

Title	メタボロミクスを用いた株改良法の開発および酵母1-ブタノール耐性への適用
Author(s)	Teoh, Shao Thing
Citation	大阪大学, 2016, 博士論文
Version Type	VoR
URL	<a href="https://doi.org/10.18910/55994">https://doi.org/10.18910/55994</a>
rights	
Note	

*Osaka University Knowledge Archive : OUKA*

<https://ir.library.osaka-u.ac.jp/>

Osaka University

博士学位論文

メタボロミクスを用いた株改良法の開発および  
酵母 1-ブタノール耐性への適用

TEOH SHAO THING

2016年1月

大阪大学大学院工学研究科

# 目次

第1章.....	3
緒論.....	3
1.1. 表現型向上および株改良の戦略.....	3
1.1.1. 合理的な手法.....	3
1.1.2. ランダムな手法.....	3
1.1.3. 半合理的な手法.....	4
1.2. メタボロミクスの株改良における位置づけ.....	6
1.2.1. メタボロミクスの概要.....	6
1.2.2. 比較研究におけるメタボロミクスの長所.....	6
1.3. 研究目的.....	7
1.4. 博士論文の構成.....	8
第2章.....	9
メタボロミクスを用いた酵母1-ブタノール耐性関与遺伝子の探索.....	9
2.1. 緒言.....	9
2.2. 実験方法.....	12
2.2.1. 実験材料.....	12
2.2.2. 1-ブタノール耐性測定.....	13
2.2.3. メタボローム解析用サンプル準備.....	13
2.2.4. GC/MS 分析.....	14
2.2.5. データ処理.....	15
2.2.6. 多変量解析.....	15
2.3. 結果と考察.....	15
2.3.1. 変異株のスクリーニングおよび1-ブタノール耐性測定.....	15
2.3.2. ガスクロマトグラフィ-質量分析 (GC/MS) を用いたメタボローム解析.....	19
2.3.3. 直交潜在構造投影法 (Orthogonal Projections to Latent Structures ; OPLS) による回帰モデル構築.....	21
2.3.4. OPLS モデル結果に基づいた新規1-ブタノール耐性株の推定.....	26
2.3.5. 新規変異株の耐性測定およびメタボローム解析.....	28
2.4. 小括.....	32
第3章.....	33
RANSAC アルゴリズムに基づく新規メタボロミクスデータマイニング法の開発.....	33
3.1. 緒言.....	33
3.2. 実験方法.....	36
3.2.1. データセット.....	36

3.2.2. Random Sample Consensus-Partial Least Squares (RANSAC-PLS)アルゴリズム	36
3.2.3. 階層クラスタ解析(Hierarchical Cluster Analysis, HCA)	37
3.2.4. その他の多変量解析	37
3.2.5. 耐性測定および GC/MS によるメタボローム解析	37
3.3. 結果と考察	37
3.3.1. RANSAC-PLS の構築	37
3.3.2. RANSAC-PLS に基づく耐性関連代謝物の推定	40
3.3.3. RANSAC-PLS より得られた重要代謝物に基づく遺伝子改変ターゲットの提案	44
3.3.4. 新規変異株の実験的検証	45
3.3.5. シミュレーションデータを用いた RANSAC-PLS 手法の再確認	48
3.4. 小括	50
第4章	52
総括	52
謝辞	54
引用文献	55
付録	64
発表論文	92
第一著者論文	92
共著者論文	92
学会発表	93

# 第 1 章

## 緒論

近年、DNA 組換え技術やバイオプロセス工学の発展により、食品、医薬品、バイオ燃料など、様々な有用物質は微生物による生産が可能になった。しかし、競争力のある工業生産に至るまでには、宿主微生物の更なる改善が必要である。微生物の増殖速度、物質生産速度、収率、副産物の抑制、種々のストレスへの耐性など、様々な表現型が物質生産工程に影響するため、それらの向上が生物学における重要なゴールである。工業用微生物の株改良には、様々な戦略および技術が存在するが、用いられる手法は次に述べる 3 つのクラスに大別することができる。

### 1.1. 表現型向上および株改良の戦略

#### 1.1.1. 合理的な手法

従来、株の改良には合理的な手法、つまり目的表現型を支配する遺伝的機構や代謝経路についての情報を用いて遺伝子改変ターゲットを選択するような手法がよく用いられた[1], [2]。しかし、ある表現型の基となる機構について知るためには、遺伝学解析、分子生物学実験、生物化学分析など、複雑かつ時間のかかる実験を積み上げて機構解明の基礎研究を行わねばならない。また、十分に研究されていない表現型については、このような基礎研究による情報が得られない可能性がある。そのため、合理的な株改良は、常に機構解明の研究より遅れている。その他、「耐性」のような複数の遺伝子機構や代謝経路により影響されている複雑な表現型については、ターゲット遺伝子の候補が多すぎるため、その間の相互作用まで考慮した株改良は非常に困難である[3], [4]。

#### 1.1.2. ランダムな手法

一方、実験室進化[5]、突然変異誘発[6]、転写因子変異誘発[3]、またはゲノムシャッフリング[7]といった手法では、機構解明の必要性を回避できる[2]。これらの戦略は、起点としてある親株からランダムな変異を起し、遺伝的多様性のある変異体集団を発生させることを利用している。その変異体集団の中では、確率的に一部の変異体は有利な変異を獲得し、目的表現型が親株に比べて向上することがある。また、変異体の集団

をスクリーニングすれば、これらの株を単離し使用することができる。しかしランダムな手法は、どの遺伝子やタンパク質の違いが表現型に寄与したかという情報を提供しないため、優良株が得られてもその更なる改良は難しいとされる[1]。その上、目的表現型によっては優良な変異体を単離するスクリーニング実験が困難な場合がある[4]。例えば、[8]ではエタノール生産菌である *Kluyveromyces marxianus* の UV による突然変異誘発の後、25,200 もの変異体の発酵培養を行い各々のエタノール生産量を測定している。

### 1.1.3. 半合理的な手法

近年では、オミクス技術の登場により、様々な株にある遺伝子の総体（ゲノム）や転写産物の総体（トランスクリプトーム）の網羅的かつハイスループットな分析が可能となった。また、株間の幅広い比較により、検出された差異に基づく候補遺伝子の改変を提案するような戦略も次第に行われるようになってきている[2], [9]–[11]。このような株間の網羅的比較に基づく株改良手法をここでは「半合理的な手法」と呼ぶ。半合理的な手法は、表現型の異なる株の間に生じた遺伝子やタンパク質の違いのみを情報とするので、ランダム手法と同様に表現型機構の解明を必要としない。また、合理的な手法と同様に、半合理的な手法で提案した遺伝子改変の中で、実際に表現型の向上につながる遺伝子改変は、更なる株改良に有用な情報を与える[12]。例えば、エタノールストレスにさらされた酵母細胞とストレスフリー条件下の酵母細胞とのトランスクリプトーム解析による比較により、トリプトファン<sup>1</sup>の生合成および取り込み関連遺伝子はエタノール耐性に寄与していると報告された[13]。この研究例では、トリプトファン<sup>1</sup>の耐性機構への関与は詳しく調べていないが、比較の結果のみで、酵母エタノール耐性を向上させるような遺伝子ターゲットを提案することができた。また、半合理的な手法はランダム手法と組み合わせることも可能である。例えば、ランダム手法で得られアルコール耐性体と、その親株とゲノム比較により、酵母のエタノールおよびイソブタノール耐性を向上させる遺伝子過剰発現ターゲットの発見が報告された[12]。また、大腸菌ではイソブタノール耐性株と親株とのゲノム比較により中鎖アルコール耐性賦与に有効な遺伝子破壊ターゲットを発見することができた[14], [15]。それから、全ゲノムを包括した遺伝子変異株ライブラリーの網羅的な耐性測定による耐性関連遺伝子の探索法も、酵母のエタノール耐性[16]や大腸菌の1-ブタノール耐性[17]に適用されている。これからも、科学技術の進歩により定量的オミクスデータを更にハイスループットで取得可能となるにしたがって、

半合理的手法はますます重要になっていくと期待される。

以上、合理的手法、ランダム手法および半合理的手法の株改良法について Table 1-1 でまとめる。

**Table 1-1. 微生物改良における合理的手法，半合理的手法，ランダム手法の比較**

合理的	半合理的	ランダム
目的表現型を支配する遺伝子機構や代謝経路についての解明情報に基づく遺伝子改変ターゲットの選択	異なる株や培養条件の細胞のオミクスレベル解析および比較に基づく遺伝子改変ターゲットの提案	ランダムな変化による多様性のある変異体の発生および優良株の単離
(例) 熱耐性向上のためのヒートショックタンパク質の過剰発現, アルコール耐性向上のためのアルコール排出ポンプタンパク質の発現	(例) 全ゲノム比較, 全トランスクリプトーム比較	(例) 実験室進化, 転写因子突然変異誘発
機構についての解明情報に依存する	機構についての解明情報に依存しない	機構についての解明情報に依存しない
優良株の取得は更なる改良に有用な情報を提供する	優良株の取得は更なる改良に有用な情報を提供する	優良株の取得は更なる改良に有用な情報を提供しない
機構の解明に様々な基礎研究の手法が必要であり, 煩雑かつ時間がかかる	網羅的でハイスループットなオミクス解析技術が必要	数多くの変異体から優良株を回収するためのスクリーニングが必要

## 1.2. メタボロミクスの株改良における位置づけ

### 1.2.1. メタボロミクスの概要

分子生物学のセントラルドグマにより、遺伝情報の流れは遺伝子 (DNA) から始まり、転写産物 (mRNA) に転写され、さらにタンパク質に翻訳され、様々な生物学的機能を果たす。細胞内の遺伝子、転写産物、タンパク質の総体は、それぞれゲノム、トランスクリプトーム、プロテオームと言われ、その網羅的解析に関わる研究分野はゲノミクス、トランスクリプトミクス、プロテオミクスと呼ばれる。また、代謝物がタンパク質の働きにより生産、消費、交換されることにより、近年では代謝物が遺伝情報の流れの究極の段階と見なされ、タンパク質に続くオミクス層を形成するとされてきた[18]。また、前述のオミクス層と同様、細胞内の代謝物の総体は「メタボローム」と言い、その網羅的解析を中心とした研究分野は「メタボロミクス」と呼ばれる。メタボロミクスはゲノミクスやトランスクリプトミクスに類似したところが多いが、いくつかの点で特徴を持っている。その一つは対象となる分子の数である。一般的に、生体に存在する代謝物の種類は、その生体の遺伝子数より遥かに少ない。例えば、出芽酵母 (*Saccharomyces cerevisiae*) の 6,000 近くの遺伝子に対し、代謝物は 600~1,000 と推定されている[19]。一方、ゲノミクス、トランスクリプトミクス、プロテオミクスは、それぞれ一つの分析プラットフォームで、その対象分子のほぼ全体を検出・計測することができる。しかし、メタボロミクスでは代謝物の化学的多様性により、一つの分析装置で全ての代謝物を検出・測定することは、もっとも簡単な細菌細胞についても[20]現在では不可能である[21]。また、遺伝子、転写産物、およびタンパク質は、一般に一对一の関係が示されるが、代謝物はそれぞれ、複数の代謝反応に関わっている[22]。したがって、ある代謝物存在量の変化について上流にさかのぼり、どの関連遺伝子またはタンパク質が変化したかを調べることは容易ではない。最後に、メタボロミクスデータは、変数 (代謝物) が少なく、定量的に扱うことができる[23]。すなわち、ゲノミクス[14], [15], [24] では一般的に遺伝子の有無を、トランスクリプトミクス[25]–[27]やプロテオミクス[28]–[30]では倍率変化が閾値を超える遺伝子を、定性的情報として用いられることが多いが、メタボロミクスでは代謝物の相対定量値そのものが情報を持っているとされる。

### 1.2.2. 比較研究におけるメタボロミクスの長所

DNA 配列から表現型までの遺伝情報の流れは、転写調節、転写後調節、翻訳後調節



などにより干渉を受けている。したがって、ゲノミクス、トランスクリプトミクスにより遺伝子や転写産物レベルで細胞間の違いが検出されても、その違いが最終の表現型に反映されない可能性が大きい[31]。一方、メタボロミクスは表現型に近いオミクス層、つまりメタボロームを形成する代謝物を網羅的に解析する。このような特徴からメタボロームの比較は、基礎研究、応用研究のいずれにも有用性を示してきた[32], [33]。また、転写産物やタンパク質が遺伝情報の流れにおける媒体であるのに対し、代謝物は様々な細胞制御プロセスを経た最終の遺伝子産物であるため、その存在量の変化は生物系の遺伝的または環境的变化に対する究極な反応を表しているといえる[34], [35]。したがって、メタボロームに起こる変化は、注目した表現型と高い相関を示すと考えられる。代謝物と表現型の関係が密接に関連しているため、メタボロームの比較により注目の表現型に関連する代謝物が発見されることが期待できる。このことから、メタボロミクスは表現型に寄与する様々な細胞内機構の基礎研究のみならず、前述の半合理的株改良手法のような応用研究においても有力な技術であると考えられる。

### 1.3. 研究目的

前述したように、半合理的株改良法は、合理的手法の長所（更なる改良のための情報を提供すること）およびランダム手法の長所（複雑な機構解明を前提条件としないこと）の両方を持ち合わせる強力な手法である。また、メタボロミクスは株改良において有用であると考えられる。それにも関わらず、メタボロミクスを用いた株改良の研究報告は少なかった。また比較メタボロミクスを用いた場合でも、それを単独な手法、あるいは主要な手法としたわけではなく、他の機能的ゲノム解析と併用した補足技術とされることが多い。そして、メタボロミクスを用いた多くの研究例では、株改良といった応用研究ではなく機構の解明、つまり基礎研究のために株の比較を行うのみである。このような状況から、メタボロミクスを用いた株改良の研究例を立てて、メタボロミクスの応用分野に貢献する必要があると感じた。以上を踏まえて、本博士論文では次の研究目的を設定した。

- (1) 株改良を目的としたメタボロミクスによる改変候補遺伝子の探索法の構築
- (2) 半合理的株改良法におけるメタボロミクスの有用性の証明

## 1.4. 博士論文の構成

以上の研究目的を達成するために、メタボロミクスの特徴を利用した実験的手法を構築した。本実験的手法を簡潔に述べると、種々の表現型性能の異なる株を用いて、その表現型の測定およびメタボローム解析による代謝物定量を行い、それから多変量解析を用いて表現型と相関を持つ重要代謝物を特定する。そして、得られた標的代謝物量を調節するための遺伝子改変ターゲットを提案するというものである。

第2章では本手法の概念実証に、出芽酵母の1-ブタノール耐性を対象とした株改良を試みた。また、その手法または戦略において回帰モデルの構築が中心となる。しかし、本研究を遂行するうちに従来の線形回帰モデル構築法の短所が明らかになった。そこで、第3章では従来法の問題点を回避するため、画像処理の分野で用いられる **Random Sample Consensus (RANSAC)** アルゴリズムを応用し、新規なデータマイニング手法を構築し検証した。

第2章と第3章の研究内容を合わせることにより、メタボロミクスの半合理的株改良における有用性を示した。すなわち、新たな手法とアイデアで株改良への応用に貢献したといえる。

## 第 2 章

### メタボロミクスを用いた酵母 1-ブタノール耐性関与遺伝子の探索

#### 2.1. 緒言

第 1 章で述べたように、ハイスループットなオミクス技術の発展により、株間での生体分子の比較を行うことで表現型向上に資する寄与する遺伝子改良ターゲットの提案に基づく半合理的株改良法が可能になった[10], [11]. このような手法は表現型の基になる機構の解明を前提条件とせず、従来の合理的改良法の主な短所[9] を回避できると期待される. また、メタボロミクスは生体内の低分子代謝物を解析対象とし[34], [35], 半合理的手法に応用できると考えられる[33]. メタボロミクスは種々の微生物表現型の解明、または向上に用いられた研究例が多く[36]–[39], メタボロミクスを用いた株改良法の有効性が示唆されている.

半合理的株改良において、例えば「感受株」と「耐性株」というような 2 種類のサンプルのみを比較する場合、その比較は容易である. しかし、2 種類のみ比較した場合、その間に検出された違いの全てが目的表現型に関与しているわけではないため、偽陽性率が高くなる[40], [41]. 一方、複数株の比較により検出された目的表現型(例えば耐性)と関連している代謝物の信頼性は高くなる[42]が、株数が多い場合は 2 株ずつの相互比較は複雑になる. そこで、メタボロミクスデータの定量的性質を利用し、代謝物量を説明変数に、目的表現型を応答変数とし、回帰モデルを構築する戦略が使用可能である[43]. 回帰モデルにおいて、重要な代謝物、つまり大きい相関係数を持った代謝物は、遺伝子改変ターゲットの探索に用いられる. 特に、メタボロミクスデータは再現性が高く、ダイナミックレンジが広く、回帰モデルの量的説明変数としての使用に適していると考えられる. また、代謝物と表現型との密接な関係により、様々な値を取りうる量的表現型に対して、関連の代謝物が相関を示すように連続の値を取ることが予測できる. そのため、従来のトランスクリプトミクスは「健康体と病体」、「処理前と処理後」、「感受株と耐性株」のような異なるクラスのサンプル比較や分類に用いられていたに対し、メタボロミクスは脊椎動物発生段階[44]や酵母分裂寿命[43] といった量的表現型の定量的予測モデルに用いられた例がある. また、モデルの構築には、多重共線性の問題を解決できモデル解釈が比較的容易な[45], [46]部分最小二乗法 (Partial Least

Squares ; 潜在構造投影法 Projection to Latent Structures または PLS と呼ぶ[47]) が広く用いられる。

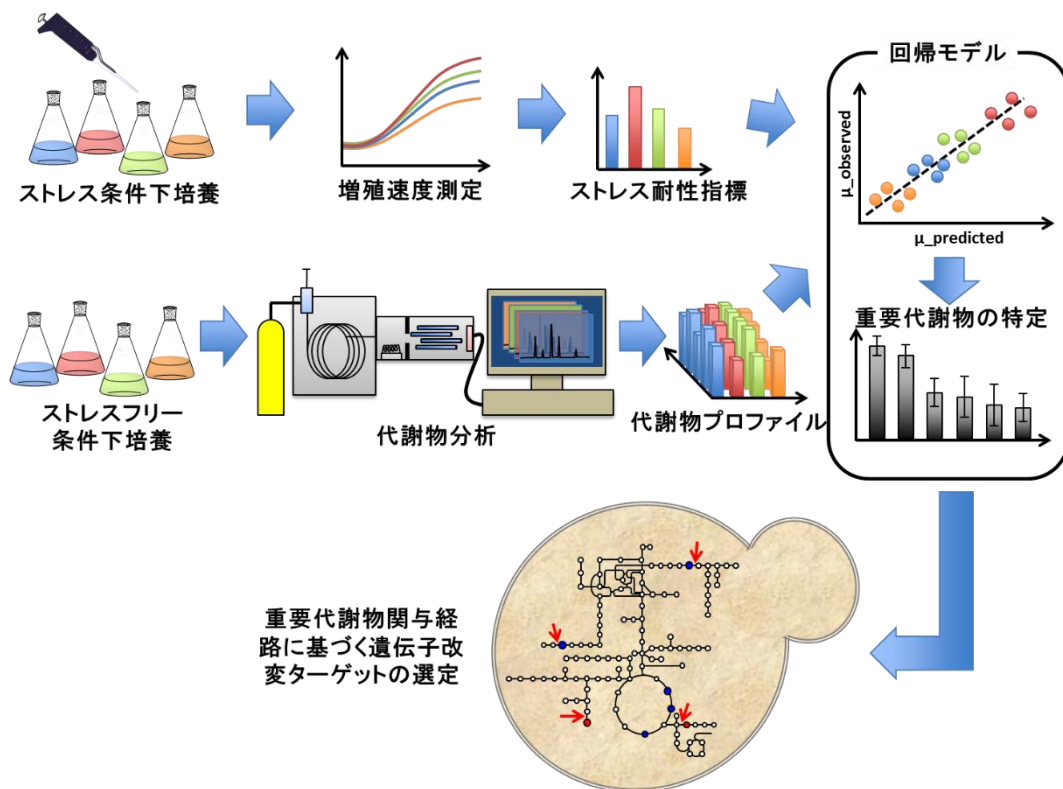
さらに、以上のような回帰モデルの詳しい解析により、目的表現型と関係のある代謝物およびその関連代謝経路の推測ができ、その情報を利用して表現型の向上を目指した新規遺伝子改変ターゲットの提案が可能となる。例えば[43]では、寿命関与遺伝子の変異株より得られた代謝物プロファイルを用いて、寿命の回帰モデルを構築した後、寿命関連代謝物の特定を行い、それらの代謝物に基づき新たな変異株を提案することにより、寿命の遺伝子機構についての解明情報を用いることなく、新たな長寿命株を取得可能であることが示された。以上より、メタボロミクスは表現型向上を目指した遺伝子改変ターゲットの候補選択に有用であることが示され、本研究では工業用微生物の様々な重要表現型（増殖速度、生産速度、耐性など）の向上についても同手法が有効であると考えている。

本章では、株改良に寄与するメタボロミクスを用いた改変候補遺伝子の探索法を構築すること目的とした。また、その概念実証に、出芽酵母の 1-ブタノール耐性を向上させることを目的表現型として選んだ。従来のバイオ燃料のエタノールに比べ、1-ブタノールのような中鎖アルコールは細胞毒性が高く[1]、1-ブタノールによる細胞膜の破壊、エネルギー代謝や栄養素輸送といった細胞過程の干渉、タンパク質の変性、DNA や脂質の酸化ダメージ、RNA の分解など、様々な悪影響が示されている[48]。なお、その測定可能な結果として、比増殖速度、基質取り込み速度、終細胞濃度、化合物生産速度および収率、様々な量的表現型の低下が挙げられる。一般に、耐性という概念とは細胞のそれらの効果に抵抗する能力のことを表している。また、宿主微生物の増殖速度の低下は、1-ブタノールの生産工程における重要な障害の一つであると挙げられている[49]。なお、*Clostridium acetobutylicum* におけるグルタチオン過剰生産[50]および出芽酵母の転写因子 eIF2B の点変異[51]による耐性向上はそれぞれ宿主微生物の 1-ブタノール生産性能を向上させたと報告されている。ゆえに、微生物によるバイオ燃料生産工程において産物であるアルコールへの耐性は重要で改善すべき表現型であると考えられる。以上を踏まえ、本研究では増殖速度に注目し、1-ブタノールストレス条件下での増殖速度をその株の 1-ブタノール耐性の指標とした。なお、本研究に用いる微生物に関しては、(1) 重要な工業用微生物である、(2) ゲノム情報が豊富なモデル微生物である、(3) 遺伝子改変技術が確立されている[52]という点で、出芽酵母 (*Saccharomyces cerevisiae*) を実験材料として選択した。また、出芽酵母に関しては

酵母の遺伝子破壊株コレクションが既に存在しているため、メタボロミクスによる比較をする対象の様々な株が容易に入手できるという点も、概念実証の対象に用いるメリットであると考えられた。

これまで、酵母のエタノール耐性について多くの研究が行われてきたが、バイオ燃料としての1-ブタノールは近年着目されたものであり、その耐性についての研究はまだ盛んではない。[24], [53]では、一般の実験室と進化により得られた耐性株をゲノム解析で比較し、その結果、主にRPN4, RTG1, YLR224Wが代表するユビキチン-プロテアソーム系によるタンパク質分解経路が耐性に関与していると示唆された。また、[54]では1-ブタノール感受性株と耐性株をプロテオーム解析で比較したが、ミトコンドリア活性、グリセロール生合成、総合ストレス応答が耐性に関与していると示唆され、種々の候補遺伝子の中GPP2（グリセロールホスファターゼ）の過剰発現は耐性の向上を導いたと報告されている。本研究では、様々な株をメタボローム解析による比較を行い、過去とは違う耐性関連遺伝子を見つけ出すことを試みた。

本章で用いた研究概略をFig. 2-1に示した。最初に、種々の酵母変異株について1-ブタノール添加ありの条件で液体培養し、その比増殖速度を1-ブタノール耐性の指標とした。同時に、同株を1-ブタノール添加無し（ストレスフリー）条件で培養および細胞回収し、メタボローム解析を行い、代謝物プロファイルを取得した。代謝物情報と耐性値を回帰モデル構築に供し、モデルに基づき耐性と相関関係を示す重要代謝物を特定した。これらの重要代謝物は耐性に直接関与しているという仮説のもと、関連代謝経路を参照し、代謝物量を調節するための遺伝子改変ターゲットを提案した。最後に、提案の遺伝子改変を持つ新規変異株を入手し、その1-ブタノール耐性とメタボローム解析を行うことで、モデルの予測を検証し、本戦略の有効性を証明した。



**Fig. 2-1.** 本研究に用いたメタボロミクスに基づいた株改良戦略

種々の酵母変異株について1-ブタノール添加ありの条件での比増殖速度を1-ブタノール耐性の指標とし、同株を1-ブタノール添加無し（ストレスフリー）条件で回収した細胞の代謝物プロファイルを用いて回帰モデルを構築した。モデルに基づき耐性と相関関係を示す重要代謝物を特定し、その関連代謝経路の参照より、代謝物量を調節するための遺伝子改変ターゲットを提案した。最後、提案した遺伝子改変を持つ新規変異株について1-ブタノール耐性とメタボローム解析を行い、モデルの予測を検証した。

## 2.2. 実験方法

### 2.2.1. 実験材料

本研究では、出芽酵母 BY4742 (MAT $\alpha$  leu2 $\Delta$ 0 lys2 $\Delta$ 0 ura3 $\Delta$ 0 his3 $\Delta$ 1)及び、Saccharomyces Genome Deletion Project [55]により構築され、EUROSCARF コレクション ([http://web.uni-frankfurt.de/fb15/mikro/euroscarf/stra\\_des.html](http://web.uni-frankfurt.de/fb15/mikro/euroscarf/stra_des.html)) から提供されている BY4742 の致死性でない遺伝子破壊株を使用した。液体培養前の保存用平板培地には、YPD プレート (10 g/L BactoTM yeast extract (BD, NJ, USA), 20 g/L BactoTM peptone (BD, NJ, USA), 20 g/L D-glucose (Nacalai Tesque, Kyoto, Japan),

