顕微鏡画像取得と網羅的な分子遺伝学的研究(生物グループ)

グループリーダー 大矢 禎一(東京大学 大学院新領域創成科学研究科 先端生命科学専攻)

5. 参考文献

- [1] M. Ohtani, A. Saka, F. Sano, Y. Ohya, and S. Morishita. Development of Image Processing Program for Yeast Cell Morphology, *J. Bioinfo. Comp. Biol.*1(4): 695-709 (2004)
- [2] T. L. Saito, M. Ohtani, H. Sawai, F. Sano, A. Saka, D. Watanabe, M. Yukawa, Y. Ohya, and S. Morishita. SCMD: *Saccharomyces Cerevisiae* Morphological Database. *Nucl. Acids. Res.* 32: D319-D322 (2004)
- [3] M. Suzuki, R. Igarashi, M. Sekiya, T. Utsugi, S. Morishita, M. Yukawa, Y. Ohya. Dynactin is involved in a checkpoint to monitor cell wall synthesis in *Saccharomyces cerevisiae*. *Nature Cell Biology* 6, 861 - 871 (2004)