20 g/L agar (Nacalai Tesque, Kyoto, Japan)) を使用した。前々培養，前培養，本培養には酵母完全培地 (6.7 g/L Difco™ yeast nitrogen base without amino acids (BD, NJ, USA), 1.92 g/L yeast synthetic drop-out media supplement without uracil (Sigma-Aldrich, MO, USA), 76 mg/L uracil (Sigma-Aldrich, MO, USA) , 20 g/L D-glucose) を使用した。

#### 2.2.2.1 - ブタノール耐性測定

-80°C で保存された酵母のグリセロールストックを YPD プレートへ植菌し，30°C で 2 日間静置培養した。プレート上に発生した菌体を更に新しい YPD プレートに植え継ぎ 30°C で 1 日静置培養した。新しい YPD プレート上に生じたシングルコロニーを 15 mL 容量のファルコンチューブ (BD, NJ, USA) 中の 1.5 mL の SC 培地に植菌し，200 rpm で振とう培養 (Bio-Shaker BR-40LF (Taitec, Tokyo, Japan)) を行い，前々培養とした。前々培養液から  $OD_{600}=0.01$  となるように，シリコン栓 (Shin-Etsu Polymer, Tokyo, Japan) を付けた 50 mL 容量のファルコンチューブ (BD, NJ, USA) 中の 5 mL の SC 培地に移し替え，前培養を開始し，13 時間振とう培養を行った。前培養液から  $OD_{600}=0.15$  となるようにシリコン栓を付けた 50 mL のファルコンチューブ中の 6 mL の SC 培地に移し替え，本培養を開始した。また，ストレス耐性測定のために本培養液に 1% または 1.5% (v/v) 1-ブタノールを添加した。本培養において，継時的に培養液の  $OD_{600}$  を iMark microplate reader (Bio-Rad, CA, USA) を用いて測定し，対数増殖期に入る範囲で測定した 4 点を利用し  $OD_{600}$  値の自然対数  $\log(OD_{600})$  対時間のプロットにフィットした直線の傾きを比増殖速度  $\mu$  とした。また，1-ブタノールストレス下での比増殖速度  $\mu_{\text{stress}}$  を耐性の指標とした。

#### 2.2.3. メタボローム解析用サンプル準備

メタボローム解析は[56]で報告された手順に似た手順で行った。培養手順は，耐性測定と同じ手順を用いた。本培養は  $OD_{600}=0.15$  で開始した後， $OD_{600}$  を iMark microplate reader で追跡し， $OD_{600}=2$  となる時間で培養液 5 mL をフィルター (直径 25 mm, ポアサイズ 0.45  $\mu\text{m}$  のナイロン膜フィルター (Millipore, MA, USA)) を用いて回収した。内部標準物質として Ribitol (Wako, Osaka, Japan) 2  $\mu\text{g}/\text{mL}$  を添加した抽出液 1 mL (methanol:water:chloroform = 5:2:2) (methanol, chloroform (Chameleon Reagent, Osaka, Japan), distilled water (Wako, Osaka, Japan)) の入った 2 mL のマイクロチ

ューブ (Eppendorf, Hamburg, Germany) にフィルターを入れ、直ちに液体窒素につけクエンチングを行った。また、クエンチングまでの回収手順は 30 秒以内で完了するようにした。抽出まで回収したサンプルは $-80^{\circ}\text{C}$  で保存した。

回収したサンプルを  $4^{\circ}\text{C}$ 、1200 rpm の条件で 30 分間 Thermomixer Comfort (Eppendorf, Hamburg, Germany) を用いてインキュベートした。850  $\mu\text{L}$  の抽出液を新しいマイクロチューブに移し、500  $\mu\text{L}$  の超純水 (Wako, Osaka, Japan) を加えてかく拌した後、16000 rcf,  $4^{\circ}\text{C}$  で 3 分間遠心分離を行った。上清 175  $\mu\text{L}$  を新しいマイクロチューブに移し、VC-96R Spin Dryer Standard (Taitec, Tokyo, Japan) を用いて遠心濃縮を 2 時間行い、メタノールを除去した。その後、VD-800F Freeze Dryer (Taitec, Tokyo, Japan) を用いて一晩凍結乾燥を行った。

サンプルの誘導体化にはオキシム化とシリル化を行った。オキシム化は 40  $\mu\text{L}$  の pyridine (10 mg/mL methoxyamine hydrochloride を含む) を加え、 $30^{\circ}\text{C}$  下で攪拌しながら 90 分間行った。シリル化は 50  $\mu\text{L}$  の MSTFA を加え、 $37^{\circ}\text{C}$  下で攪拌しながら 30 分間行った。誘導体化されたサンプルは 24 時間のうちに GC/MS 分析に供した。

#### 2.2.4. GC/MS 分析

GC/MS 分析には GC-2010 Plus gas chromatograph (島津製作所, Kyoto, Japan), AOC-20is series injector/autosampler (島津製作所) および GCMS-QP2010 Ultra mass spectrometer (島津製作所) を用いた。カラムには内径 0.25 mm, 長さ 30 m の fused silica capillary column coated with 0.25  $\mu\text{m}$  InertCap 5MS/NP (GL Sciences, Tokyo, Japan) を用いた。Front inlet の温度は  $230^{\circ}\text{C}$  に設定した。ヘリウムガスの流量は 1.12 mL/min に設定した。カラムオーブンは 2 分間、 $80^{\circ}\text{C}$  を維持し、続いて  $^{\circ}\text{C}$  から  $330^{\circ}\text{C}$  まで  $15^{\circ}\text{C}/\text{min}$  の条件で昇温させ、その後  $330^{\circ}\text{C}$  を 6 分間維持させた。Transfer line および ion-source の温度はそれぞれ  $250^{\circ}\text{C}$  と  $200^{\circ}\text{C}$  に設定した。MS は 85-500 m/z の範囲を 20Hz でデータ取得した。

また、化合物ピークの保持係数を計算するために、C8-20 アルカンの混合液、C21-40 アルカンの混合液およびピリジンのミックスを準備し、各アルカンピークの保持時間を測定した。なお、酵母サンプルの分析中、6 サンプルごとにピリジンのブランクサンプルを注入し、カラムブリードやキャリーオーバーが起こっていないことを確認した。



### 2.2.5. データ処理

GC/MS より得られた生データは GCMSsolution (島津製作所, Kyoto, Japan)にて netCDF フォーマット(\*.cdf)に変換し, MetAlign[57]を用いてピーク検出, ベースライン補正, 保持時間補正 (アラインメント) を行った. 続いて, MSClust[58]を用い, 同一化合物に由来していると推測されたピークを同じ保持時間に統一した. その後, データ行列は Aloutput2 ver.1.29 [59]に供し, ピークの自動同定を行い, 最後に代謝物定量値は内部標準物質である Ribitol にて標準化した.

### 2.2.6. 多変量解析

得られた代謝物定量値のデータ行列は SIMCA-P+ version 12 (Umetrics, Umeå, Sweden) に供し多変量解析を行った. データは Unit variance (UV)でスケージングした後, 代謝物定量値を説明変数, 耐性指標値を応答変数として OPLS 法により回帰モデルを構築した.

## 2.3. 結果と考察

### 2.3.1. 変異株のスクリーニングおよび 1 - ブタノール耐性測定

本研究では, 出芽酵母 BY4742 株および, それを親株とした 107 の単一遺伝子破壊株を実験材料とした(付録の Table A1 を参照). 破壊株の選択にあたり, Saccharomyces Genome Database[60]で「転写因子」とされる遺伝子の破壊株を本研究では選択した. 転写因子は様々な代謝経路や遺伝子系の調節に関与しており, その一遺伝子の破壊で多くの遺伝子発現に効果を与え, 広範囲な代謝に影響することが予想される. したがって, 転写因子破壊株群より, 様々な代謝プロファイルおよび耐性性能を持つ株が得られると期待される[61]. そして, 構築された回帰モデルの精度向上には, 測定エラーに比べ説明変数および応答変数の分散を大きくすることが必要であるので, 幅広い耐性性能を示し得る転写因子破壊株の株群を用いることが有利であると考えられる.

本実験に入る前に, 実験対象の株数を予備実験により絞り込んだ. 予備実験では, 親株 BY4742 および 107 の変異株について, 1-ブタノール添加あり (ストレス) と無し (ストレスフリー) の条件下で増殖速度を  $N=1$  で測定し, 以下の条件を満たす変異株を選択した.

- ① Saccharomyces Genome Database に掲載された情報により, 代謝に関連する転写因子の破壊株である (33 株)

- ② 親株と似たストレスフリー条件下比増殖速度を示す (BY4742 との差が全体の測定値標準偏差より小さい ; 90 株)
- ③ 幅広い範囲のストレス条件下比増殖速度を示す

**Table 2-1. メタボローム解析および回帰モデル構築に用いた変異株**

<b>Strain</b>	<b>Description in Saccharomyces Genome Database</b>
<i>aro80Δ</i>	Zinc finger transcriptional activator of the Zn2Cys6 family; activates transcription of aromatic amino acid catabolic genes in the presence of aromatic amino acids
<i>azf1Δ</i>	Zinc-finger transcription factor, involved in induction of CLN3 transcription in response to glucose; genetic and physical interactions indicate a possible role in mitochondrial transcription or genome maintenance
<i>bas1Δ</i>	Myb-related transcription factor involved in regulating basal and induced expression of genes of the purine and histidine biosynthesis pathways; also involved in regulation of meiotic recombination at specific genes
<i>dal80Δ</i>	Negative regulator of genes in multiple nitrogen degradation pathways; expression is regulated by nitrogen levels and by Gln3p; member of the GATA-binding family, forms homodimers and heterodimers with Deh1p
<i>gat2Δ</i>	Protein containing GATA family zinc finger motifs; similar to Gln3p and Dal80p; expression repressed by leucine
<i>gcn4Δ</i>	Basic leucine zipper (bZIP) transcriptional activator of amino acid biosynthetic genes in response to amino acid starvation; expression is tightly regulated at both the transcriptional and translational levels
<i>leu3Δ</i>	Zinc-knuckle transcription factor, repressor and activator; regulates genes involved in branched chain amino acid biosynthesis and ammonia assimilation; acts as a repressor in leucine-replete conditions and as an activator in the presence of alpha-isopropylmalate, an intermediate in leucine biosynthesis that accumulates during leucine starvation
<i>lys14Δ</i>	Transcriptional activator involved in regulation of genes of the lysine biosynthesis pathway; requires 2-amino adipate semialdehyde as co-inducer
<i>mks1Δ</i>	Pleiotropic negative transcriptional regulator involved in Ras-CAMP and lysine

---

	biosynthetic pathways and nitrogen regulation; involved in retrograde (RTG) mitochondria-to-nucleus signaling
<b><i>mot3Δ</i></b>	Transcriptional repressor and activator with two C2-H2 zinc fingers; involved in repression of a subset of hypoxic genes by Rox1p, repression of several DAN/TIR genes during aerobic growth, and repression of ergosterol biosynthetic genes in response to hyperosmotic stress; contributes to recruitment of the Tup1p-Cyc8p general repressor to promoters; involved in positive transcriptional regulation of CWP2 and other genes; can form the [MOT3+] prion
<b><i>oaf1Δ</i></b>	Oleate-activated transcription factor, acts alone and as a heterodimer with Pip2p; activates genes involved in beta-oxidation of fatty acids and peroxisome organization and biogenesis
<b><i>put3Δ</i></b>	Transcriptional activator of proline utilization genes, constitutively binds PUT1 and PUT2 promoter sequences as a dimer and undergoes a conformational change to form the active state; differentially phosphorylated in the presence of different nitrogen sources; has a Zn(2)-Cys(6) binuclear cluster domain
<b><i>rsf2Δ</i></b>	Zinc-finger protein involved in transcriptional control of both nuclear and mitochondrial genes, many of which specify products required for glycerol-based growth, respiration, and other functions
<b><i>sip4Δ</i></b>	C6 zinc cluster transcriptional activator that binds to the carbon source-responsive element (CSRE) of gluconeogenic genes; involved in the positive regulation of gluconeogenesis; regulated by Snf1p protein kinase; localized to the nucleus
<b><i>sko1Δ*</i></b>	Basic leucine zipper transcription factor of the ATF/CREB family; forms a complex with Tup1p and Cyc8p to both activate and repress transcription; cytosolic and nuclear protein involved in osmotic and oxidative stress responses
<b><i>stp2Δ</i></b>	Transcription factor, activated by proteolytic processing in response to signals from the SPS sensor system for external amino acids; activates transcription of amino acid permease genes
<b><i>thi2Δ</i></b>	Transcriptional activator of thiamine biosynthetic genes; interacts with regulatory factor Thi3p to control expression of thiamine biosynthetic genes with respect to thiamine availability; acts together with Pdc2p to respond to

---

---

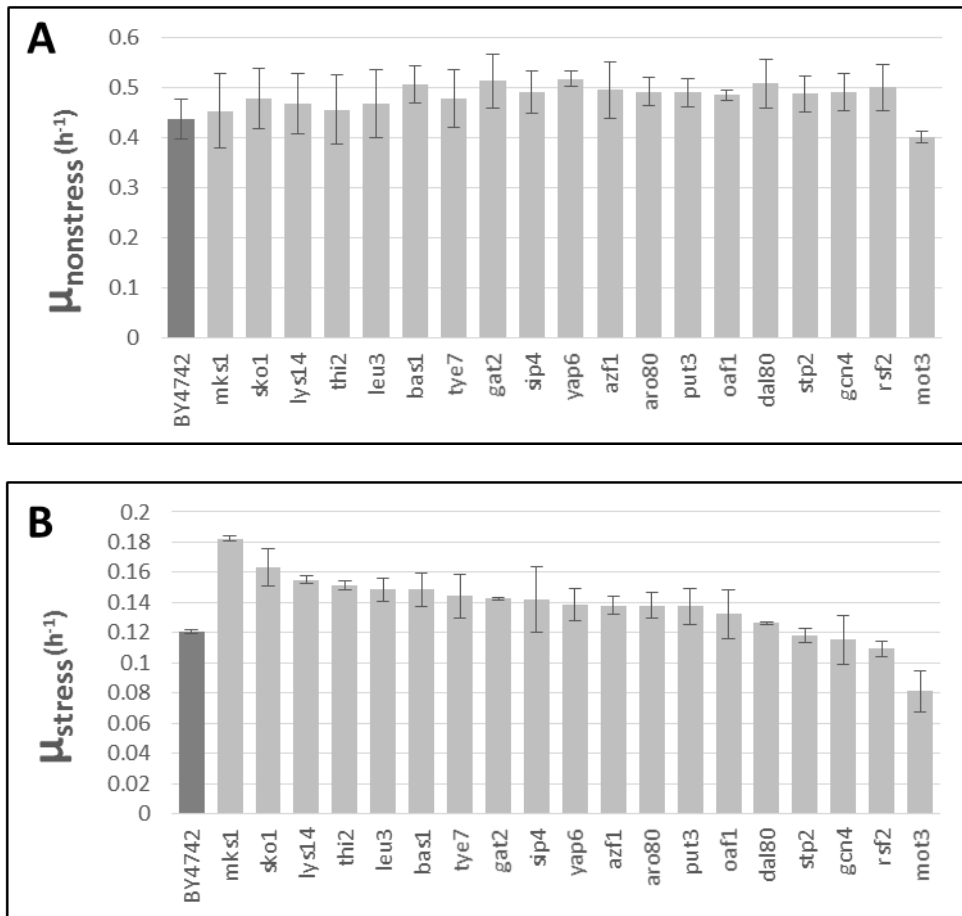
thiaminediphosphate demand, possibly as related to carbon source availability;  
zinc finger protein of the Zn(II)2Cys6 type

***tye7Δ*** Serine-rich protein that contains a basic-helix-loop-helix (bHLH) DNA binding motif; binds E-boxes of glycolytic genes and contributes to their activation; may function as a transcriptional activator in Ty1-mediated gene expression

***yap6Δ*** Basic leucine zipper (bZIP) transcription factor; physically interacts with the Tup1-Cyc8 complex and recruits Tup1p to its targets; overexpression increases sodium and lithium tolerance; computational analysis suggests a role in regulation of expression of genes involved in carbohydrate metabolism

---

以上の条件で、メタボローム解析に用いる変異株は Table 2-1 の 19 株に絞ることができた。まず、この 19 株についてストレス無し条件下での増殖速度  $\mu_{\text{nonstress}}$  を測定し、ストレス無し条件下では各株の増殖速度が大きく変わっていないことを確認した。Fig. 2-1A で示されるように、*mot3Δ*以外の株はほとんど同じ増殖速度を示した。また、*mot3Δ*株を含んでも様々な株 (Fig. 2-1 に示される BY4742 は除く) の  $\mu_{\text{nonstress}}$  の平均値は  $0.482 \text{ h}^{-1}$  で標準偏差は  $0.0261 \text{ h}^{-1}$  であって、標準偏差と平均値の比、つまり相対標準偏差は 0.0540 (5.40%) だったので、様々な株の  $\mu_{\text{nonstress}}$  は平均から大きく変わらないと言える。次に、以上 19 株は 1.5% (v/v) 1-ブタノール濃度の培養条件下で改めストレス条件下増殖速度  $\mu_{\text{stress}}$  を測定した。Fig. 2-2B で示されるように、このストレス条件下では、様々な株は幅広い比増殖速度を示し、測定値の範囲は  $0.0810 \text{ h}^{-1}$  (*mot3Δ*) から  $0.182 \text{ h}^{-1}$  (*mks1Δ*) までであった。また、様々な株の  $\mu_{\text{stress}}$  の平均  $0.137 \text{ h}^{-1}$  に対し、標準偏差は  $0.0212 \text{ h}^{-1}$ 、相対標準偏差は 0.154 (15.4%) であった。以上のそれぞれ計算値より、選ばれた 19 株の株群は互いに類似したストレス無し条件下増殖速度を示しながら幅広いストレス条件下増殖速度を示すことが確認された。また、緒論で述べられたように、各株のストレス条件下増殖速度 ( $\mu_{\text{stress}}$ ) は、その株の 1-ブタノールストレス耐性の指標値とした。



**Fig. 2-2. メタボローム解析に用いた 19 株の増殖速度測定結果**

(A) メタボローム解析に使用された 19 株および親株のストレス無し条件下での比増殖速度 ( $\mu_{\text{nonstress}}$ ). 各株について N=4 測定の平均値を表し, エラーバーはその株の測定値標準偏差を表している. (B) 同じ株群の 1.5% (v/v) 1-ブタノール濃度における比増殖速度 ( $\mu_{\text{stress}}$ ). 各株について N=2 測定の平均値を表し, エラーバーは標準偏差を表している. 測定値は付録の Table A3 に表示される.

### 2.3.2. ガスクロマトグラフィ - 質量分析 (GC/MS) を用いたメタボローム解析

様々な増殖速度を示した株のメタボローム解析による比較は, メタボロームが増殖速度によって変動することにより複雑になると考えられる [62]. しかし, 耐性測定の結果から明らかであるように, ストレス条件下では様々な変異株が幅広い増殖速度を示していることがわかる. したがって, ストレス条件下でのメタボロームを株の比較に用いた場合, 増殖速度の影響によるバイアスが生じうると考えられた. 一方, 生体細胞では遺伝子の機能欠損など, 代謝フラックスを変えようとする摂動を受けた場合, 代謝物存在

量が変化することにより、フラックスの変動が抑えられるような、代謝フラックスの恒常性システムを持っている[63]. このことから、ストレスフリー条件のメタボロームも、ストレス条件下の増殖速度を反映する情報を持っていると考え、本研究ではストレスフリー条件下で回収した細胞をメタボローム解析に用いることにした.

出芽酵母の 6000 近くの遺伝子数に比べ、その独特な代謝物の数はたった 600 前後であると推測されている[63]. しかし第 1 章で述べたように、種々の代謝物は幅広い化合物の多様性を示し、単一の分析プラットフォームでの網羅的分析は現在不可能である. したがって、本研究では分析対象を主に親水性代謝物に限定した. その理由は、親水性代謝物の抽出、誘導体化および分析のプロトコールが確立されていること[64], また親水性代謝物の情報は生物の生理的状态または表現型(例えば、ゼブラフィッシュの発生段階[44]と酵母株の分裂寿命[43])の予測に使用できることが示されているからである. なお、本研究の分析プラットフォームには、ガスクロマトグラフィ-質量分析(GC/MS)を用いた. GC/MS は、再現性が高く、多くのピークが取得できる、保持係数(Retention Index, RI) およびマススペクトルでのライブラリー参照によるピーク同定が簡単である、などの利点が挙げられる[59]. これらの利点により、GC/MS は偏向の少ない非標的(ノンターゲット; non-targeted)分析を行っても、多くのピークの検出、同定および定量が可能であると期待された.

次に、以上 19 株についてメタボローム解析を行った. ここで、解析目的は株間の比較であるため、代謝物の絶対定量をする必要がなく、各株について代謝物の相対定量を行った. サンプル準備および分析の詳細は実験方法に記載した. また、各株については N=4 (異なるシングルコロニーから) サンプルを用意した. 付録 Table A4 に記載されたパラメーターでデータ処理後、50 の同定化合物と 15 の未同定化合物、全部で 65 の化合物ピークの定量値を格納したデータ行列が得られた. 同定された化合物はアミノ酸、有機酸、核酸、糖など代謝物クラスを占めた (Table 2-2). なお、回帰モデルに供したデータ行列に格納された 65 化合物の詳細は付録の Table A5 で表示される.

**Table 2-2. GC/MS 分析より同定できた代謝物**

<u>アミノ酸および関連代謝物</u>		
Alanine	Glycine	Serine
Asparagine	Isoleucine	Threonine
Aspartic acid	Leucine	Tryptophan

Cystathionine	Lysine	Tyrosine
Cystine	Phenylalanine	Valine
Glutamic acid	Proline	
Glutamine	Pyroglutamic acid	
<u>解糖系+TCA サイクル関連代謝物</u>		
Isocitric acid / Citric acid	Malic acid	Oxalacetic acid / Pyruvate
<u>核酸および関連代謝物</u>		
Orotic acid	Uracil	
<u>尿素サイクルおよび関連代謝物</u>		
Ornithine	Urea	
<u>ポリアミン</u>		
Cadaverine	Spermidine	
<u>その他</u>		
2-Aminoadipic acid	Lactic acid	Plamitic acid(16:0)
4-Aminobenzoic acid	n-Propylamine	Quinolinic acid
Glucarate	Phthalic acid	Stearic acid(17:0)

代謝物対がスラッシュ（ / ）を挟んで記載される場合は，その代謝物は本研究の分析システムで分離できなかったことを意味する．なお，化合物同定の詳細，および実際に OPLS モデルに供したピークの情報は付録の Table A5 に表示される．

### 2.3.3. 直交潜在構造投影法（Orthogonal Projections to Latent Structures ; OPLS）による回帰モデル構築

メタボロミクスデータの解析には様々な手法が存在する．連続の値を取る応答変数の回帰分析には，直交潜在構造投影法（Orthogonal Projections to Latent Structures ; OPLS） [65]が広く使用されている．OPLS 法は，高次元の説明変数データを，その分

散を良く説明するような数少ない潜在変数（または‘成分’）に要約する。また、その一つである予測成分は、応答変数と最大の相関を持つ方向に説明変数空間内でベクトル軸が構築され、各サンプルをその軸へ投影した値を用いることで、応答変数の回帰分析をすることができる。これが、OPLS 回帰モデルの構築である。また、予測成分以外の成分は応答変数と相関しないので直交成分と呼び、それによって説明されるデータの分散は、応答変数（耐性）と関係のない因子に該当すると考えられる。そのため OPLS 法は、メタボローム解析の段階に起こりうる系統的誤差と確率的誤差を取り除くに有用であるともいえる。なお OPLS 法を応用した回帰モデル構築例として、緑茶品質の階級 [64]、動物の発生段階 [44]、酵母株の寿命 [43] などが報告されている。本研究では OPLS 法を 1 - ブタノール耐性の回帰モデル構築に用いた。

OPLS 回帰モデル構築にあたり、75 サンプルからの 65 代謝物ピークの相対定量値を説明変数に、その株の耐性値を応答変数とした。構築されたモデルは、4 つの潜在変数（1 つの予測成分および 3 つの直交成分）から構成された。ここで、モデルを評価するために以下 3 つの指標値を用いた。

$$RMSEE = \sqrt{\frac{\sum(y_{obs} - y_{pred})^2}{N}}$$

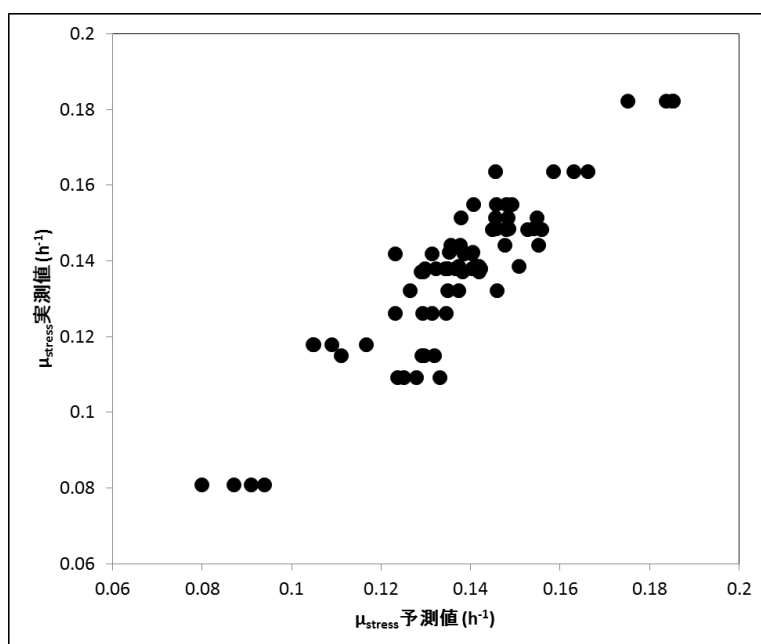
$$R^2Y = 1 - \frac{SSY_A}{SSY_0} = 1 - \frac{\sum(y_{residual} - \bar{y}_{residual})^2}{\sum(y_{obs} - \bar{y})^2}$$

$$Q^2Y = 1 - \frac{PRESS}{SSY_0} = 1 - \frac{\sum(y_{obs} - y_{predcv})^2}{\sum(y_{obs} - \bar{y})^2}$$

RMSEE (Root Mean Squared Error of Estimation) は、説明変数（ここで代謝物定量値）からモデルにより計算された応答変数（耐性）の予測値と、応答変数の実測値の差、つまりモデルのエラーを表している。また、RMSEE は応答変数と同じ次元、つまり  $h^{-1}$  になる。今回得られた RMSEE 値は  $0.00898 h^{-1}$  であり、応答変数ベクトルの標準偏差  $\sigma_y$  ( $0.0213 h^{-1}$ ) と比べて小さい ( $RMSEE / \sigma_y = 0.421$ ) ので、モデルは十分に正確であると考えられる。一方、 $R^2Y$  はモデルにより説明された  $y$  分散の割合を表している。上の式では、 $SSY$  は Y Sum of Squares、つまり平方和を意味していて、 $y$  変数の平方和は  $SSY_0$  で表し、A 個の潜在変数により説明された分散を引き算した後の  $y$  残差平方和は  $SSY_A$  で表しているが、モデルにより応答変数  $y$  が良く説明されるほど、 $SSY_A$  は 0 に近づき、 $R^2Y$  は 1 に近づく。今回得られたモデルでは  $R^2Y = 0.835$  なので、



$y$ 分散 ( $y$ の変動の傾向) がモデルにより良く説明されたと考えられる. そして,  $Q^2Y$  はモデルの堅牢性, つまりクロスバリデーションを行った時の  $y$  予測値 (計算値) の安定性を表している. クロスバリデーションとは, モデルの一部 (一般的に 1 / 7 のサンプル数) をモデル構築の段階から外して, モデルを再構築し, このように構築されたサブモデルにより先ほど外されたサンプルについて  $y$  の予測値を計算する, という操作のことを指すが, 一般的には各サンプルが一度外されて予測値を計算されるように, データの 1 / 7 ごとにクロスバリデーションが自動的に行われている. なお, 上の式で表されるように, PRESS (Predicted Sum of Squares) とは各サンプルの予測エラーの平方和のことを表し, PRESS と  $y$  変数平方和  $SSY_0$  の比率を 1 から引いた値は  $Q^2Y$  として表される. 本モデルでは  $Q^2Y=0.554$  で, クロスバリデーション下での予測エラーの平方和は  $y$  変数の平方和より半分以下小さかったため堅牢性が十分に高いと判断した.

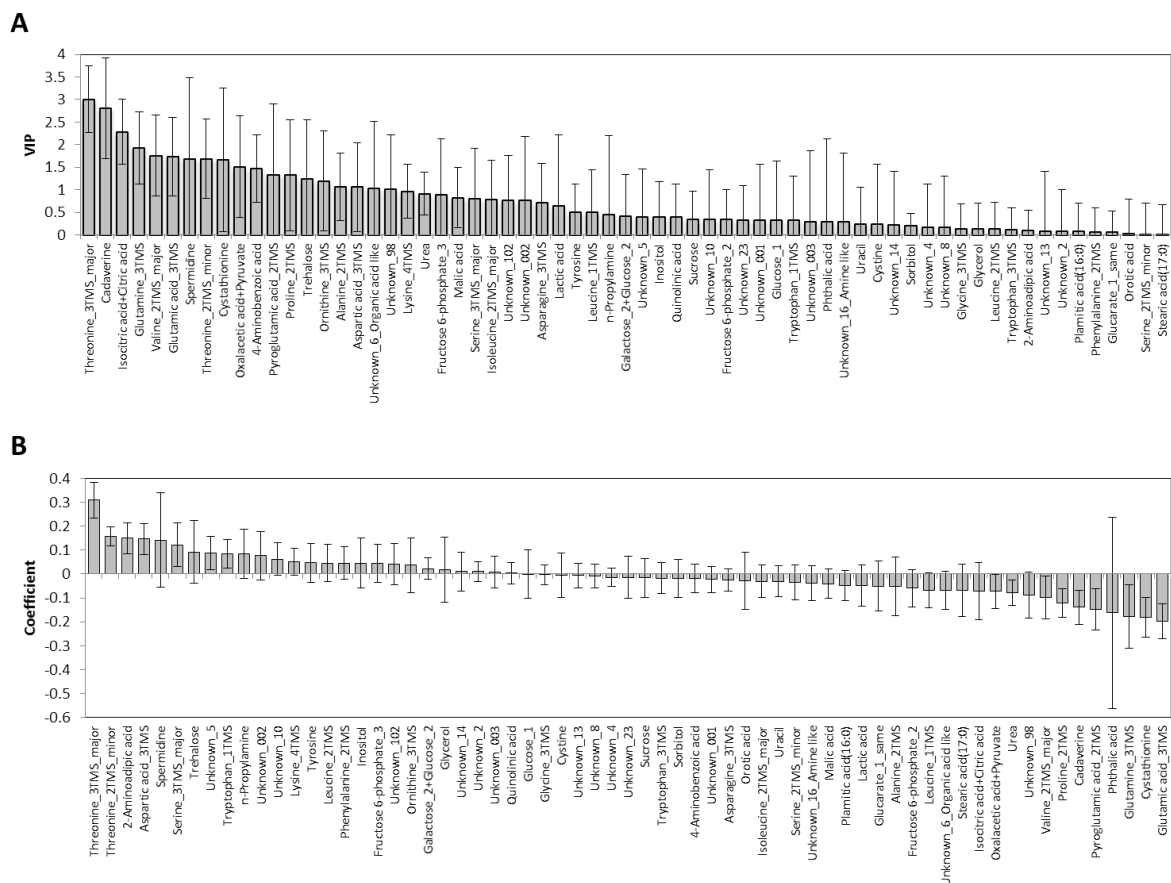


**Fig. 2-3. OPLS モデルの実測値対予測値プロット**

縦軸は株の耐性実測値( $\mu_{\text{stress\_Predicted}}$ )を示し, 横軸は各サンプルの代謝物プロファイルから計算した耐性予測値( $\mu_{\text{stress\_Observed}}$ )を表した. 1 株について N=4 サンプルのメタボローム解析を行った.

続いて, 構築された OPLS モデルから, 耐性と相関関係を示す重要代謝物の特定を行った. そのためには, それぞれの代謝物について計算された Variable Importance in

the Projection (VIP)値と PLS coefficient を用いた(Fig. 2-4). 単一応答変数の OPLS において, VIP 値は元のデータ行列における代謝物と耐性との相関の大きさを表す値である. 一方, 応答変数と相関のない直交成分により説明された分散を説明変数データから引き算した後, データの残差行列においては各代謝物の耐性との相関が違ってくるが, そのときの相関係数は PLS coefficient により表される. なお, Coefficient は相関の方向 (正または負) を示しており, 代謝物と耐性との関係をより詳細に表している. つまり, Coefficient は OPLS モデルの包括的な情報を有すると言えるが, モデルによっては直交成分の数が多すぎる場合, 小さな確率的誤差などにより代謝物の Coefficient が大きく変わることがある. VIP 値と Coefficient のそれぞれに有用な情報が含まれているため, ここで両方とも耐性関連代謝物の推定に用いることにした. また, Fig. 2-4 の各プロットから, 値が大きく, エラーバーの小さい代謝物を特定し重要代謝物とした.



**Fig. 2-4. OPLS モデルによる各代謝物のスコア**

(A) Variable Importance in the Projection (VIP). (B) PLS Coefficients. 各グラフのエラーバーはクロスバリデーションにより計算された標準誤差を示す.

Fig. 2-4A で表示されるように, AOutput により同定された化合物ピークには Threonine\_3TMS\_major, Cadavarine, Isocitric acid+citric acid, Glutamine\_3TMS, Valine\_2TMS\_major, Glutamic acid\_3TMS, Threonine\_2TMS\_minor が高い VIP 値を示した. 一方, Fig. 2-4B より, Coefficient の大きな化合物ピークには Threonine\_3TMS\_major, Threonine\_2TMS\_minor, 2-Aminoadipic acid, Aspartic acid\_3TMS (Coefficient の正のもの), および Glutamic acid\_3TMS, Cystathionine (Coefficient の負のもの) が挙げられる. この際, 各化合物の VIP または Coefficient 値の大きさだけでなく, そのエラーバーも考慮した. エラーバーについては, ここではモデルクロスバリデーションの結果により計算された標準誤差を表している. エラーバーが小さい場合は, 一部のデータをモデルから除外してもその化合物の VIP または Coefficient はあまり変わらない, つまりその値が安定していることを示している. また, 複数の化合物が同一の代謝物に該当する (例えば Threonine\_2TMS\_minor と Threonine\_3TMS\_major とは, スレオニンに TMS 基がそれぞれ 2 個と 3 個付いた誘導体である) 場合もあった. このように, 耐性と相関を示した重要代謝物のリストを Table 2-3 のように作成した.

Table 2-3. OPLS モデルより特定された 1 - ブタノール耐性関連代謝物の候補

Metabolite	High VIP	Large Coeff	Correlation
2-Aminoadipic acid		✓	Positive
Aspartic acid		✓	Positive
Cadavarine	✓		Negative
Cystathionine		✓	Negative
Glutamic acid	✓	✓	Negative
Glutamine	✓		Negative
Isocitric acid+citric acid	✓		Negative
Threonine	✓	✓	Positive
Valine	✓		Negative

### 2.3.4. OPLS モデル結果に基づいた新規 1-ブタノール耐性株の推定

次に、Yeast Biochemical Pathway Database (<http://pathway.yeastgenome.org>)を用いて、Table 2-3 の候補代謝物が多く含まれる代謝経路を探索した。その結果、重要代謝物がスレオニン生合成経路および TCA サイクルに多く含まれた。特に Fig. 2-5 の部分的代謝マップで表示されているように、TCA サイクルの入り口の近くに耐性と負に相関する代謝物が多く、なお、スレオニン生合成経路に耐性と正に相関した代謝物が多かった。

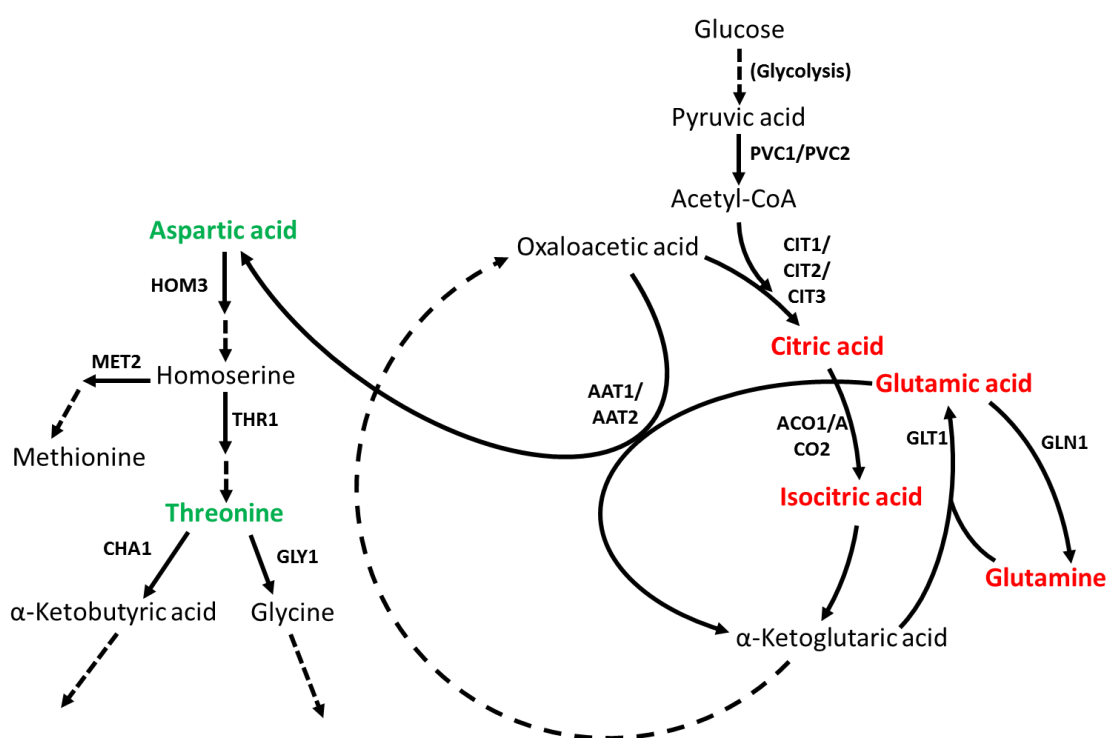


Fig. 2-5. スレオニン合成経路と TCA サイクルに含まれる重要代謝物およびその関連遺伝子

耐性と正相関の代謝物 (Threonine, Aspartic acid) は緑で示し、耐性と負相関の代謝物 (Citric acid, Glutamic acid, Glutamine) は赤で示した。

2.3.2 節でも述べられたように、遺伝子の機能欠損などは生体細胞の増殖速度を含む代謝フラックスを変えようとし、代謝物存在量の変化により様々な重要なフラックスの変動が抑えられていると考えられている[63]。ただし、1-ブタノールの暴露など環境ストレスを受けた場合、このような増殖フラックスの維持ができなくなり、増殖速度が低下してしまう可能性がある。このことは、 $\mu_{\text{nonstress}}$  の似たような株が 1-ブタノールスト

レス条件下で  $\mu_{\text{stress}}$  が異なってくる原因の一つと考えられる。また、本研究ではフラックスを維持している代謝物の変動はストレスフリー条件下でも生じているため、 $\mu_{\text{stress}}$  という細胞の潜在的な能力はストレスフリー代謝物プロファイルにも反映されていると仮定している。この仮説からスレオニンの重要代謝物としての意味を説明しようとする。[66]では、*in silico* シミュレーションで大腸菌の代謝フラックス解析を行ったところ、代謝フラックスの高い、環境ストレスなどによる摂動に抵抗し、増殖恒常性を支えるような代謝反応群が報告されている。この代謝反応群のことをハイフラックス代謝基幹 (High-flux backbone, HFB) と呼ばれ、[67], [68]では、HFB の出芽酵母における存在と、その hydroxyurea 耐性への関与が示された。更に、スレオニン生合成経路は、正にその HFB の一部であると推定されている。ゆえに、スレオニンの 1-ブタノール耐性への関与は HFB の働きと関わっている可能性が示唆される。

一方、TCA サイクルは、細胞内エネルギー代謝の中心であると同時に、様々な代謝物の前駆体を提供する代謝ハブと言える[69]。一般的に、アルコール暴露は細胞のエネルギー代謝性能を低下させる効果が知られている[48]。また、エタノール耐性を対象にした研究では、TCA サイクルの上方調節は耐性に正に寄与したと報告された[70]–[72]。TCA サイクルの高い活性は、ストレス条件下でのエネルギー代謝の維持に寄与できるのでこれらの研究報告は一致していると考えられる。しかし、本研究では TCA サイクルは 1-ブタノール耐性と負に相関している結果が得られた。1-ブタノールを含むアルコールストレスは細胞内で酸化ダメージを引き起こす効果も知られていて[48]、また、TCA サイクルにより酸化ダメージが発生し、過去研究により TCA サイクル活性の抑制は抗生物質により引き起こされる酸化ダメージの緩和に寄与しているとも報告されている[73], [74]。以上より、TCA サイクル活性の低下は 1-ブタノール耐性にも寄与している可能性が示唆された。なお、アルコールによってはエネルギー代謝の維持と酸化ダメージの低減との相対的重要性が異なり、TCA サイクルの耐性への寄与の正負はそれにより決まることが考えられる。

これまで特定された重要代謝物は、耐性とは相関関係のみ示されていた。そこで、これらの代謝物が実際に耐性と因果関係で結び付けられるかを検証するために、遺伝子改変による代謝物量の調節を試みた。一方、過去の研究より、目的表現型と相関を持っている代謝物を遺伝子改変にて調節することで、表現型の向上が可能であることが示唆されている[43]。ゆえに、Fig. 2-5 の重要代謝物と耐性との関係を新たな変異株の提案により検証することにした。スレオニン経路に関しては、アスパラギン酸からスレオニン

への流れを増大させるために *MET2* 遺伝子の破壊, および, スレオニン分解を阻止し, その蓄積を高めるための *CHA1* 破壊を提案した. また, TCA サイクルに関しては, クエン酸・イソクエン酸の蓄積量を下げることのための *CIT1-3* 遺伝子の破壊を提案した. さらに, TCA サイクルとスレオニン合成経路を結ぶ, *AAT1-2* 遺伝子によるアミノ基転移反応を阻止するための遺伝子破壊も試みた. それぞれのターゲット遺伝子については, Table 2-4 のような単一遺伝子の破壊株を取得した.

**Table 2-4. OPLS モデルの重要代謝物に基づいて推定された耐性株**

Strain	ターゲット遺伝子によりコードされる酵素	遺伝子破壊の目的
<i>cha1Δ</i>	L-serine/L-threonine deaminase	α-Ketobutyric acid への分解を阻止することによりスレオニン蓄積量を上げる
<i>met2Δ</i>	L-homoserine-O-acetyltransferase	ホモセリンからメチオニン生合成への流れ代謝物の流れを阻止することによりスレオニン合成への代謝を増大し, スレオニン量を向上させる
<i>cit1Δ</i>	Citrate synthase	クエン酸・イソクエン酸を減少させる
<i>cit2Δ</i>	Citrate synthase	クエン酸・イソクエン酸を減少させる
<i>cit3Δ</i>	Citrate synthase	クエン酸・イソクエン酸を減少させる
<i>aat1Δ</i>	Aspartate aminotransferase	オキサロ酢酸とグルタミン酸とのアミノ基転移反応を阻止することによりスレオニン合成経路と TCA サイクルに同時に影響を及ぼす
<i>aat2Δ</i>	Aspartate aminotransferase	オキサロ酢酸とグルタミン酸とのアミノ基転移反応を阻止することによりスレオニン合成経路と TCA サイクルに同時に影響を及ぼす

### 2.3.5. 新規変異株の耐性測定およびメタボローム解析

このように取得した新たな変異株について耐性測定を行った. これまで使用した変異株は, *KanMX* マーカー遺伝子の挿入によりターゲット遺伝子を破壊するという手法により構築された遺伝子破壊株だったため, *KanMX* マーカー耐性に対する影響を取り除くために, (BY4742 では) 元々機能しない *HIS3* 遺伝子[16] に *KanMX* を挿入して構築した *his3Δ* を同じ EUROSCARF コレクションから入手しコントロール株として使

用した。

まず、それぞれの株についてストレス無し条件下での増殖速度を測定した。Fig. 2-6Aで示されるように、今回得られた新株は全てコントロール株 *his3Δ* に近い  $\mu_{\text{nonstress}}$  を示した。*cit3Δ* 株については増殖速度が低くなっていると見えたが統計的判定による *his3Δ* との有意差は示されなかった。一方、Fig. 2-6B で表示されるように、提案された7株の中、*cha1Δ*、*met2Δ*、および *cit2Δ* の3株はコントロールの *his3Δ* より高い耐性を示した。また、*met2Δ* および *cit2Δ* の耐性指標値  $\mu_{\text{stress}}$  は、Student の *t* 検定により *his3Δ* と有意に差を示していることが検証された (*p*-value < 0.05, 付録の Table A6 を参照)。

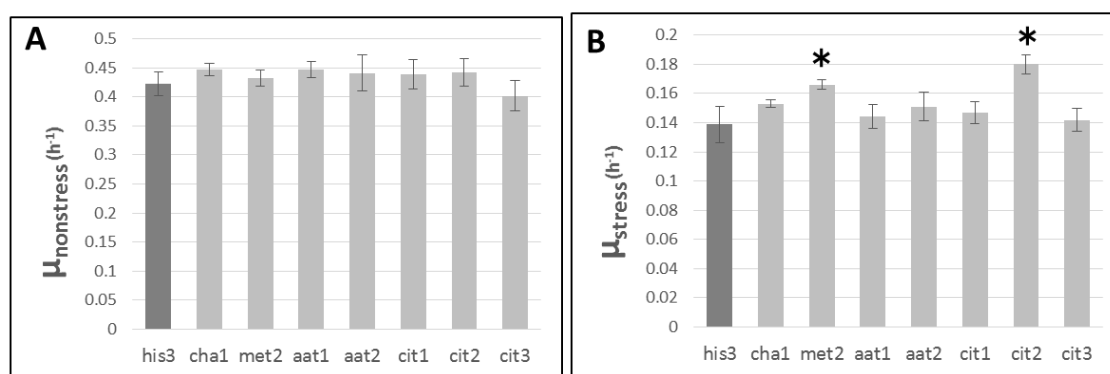


Fig. 2-6. OPLS モデルに基づいて提案された変異株の耐性測定結果

OPLS モデルに基づいて取得した新規変異株の (A) ストレス無し条件下比増殖速度 ( $\mu_{\text{nonstress}}$ ), および (B) 1.5% (v/v) 1-ブタノール濃度における比増殖速度 ( $\mu_{\text{stress}}$ ). それぞれのグラフは N=4 測定値の平均を表し, エラーバーは測定値標準偏差を表す. コントロール株 *his3Δ* に比べ有意差 (*p*-value < 0.05) を示した株には \* を付けた. 測定値は付録の Table A6 に表示される.

以上の新規変異株についてメタボローム解析を行った. 既に述べたように, *cha1Δ* および *met2Δ* はスレオニンを上げるための変異株であり, Fig. 2-7A で示されるように, この2株は実際に *his3Δ* より高い Threonine\_3TMS\_Major ピークの相対定量値を示し, その違いは統計的な有意差であった (*p*-value < 0.05). OPLS モデルではスレオニンが正の Coefficient を持った (Fig. 2-4B) ので, 本結果はモデルの予測と一致したと言える. 一方, クエン酸・イソクエン酸を減らすための種々変異株の中では *cit2Δ* のみ該当化合物ピーク (Isocitric acid+citric acid) が顕著に減少し, *his3* と比べて有意差 (*p*-value < 0.05) を示した (Fig. 2-7B). 本株は Fig. 2-6 で高い耐性を示したので,

この結果も、モデルで該当化合物が負の Coefficient を持ったことと一致した。また、本結果は、使用した培養条件では BY4742 株の *CIT* 遺伝子の中、*CIT2* のみ有意に活性を持っているということを表している。

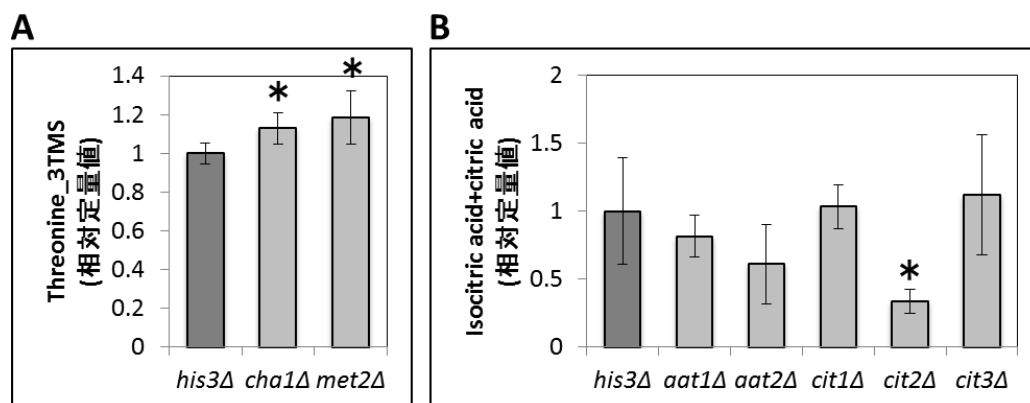


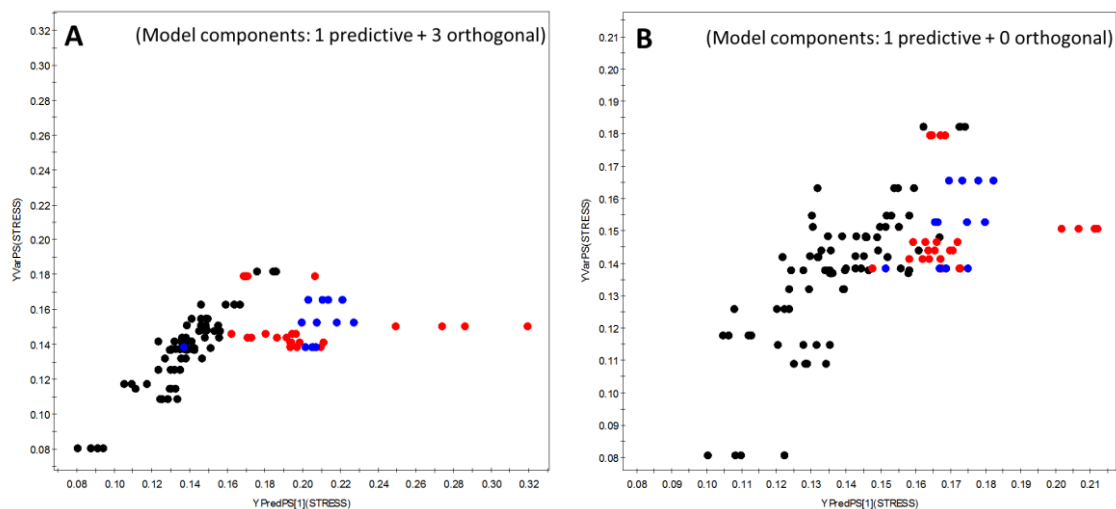
Fig. 2-7. 新規変異株の代謝物相対定量結果

(A) *cha1Δ*, *met2Δ* およびコントロール株 *his3Δ* の Threonine\_3TMS 相対定量値。(B) *cit1-3Δ*, *aat1-2Δ* およびコントロール株 *his3Δ* の Isocitric acid + Citric acid 相対定量値。各代謝物のピーク強度は同サンプルの内部標準物 (リビトール) ピーク強度で除し、相対定量値として表示した。また、本グラフでは N=4 サンプルの平均値を取り、*his3Δ* に比べての相対値に換算して表示した。エラーバーは測定値の標準偏差を表す。コントロール株 *his3Δ* に比べ有意差 ( $p$ -value < 0.05) を示した株には \* を付けた。

以上より、OPLS モデルはメタボロームデータと 1-ブタノールストレス条件下での増殖速度との間に関係づけることができると示された。なお、その関係は新たに取得した代謝酵素変異株についても成り立つことが証明できた。このことから、新規変異株の高い耐性は、モデルにより示された重要代謝物量の変化によると言える。すなわち、異なる株のメタボローム比較は、1-ブタノール耐性の向上に寄与する遺伝子改変ターゲット選択に有用であることが示された。

最後に、本小節で得られた新株を 2.3.3 節で構築された OPLS モデルにフィットし、適合度を調べた。ここで、フィットとはモデルを用いて新株の代謝物情報から耐性の予測値を計算し、その実測値と比較することを指す。しかし、この場合は Fig. 2-8A で示されるように新株は OPLS モデルと高い適合度を示さなかった。





**Fig. 2-8. 新株を OPLS モデルにフィットしてみた結果**

(A) 2.3.3 節で構築された OPLS モデルに新しい株をフィットした結果. (B) 1 予測成分のみで構築された OPLS モデルに新しい株をフィットした結果. 黒色の点は, モデル構築に寄与した転写因子変異株 19 株のデータを表し, 青色の点は, スレオニン代謝関連変異株 (*cha1Δ*及び *met2Δ*) を表し, 赤色の点は, TCA サイクル関連変異株 (*cit1-3Δ*, *aat1-2Δ*) を表している. 青色と赤色のデータはそれぞれ異なる培養・分析・データ処理バッチにより得られた. なお, どのモデルにおいても新株のデータはモデル構築に寄与していない.

前に構築された OPLS モデルはオーバーフィットしている可能性があるため, モデルの潜在変数を減らし, 1 つの説明成分のみのモデルを構築してみた. この新しいモデルの構造は前より簡単になっているため, Fig. 2-8B で示されているように, 新株の予測値は実測値に近づいた. このことから, 分析・データ処理バッチの違いによる系統誤差は, 新しく取得した代謝物データのモデルとの適合を低下されたと考えられる. 本研究では, OPLS モデルを重要代謝物の探索のために構築したので異なるバッチによる系統誤差はあまり問題にならなかったが, モデルを新株の耐性予測に用いたのであれば, 系統誤差を低減するようデータ処理を行わなければならないと示された. なお, Fig. 2-8 の一番右で赤色で示されているサンプルはモデルから極端に離れている. これらのサンプルは全て *aat2Δ* 株に属し, 単なる系統誤差ではなくそもそもモデルにより説明されていない代謝物-耐性の相関パターンを示していると考えられる. 以上の結果より, 種々の微生物変異株を含むデータセットでは全ての株が同じ相関パターンに属せず, 実は複数の相関パターンが存在し得ることが示唆された.

## 2.4. 小括

本章では、メタボロミクスを用いた株改良戦略を提案し、概念実証のために酵母の1-ブタノールストレスを対象とし、提案の戦略を実行することにより新規な耐性が取得できた。緒論の1.1.3節でも述べたように、遺伝子や転写産物に基づいた株間の比較を用いた耐性の研究は従来行われていたが、代謝物を比較対象にした研究例は少なく、代謝物量の違いから耐性向上に寄与する遺伝子を有効に推測した点で、本研究は新規性を示したと言える。また、従来の研究例では2株または少数の株の相互比較を行っていたが、本章の緒言でも述べたように、多くの株を比較することにより偽陽性を抑え、有効な違いを見出すと考えられるので、本研究では代謝物情報の定量的な性質を生かし、耐性を応答変数とした回帰モデルの構築により数多くの株の同時比較を実現させた。本研究結果により、半合理的株改良法におけるメタボロミクスの有用性が証明されたと言える。

また本研究では、ストレスフリー条件で取得したメタボロームデータをストレス条件で測定した耐性のモデル構築に用いた。本研究成果により、ストレスフリーな標準培養条件下で多くの変異株についてメタボローム解析を行い、代謝物データを収集すれば、この一度収集した代謝物データを基に様々な表現型に対して回帰モデルを構築し、株改良の遺伝子改変を提案する可能性が示された。このような代謝物データの再利用可能性は、従来法と比べて本手法の大きな長所であると考えられる。

しかし、このような代謝物データセットの構築および運用に際しては一つの問題が挙げられる。それは、大きなデータセット内には、代謝物と目的表現型との相関パターンは全てのサンプルに一致せず、複数サンプルサブセットに異なる代謝物相関パターンが現れる可能性があるということである。その場合、代謝物データの全体を一度に回帰モデルに供すると、様々な相関パターンが相殺しモデル性能が下がり、他に有意な相関を示す重要代謝物が示されない可能性がある。この問題点は、本博士論文第3章で述べる方法により解決を試みた。

# 第 3 章

## RANSAC アルゴリズムに基づく新規メタボロミクスデータマイニング法の開発

### 3.1. 緒言

前章で述べたように、増殖速度のような量的表現型について、PLS 法による回帰モデルの構築はその表現型と相関関係を持つ代謝物の探索に有用である。しかし、2 章で構築した OPLS モデル (Fig. 2-3, Fig. 2-4) を詳しく見ると、PLS 法およびその一種である OPLS 法によるモデル構築において改良すべき点が明らかになった。まず、PLS 法を含む一般的な線形回帰モデルの構築にあたり、説明変数と応答変数との間に示される相関関係が全てのサンプルにおいて同じであるという前提条件がある。しかし、全体的な相関パターンに一致しない外れ値 (アウトライヤー) や、異なる相関パターンを示すサンプルサブセットがデータセットに含まれた場合、大きな予測誤差や堅牢性の低下により優れたモデルが構築できない (Fig. 3-1)。

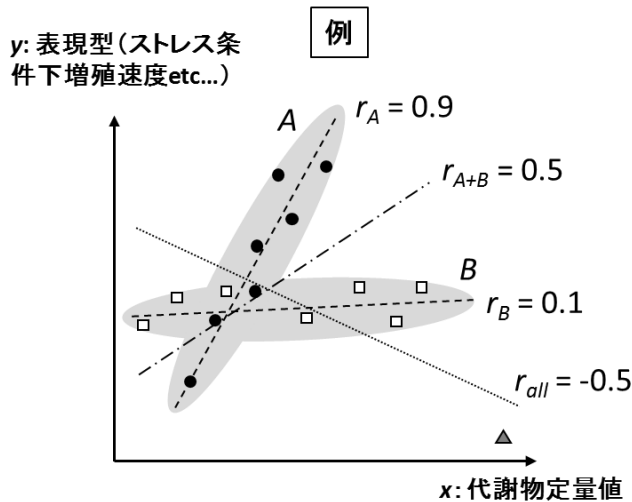


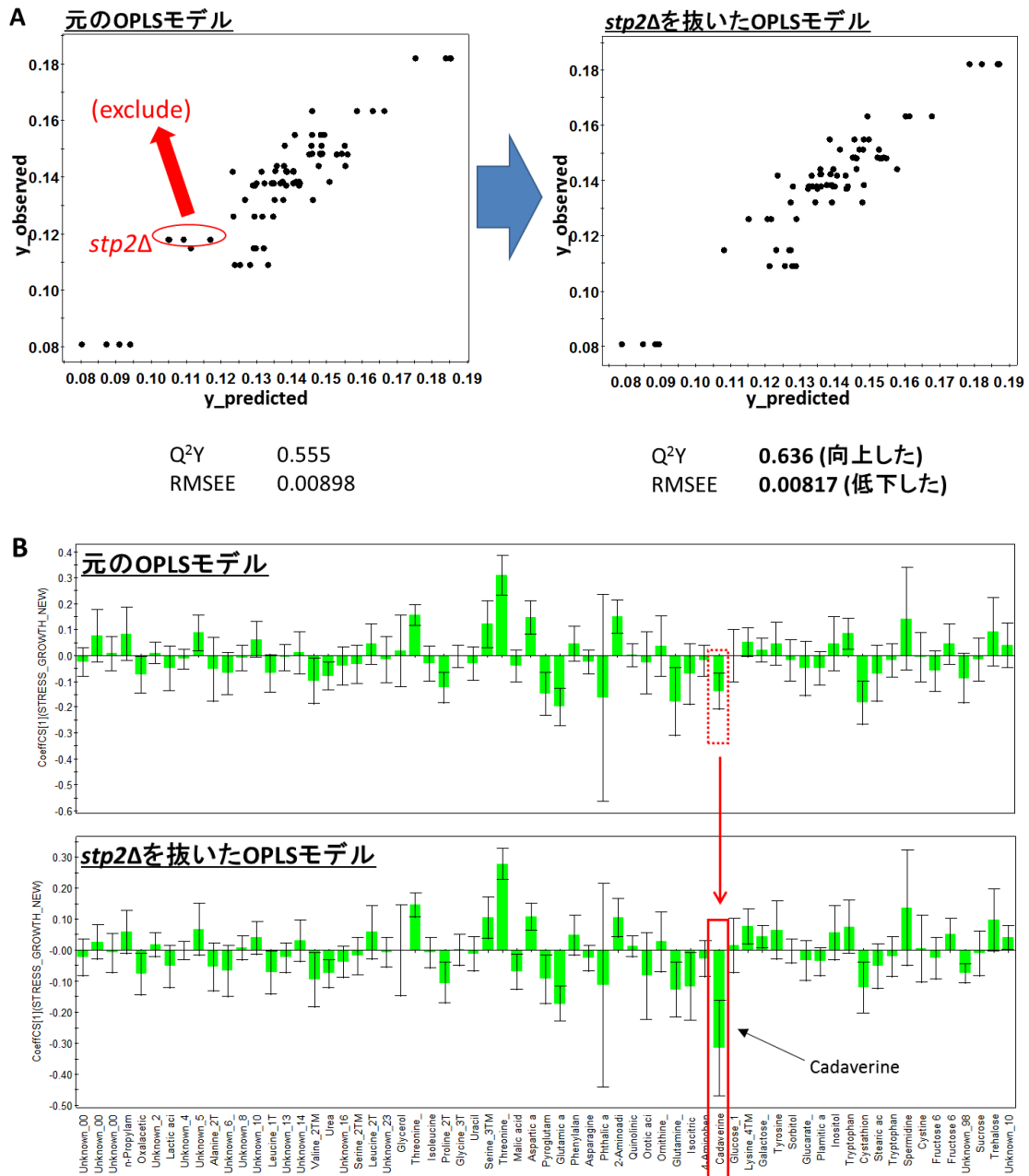
Fig. 3-1. 異なる相関パターンに従ったサンプルがデータセットに含まれた場合の回帰モデル構築の難儀

全てのサンプルが共通の相関パターンを示さない場合は、線形回帰モデルの結果が大幅に変わることがある。サブセット A (黒色の丸) では、 $x$  と  $y$  が正相関している ( $r_A = 0.9$ ) に対し、サブセット B (白色の四角形) では、 $x$  と  $y$  は高い相関を示していない ( $r_B = 0.1$ )。A と B のサン

ルを同じ回帰モデルに供した場合は、得られた相関係数が A と B のどちらも正確に表せない ( $r_{A+B}=0.5$ )。また、A、B のどちらにも属さないアウトライヤー (灰色の三角形) が混在すると、回帰モデルは更に歪まれ、ほとんどのデータには適応しないような相関関係を示す ( $r_{all} = -0.5$ )。

しかし、同じデータセットに複数の相関パターンが存在するという現象は、微生物のメタボロミクスの場合に特に可能性が高いと考えられる。例えば、異なる遺伝的背景の株、または異なる機能クラスに属する遺伝子の変異株が同じデータセットに含まれている場合は、そのサンプルのサブセットがそれぞれ異なる相関パターンを持ち、それらが相殺し合うことにより意味のあるモデル結果が得られないということが考えられる。実例として、[72]では、エタノール耐性の PLS モデルを酵母の転写因子変異株から構築した場合、代謝酵素変異株の情報はそのモデルに正確に説明できなかった結果が挙げられる。以上より、数多くの株が含まれ、その中に複数の相関パターンが存在するような大規模のデータセットを扱った場合は、良い PLS モデルが構築できず提案した半合理的戦略が使用できない可能性がある。

回帰モデル構築に供するサンプルを、全体傾向に一致したサンプルに限定することによりモデル性能が向上する実例を次のように示す。第 2 章 (2.3.3 節) では、75 サンプルの 65 代謝物定量値を説明変数として OPLS モデルを構築したが、そのデータセットから、モデル適合度の低い株 (例えば *stpΔ*) の 4 サンプルを取り除いてみた (Fig. 3-2A)。その結果、残りの 71 サンプルの代謝物情報および耐性値で改めて構築した OPLS モデルは、前より低い予測誤差 (RMSEE) および高い堅牢性 ( $Q^2Y$ ) を示した。また、元のモデルと新しいモデルの代謝物 Coefficient プロットを比較すると、Coefficient のプロファイルが変化しており、例えば前は Coefficient が小さかった Cadaverine は、負の方向に大きな Coefficient を示した (Fig. 3-2B)。このことより、データセットから全体傾向に一致しない、あるいは一致したサンプルの取捨選択は、モデル性能の向上および新たな重要代謝物の特定につながる可能性が示された。



**Fig. 3-2. 適合の良くないサンプルの除去による OPLS モデル性能向上**

(A) 第 2 章で構築した OPLS モデルおよび本章で *stp2Δ* のサンプルを取り除いて構築したモデルの Observed vs Predicted プロットの比較. 一部のサンプルを除いたことにより, 予測誤差を表す RMSEE が小さくなり, モデル堅牢性を表す Q<sup>2</sup>Y が高くなった. (B) 元のモデルと *stp2Δ* サンプル抜きのモデルの Coefficients 比較. 各グラフのエラーバーはクロスバリデーションにより計算された標準誤差を示す. *stp2Δ* サンプルの除外により Cadaverine は元より大きい Coefficient を示した.

以上を踏まえ、様々な変異株の比較を中心にした半合理的改良に際しては、データセットから、相関パターン的一致したサンプルを選択し PLS モデル構築に供する利点があると考えられる。しかし、このようなサンプルの取捨選択は解析者の主観で行うのではなく、**客観性を有する手法**で行われなければならない。

本章では、その選択手法の開発を研究課題とした。また、研究課題に対しては、画像処理の分野で開発された Random Sample Consensus (RANSAC)[75]という、サンプルサブセット選択のアルゴリズムを用いることにした。具体的に、RANSAC と PLS を組み合わせ、RANSAC の概念を基に PLS モデル構築に供するサンプルサブセットを選択する RANSAC-PLS アルゴリズムを構築することを目標とした。なお、RANSAC-PLS を構築の後、第 2 章でデータセットに適用し、それにより新たな 1-ブタノール耐性関連代謝物を探索し、その検証に新規耐性株の提案を試みた。

## 3.2. 実験方法

### 3.2.1. データセット

本研究では、第 2 章で得られたデータセットを RANSAC-PLS 実行対象に用いた。データセットには、19 の酵母転写因子変異株から得られた 75 サンプルの、65 説明変数（代謝物定量値）および 1 応答変数（1-ブタノール耐性測定値）が含まれた。代謝物定量値には、予め内部標準物（Ribitol）で標準化した相対定量値を用いた。

### 3.2.2. Random Sample Consensus-Partial Least Squares (RANSAC-PLS)アルゴリズム

計算手順は R 言語バージョン 3.2.1 (<https://www.r-project.org/>)を用いて実装した。データセットについて、まずオートスケーリング（各列のセンタリングおよび標準偏差でのスケーリング）を行った。次に Orthogonal Signal Correction (OSC) [76]を用いて、データセット全体に対して OPLS 法と同じような直交成分を構築し、応答変数と関係のない直交分散を除いた。OSC の関数は(<https://gist.github.com/dgrapov/5166570>)より得られたコードを参照に構築した。OSC を経た後のデータ残差行列は RANSAC-PLS に供した。RANSAC アルゴリズムについては[21]を参考にし、更に、各グループで構築されたサブモデルに対してイン라이어数ではなく、イン라이어数と  $Q^2Y$  値の積として得られる‘スコア’（score = inlier num/total samples \*  $Q^2Y$ ）で評価した。RANSAC-PLS 内の PLS サブモデルの構築には、R 拡張パッケージ ‘pls’ [77]より得られた *pls* 関数を用いた。

### 3.2.3. 階層クラスタ解析(Hierarchical Cluster Analysis, HCA)

HCA は R 言語の *hclust* 関数を用いて, RANSAC-PLS より得られたモデルから抽出した PLS Coefficients の行列(1001 models × 65 coefficients)について行った. ここでは 1 行が 1 モデルに対応するように, モデルを ‘サンプル’ として扱い, Coefficients の似たようなサンプル同士を集約した. なお, 距離には Pearson 相関係数を用い, 連結法には Ward のメソッドを用いた. また, HCA 結果をヒートマップで表示し, 顕著に分かれているクラスターを目視で特定した.

### 3.2.4. その他の多変量解析

RANSAC-PLS 以外の PLS モデル構築には SIMCA-P+ v12 (Umterics, Umeå, Sweden)を用いた.

### 3.2.5. 耐性測定および GC/MS によるメタボローム解析

全ての実験手順は第 2 章 (2.2 節) に記載された手順と同様に行った.

## 3.3. 結果と考察

### 3.3.1. RANSAC-PLS の構築

RANSAC の一般的な概念では, データセットから少数のサンプルを用いてサブモデルを構築し, 全サンプルからサブモデルに適合したサンプルをイン라이어として特定し, イン라이어数の評価よりサブモデルを保留または棄却する[75].

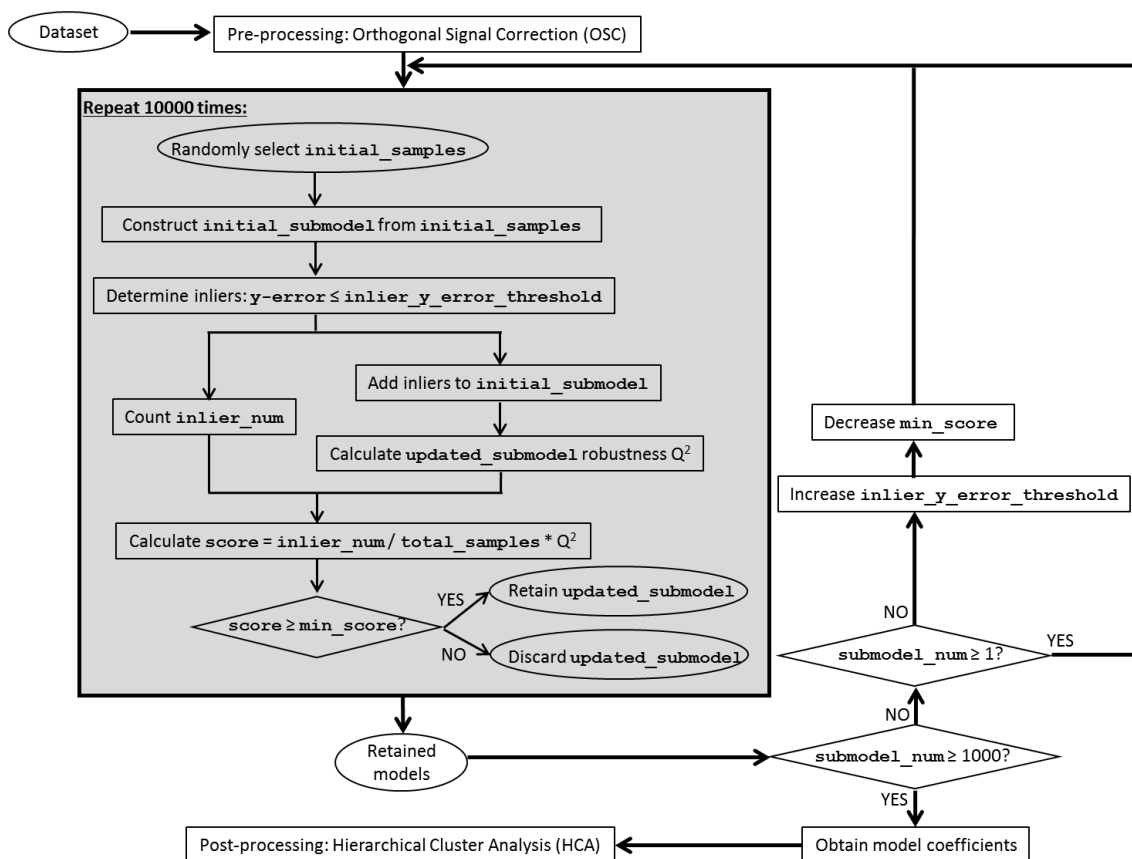


Fig. 3-3. RANSAC-PLS のアルゴリズムおよびデータマイニング法の流れ

灰色の箱は RANSAC-PLS アルゴリズムの部分を表している。まず、応答変数と相関のない直交変分は Orthogonal Signal Correction (OSC)により取り除く。次に、PLS サブモデルを構築し、全サンプルのサブモデルからの距離を計算し、距離が決められた閾値以内のサンプルをイン라이어とし、イン라이어数およびイン라이어を含んだサブモデルの  $Q^2Y$  値よりサブモデルスコアを計算し、スコアが決められた最小値より高い場合はサブモデルを保存し、最小値より低い場合は棄却するという手順を繰り返す。10000 モデルごとに、保存されたモデルの総数を評価し、0 の場合は各閾値を操縦している‘厳しさ’のパラメータ (stringency) を下げる、1000 以上の場合では、RANSAC-PLS を終了する。RANSAC-PLS より得られたモデルの Coefficients を行列にし、HCA より Coefficients プロファイルが似たモデル同士をクラスタリングした。本計算手順に用いたパラメータや設定値は付録の Table A8 に記載された。

従来の RANSAC では、サブモデルの評価にイン라이어数を用いる。ここで、イン라이어とは、サブモデルによる予測誤差が一定の閾値より低いものとされる。また、サブモデルのイン라이어数が予め決めた最小値より高い場合、サブモデルを‘有効’ (ある程度データセットと適合性を示す[75]) とし、そのサブモデルを保存する。一方、



イン라이어数が最小値より低い場合はそのサブモデルを棄却する。しかし、本研究では RANSAC アルゴリズムの中で PLS モデルを構築しているので、PLS サブモデルの堅牢性も考慮することを目的とし、単純なイン라이어数でなくモデル ‘score’ を次のように定義し、モデル評価に用いた：

$$\text{score} = \text{inlier num} / \text{total samples} * Q^2 Y_{\text{submodel with inliers included}}$$

ケモメトリクスおよびメタボロミクスデータには、目的の応答変数と相関のない傾向パターン、つまり直交変分 (Orthogonal variation) が生じることが知られている [65], [78]。例えば、実験操作における確率的誤差、サンプル保存における代謝物変化、検出器の経時的劣化など、様々な原因により一部の代謝物が同じ傾向で変動することが考えられる。このような直交変分は応答変数の予測に有用ではないが、その存在は回帰分析に影響するかもしれない。したがって、全データセットに対して、Orthogonal Signal Correction (OSC) [76] を用いて、OPLS 法と同じような直交成分をデータセット全体に対して計算し、それにより説明される分散をデータから除くことをデータの前処理として行った。本研究のデータセットについては、OSC により計算された第 1 成分、第 2 成分はそれぞれデータ分散の 17.31% と 12.69% (付録 Table A7) を説明し、有意な OSC 成分とした。また、第 3 成分により説明された分散は 5.36% であり、それ以降の成分は更に小さな割合の分散を説明したので、適用する OSC 成分の数を 2 とした。なお、OSC により直交変分が取り除かれたデータセットについては、PLS モデル構築の際、計算される PLS 成分は応答変数と最大に相関すると予想される [65]。ゆえに、RANSAC 内の PLS サブモデル構築に用いる PLS 成分の数は 1 に、簡易に設定することができる。

OSC を経た後の残差行列は RANSAC-PLS に供した。RANSAC-PLS の他のパラメータについては次のように定めた。各繰り返しのサブモデルに用いるサンプル (inlier\_samples) 数は、全サンプル数 (75) の 1/10、つまり 8 (75/10=7.5 を四捨五入した) にした。サンプルをイン라이어とするかどうかを決めるための、イン라이어予測誤差閾値 (inlier\_y\_error\_threshold)、およびサブモデルを保存するか棄却するかを決めるためのスコア最小値 (min\_score) は、‘厳しさ’ の指標値 stringency により操作されるよう次のように設定した：

$$\text{inlier } y\text{-error threshold} = (0.5 - 0.25 * \text{stringency}) * \text{s.d.}_y$$

$$\text{min. score} = 1 * \text{stringency}$$

なお、inlier y-error threshold は予測誤差の大きさを簡単に解釈できるよう、応答変数の標準偏差の倍数として表現した。

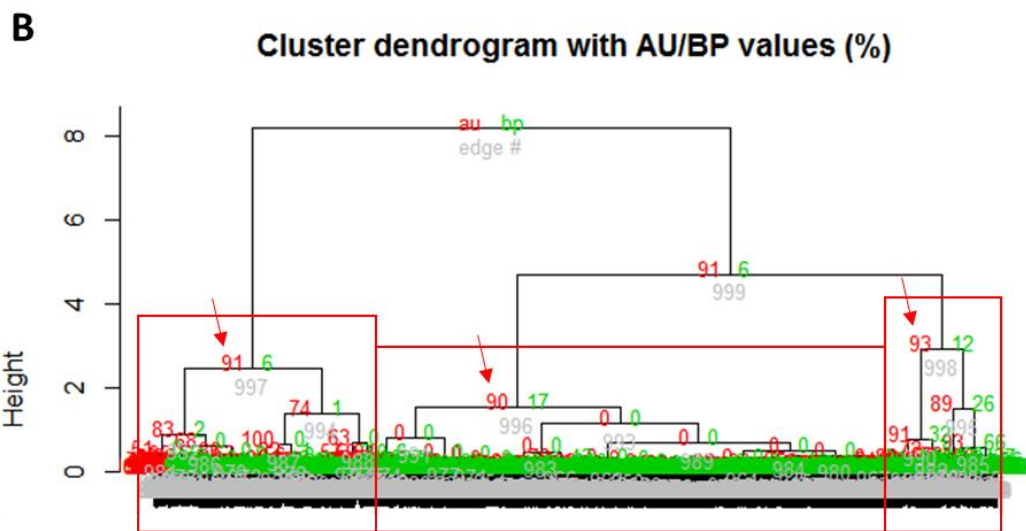
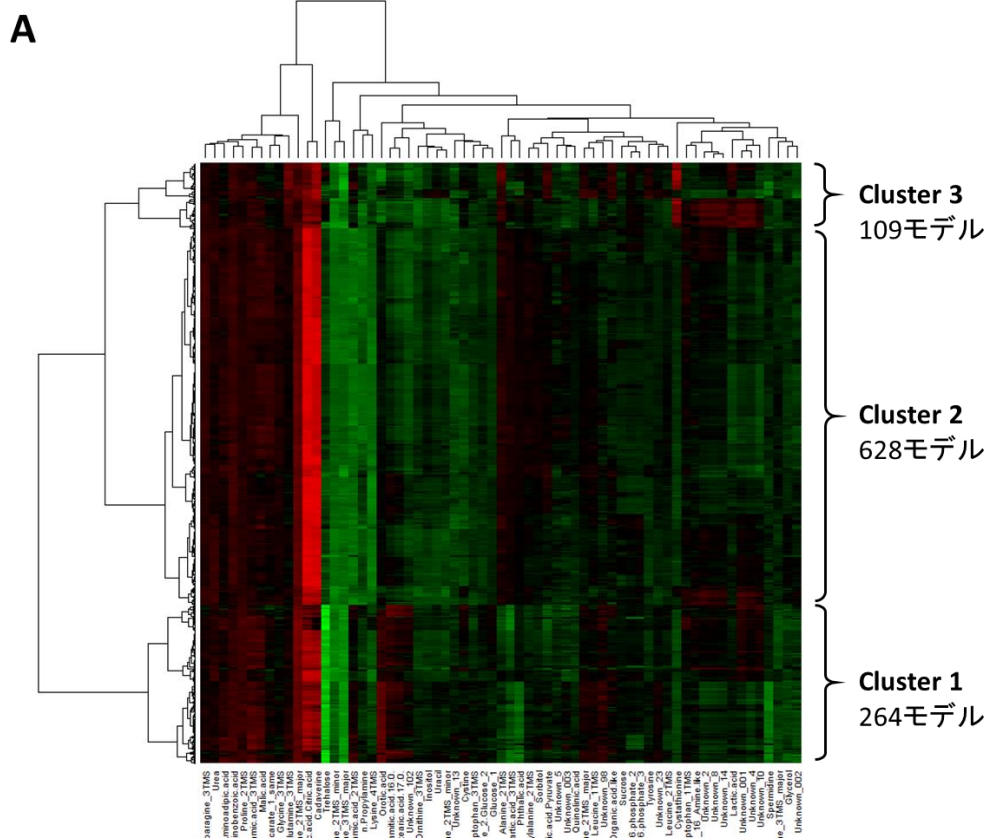
Stringency パラメータの最初値は 1 とし、10000 繰り返しおきに、サブモデルが保存されたかどうかをチェックする。また、保存されたサブモデル数が 0 の場合は、stringency の現在値に×0.9 を掛けて、少し低くなった stringency 値で次の 10000 回を実行するような仕組みを構築した。以上 2 つの式から分かるように、stringency が低くなるほど、inlier\_y\_error\_threshold は 0.25 から 0.5 に近づき、min\_score は 1 から 0 に近づく。つまり、イン라이어として認める条件およびサブモデルを保存する条件は共に緩くなり、次の 10000 回の試行ではサブモデルが保存される確率が上がる。このように、なるべく厳しくしながら、モデルが保存されるように RANSAC パラメータを設定することにより、保存されたモデルの‘有効性’を最大化できると考えられる。

最後に、取得するモデル数に関しては 10000 繰り返しごとに、それまで保存されたモデルの総数をチェックし、総モデル数が 1000 以上であれば RANSAC-PLS を終了するように設定した。

以上、説明した RANSAC-PLS の流れは Fig. 3-3 に示した。なお、用いたパラメーターなどは付録の Table A8 にまとめた。

### 3.3.2. RANSAC-PLS に基づく耐性関連代謝物の推定

第 2 章で得られたデータセットを入力データとし、上記で説明したパラメーターを用いて RANSAC-PLS を実行した。その結果、1001 モデルが得られた。また、得られたモデルの平均イン라이어数 43.57 である。次に、得られたモデルにおける重要代謝物（耐性と相関の高い代謝物）の特定を行った。第 2 章では、代謝物の VIP 値と Coefficient の両方を用いて重要代謝物を探索したが、VIP 値には相関の正負の情報が含まれないので、以下説明するような多くのモデルの同時比較に不便であると考えられ、なお、一般的に VIP の高い代謝物は大きい Coefficient を示しているので本章では Coefficient のみについて検討した。しかし、1001 ものモデルについて、各々の Coefficient プロットから重要代謝物を特定することはあまりにも煩雑すぎる。そこで、階層クラスタ解析（Hierarchical Cluster Analysis, HCA）を行い、Coefficient プロットが似ているモデル同士を集める手段を取ることにした。1001 モデル×65 代謝物 Coefficients の行列を HCA に供し、その結果を Fig. 3-4A のようなヒートマップに表示した。

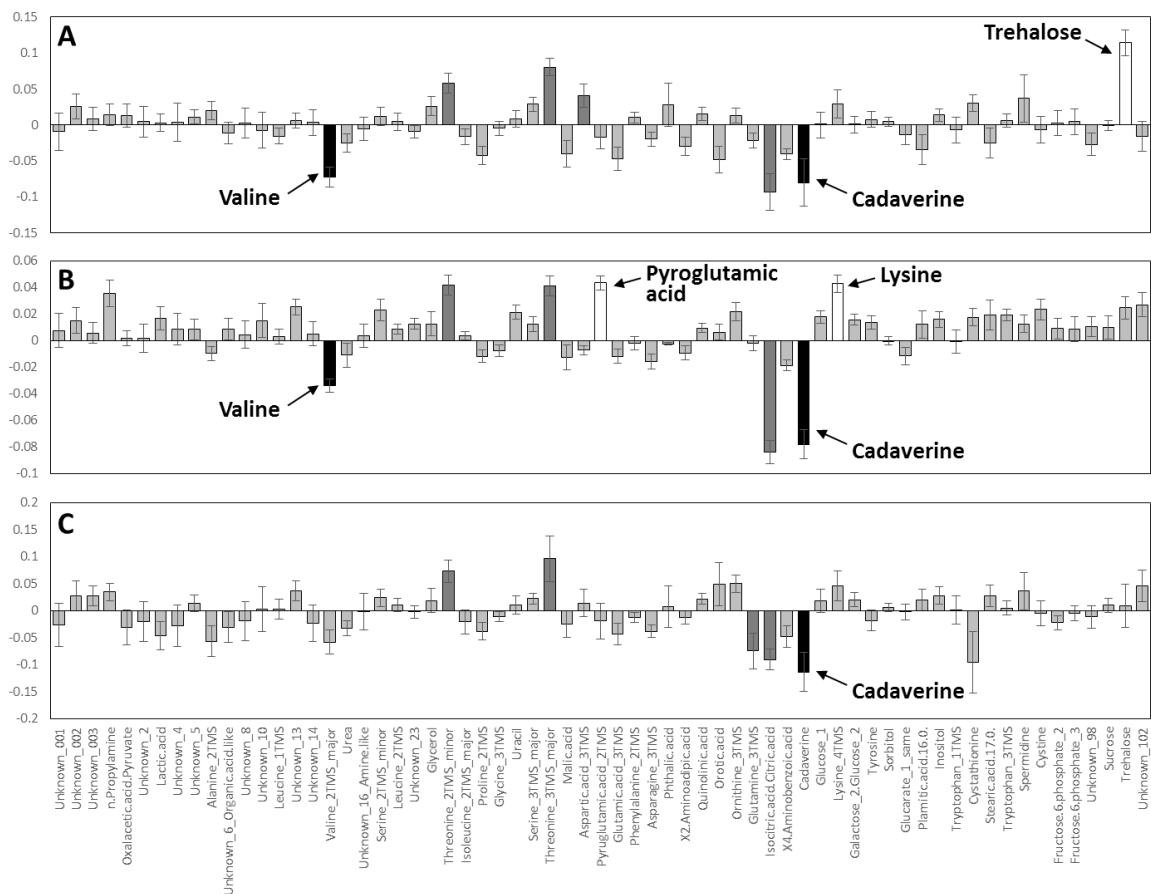


**Fig. 3-4. RANSAC-PLS モデルより得られた PLS Coefficients の HCA 結果**

(A) HCA 結果をヒートマップにより示した. 1セルは1モデルの1代謝物 Coefficient に対応し, その色は負 (赤) から正 (緑) まで Coefficient の値を表している. Coefficient がゼロに近い場合は黒で示される. また, 各行は1モデルの Coefficient に対応し, 各列は1代謝物の Coefficient に対応している. また, ヒートマップの目視により適切な高さで系統樹を切り, ク

ラスターを分割した。(B) ヒートマップに対応している系統樹および ‘pvclust’ により計算された各クラスターの存在確率 (赤の数字)。クラスター存在確率は、一部のデータだけを使用し HCA を再実行した時に同じサンプルがそのクラスターに入る確率を表している。pvclust の結果より、前で特定された 3 クラスターは (赤い矢印を付けたように) それぞれ 91 %, 90 %, 93 % の確率を示し、本データで実際に存在しているクラスターであることが示された。

ヒートマップの目視により、得られた 1001 モデルの Coefficient プロットは大きく 3 パターンに分かれていると見られる。ゆえに、Fig. 3-4A で示されたようにモデルを 3 つのクラスターに分割した。それぞれのクラスターには、264, 628, 109 モデルが含まれている。なお、それぞれのクラスターについて、‘pvclust’ という R パッケージを用いて ‘クラスター存在確率’ を計算した。クラスター存在確率は、一部のデータだけを使って HCA を再実行することを繰り返す作業 (ここで resampling または再サンプリングと呼ぶ) により、そのクラスターに同じサンプル (つまり PLS モデル) が入る確率を百分率で表しているが、100 に近いほどそのクラスターが ‘安定’ であると考えられる。また、Fig. 3-4B で示されるようにヒートマップで特定した 3 つのクラスターはそれぞれ 91 %, 90 %, 93 % の存在確率を示したので、本データで実際に存在し、PLS モデルを分けるために有用なクラスターであると考えられた。続いて、各クラスターの重要代謝物を特定するために、クラスター内のモデルで代謝物ごとに Coefficient の平均値を取り、Fig. 3-5 のような平均 Coefficients プロファイルを構築した。



**Fig. 3-5. 各 HCA クラスターの平均 PLS coefficients プロファイル**

各クラスターに含まれるモデルの PLS coefficients を平均したプロット。(A) クラスター1 の平均 Coefficients. (B) クラスター2 の平均 Coefficients. (C) クラスター3 の平均 Coefficients. エラーバーはクラスター内 Coefficient 値の標準偏差を示す. 平均 Coefficient が大きく, エラーバーが比較的に小さな代謝物は重要代謝物とし, 白 (正相関) または黒 (負相関) で表示した. なお, 第 2 章で既に検討した代謝物は濃い灰色で示した.

Fig. 3-5 の平均 Coefficients プロットより, Coefficient が大きくエラーバー (モデル内での標準偏差) の小さな重要代謝物を特定した. 前章で既に調べたスレオニン (Threonine\_2TMS, Threonine\_3TMS) およびクエン酸・イソクエン酸 (Isocitric acid+Citric acid) は, どのクラスターにおいても大きい Coefficient を示した. これらの代謝物は, 全てのサンプル (もしくはほとんどのサンプル) において耐性と相関を示しているため, 前章で RANSAC-PLS を行わなくても重要代謝物として検出できたことが本結果から明らかになった. また, それ以外にも, 各クラスターは特有の重要代謝物を示した. 例えば, Cluster 1 では Trehalose, Cluster 2 では Lysine\_4TMS と

Pyroglutamic\_acid\_2TMS が挙げられる。ここで、各 HCA クラスタに含まれているモデルは、ある独立したサブセットのサンプルからできていると考えられる。また、各クラスタに対応するサンプルサブセットでは異なる代謝物が独自に応答変数（耐性）と高い相関を示していることが考えられる。例えば、Pyroglutamic\_acid\_2TMS は、Cluster 2 では正の Coefficient を持ったが、他のクラスタでは逆に小さい負の Coefficient を持っているため、この代謝物は Cluster 2 に該当するサンプルサブセットに限り、耐性と高い正相関を示したと言える。一方、Cadaverine や Valine\_2TMS\_major のような化合物は、3 つの全クラスタで大きな Coefficient を示した。それらの化合物ピークに関しては、多くのサンプルでは耐性と相関関係を示しているが、一部のサンプルでは異なる相関パターンに従ったため、第 2 章の Fig. 2-4B では大きい Coefficient を持たなかったことが考えられる。つまり、RANSAC-PLS により、多数のサンプルと一致しない傾向を持つサンプルを除くことができたと言える。以上特定された重要代謝物は Table 3-1 にまとめた。

**Table 3-1. RANSAC-PLS より推定された 1 - ブタノール耐性関連代謝物および提案された変異株**

候補代謝物	耐性との相関	提案した変異株
トレハロース	正	<b><i>ath1Δ</i></b> , <b><i>nth1Δ</i></b> (トレハロースを増やすため)
バリン	負	<b><i>bat1Δ</i></b> , <b><i>bat2Δ</i></b> (バリンを減らすため)
カダベリン	負	<b><i>spe1Δ</i></b> (カダベリンを減らすため)
リジン	正	<b><i>spe1Δ</i></b> (リジンを増やすため)
ピログルタミン酸	正	<b><i>oxp1Δ</i></b> (ピログルタミン酸を増やすため)

耐性測定およびメタボローム解析による検証を行った株は**太字**で表した。入手できなかった株または本研究の手法で培養できなかった株は下線で示した。

### 3.3.3. RANSAC-PLS より得られた重要代謝物に基づく遺伝子改変ターゲットの提案

以上より、耐性関連と推定される代謝物の候補が得られた。続いて、候補となる重要代謝物に対し、第 2 章で行われたような遺伝子改変ターゲットを提案し耐性関連を検証した。各代謝物に関連する代謝反応およびそれに関与する遺伝子は Yeast Pathways Database (<http://pathway.yeastgenome.org>)で検索を行った。また、耐性が関連代謝物と連動して変化するという仮説の下で、Table 3-1 の各代謝物の量を耐性が向上する方

向に調節するための遺伝子改変ターゲットを推定することにした。

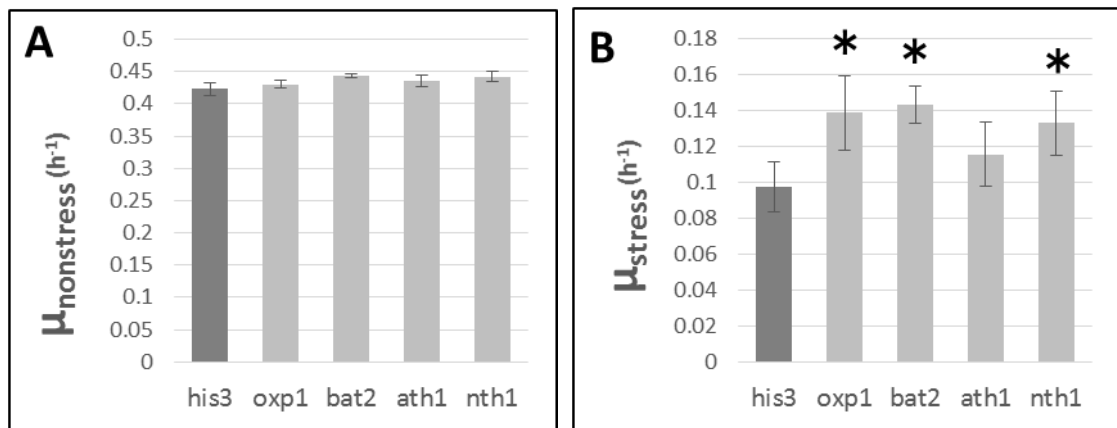
すなわちピログルタミン酸、トレハロース、リジンは増加させるべき代謝物とし、一方バリンとカダベリンは減少させるべき代謝物と考え遺伝子改変を提案した。ピログルタミン酸（正相関）は、*OXPI*によりコードされるオキソプロリナーゼ（オキソプロリンつまりピログルタミン酸の分解酵素）の欠損により増加することが予測される[79]。また、トレハロース（正相関）は *ATH1* よりコードされる酸性条件のトレハラーゼ（トレハロース分解酵素）[80]および *NTH1*によりコードされる中性条件トレハラーゼ[81]により分解される。したがって、トレハロースの蓄積を目指した遺伝子改変として、これら 2 つの遺伝子破壊を提案した。一方、バリン（負相関）は *BAT1* および *BAT2*によりコードされるアミノ基移転酵素の働きにより  $\alpha$ -ケトイソ吉草酸から合成されることが知られている[82]。また、該当反応は可逆反応であり、*BAT* 遺伝子破壊によりバリン量が増加し、それに連れて耐性が変化するかどうかを調べるために *BAT* 遺伝子の破壊も提案した。最後に、リジン（正相関）とカダベリン（負相関）については、それぞれリジン脱炭酸酵素により媒介される代謝反応の基質と産物であり、この酵素反応の欠損はリジンの蓄積、カダベリンの減少、および耐性の向上につながると予測された。しかし、出芽酵母においては特異的なリジン脱炭酸酵素遺伝子は知られていない。そこで、*SPE1*によりコードされるオルニチン脱炭酸酵素はリジン脱炭酸酵素活性も持つという研究報告[83], [84]を参考に、この遺伝子の破壊を提案した。以上、提案された遺伝子改変について、該当変異株は Table 3-1 の右側に記載した。

#### 3.3.4. 新規変異株の実験的検証

前章と同様に、*bat1Δ*を除き、Table 3-1 で提案した変異株は EUROSCARF コレクション([http://web.uni-frankfurt.de/fb15/mikro/euroscarf/stra\\_des.html](http://web.uni-frankfurt.de/fb15/mikro/euroscarf/stra_des.html))より取得した。*bat1Δ*に関しては、コレクションのリストに掲載されていなかったため、その取得を諦めた。さらに、*spe1Δ*は本研究に用いられた酵母完全液体培地で培養できないことがわかったため、検証実験から除外した。また、前章で説明したように、*his3Δ*は *KanMX* マーカーの影響を取り除くためにコントロール株として取得した。

続いて、第 2 章実験方法 (2.2.2 節) で述べた手順にて本章で得られた新株について増殖速度測定を行った。Fig. 3-6A で示されるように、新たに提案した株は全て *his3Δ*に近い比増殖速度 ( $\mu_{\text{nonstress}}$ ) を示した。一方、1-ブタノール添加ありのストレス条件下では、新株の比増殖速度 ( $\mu_{\text{stress}}$ ) は大きく異なった (Fig. 3-6B および付録 Table A9)。

測定結果より、提案した4株 (*oxp1Δ*, *bat2Δ*, *ath1Δ*, *nth1Δ*) のうち3株 (*oxp1Δ*, *bat2Δ*, *nth1Δ*) がコントロール株の *his3Δ* より有意に高い (Student の *t* 検定による  $p$ -value < 0.05) 耐性指標値  $\mu_{\text{stress}}$  を示した。



**Fig. 3-6. RANSAC-PLS に基づいて提案された変異株の増殖速度測定結果.**

RANSAC-PLS 結果に基づいて取得した新規変異株の (A) ストレス無し条件下での増殖速度 ( $\mu_{\text{nonstress}}$ ), および (B) 1.5% (v/v) 1-ブタノールストレス条件下における比増殖速度 ( $\mu_{\text{stress}}$ ). 本グラフは N=4 測定値の平均を表し, エラーバーは測定値標準偏差を表す. コントロール株 *his3Δ* より有意に高い耐性を示した株には\*マークを付けた ( $p$ -value < 0.05). 測定値は付録の Table A9 に表示される.

最後の検証段階として, 以上得られた変異株についてメタボローム解析を行い, 代謝物量が予測通りに変化したかどうかを確認した. まず, *oxp1Δ* 株とコントロールの *his3Δ* の Pyroglutamic acid 相対定量値を比べた (Fig. 3-7A). ここでは, この代謝物の測定誤差が大きく, N 数が少なかったため *oxp1Δ* と *his3Δ* との有意差判定ができなかった ( $p$ -value = 0.2055). 本実験では N=4 の測定により代謝物を定量したが, Pyroglutamic acid については有意差を示すためにより多くの実験繰り返し数が必要と考えられる. しかし, Fig. 3-7A の結果から, *oxp1Δ* のピログルタミン酸は *his3Δ* より高く蓄積していることが示唆された. また, その増加は実際に 1-ブタノール耐性に寄与したかもしれないと考えられる. 次に2つのトレハラーゼ遺伝子破壊株に関しては, *ath1Δ* と *nth1Δ* の両方とも *his3Δ* と有意差のある Trehalose 定量値を示したが, *nth1Δ* では Trehalose がコントロール株と比較して著しく高く, *ath1Δ* では Trehalose が増加したが *nth1Δ* ほど著しい上昇ではなかった (Fig. 3-7B). この傾向は, *nth1Δ* は有意な耐性の向上を



示し、*ath1Δ* では耐性が有意に向上しなかった耐性測定結果 (Fig. 3-6B) に一致している。*ATH1* は主に酸性条件下で活性を持つと報告されており [80]、本研究で用いた中性に近い完全培地では *ATH1* が元々活性を持たないため、その遺伝子破壊の効果が小さくなったと考察している。

以上より、ピログルタミン酸とトレハロースについては耐性との関係が検証できた。しかしバリンは、その量を減少させるための *bat2Δ* は高い耐性を示したことに對し、代謝物定量から Valine 定量値は実際には変化しなかったことが分かった (Fig. 3-7C)。この理由として、*BAT2* はバリンを含む分枝鎖アミノ酸の分解でなく、主に合成に働いている酵素をコードしている [85] ことから説明可能であると考えている。モデルでは Valine の定量値が耐性と負に相関していたが、ここでは *bat2Δ* によるバリン生合成経路の摂動はその代謝物の蓄積量を変化させなくても耐性を向上させた。この場合は、バリンその代謝物自体が耐性に影響を及ぼすよりも、その関連代謝経路が耐性と何等かの関係で連結していると考えられる。また、本研究ではバリンの合成経路と 1-ブタノール耐性との関係を解明しなくても、半合理的に *bat2Δ* を提案したことにより耐性が上がったという結果より、本法は表現型の向上に有用であることが更に実証された。

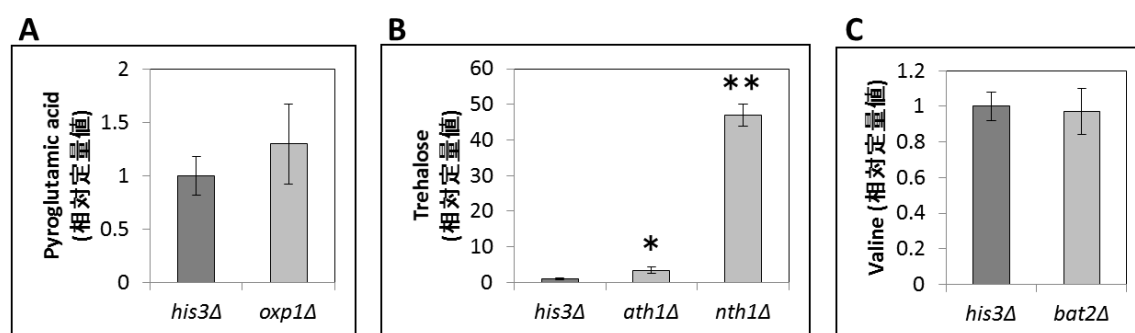


Fig. 3-7. 新規変異株の代謝物相対定量結果

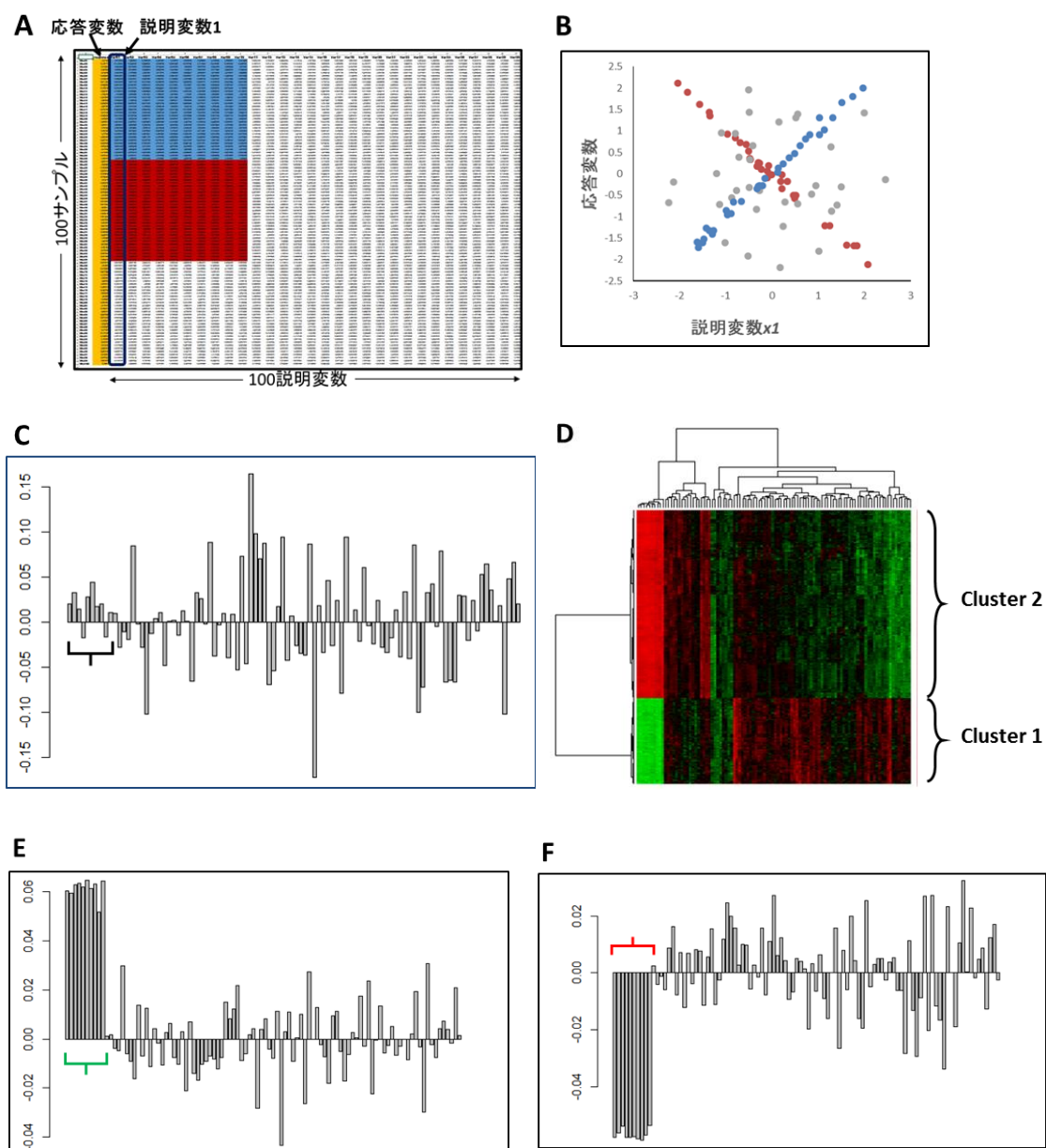
(A) *oxp1Δ* およびコントロール株 *his3Δ* の Pyroglutamic acid 相対定量値。(B) *ath1Δ*, *nth1Δ* およびコントロール株 *his3Δ* の Trehalose 相対定量値。(C) *bat2Δ* およびコントロール株 *his3Δ* の Valine 相対定量値。各代謝物のピーク強度は同サンプルの内部標準物質 (リビトール) のピーク強度で除し、相対定量値として表した。また、本グラフでは N=4 サンプルの平均値を取り、*his3Δ* に対する相対値に換算して表示した。エラーバーは測定値の標準偏差を表す。なお、コントロール株 *his3Δ* より有意に異なった ( $p$ -value < 0.05) 代謝物定量値は\*マークを付けた。

全体的に、予測された代謝物と耐性との相関関係は全ての新規耐性株で示されていないかったが、提案した4株のうち、3株は高い耐性を示したので、本手法は十分に有用であると判断した。また、本章で挙げられた耐性関連代謝物の中、トレハロースは様々なストレス耐性に関与していると報告されている[86]–[90]が、ピログルタミン酸についてはストレス耐性またはストレス反応への関与について研究報告はこれまでなかった。したがって、前章で構築したメタボロミクスに基づいた半合理的改良法および本章で開発した RANSAC-PLS を用いたデータマイニング法の組み合わせにより、注目の表現型に関与している代謝物の特定、つまりその表現型についての新たな知見を得ることができると示された。

### 3.3.5. シミュレーションデータを用いた RANSAC-PLS 手法の再確認

本研究で開発された RANSAC-PLS 解析法の実行においては、いくつかユーザーにより決めなければならないパラメータがある。それは、(1) 繰り返しを開始する際にランダムに選ぶ初期サンプル (initial samples) の数、(2) サンプルをインライヤーとして判断するための閾値 *inlier y-error threshold*、および(3) サブモデルを保存するかどうかを決めるためのスコア最小値 *min. score* である。3.3.1 節で説明したように、本研究では(2)と(3)を *stringency* (‘厳しさ’) というパラメータで支配し、*stringency* を高い値からモデルが保存される範囲まで徐々に下げ、モデルの保存を許しながらなるべく厳しい条件に調節することにより保存されるモデルの良さをある程度担保するような戦略を取った。また、(1) 初期サンプルの数については、2以上でなければならないが、Fig. 3-1 から分かるように、初期に選ぶサンプルの数が多いほど異なる相関パターンに属するサンプルが同時に選ばれてしまい、構築された一時的サブモデルがどの相関パターンにも適合していない可能性が高くなる。ゆえに、本研究では総サンプル数の 1/10、つまり  $75 \div 10 = 7.5$ 、四捨五入で 8 にしてみた。

このようなパラメータを用いたら、RANSAC-PLS が予想通りに相関パターンの異なるサンプルを見分けることができる更なる検証のために、シミュレーションデータセットに RANSAC-PLS を適用した。このシミュレーションデータセットは、100 行 (サンプル)  $\times$  101 列 (1 応答変数 + 100 説明変数) を含む行列として作成し、各セルにガウス分布に従った乱数を格納した。ただし、10 個の説明変数については乱数ではなく、33 サンプルでは応答変数と正相関し、33 サンプルでは負相関するような値を格納し、残りの 34 サンプルでは乱数のままにして ‘模擬重要代謝物’ とした (Fig. 3-8A)。



**Fig. 3-8. シミュレーションデータによる RANSAC-PLS の更なる検証**

(A) シミュレーションデータセットの概念図。作成されたデータセットには 100 行×101 列あって、各行は一つの‘サンプル’に対応し、最初の列は応答変数（‘耐性’）に対応し、各列は一つの説明変数（‘代謝物’）に対応している。なお、応答変数およびほとんどの説明変数はガウス分布に従うように値をランダムに格納した。青色と赤色の部分では、33 サンプルずつについて説明変数  $x_1 \sim x_{10}$  が応答変数とそれぞれ正相関と負相関を取るように値を決めた。この 10 個の値が特別に決められた説明変数は‘模擬重要代謝物’とした。(B) ‘模擬重要代謝物’の一つ、 $x_1$  について応答変数との関係を示す。青色のサンプルは一つ目の相関パターンに属し、つまり応答

変数と正に相関している。赤色のサンプルは二つ目の相関パターンに属し、つまり応答変数と負に相関している。また、灰色のサンプルでは  $x_1$  は応答変数と相関していない。(C) シミュレーションデータを普通の PLS モデル構築に供した時の PLS Coefficient プロット。  $x_1 \sim x_{10}$  は、異なる相関パターンのサンプルによる影響で Coefficient が小さくなり重要代謝物として特定できなかった。(D) シミュレーションデータを RANSAC-PLS に供し、得られたモデルの Coefficient を HCA で解析した結果。2通りの Coefficient パターンが明らかに見られ、モデルがパターンに従ってきれいに分離できた。(E) Cluster 1 の平均 Coefficient および (F) Cluster 2 の平均 Coefficient。それぞれのクラスタにおいて、‘模擬重要代謝物’は明らかに応答変数と正または負の相関を示した。

このシミュレーションデータセットの 100 説明変数のうち、90 個は完全にランダムな値を取っているのと同じくランダムな値により作成した応答変数ベクトルと無相関であると考えられる。一方、上述の‘模擬重要代謝物’は Fig. 3-8B で示されているように応答変数に依存しているが、その依存の仕方(相関パターン)は3通り(正、負、無)存在している。このようなデータセットを全て PLS モデルに供した時に、良いモデルは構築できず、Fig. 3-8C で示されるように PLS Coefficient では‘模擬重要代謝物’を特定することができなかった。一方、RANSAC-PLS の適用により得られたモデルの Coefficient を 3.3.2 節で述べられたように HCA に供した時、Fig. 3-8D のヒートマップで示されるようにモデルは明らかに2クラスタに分かれた。また、各クラスタについて平均 Coefficient を算出した時、‘模擬重要代謝物’が応答変数と正相関しているサンプルを含んだモデルのクラスタ平均 Coefficient プロット (Fig. 3-8E) および応答変数と負相関しているサンプルを含んだモデルのクラスタ平均 Coefficient プロット (Fig. 3-8F) に、正相関のパターンと負相関のパターンがそれぞれ明らかに表示されていた。この結果より、相反している相関パターンが存在し、従来の PLS モデルでは重要代謝物を特定できない場合でも、本研究で用いた RANSAC-PLS 手法およびそのパラメータ設定方法では、それぞれの相関パターンにおける重要代謝物と応答変数(目的表現型)との正しい相関関係を特定できると証明された。

### 3.4. 小括

PLS のような線形回帰分析はサンプルの一致した相関パターンを前提条件としている。また、メタボロミクスデータにはアウトライヤーや、複数の相関パターンの存在に

よりその条件が満たされず、良いモデルが構築できない場合がある。本章では **Random Sample Consensus (RANSAC)** と **PLS** を組み合わせた **RANSAC-PLS** アルゴリズムを新規データマイニング法に採用し、前章で得られたデータセットに適用した。適用した結果、従来法では得られなかった耐性関連代謝物候補が特定され、これらの代謝物に基づいて新たな耐性株を取得すること、および 1-ブタノール耐性機構についての新たな知見(例えばピログルタミン酸の耐性への関与)を得ることができた。本研究結果より、**RANSAC-PLS** は、メタボロミクスに基づいた半合理的株改良法におけるデータマイニングに有用であることが実証された。

## 第4章

### 総括

物質生産工程に用いられる微生物宿主の様々な表現型の中には、増殖速度、生産速度といった量的表現型の例が多く挙げられる。したがって、それらの半合理的戦略による株改良においては、メタボローム解析データの定量的性質が特に有用であると考えられる。本博士論文研究では、メタボロミクスを用いた株改良の戦略を提案し、比較メタボロミクスに基づいた実験的手法および、従来法より本戦略に適した計算手法を開発した。本戦略は、向上させたい量的表現型に対して、主として（1）代謝物からの回帰モデル構築、（2）モデルからの重要代謝物特定、および（3）重要代謝物に基づく遺伝子改変ターゲットの提案という3段階から構成される。なお、本戦略はメタボロミクスの特徴を有効に発揮しており、第2章、第3章で開発および検証された手法と合わせて、上記の量的表現型の改良に新規性のある有用な株改良法のフレームワークになると考えられる。また、本フレームワークの有用性について、本研究では便宜上、既存の変異株コレクションから適切な株を取得して検証したが、分子生物学的手法による変異株を新しく構築する場合にも十分応用可能であると考えられる。

本研究により、半合理的な株改良戦略が提案され、1-ブタノール耐性株の有効な選抜により戦略の有用性が示されたが、本戦略は従来のランダム法、合理的な手法などと組み合わせることで、より効果的に微生物生育を目指すことが可能である。例えば、緒論の1.1.2節で述べられたように、実験室進化や突然変異の誘発など、つまりランダム法により表現型が望む方向に変化している優良株を取得することができる。そこで、ランダム法により得られた変異株を比較対象に半合理的な手法を駆使し、メタボローム解析により表現型と関連した代謝物を検出すれば、更なる改良のための遺伝子改変が提案できる。そして、試した遺伝子改変ターゲットのうち、有効な改変により実際に目的表現型に関与している遺伝子が示されるので、それらの遺伝子に対し機構の解明研究を行い、機構について知見を得た上で合理的な手法により更なる株改良を行うことができると考えられる。本研究により、メタボロミクスの応用研究にける有望な研究結果が挙げられた。メタボロミクスとは比較的新しい技術分野であり、改善が期待されるところが多く残っている。例えば、スループットを上げるための培養・サンプリングの自動化、分析の選択性や感度、分析結果の再現性など、この技術分野の進歩を目指した研究開発が様々な

アプローチで盛んに行われている。またメタボロミクスを応用する際、有効な分析プラットフォームや様々なパラメーターの適切な設定法など、技術的に困難となるところがあることも事実である。それにも関わらず、メタボロミクス技術は、独立した技術としても、現在の状態で十分に利用できること、本研究結果により示すことができた。

本研究で提案された戦略は、データ駆動型の戦略である。このような性質から、本戦略を実行する際の成功率は、用いられるデータの種類、量、および質に強く依存していると考えられる。そのため、本研究結果は用いた株の種類や、分析プラットフォームにより影響されている可能性が大きい。本研究では、幅広い代謝物プロファイルおよび1-ブタノール耐性を示す株を得るために、転写因子と想定された遺伝子の変異株を用いたが、別の可能性として、代謝酵素の変異株群を用いることも考えられる。代謝酵素の場合は、遺伝子と代謝物の関連付けがより簡単にでき、モデルの重要代謝物から新規変異株を容易に提案できるという利点が考えられる。一方、代謝酵素の破壊株は、代謝物生合成機能を完全に欠損する可能性があり、そのため生じる栄養要求性は様々な株の培養手順を複雑にさせる恐れがある。最後に本研究で用いた分析プラットフォームに関して、再現性および分離性能の高いGC/MSを用いたが、GC/MSでは検出・同定できない代謝物を定量するためには、LC/MSまたはCE/MSの使用も検討できると考えられる。

本研究では、メタボロミクスの株改良における応用例を挙げたが、概念実証の対象は酵母の1-ブタノール耐性の一つのみであった。本研究により提案された戦略の幅広い適用性、つまりその一般性を更に検証するためには、他の微生物、あるいは他の表現型の向上への適用の試みが必要と考えられる。様々な微生物宿主や表現型への適用により、本手法の更なる改善および発展が期待される。

なお、本研究ではEUROSCARFという酵母遺伝子破壊株のコレクションを用いて代謝物の比較により半合理的戦略を駆使した。つまり、この半合理的戦略は、既存の変異株コレクションを利用する方法を提供している。また、本研究結果により、ストレス無し条件下の代謝物情報はストレス条件下の耐性（菌株の潜在的性能）を反映することが可能であると示された。以上の事実を踏まえて、本手法の将来への拡張として、既存の変異株コレクションをメタボローム解析に供し、ある“標準条件”（つまりストレス無し）のメタボロームデータを蓄積しておくメリットが強く感じられる。このようなデータライブラリーの構築により、本研究の手法は様々な表現型の向上に適用できると期待される。

## 謝辞

本研究を遂行し、終始格別なるご指導とご鞭撻を賜りました大阪大学大学院工学研究科 福崎英一郎教授に謹んで感謝の意を表します。

本論文の審査において、貴重なご指導とご助言を賜りました大阪大学大学院工学研究科 清水浩教授、大阪大学大学院工学研究科 紀ノ岡正博教授に心から厚く感謝申し上げます。

本研究を進めるにあたり、日頃から貴重なご助言と励ましの言葉を賜り、研究環境の整備や実験サポートの面で貴重なご尽力を頂きました大阪大学大学院工学研究科 新間秀一准教授、小野比佐好助教に深く感謝申し上げます。

本研究を行うにあたり、出芽酵母遺伝学や生理学について深いディスカッションをさせていただき、変異株の構築および提供により本研究へご協力を賜り、また貴重なご助言や励ましの言葉を賜りました長浜バイオ大学バイオサイエンス学部バイオサイエンス学科 向由起夫准教授に深く感謝申し上げます。

本研究を進めるにあたり、解析の議論や論文投稿において有用なご助言とご指導を頂きました崇城大学生物生命学部応用微生物工学科 中山泰宗准教授、大阪大学大学院工学研究科 Sastia Prama PUTRI 特任助教、大阪大学大学院工学研究科 Walter Alvarez LAVINA 博士に心から厚く感謝申し上げます。

本研究を行うに際し、相談や議論をしていただき、実験の指導や補助、事務手続きをしてくださいました大阪大学大学院工学研究科 生物資源工学領域の諸先輩、学生諸氏、スタッフの皆様に感謝の意を表します。

これまで私の学生生活を経済的に支援し、大阪大学への留学を可能にして下さいました日本文部科学省、佐藤陽国際奨学金財団に、心から感謝の意を表します。

最後に、これまで応援し続けてくれた多くの友人、そして常に支えてくれた家族に深い感謝の意を表して謝辞と致します。



## 引用文献

- [1] C. R. Fischer, D. Klein-Marcuschamer, and G. Stephanopoulos, "Selection and optimization of microbial hosts for biofuels production," *Metab. Eng.*, vol. 10, no. 6, pp. 295–304, Nov. 2008.
- [2] G. Skretas and F. N. Kolisis, "Combinatorial approaches for inverse metabolic engineering applications," *Comput. Struct. Biotechnol. J.*, vol. 3, p. e201210021, 2012.
- [3] H. Alper, J. Moxley, E. Nevoigt, G. R. Fink, and G. Stephanopoulos, "Engineering Yeast Transcription Machinery for Improved Ethanol Tolerance and Production," *Science*, vol. 314, no. 5805, pp. 1565–1568, Dec. 2006.
- [4] C. N. S. Santos and G. Stephanopoulos, "Combinatorial engineering of microbes for optimizing cellular phenotype," *Curr. Opin. Chem. Biol.*, vol. 12, no. 2, pp. 168–176, Apr. 2008.
- [5] U. Sauer, "Evolutionary engineering of industrially important microbial phenotypes," *Adv. Biochem. Eng. Biotechnol.*, vol. 73, pp. 129–169, 2001.
- [6] J. L. Adrio and A. L. Demain, "Genetic improvement of processes yielding microbial products," *FEMS Microbiol. Rev.*, vol. 30, no. 2, pp. 187–214, Mar. 2006.
- [7] T. Snoek, M. Picca Nicolino, S. Van den Bremt, S. Mertens, V. Saels, A. Verplaetse, J. Steensels, and K. J. Verstrepen, "Large-scale robot-assisted genome shuffling yields industrial *Saccharomyces cerevisiae* yeasts with increased ethanol tolerance," *Biotechnol. Biofuels*, vol. 8, p. 32, 2015.
- [8] Z. Pang, J. Liang, X. Qin, J. Wang, J. Feng, and R. Huang, "Multiple induced mutagenesis for improvement of ethanol production by *Kluyveromyces marxianus*," *Biotechnol. Lett.*, vol. 32, no. 12, pp. 1847–1851, Dec. 2010.
- [9] S. Y. Lee, D.-Y. Lee, and T. Y. Kim, "Systems biotechnology for strain improvement," *Trends Biotechnol.*, vol. 23, no. 7, pp. 349–358, Jul. 2005.
- [10] S. J. Park, S. Y. Lee, J. Cho, T. Y. Kim, J. W. Lee, J. H. Park, and M.-J. Han, "Global physiological understanding and metabolic engineering of microorganisms based on omics studies," *Appl. Microbiol. Biotechnol.*, vol. 68, no. 5, pp. 567–579, Sep. 2005.
- [11] K. Yoshikawa, C. Furusawa, T. Hirasawa, and H. Shimizu, "Design of Superior Cell Factories Based on Systems Wide Omics Analysis," in *Systems Metabolic Engineering*, C. Wittmann and S. Y. Lee, Eds. Springer Netherlands, 2012, pp. 57–81.
- [12] M.-E. Hong, K.-S. Lee, B. J. Yu, Y.-J. Sung, S. M. Park, H. M. Koo, D.-H. Kweon, J.

- C. Park, and Y.-S. Jin, "Identification of gene targets eliciting improved alcohol tolerance in *Saccharomyces cerevisiae* through inverse metabolic engineering," *J. Biotechnol.*, vol. 149, no. 1–2, pp. 52–59, Aug. 2010.
- [13] T. Hirasawa, K. Yoshikawa, Y. Nakakura, K. Nagahisa, C. Furusawa, Y. Katakura, H. Shimizu, and S. Shioya, "Identification of target genes conferring ethanol stress tolerance to *Saccharomyces cerevisiae* based on DNA microarray data analysis," *J. Biotechnol.*, vol. 131, no. 1, pp. 34–44, Aug. 2007.
- [14] S. Atsumi, T.-Y. Wu, I. M. P. Machado, W.-C. Huang, P.-Y. Chen, M. Pellegrini, and J. C. Liao, "Evolution, genomic analysis, and reconstruction of isobutanol tolerance in *Escherichia coli*," *Mol. Syst. Biol.*, vol. 6, p. 449, Dec. 2010.
- [15] J. J. Minty, A. A. Lesnfsky, F. Lin, Y. Chen, T. A. Zaroff, A. B. Veloso, B. Xie, C. A. McConnell, R. J. Ward, D. R. Schwartz, J.-M. Rouillard, Y. Gao, E. Gulari, and X. N. Lin, "Evolution combined with genomic study elucidates genetic bases of isobutanol tolerance in *Escherichia coli*," *Microb. Cell Factories*, vol. 10, p. 18, 2011.
- [16] K. Yoshikawa, T. Tanaka, C. Furusawa, K. Nagahisa, T. Hirasawa, and H. Shimizu, "Comprehensive phenotypic analysis for identification of genes affecting growth under ethanol stress in *Saccharomyces cerevisiae*," *FEMS Yeast Res.*, vol. 9, no. 1, pp. 32–44, Feb. 2009.
- [17] L. H. Reyes, M. P. Almario, and K. C. Kao, "Genomic Library Screens for Genes Involved in n-Butanol Tolerance in *Escherichia coli*," *PLoS ONE*, vol. 6, no. 3, p. e17678, Mar. 2011.
- [18] G. J. Patti, O. Yanes, and G. Siuzdak, "Innovation: Metabolomics: the apogee of the omics trilogy," *Nat. Rev. Mol. Cell Biol.*, vol. 13, no. 4, pp. 263–269, Apr. 2012.
- [19] M. J. Herrgård, N. Swainston, P. Dobson, W. B. Dunn, K. Y. Arga, M. Arvas, N. Blüthgen, S. Borger, R. Costenoble, M. Heinemann, M. Hucka, N. Le Novère, P. Li, W. Liebermeister, M. L. Mo, A. P. Oliveira, D. Petranovic, S. Pettifer, E. Simeonidis, K. Smallbone, I. Spasić, D. Weichart, R. Brent, D. S. Broomhead, H. V. Westerhoff, B. Kirdar, M. Penttilä, E. Klipp, B. Ø. Palsson, U. Sauer, S. G. Oliver, P. Mendes, J. Nielsen, and D. B. Kell, "A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology," *Nat. Biotechnol.*, vol. 26, no. 10, pp. 1155–1160, Oct. 2008.
- [20] D. A. Bender, "The promise of metabolomics," *J. Sci. Food Agric.*, vol. 85, no. 1, pp. 7–9, Jan. 2005.
- [21] W. B. Dunn, D. I. Broadhurst, H. J. Atherton, R. Goodacre, and J. L. Griffin, "Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy," *Chem. Soc. Rev.*, vol. 40, no. 1, pp.

- 387–426, Jan. 2011.
- [22] J. Smedsgaard and J. Nielsen, “Metabolite profiling of fungi and yeast: from phenotype to metabolome by MS and informatics,” *J. Exp. Bot.*, vol. 56, no. 410, pp. 273–286, Jan. 2005.
- [23] M. Ala-Korpela, A. J. Kangas, and P. Soinen, “Quantitative high-throughput metabolomics: a new era in epidemiology and genetics,” *Genome Med.*, vol. 4, no. 4, p. 36, Apr. 2012.
- [24] D. González-Ramos, M. van den Broek, A. J. van Maris, J. T. Pronk, and J.-M. G. Daran, “Genome-scale analyses of butanol tolerance in *Saccharomyces cerevisiae* reveal an essential role of protein degradation,” *Biotechnol. Biofuels*, vol. 6, no. 1, p. 48, Apr. 2013.
- [25] D. Rossouw, T. Næs, and F. F. Bauer, “Linking gene regulation and the exo-metabolome: A comparative transcriptomics approach to identify genes that impact on the production of volatile aroma compounds in yeast,” *BMC Genomics*, vol. 9, no. 1, p. 530, Nov. 2008.
- [26] “Comparative analysis of transcriptional responses to saline stress in the laboratory and brewing strains of *Saccharomyces cerevisiae* with DNA microarray - Springer.”
- [27] H. Chong, H. Geng, H. Zhang, H. Song, L. Huang, and R. Jiang, “Enhancing *E. coli* isobutanol tolerance through engineering its global transcription factor cAMP receptor protein (CRP),” *Biotechnol. Bioeng.*, vol. 111, no. 4, pp. 700–708, Apr. 2014.
- [28] B. Blagoev, S.-E. Ong, I. Kratchmarova, and M. Mann, “Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics,” *Nat. Biotechnol.*, vol. 22, no. 9, pp. 1139–1145, Sep. 2004.
- [29] K. Sivagnanam, V. G. S. Raghavan, M. Shah, R. L. Hettich, N. C. Verberkmoes, and M. G. Lefsrud, “Shotgun proteomic monitoring of *Clostridium acetobutylicum* during stationary phase of butanol fermentation using xylose and comparison with the exponential phase,” *J. Ind. Microbiol. Biotechnol.*, vol. 39, no. 6, pp. 949–955, Jun. 2012.
- [30] B. Han, V. Ujor, L. B. Lai, V. Gopalan, and T. C. Ezeji, “Use of Proteomic Analysis To Elucidate the Role of Calcium in Acetone-Butanol-Ethanol Fermentation by *Clostridium beijerinckii* NCIMB 8052,” *Appl. Environ. Microbiol.*, vol. 79, no. 1, pp. 282–293, Jan. 2013.
- [31] G. W. Birrell, J. A. Brown, H. I. Wu, G. Giaever, A. M. Chu, R. W. Davis, and J. M. Brown, “Transcriptional response of *Saccharomyces cerevisiae* to DNA-damaging agents does not identify the genes that protect against these agents,” *Proc. Natl.*

- Acad. Sci. U. S. A.*, vol. 99, no. 13, pp. 8778–8783, Jun. 2002.
- [32] S. Rochfort, “Metabolomics reviewed: a new ‘omics’ platform technology for systems biology and implications for natural products research,” *J. Nat. Prod.*, vol. 68, no. 12, pp. 1813–1820, Dec. 2005.
- [33] S. P. Putri, Y. Nakayama, F. Matsuda, T. Uchikata, S. Kobayashi, A. Matsubara, and E. Fukusaki, “Current metabolomics: Practical applications,” *J. Biosci. Bioeng.*, vol. 115, no. 6, pp. 579–589, Jun. 2013.
- [34] O. Fiehn, “Metabolomics — the link between genotypes and phenotypes,” in *Functional Genomics*, C. Town, Ed. Springer Netherlands, 2002, pp. 155–171.
- [35] E. Fukusaki and A. Kobayashi, “Plant metabolomics: potential for practical operation,” *J. Biosci. Bioeng.*, vol. 100, no. 4, pp. 347–354, Oct. 2005.
- [36] M. Braaksma, S. Bijlsma, L. Coulier, P. J. Punt, and M. J. van der Werf, “Metabolomics as a tool for target identification in strain improvement: the influence of phenotype definition,” *Microbiology*, vol. 157, no. 1, pp. 147–159, Jan. 2011.
- [37] R. H. Nugroho, K. Yoshikawa, and H. Shimizu, “Metabolomic analysis of acid stress response in *Saccharomyces cerevisiae*,” *J. Biosci. Bioeng.*, vol. 120, no. 4, pp. 396–404, Oct. 2015.
- [38] J. Wang, L. Chen, X. Tian, L. Gao, X. Niu, M. Shi, and W. Zhang, “Global Metabolomic and Network analysis of *Escherichia coli* Responses to Exogenous Biofuels,” *J. Proteome Res.*, vol. 12, no. 11, pp. 5302–5312, Nov. 2013.
- [39] Y. Wang, M. Shi, X. Niu, X. Zhang, L. Gao, L. Chen, J. Wang, and W. Zhang, “Metabolomic basis of laboratory evolution of butanol tolerance in photosynthetic *Synechocystis* sp. PCC 6803,” *Microb. Cell Factories*, vol. 13, no. 1, p. 151, Nov. 2014.
- [40] M. Beckmann, D. P. Enot, D. P. Overly, and J. Draper, “Representation, Comparison, and Interpretation of Metabolome Fingerprint Data for Total Composition Analysis and Quality Trait Investigation in Potato Cultivars,” *J. Agric. Food Chem.*, vol. 55, no. 9, pp. 3444–3451, May 2007.
- [41] C. Böttcher, L. Westphal, C. Schmotz, E. Prade, D. Scheel, and E. Glawischnig, “The Multifunctional Enzyme CYP71B15 (PHYTOALEXIN DEFICIENT3) Converts Cysteine-Indole-3-Acetonitrile to Camalexin in the Indole-3-Acetonitrile Metabolic Network of *Arabidopsis thaliana*,” *Plant Cell*, vol. 21, no. 6, pp. 1830–1845, Jun. 2009.
- [42] G. J. Patti, R. Tautenhahn, and G. Siuzdak, “Meta-analysis of untargeted metabolomic data from multiple profiling experiments,” *Nat. Protoc.*, vol. 7, no. 3, pp. 508–516, Mar. 2012.

- [43] R. Yoshida, T. Tamura, C. Takaoka, K. Harada, A. Kobayashi, Y. Mukai, and E. Fukusaki, “Metabolomics-based systematic prediction of yeast lifespan and its application for semi-rational screening of ageing-related mutants,” *Aging Cell*, vol. 9, no. 4, pp. 616–625, 2010.
- [44] S. Hayashi, S. Akiyama, Y. Tamaru, Y. Takeda, T. Fujiwara, K. Inoue, A. Kobayashi, S. Maegawa, and E. Fukusaki, “A novel application of metabolomics in vertebrate development,” *Biochem. Biophys. Res. Commun.*, vol. 386, no. 1, pp. 268–272, Aug. 2009.
- [45] A.-L. Boulesteix and K. Strimmer, “Partial least squares: a versatile tool for the analysis of high-dimensional genomic data,” *Brief. Bioinform.*, vol. 8, no. 1, pp. 32–44, Jan. 2007.
- [46] J. M. Fonville, S. E. Richards, R. H. Barton, C. L. Boulange, T. M. D. Ebbels, J. K. Nicholson, E. Holmes, and M.-E. Dumas, “The evolution of partial least squares models and related chemometric approaches in metabonomics and metabolic phenotyping,” *J. Chemom.*, vol. 24, no. 11–12, pp. 636–649, Nov. 2010.
- [47] S. Wold, M. Sjöström, and L. Eriksson, “PLS-regression: a basic tool of chemometrics,” *Chemom. Intell. Lab. Syst.*, vol. 58, no. 2, pp. 109–130, Oct. 2001.
- [48] S. A. Nicolaou, S. M. Gaida, and E. T. Papoutsakis, “A comparative view of metabolite and substrate stress and tolerance in microbial bioprocessing: From biofuels and chemicals, to biocatalysis and bioremediation,” *Metab. Eng.*, vol. 12, no. 4, pp. 307–331, Jul. 2010.
- [49] M. Dunlop, “Engineering microbes for tolerance to next-generation biofuels,” *Biotechnol. Biofuels*, vol. 4, no. 1, p. 32, Sep. 2011.
- [50] L. Zhu, H. Dong, Y. Zhang, and Y. Li, “Engineering the robustness of *Clostridium acetobutylicum* by introducing glutathione biosynthetic capability,” *Metab. Eng.*, vol. 13, no. 4, pp. 426–434, Jul. 2011.
- [51] R. Swidah, H. Wang, P. J. Reid, H. Z. Ahmed, A. M. Pisanelli, K. C. Persaud, C. M. Grant, and M. P. Ashe, “Butanol production in *S. cerevisiae* via a synthetic ABE pathway is enhanced by specific metabolic engineering and butanol resistance,” *Biotechnol. Biofuels*, vol. 8, no. 1, p. 97, Jul. 2015.
- [52] E. J. Steen, R. Chan, N. Prasad, S. Myers, C. J. Petzold, A. Redding, M. Ouellet, and J. D. Keasling, “Metabolic engineering of *Saccharomyces cerevisiae* for the production of n-butanol,” *Microb. Cell Factories*, vol. 7, no. 1, p. 36, Dec. 2008.
- [53] A. M. Zaki, T. T. Wimalasena, and D. Greetham, “Phenotypic characterisation of *Saccharomyces* spp. for tolerance to 1-butanol,” *J. Ind. Microbiol. Biotechnol.*, vol. 41, no. 11, pp. 1627–1636, Nov. 2014.

- [54] P. Ghiaci, J. Norbeck, and C. Larsson, "Physiological adaptations of *Saccharomyces cerevisiae* evolved for improved butanol tolerance," *Biotechnol. Biofuels*, vol. 6, no. 1, p. 1, Jul. 2013.
- [55] G. Giaever, A. M. Chu, L. Ni, C. Connelly, L. Riles, S. Véronneau, S. Dow, A. Lucau-Danila, K. Anderson, B. André, A. P. Arkin, A. Astromoff, M. E. Bakkoury, R. Bangham, R. Benito, S. Brachat, S. Campanaro, M. Curtiss, K. Davis, A. Deutschbauer, K.-D. Entian, P. Flaherty, F. Foury, D. J. Garfinkel, M. Gerstein, D. Gotte, U. Güldener, J. H. Hegemann, S. Hempel, Z. Herman, D. F. Jaramillo, D. E. Kelly, S. L. Kelly, P. Kötter, D. LaBonte, D. C. Lamb, N. Lan, H. Liang, H. Liao, L. Liu, C. Luo, M. Lussier, R. Mao, P. Menard, S. L. Ooi, J. L. Revuelta, C. J. Roberts, M. Rose, P. Ross-Macdonald, B. Scherens, G. Schimmack, B. Shafer, D. D. Shoemaker, S. Sookhai-Mahadeo, R. K. Storms, J. N. Strathern, G. Valle, M. Voet, G. Volckaert, C. Wang, T. R. Ward, J. Wilhelmy, E. A. Winzeler, Y. Yang, G. Yen, E. Youngman, K. Yu, H. Bussey, J. D. Boeke, M. Snyder, P. Philippsen, R. W. Davis, and M. Johnston, "Functional profiling of the *Saccharomyces cerevisiae* genome," *Nature*, vol. 418, no. 6896, pp. 387–391, Jul. 2002.
- [56] Z. Hashim, S. T. Teoh, T. Bamba, and E. Fukusaki, "Construction of a metabolome library for transcription factor-related single gene mutants of *Saccharomyces cerevisiae*," *J. Chromatogr. B*, vol. 966, pp. 83–92, Sep. 2014.
- [57] A. Lommen, "MetAlign: Interface-Driven, Versatile Metabolomics Tool for Hyphenated Full-Scan Mass Spectrometry Data Preprocessing," *Anal. Chem.*, vol. 81, no. 8, pp. 3079–3086, Apr. 2009.
- [58] Y. M. Tikunov, S. Laptinok, R. D. Hall, A. Bovy, and R. C. H. de Vos, "MSClust: a tool for unsupervised mass spectra extraction of chromatography-mass spectrometry ion-wise aligned data," *Metabolomics*, vol. 8, no. 4, pp. 714–718, Aug. 2012.
- [59] H. Tsugawa, Y. Tsujimoto, M. Arita, T. Bamba, and E. Fukusaki, "GC/MS based metabolomics: development of a data mining system for metabolite identification by using soft independent modeling of class analogy (SIMCA)," *BMC Bioinformatics*, vol. 12, no. 1, p. 131, May 2011.
- [60] S. S. Dwight, M. A. Harris, K. Dolinski, C. A. Ball, G. Binkley, K. R. Christie, D. G. Fisk, L. Issel-Tarver, M. Schroeder, G. Sherlock, A. Sethuraman, S. Weng, D. Botstein, and J. M. Cherry, "Saccharomyces Genome Database (SGD) provides secondary gene annotation using the Gene Ontology (GO)," *Nucleic Acids Res.*, vol. 30, no. 1, pp. 69–72, Jan. 2002.
- [61] H. Goodarzi, B. D. Bennett, S. Amini, M. L. Reaves, A. K. Hottes, J. D. Rabinowitz,

- and S. Tavazoie, “Regulatory and metabolic rewiring during laboratory evolution of ethanol tolerance in *E. coli*,” *Mol. Syst. Biol.*, vol. 6, no. 1, Jun. 2010.
- [62] J. Schaub and M. Reuss, “In vivo dynamics of glycolysis in *Escherichia coli* shows need for growth-rate dependent metabolome analysis,” *Biotechnol. Prog.*, vol. 24, no. 6, pp. 1402–1407, Nov. 2008.
- [63] L. M. Raamsdonk, B. Teusink, D. Broadhurst, N. Zhang, A. Hayes, M. C. Walsh, J. A. Berden, K. M. Brindle, D. B. Kell, J. J. Rowland, H. V. Westerhoff, K. van Dam, and S. G. Oliver, “A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations,” *Nat. Biotechnol.*, vol. 19, no. 1, pp. 45–50, 2001.
- [64] W. Pongsuwan, E. Fukusaki, T. Bamba, T. Yonetani, T. Yamahara, and A. Kobayashi, “Prediction of Japanese Green Tea Ranking by Gas Chromatography/Mass Spectrometry-Based Hydrophilic Metabolite Fingerprinting,” *J. Agric. Food Chem.*, vol. 55, no. 2, pp. 231–236, Jan. 2007.
- [65] J. Trygg and S. Wold, “Orthogonal projections to latent structures (O-PLS),” *J. Chemom.*, vol. 16, no. 3, pp. 119–128, Mar. 2002.
- [66] E. Almaas, B. Kovács, T. Vicsek, Z. N. Oltvai, and A.-L. Barabási, “Global organization of metabolic fluxes in the bacterium *Escherichia coli*,” *Nature*, vol. 427, no. 6977, pp. 839–843, Feb. 2004.
- [67] J. Hartman and N. Tippery, “Systematic quantification of gene interactions by phenotypic array analysis,” *Genome Biol.*, vol. 5, no. 7, p. R49, Jun. 2004.
- [68] J. L. Hartman, “Buffering of deoxyribonucleotide pool homeostasis by threonine metabolism,” *Proc. Natl. Acad. Sci.*, vol. 104, no. 28, pp. 11700–11705, Jul. 2007.
- [69] J. M. Berg, J. L. Tymoczko, and L. Stryer, “The Citric Acid Cycle,” 2002.
- [70] D. Stanley, A. Bandara, S. Fraser, P. j. Chambers, and G. a. Stanley, “The ethanol stress response and ethanol tolerance of *Saccharomyces cerevisiae*,” *J. Appl. Microbiol.*, vol. 109, no. 1, pp. 13–24, Jul. 2010.
- [71] S. Liu and N. Qureshi, “How microbes tolerate ethanol and butanol,” *New Biotechnol.*, vol. 26, no. 3–4, pp. 117–121, Oct. 2009.
- [72] E. Ohta, Y. Nakayama, Y. Mukai, T. Bamba, and E. Fukusaki, “Metabolomic approach for improving ethanol stress tolerance in *Saccharomyces cerevisiae*,” *J. Biosci. Bioeng.*
- [73] V. C. Thomas, L. C. Kinkead, A. Janssen, C. R. Schaeffer, K. M. Woods, J. K. Lindgren, J. M. Peaster, S. S. Chaudhari, M. Sadykov, J. Jones, S. M. M. AbdelGhani, M. C. Zimmerman, K. W. Bayles, G. A. Somerville, and P. D. Fey, “A Dysfunctional Tricarboxylic Acid Cycle Enhances Fitness of *Staphylococcus*

- epidermidis During  $\beta$ -Lactam Stress,” *mBio*, vol. 4, no. 4, pp. e00437–13, Aug. 2013.
- [74] M. A. Kohanski, D. J. Dwyer, B. Hayete, C. A. Lawrence, and J. J. Collins, “A Common Mechanism of Cellular Death Induced by Bactericidal Antibiotics,” *Cell*, vol. 130, no. 5, pp. 797–810, Sep. 2007.
- [75] M. A. Fischler and R. C. Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography,” *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [76] M. Blanco, J. Coello, I. Montoliu, and M. A. Romero, “Orthogonal signal correction in near infrared calibration,” *Anal. Chim. Acta*, vol. 434, no. 1, pp. 125–132, Apr. 2001.
- [77] B.-H. Mevik and R. Wehrens, “The pls package: principal component and partial least squares regression in R,” *J. Stat. Softw.*, vol. 18, no. 2, pp. 1–24, 2007.
- [78] S. Wiklund, E. Johansson, L. Sjöström, E. J. Mellerowicz, U. Edlund, J. P. Shockcor, J. Gottfries, T. Moritz, and J. Trygg, “Visualization of GC/TOF-MS-Based Metabolomics Data for Identification of Biochemically Interesting Compounds Using OPLS Class Models,” *Anal. Chem.*, vol. 80, no. 1, pp. 115–122, Jan. 2008.
- [79] W. Lu, M. F. Clasquin, E. Melamud, D. Amador-Noguez, A. A. Caudy, and J. D. Rabinowitz, “Metabolomic Analysis via Reversed-Phase Ion-Pairing Liquid Chromatography Coupled to a Stand Alone Orbitrap Mass Spectrometer,” *Anal. Chem.*, vol. 82, no. 8, pp. 3212–3221, Apr. 2010.
- [80] P. Alizadeh and D. J. Klionsky, “Purification and biochemical characterization of the ATH1 gene product, vacuolar acid trehalase, from *Saccharomyces cerevisiae*,” *FEBS Lett.*, vol. 391, no. 3, pp. 273–278, Aug. 1996.
- [81] M. Kopp, H. Müller, and H. Holzer, “Molecular analysis of the neutral trehalase gene from *Saccharomyces cerevisiae*,” *J. Biol. Chem.*, vol. 268, no. 7, pp. 4766–4774, Mar. 1993.
- [82] G. Kispal, H. Steiner, D. A. Court, B. Rolinski, and R. Lill, “Mitochondrial and Cytosolic Branched-chain Amino Acid Transaminases from Yeast, Homologs of the myc Oncogene-regulated Eca39 Protein,” *J. Biol. Chem.*, vol. 271, no. 40, pp. 24458–24464, Oct. 1996.
- [83] D. Walters and T. Cowley, “Polyamine metabolism in *Saccharomyces cerevisiae* exposed to ethanol,” *Microbiol. Res.*, vol. 153, no. 2, pp. 179–184, Aug. 1998.
- [84] D. R. Walters and T. Cowley, “Formation of cadaverine derivatives in *Saccharomyces cerevisiae*,” *FEMS Microbiol. Lett.*, vol. 145, no. 2, pp. 255–259, Dec. 1996.
- [85] M. Colón, F. Hernández, K. López, H. Quezada, J. González, G. López, C. Aranda,



- and A. González, “Saccharomyces cerevisiae Bat1 and Bat2 Aminotransferases Have Functionally Diverged from the Ancestral-Like Kluyveromyces lactis Orthologous Enzyme,” *PLoS ONE*, vol. 6, no. 1, p. e16099, Jan. 2011.
- [86] M. Iordachescu and R. Imai, “Trehalose Biosynthesis in Response to Abiotic Stresses,” *J. Integr. Plant Biol.*, vol. 50, no. 10, pp. 1223–1229, Oct. 2008.
- [87] K. Jia, Y. Zhang, and Y. Li, “Systematic engineering of microorganisms to improve alcohol tolerance,” *Eng. Life Sci.*, vol. 10, no. 5, pp. 422–429, 2010.
- [88] L. H. Reyes, A. S. Abdelaal, and K. C. Kao, “Genetic Determinants for n-Butanol Tolerance in Evolved Escherichia coli Mutants: Cross Adaptation and Antagonistic Pleiotropy between n-Butanol and Other Stressors,” *Appl. Environ. Microbiol.*, vol. 79, no. 17, pp. 5313–5320, Sep. 2013.
- [89] H. Tapia, L. Young, D. Fox, C. R. Bertozzi, and D. Koshland, “Increasing intracellular trehalose is sufficient to confer desiccation tolerance to Saccharomyces cerevisiae,” *Proc. Natl. Acad. Sci.*, vol. 112, no. 19, pp. 6122–6127, May 2015.
- [90] Y. Yoshiyama, K. Tanaka, K. Yoshiyama, M. Hibi, J. Ogawa, and J. Shima, “Trehalose accumulation enhances tolerance of Saccharomyces cerevisiae to acetic acid,” *J. Biosci. Bioeng.*, vol. 119, no. 2, pp. 172–175, Feb. 2015.

## 付録

Table A1. 第2章に用いた転写因子単一遺伝子破壊株

Strain	I	II	Description in <i>Saccharomyces Genome Database</i>
<i>ace1/cup2</i>	1		Copper-binding transcription factor; activates transcription of the metallothionein genes CUP1-1 and CUP1-2 in response to elevated copper concentrations
<i>ace2</i>			Transcription factor required for septum destruction after cytokinesis; phosphorylation by Cbk1p blocks nuclear exit of Ace2p during the M-to-G1 transition, causing its specific localization to daughter cell nuclei, and also increases Ace2p activity; phosphorylation by Cdc28p and Pho85p prevents nuclear import during cell cycle phases other than cytokinesis; part of the RAM network that regulates cellular polarity and morphogenesis
<i>ada2</i>			Transcription coactivator, component of the ADA and SAGA transcriptional adaptor/HAT (histone acetyltransferase) complexes
<i>adr1</i>			Carbon source-responsive zinc-finger transcription factor, required for transcription of the glucose-repressed gene ADH2, of peroxisomal protein genes, and of genes required for ethanol, glycerol, and fatty acid utilization
<i>aft1</i>	1		Transcription factor involved in iron utilization and homeostasis; binds the consensus site PyPuCACCCPu and activates the expression of target genes in response to changes in iron availability; in iron-replete conditions activity is negatively regulated by Grx3p, Grx4p, and Fra2p, which regulate Aft1p translocation from the nucleus to the cytoplasm
<i>aft2</i>	1		Iron-regulated transcriptional activator; activates genes involved in intracellular iron use and required for iron homeostasis and resistance to oxidative stress; similar to Aft1p
<i>arg80</i>			Transcription factor involved in regulation of arginine-responsive genes; acts with Arg81p and Arg82p

---

<i>arg81/argr2</i>		Zinc-finger transcription factor of the Zn(2)-Cys(6) binuclear cluster domain type, involved in the regulation of arginine-responsive genes; acts with Arg80p and Arg82p
<i>arg82</i>		Inositol polyphosphate multikinase (IPMK), sequentially phosphorylates Ins(1,4,5)P3 to form Ins(1,3,4,5,6)P5; also has diphosphoinositol polyphosphate synthase activity; regulates arginine-, phosphate-, and nitrogen-responsive genes
<i>aro80</i>	1	Zinc finger transcriptional activator of the Zn2Cys6 family; activates transcription of aromatic amino acid catabolic genes in the presence of aromatic amino acids
<i>arr1</i>	1	Transcriptional activator of the basic leucine zipper (bZIP) family, required for transcription of genes involved in resistance to arsenic compounds
<i>arr1</i>	1	Transcriptional activator of the basic leucine zipper (bZIP) family, required for transcription of genes involved in resistance to arsenic compounds
<i>azf1</i>	1	Zinc-finger transcription factor, involved in induction of CLN3 transcription in response to glucose; genetic and physical interactions indicate a possible role in mitochondrial transcription or genome maintenance
<i>bas1</i>	1	Myb-related transcription factor involved in regulating basal and induced expression of genes of the purine and histidine biosynthesis pathways; also involved in regulation of meiotic recombination at specific genes
<i>bas2/grf10/pho2</i>	1	Homeobox transcription factor; regulatory targets include genes involved in phosphate metabolism; binds cooperatively with Pho4p to the PHO5 promoter; phosphorylation of Pho2p facilitates interaction with Pho4p
<i>cad1/yap2</i>	1 1	AP-1-like basic leucine zipper (bZIP) transcriptional activator involved in stress responses, iron metabolism, and pleiotropic drug

---

---

		resistance; controls a set of genes involved in stabilizing proteins; binds consensus sequence TTACTAA
<i>cat8</i>		Zinc cluster transcriptional activator necessary for derepression of a variety of genes under non-fermentative growth conditions, active after diauxic shift, binds carbon source responsive elements
<i>cbf1</i>		Dual function helix-loop-helix protein; binds the motif CACRTG present at several sites including MET gene promoters and centromere DNA element I (CDEI); affects nucleosome positioning at this motif; associates with other transcription factors such as Met4p and Isw1p to mediate transcriptional activation or repression; associates with kinetochore proteins and required for efficient chromosome segregation
<i>cha4</i>	1	DNA binding transcriptional activator, mediates serine/threonine activation of the catabolic L-serine (L-threonine) deaminase (CHA1); Zinc-finger protein with Zn[2]-Cys[6] fungal-type binuclear cluster domain
<i>cin5</i>	1	Basic leucine zipper (bZIP) transcription factor of the yAP-1 family; physically interacts with the Tup1-Cyc8 complex and recruits Tup1p to its targets; mediates pleiotropic drug resistance and salt tolerance; nuclearly localized under oxidative stress and sequestered in the cytoplasm by Lot6p under reducing conditions
<i>crz1</i>	1	Transcription factor that activates transcription of genes involved in stress response; nuclear localization is positively regulated by calcineurin-mediated dephosphorylation
<i>cst6</i>	1	Basic leucine zipper (bZIP) transcription factor, in ATF/CREB family; mediates transcriptional activation of NCE103 (encoding carbonic anhydrase) in response to low CO <sub>2</sub> levels such as in the ambient air; proposed to be a regulator of oleate responsive genes; involved in utilization of non-optimal carbon sources and chromosome stability
<i>cup9</i>	1	Homeodomain-containing transcriptional repressor of PTR2, which encodes a major peptide transporter; imported peptides activate

---

---

		ubiquitin-dependent proteolysis, resulting in degradation of Cup9p and de-repression of PTR2 transcription
<i>dal80</i>		Negative regulator of genes in multiple nitrogen degradation pathways; expression is regulated by nitrogen levels and by Gln3p; member of the GATA-binding family, forms homodimers and heterodimers with Deh1p
<i>dal81</i>	1	Positive regulator of genes in multiple nitrogen degradation pathways; contains DNA binding domain but does not appear to bind the dodecanucleotide sequence present in the promoter region of many genes involved in allantoin catabolism
<i>dal82</i>		Positive regulator of allophanate inducible genes; binds a dodecanucleotide sequence upstream of all genes that are induced by allophanate; contains an UISALL DNA-binding, a transcriptional activation, and a coiled-coil domain
<i>ecm22</i>	1	Sterol regulatory element binding protein, regulates transcription of sterol biosynthetic genes; contains Zn[2]-Cys[6] binuclear cluster; homologous to Upc2p; relocates from intracellular membranes to perinuclear foci on sterol depletion
<i>fzf1</i>	1	Transcription factor involved in sulfite metabolism, sole identified regulatory target is SSU1, overexpression suppresses sulfite-sensitivity of many unrelated mutants due to hyperactivation of SSU1, contains five zinc fingers
<i>gal3</i>	1	Transcriptional regulator involved in activation of the GAL genes in response to galactose; forms a complex with Gal80p to relieve Gal80p inhibition of Gal4p; binds galactose and ATP but does not have galactokinase activity
<i>gal4</i>	1	DNA-binding transcription factor required for the activation of the GAL genes in response to galactose; repressed by Gal80p and activated by Gal3p

---

---

<i>gal80</i>			Transcriptional regulator involved in the repression of GAL genes in the absence of galactose; inhibits transcriptional activation by Gal4p; inhibition relieved by Gal3p or Gal1p binding
<i>gat1</i>	1		Transcriptional activator of genes involved in nitrogen catabolite repression; contains a GATA-1-type zinc finger DNA-binding motif; activity and localization regulated by nitrogen limitation and Ure2p
<i>gat2</i>			Protein containing GATA family zinc finger motifs; similar to Gln3p and Dal80p; expression repressed by leucine
<i>gcn4</i>	1	1	Basic leucine zipper (bZIP) transcriptional activator of amino acid biosynthetic genes in response to amino acid starvation; expression is tightly regulated at both the transcriptional and translational levels
<i>gcr2</i>			Transcriptional activator of genes involved in glycolysis; interacts and functions with the DNA-binding protein Gcr1p
<i>gis1</i>			JmjC domain-containing histone demethylase and transcription factor; involved in expression of genes during nutrient limitation; negatively regulates DPP1 and PHR1; activity is modulated by limited proteasome-mediated proteolysis; has a JmjC and a JmjN domain in the N-terminal region that interact, promoting Gis1p stability and proper transcriptional activity; contains transactivating domains TAD1 and TAD2 downstream of the Jmj domains and a C-terminal DNA binding domain
<i>gln3</i>	1		Transcriptional activator of genes regulated by nitrogen catabolite repression (NCR), localization and activity regulated by quality of nitrogen source
<i>gzf3</i>	1		GATA zinc finger protein and Dal80p homolog that negatively regulates nitrogen catabolic gene expression by competing with Gat1p for GATA site binding; function requires a repressive carbon source; dimerizes with Dal80p and binds to Tor1p
<i>hac1</i>	1	1	Basic leucine zipper (bZIP) transcription factor (ATF/CREB1 homolog) that regulates the unfolded protein response, via UPRE

---

---

			binding, and membrane biogenesis; ER stress-induced splicing pathway facilitates efficient Hac1p synthesis
<i>hal9</i>	1		Transcription factor that activates transcription of genes involved in stress response; nuclear localization is positively regulated by calcineurin-mediated dephosphorylation
<i>hap2</i>			Subunit of the heme-activated, glucose-repressed Hap2p/3p/4p/5p CCAAT-binding complex, a transcriptional activator and global regulator of respiratory gene expression; contains sequences sufficient for both complex assembly and DNA binding
<i>hap3</i>			Subunit of the heme-activated, glucose-repressed Hap2p/3p/4p/5p CCAAT-binding complex, a transcriptional activator and global regulator of respiratory gene expression; contains sequences contributing to both complex assembly and DNA binding
<i>hap4</i>			Subunit of the heme-activated, glucose-repressed Hap2p/3p/4p/5p CCAAT-binding complex, a transcriptional activator and global regulator of respiratory gene expression; provides the principal activation function of the complex
<i>hap5</i>			Subunit of the heme-activated, glucose-repressed Hap2/3/4/5 CCAAT-binding complex, a transcriptional activator and global regulator of respiratory gene expression; required for assembly and DNA binding activity of the complex
<i>ime1</i>			Master regulator of meiosis that is active only during meiotic events, activates transcription of early meiotic genes through interaction with Ume6p, degraded by the 26S proteasome following phosphorylation by Ime2p
<i>ino2</i>	1	1	Component of the heteromeric Ino2p/Ino4p basic helix-loop-helix transcription activator that binds inositol/choline-responsive elements (ICREs), required for derepression of phospholipid biosynthetic genes in response to inositol depletion
<i>ino4</i>		1	Transcription factor required for derepression of inositol-choline-regulated genes involved in phospholipid synthesis; forms a complex,

---

---

		with Ino2p, that binds the inositol-choline-responsive element through a basic helix-loop-helix domain
<i>ixr1</i>		Protein that binds DNA containing intrastrand cross-links formed by cisplatin, contains two HMG (high mobility group box) domains, which confer the ability to bend cisplatin-modified DNA; mediates aerobic transcriptional repression of COX5b
<i>leu3</i>	1	Zinc-knuckle transcription factor, repressor and activator; regulates genes involved in branched chain amino acid biosynthesis and ammonia assimilation; acts as a repressor in leucine-replete conditions and as an activator in the presence of alpha-isopropylmalate, an intermediate in leucine biosynthesis that accumulates during leucine starvation
<i>lys14</i>	1	Transcriptional activator involved in regulation of genes of the lysine biosynthesis pathway; requires 2-aminoadipate semialdehyde as co-inducer
<i>mac1</i>	1	Copper-sensing transcription factor involved in regulation of genes required for high affinity copper transport
<i>met28</i>	1	Basic leucine zipper (bZIP) transcriptional activator in the Cbf1p-Met4p-Met28p complex, participates in the regulation of sulfur metabolism
<i>met31</i>	1	Zinc-finger DNA-binding protein, involved in transcriptional regulation of the methionine biosynthetic genes, similar to Met32p
<i>mga1</i>		Protein similar to heat shock transcription factor; multicopy suppressor of pseudohyphal growth defects of ammonium permease mutants
<i>mig1</i>		Transcription factor involved in glucose repression; sequence specific DNA binding protein containing two Cys2His2 zinc finger motifs; regulated by the SNF1 kinase and the GLC7 phosphatase
<i>mig2</i>		Protein containing zinc fingers, involved in repression, along with Mig1p, of SUC2 (invertase) expression by high levels of glucose; binds to Mig1p-binding sites in SUC2 promoter

---



---

<i>mig3</i>	1	Probable transcriptional repressor involved in response to toxic agents such as hydroxyurea that inhibit ribonucleotide reductase; phosphorylation by Snf1p or the Mec1p pathway inactivates Mig3p, allowing induction of damage response genes
<i>mks1</i>	1	Pleiotropic negative transcriptional regulator involved in Ras-CAMP and lysine biosynthetic pathways and nitrogen regulation; involved in retrograde (RTG) mitochondria-to-nucleus signaling
<i>mot3</i>	1 1	Transcriptional repressor and activator with two C2-H2 zinc fingers; involved in repression of a subset of hypoxic genes by Rox1p, repression of several DAN/TIR genes during aerobic growth, and repression of ergosterol biosynthetic genes in response to hyperosmotic stress; contributes to recruitment of the Tup1p-Cyc8p general repressor to promoters; involved in positive transcriptional regulation of CWP2 and other genes; can form the [MOT3+] prion
<i>msn1</i>	1	Transcriptional activator involved in regulation of invertase and glucoamylase expression, invasive growth and pseudohyphal differentiation, iron uptake, chromium accumulation, and response to osmotic stress; localizes to the nucleus
<i>msn2</i>	1	Transcriptional activator related to Msn4p; activated in stress conditions, which results in translocation from the cytoplasm to the nucleus; binds DNA at stress response elements of responsive genes, inducing gene expression
<i>msn4</i>	1	Transcriptional activator related to Msn2p; activated in stress conditions, which results in translocation from the cytoplasm to the nucleus; binds DNA at stress response elements of responsive genes, inducing gene expression
<i>ngg1/ada3</i>		Transcriptional regulator involved in glucose repression of Gal4p-regulated genes; component of transcriptional adaptor and histone acetyltransferase complexes, the ADA complex, the SAGA complex, and the SLIK complex

---

---

<i>not3</i>			Subunit of the CCR4-NOT complex, which is a global transcriptional regulator with roles in transcription initiation and elongation and in mRNA degradation
<i>nrg1</i>	1		Transcriptional repressor that recruits the Cyc8p-Tup1p complex to promoters; mediates glucose repression and negatively regulates a variety of processes including filamentous growth and alkaline pH response
<i>oaf1</i>			Oleate-activated transcription factor, acts alone and as a heterodimer with Pip2p; activates genes involved in beta-oxidation of fatty acids and peroxisome organization and biogenesis
<i>opi1</i>	1		Transcriptional regulator of a variety of genes; phosphorylation by protein kinase A stimulates Opi1p function in negative regulation of phospholipid biosynthetic genes; involved in telomere maintenance
<i>pdr1</i>	1		Zinc cluster protein that is a master regulator involved in recruiting other zinc cluster proteins to pleiotropic drug response elements (PDREs) to fine tune the regulation of multidrug resistance genes
<i>pdr3</i>	1	1	Transcriptional activator of the pleiotropic drug resistance network, regulates expression of ATP-binding cassette (ABC) transporters through binding to cis-acting sites known as PDREs (PDR responsive elements); post-translationally up-regulated in cells lacking a functional mitochondrial genome
<i>pdr8</i>	1	1	Transcription factor; targets include ATP-binding cassette (ABC) transporters, major facilitator superfamily transporters, and other genes involved in the pleiotropic drug resistance (PDR) phenomenon
<i>pho23</i>			Probable component of the Rpd3 histone deacetylase complex, involved in transcriptional regulation of PHO5; affects termination of snoRNAs and cryptic unstable transcripts (CUTs); C-terminus has similarity to human candidate tumor suppressor p33(ING1) and its isoform ING3
<i>pho4</i>	1		Basic helix-loop-helix (bHLH) transcription factor of the myc-family; activates transcription cooperatively with Pho2p in response to

---

---

			phosphate limitation; binding to 'CACGTG' motif is regulated by chromatin restriction, competitive binding of Cbf1p to the same DNA binding motif and cooperation with Pho2p; function is regulated by phosphorylation at multiple sites and by phosphate availability
<i>ppr1</i>	1	1	Zinc finger transcription factor containing a Zn(2)-Cys(6) binuclear cluster domain, positively regulates transcription of URA1, URA3, URA4, and URA10, which are involved in de novo pyrimidine biosynthesis, in response to pyrimidine starvation; activity may be modulated by interaction with Tup1p
<i>put3</i>			Transcriptional activator of proline utilization genes, constitutively binds PUT1 and PUT2 promoter sequences as a dimer and undergoes a conformational change to form the active state; differentially phosphorylated in the presence of different nitrogen sources; has a Zn(2)-Cys(6) binuclear cluster domain
<i>rgt1</i>	1	1	Glucose-responsive transcription factor that regulates expression of several glucose transporter (HXT) genes in response to glucose; binds to promoters and acts both as a transcriptional activator and repressor
<i>ric1</i>		1	Protein involved in retrograde transport to the cis-Golgi network; forms heterodimer with Rgp1p that acts as a GTP exchange factor for Ypt6p; involved in transcription of rRNA and ribosomal protein genes
<i>rim101</i>	1		Transcriptional repressor involved in response to pH and in cell wall construction; required for alkaline pH-stimulated haploid invasive growth and sporulation; activated by proteolytic processing; similar to <i>A. nidulans</i> PacC
<i>rox1</i>			Heme-dependent repressor of hypoxic genes; contains an HMG domain that is responsible for DNA bending activity
<i>rpn4</i>	1		Transcription factor that stimulates expression of proteasome genes; Rpn4p levels are in turn regulated by the 26S proteasome in a negative feedback control mechanism; RPN4 is transcriptionally regulated by various stress responses

---

---

<i>rsf2</i>		Zinc-finger protein involved in transcriptional control of both nuclear and mitochondrial genes, many of which specify products required for glycerol-based growth, respiration, and other functions
<i>sfl1</i>	1	Transcriptional repressor and activator; involved in repression of flocculation-related genes, and activation of stress responsive genes; negatively regulated by cAMP-dependent protein kinase A subunit Tpk2p
<i>sin3</i>		Component of the Sin3p-Rpd3p histone deacetylase complex, involved in transcriptional repression and activation of diverse processes, including mating-type switching and meiosis; involved in the maintenance of chromosomal integrity
<i>sip4</i>		C6 zinc cluster transcriptional activator that binds to the carbon source-responsive element (CSRE) of gluconeogenic genes; involved in the positive regulation of gluconeogenesis; regulated by Snf1p protein kinase; localized to the nucleus
<i>skn7</i>	1	Nuclear response regulator and transcription factor; physically interacts with the Tup1-Cyc8 complex and recruits Tup1p to its targets; part of a branched two-component signaling system; required for optimal induction of heat-shock genes in response to oxidative stress; involved in osmoregulation
<i>sko1</i>	1	Basic leucine zipper transcription factor of the ATF/CREB family; forms a complex with Tup1p and Cyc8p to both activate and repress transcription; cytosolic and nuclear protein involved in osmotic and oxidative stress responses
<i>spt23</i>		ER membrane protein involved in regulation of OLE1 transcription, acts with homolog Mga2p; inactive ER form dimerizes and one subunit is then activated by ubiquitin/proteasome-dependent processing followed by nuclear targeting
<i>stb5</i>	1	Transcription factor, involved in regulating multidrug resistance and oxidative stress response; forms a heterodimer with Pdr1p; contains a

---

---

		Zn(II)2Cys6 zinc finger domain that interacts with a pleiotropic drug resistance element in vitro
<i>stp1</i>	1	Transcription factor, undergoes proteolytic processing by SPS (Ssy1p-Ptr3p-Ssy5p)-sensor component Ssy5p in response to extracellular amino acids; activates transcription of amino acid permease genes and may have a role in tRNA processing
<i>stp2</i>	1	Transcription factor, activated by proteolytic processing in response to signals from the SPS sensor system for external amino acids; activates transcription of amino acid permease genes
<i>sut1</i>		Transcription factor of the Zn[II]2Cys6 family involved in sterol uptake; involved in induction of hypoxic gene expression
<i>swi6</i>	1	Transcription cofactor; forms complexes with Swi4p and Mbp1p to regulate transcription at the G1/S transition; involved in meiotic gene expression; also binds Stb1p to regulate transcription at START; cell wall stress induces phosphorylation by Mpk1p, which regulates Swi6p localization; required for the unfolded protein response, independently of its known transcriptional coactivators
<i>tea1</i>		Ty1 enhancer activator required for full levels of Ty enhancer-mediated transcription; C6 zinc cluster DNA-binding protein
<i>thi2</i>	1	Transcriptional activator of thiamine biosynthetic genes; interacts with regulatory factor Thi3p to control expression of thiamine biosynthetic genes with respect to thiamine availability; acts together with Pdc2p to respond to thiaminediphosphate demand, possibly as related to carbon source availability; zinc finger protein of the Zn(II)2Cys6 type
<i>tye7</i>		Serine-rich protein that contains a basic-helix-loop-helix (bHLH) DNA binding motif; binds E-boxes of glycolytic genes and contributes to their activation; may function as a transcriptional activator in Ty1-mediated gene expression
<i>uga3</i>		Transcriptional activator necessary for gamma-aminobutyrate (GABA)-dependent induction of GABA genes (such as UGA1, UGA2,

---

---

			UGA4); zinc-finger transcription factor of the Zn(2)-Cys(6) binuclear cluster domain type; localized to the nucleus
<i>ume6</i>	1	1	Key transcriptional regulator of early meiotic genes, binds URS1 upstream regulatory sequence, couples metabolic responses to nutritional cues with initiation and progression of meiosis, forms complex with Ime1p, and also with Sin3p-Rpd3p
<i>usv1</i>	1	1	Putative transcription factor containing a C2H2 zinc finger; mutation affects transcriptional regulation of genes involved in growth on non-fermentable carbon sources, response to salt stress and cell wall biosynthesis
<i>wtm2</i>	1		Transcriptional modulator involved in regulation of meiosis, silencing, and expression of RNR genes; involved in response to replication stress; contains WD repeats
<i>xbp1</i>	1		Transcriptional repressor that binds to promoter sequences of the cyclin genes, CYS3, and SMF2; expression is induced by stress or starvation during mitosis, and late in meiosis; member of the Swi4p/Mbp1p family; potential Cdc28p substrate
<i>yap1</i>	1		Basic leucine zipper (bZIP) transcription factor required for oxidative stress tolerance; activated by H <sub>2</sub> O <sub>2</sub> through the multistep formation of disulfide bonds and transit from the cytoplasm to the nucleus; mediates resistance to cadmium
<i>yap3</i>			Basic leucine zipper (bZIP) transcription factor
<i>yap5</i>			Basic leucine zipper (bZIP) iron-sensing transcription factor
<i>yap6</i>	1	1	Basic leucine zipper (bZIP) transcription factor; physically interacts with the Tup1-Cyc8 complex and recruits Tup1p to its targets; overexpression increases sodium and lithium tolerance; computational analysis suggests a role in regulation of expression of genes involved in carbohydrate metabolism
<i>yap7</i>			Putative basic leucine zipper (bZIP) transcription factor

---

---

<i>yrm1</i>	1	Zn2-Cys6 zinc-finger transcription factor that activates genes involved in multidrug resistance; paralog of Yrr1p, acting on an overlapping set of target genes
<i>yrr1</i>	1	Zn2-Cys6 zinc-finger transcription factor that activates genes involved in multidrug resistance; paralog of Yrm1p, acting on an overlapping set of target genes
<i>zap1</i>		Zinc-regulated transcription factor; binds to zinc-responsive promoters to induce transcription of certain genes in presence of zinc, represses other genes in low zinc; regulates its own transcription; contains seven zinc-finger domains
<b>(TOTAL)</b>	<b>4</b>	<b>3</b>
	<b>5</b>	<b>3</b>

---

I – 遺伝子の記述に次のいずれかが含まれる：stress, response, tolerance, resistance

II – 遺伝子の記述に次のいずれかが含まれる：metabol\* (例えば metabolism, metabolic, metabolite) ; \*synthe\* (例えば synthesis, synthetic, biosynthesis, biosynthetic) ; catabol\* (例えば catabolism, catabolic, catabolite), transport

Table A2. 第2章に用いた転写因子変異株のストレス条件下・ストレスフリー条件下での比増殖速度測定値 (N=1)

Strain	Batc h	$\mu_{\text{non-stress}}$	$\mu_{\text{stress}}$	ratio
<i>ace2</i>	1	0.384	0.182	0.474
<i>arg80</i>	1	0.442	0.204	0.462
<i>aro80</i>	1	0.452	0.231	0.510
<i>bas1</i>	1	0.451	0.216	0.480
<i>dal80</i>	1	0.459	0.186	0.405
<i>fzf1</i>	1	0.455	0.210	0.462
<i>gcn4</i>	1	0.458	0.168	0.367
<i>gln3</i>	1	0.460	0.216	0.469
<i>gzf3</i>	1	0.441	0.183	0.415
<i>ino4</i>	1	0.286	0.122	0.427
<i>leu3</i>	1	0.444	0.187	0.422
<i>met28</i>	1	0.438	0.180	0.410
<i>mot3</i>	1	0.466	0.204	0.438
<i>oaf1</i>	1	0.460	0.216	0.469
<i>pho4</i>	1	0.444	0.200	0.450
<i>rpn4</i>	1	0.420	0.104	0.246
<i>stp1</i>	1	0.449	0.169	0.378
<i>stp2</i>	1	0.433	0.135	0.311
<i>uga3</i>	1	0.465	0.214	0.461
<i>yap6</i>	1	0.461	0.221	0.479
<b>wt-1</b>	1	0.457	0.160	0.351
<i>arg82</i>	2	0.401	0.100	0.249
<i>bas2/grf10/p</i>	2	0.422	0.220	0.520
<i>ho2</i>				
<i>cad1/yap2</i>	2	0.443	0.248	0.560
<i>dal81</i>	2	0.439	0.244	0.557
<i>hap2</i>	2	0.433	0.214	0.495



---

<i>hap3</i>	2	0.413	0.171	0.415
<i>hap5</i>	2	0.444	0.227	0.512
<i>ime1</i>	2	0.423	0.231	0.545
<i>ino2</i>	2	0.244	0.155	0.634
<i>ixr1</i>	2	0.418	0.244	0.584
<i>lys14</i>	2	0.439	0.238	0.543
<i>met31</i>	2	0.396	0.222	0.561
<i>opi1</i>	2	0.264	0.156	0.593
<i>thi2</i>	2	0.434	0.236	0.543
<b>wt-2</b>	2	0.455	0.190	0.417
<i>adr1</i>	3	0.431	0.220	0.511
<i>arr1</i>	3	0.423	0.195	0.461
<i>ecm22</i>	3	0.443	0.218	0.493
<i>hac1</i>	3	0.431	0.209	0.485
<i>hap4</i>	3	0.447	0.193	0.431
<i>mig2</i>	3	0.421	0.203	0.483
<i>mks1</i>	3	0.421	0.223	0.530
<i>nrg1</i>	3	0.414	0.209	0.504
<i>pdr1</i>	3	0.417	0.186	0.446
<i>pho23</i>	3	0.435	0.223	0.514
<i>ric1</i>	3	0.327	0.211	0.644
<i>sin3</i>	3	0.460	0.212	0.460
<i>sut1</i>	3	0.437	0.208	0.476
<i>ume6</i>	3	0.440	0.233	0.529
<b>wt-3</b>	3	0.399	0.164	0.412
<i>cat8</i>	4	0.352	0.222	0.629
<i>cha4</i>	4	0.394	0.213	0.540
<i>cin5</i>	4	0.409	0.230	0.563
<i>crz1</i>	4	0.419	0.224	0.534
<i>hal9</i>	4	0.417	0.214	0.514
<i>msn2</i>	4	0.396	0.202	0.510

---

---

<i>msn4</i>	4	0.387	0.210	0.541
<i>pdr8</i>	4	0.434	0.251	0.578
<i>ppr1</i>	4	0.419	0.265	0.633
<i>rgt1</i>	4	0.418	0.217	0.519
<i>rim101</i>	4	0.408	0.227	0.555
<i>sip4</i>	4	0.409	0.209	0.513
<i>skn7</i>	4	0.434	0.218	0.501
<i>stb5</i>	4	0.413	0.107	0.259
<i>yap1</i>	4	0.422	0.216	0.512
<i>yap3</i>	4	0.421	0.217	0.515
<b>wt-4</b>	4	0.410	0.202	0.493
<i>aft2</i>	5	0.463	0.237	0.513
<i>arr1</i>	5	0.439	0.205	0.468
<i>gcr2</i>	5	0.318	0.124	0.389
<i>msn1</i>	5	0.430	0.213	0.495
<i>pdr3</i>	5	0.443	0.214	0.484
<i>sfl1</i>	5	0.455	0.183	0.401
<i>sko1</i>	5	0.428	0.272	0.635
<i>swi6</i>	5	0.377	0.076	0.201
<i>tye7</i>	5	0.453	0.249	0.550
<i>usv1</i>	5	0.445	0.216	0.486
<i>wtm2</i>	5	0.450	0.239	0.532
<i>xbp1</i>	5	0.431	0.220	0.510
<i>yap5</i>	5	0.445	0.191	0.430
<i>yap7</i>	5	0.435	0.218	0.500
<i>yrm1</i>	5	0.454	0.205	0.452
<i>yrr1</i>	5	0.458	0.193	0.420
<b>wt-5</b>	5	0.442	0.162	0.367
<i>aft1</i>	6	0.434	0.166	0.384
<i>cbf1</i>	6	0.349	0.166	0.476
<i>cst6</i>	6	0.476	0.219	0.460

---

<i>gat1</i>	6	0.468	0.245	0.523
<i>gis1</i>	6	0.476	0.257	0.540
<i>mig1</i>	6	0.441	0.222	0.505
<i>mig3</i>	6	0.471	0.254	0.539
<i>put3</i>	6	0.471	0.229	0.486
<i>rsf2</i>	6	0.456	0.183	0.402
<i>zap1</i>	6	0.448	0.224	0.498
<b>wt-6</b>	6	0.444	0.193	0.435
<i>ace1/cup2</i>	7	0.499	0.212	0.426
<i>ada2</i>	7	0.320	0.103	0.320
<i>arg81/argr2</i>	7	0.477	0.214	0.449
<i>azf1</i>	7	0.491	0.281	0.573
<i>cup9</i>	7	0.457	0.202	0.442
<i>dal82</i>	7	0.471	0.232	0.494
<i>gal3</i>	7	0.492	0.217	0.440
<i>gal4</i>	7	0.508	0.209	0.411
<i>gal80</i>	7	0.466	0.216	0.464
<i>gat2</i>	7	0.474	0.245	0.518
<i>mac1</i>	7	0.448	0.240	0.537
<i>mga1</i>	7	0.507	0.168	0.331
<i>ngg1/ada3</i>	7	0.387	0.166	0.429
<i>not3</i>	7	0.449	0.191	0.426
<i>rox1</i>	7	0.499	0.216	0.434
<i>spt23</i>	7	0.441	0.229	0.520
<i>tea1</i>	7	0.448	0.197	0.441
<b>wt-7</b>	7	0.480	0.161	0.336

wt-1 から wt-7 まで名付けたサンプルは、参照のために各培養バッチに含まれた標準株 (BY4742).

'Ratio'は  $\mu_{\text{stress}}/\mu_{\text{non-stress}}$  比を意味する.

**Table A3.** メタボローム解析および回帰モデル構築に用いた 19 株のストレスフリー条件下および 1.5% (v/v)1 - ブタノール条件下比増殖速度

	$\mu_{\text{nonstress}}$						$\mu_{\text{stress}}$			
	n1	n2	n3	n4	Mean	S.D.	n1	n2	Mean	S.D.
mks1	0.4147	0.5525	0.4635	0.3815	<b>0.4531</b>	0.0744	0.1836	0.1811	<b>0.1823</b>	0.0018
sko1	0.4389	0.5584	0.4829	0.4250	<b>0.4763</b>	0.0601	0.1724	0.1549	<b>0.1636</b>	0.0124
lys14	0.4240	0.5446	0.4839	0.4159	<b>0.4671</b>	0.0599	0.1570	0.1532	<b>0.1551</b>	0.0027
thi2	0.4106	0.5473	0.4677	0.3971	<b>0.4557</b>	0.0684	0.1537	0.1491	<b>0.1514</b>	0.0032
leu3	0.4802	0.5427	0.4646	0.3785	<b>0.4665</b>	0.0677	0.1432	0.1539	<b>0.1485</b>	0.0076
bas1	0.4523	0.5362	0.5213	0.5145	<b>0.5061</b>	0.0370	0.1562	0.1407	<b>0.1484</b>	0.0110
tye7	0.4831	0.5377	0.4902	0.3984	<b>0.4773</b>	0.0579	0.1340	0.1545	<b>0.1443</b>	0.0145
gat2	0.5220	0.5481	0.5451	0.4335	<b>0.5122</b>	0.0537	0.1429	0.1421	<b>0.1425</b>	0.0006
sip4	0.4682	0.5476	0.4932	0.4535	<b>0.4906</b>	0.0414	0.1574	0.1267	<b>0.1421</b>	0.0217
yap6	0.5124	0.5388	0.5079	0.5071	<b>0.5165</b>	0.0150	0.1461	0.1311	<b>0.1386</b>	0.0107
azf1	0.4979	0.5645	0.4885	0.4273	<b>0.4945</b>	0.0562	0.1340	0.1422	<b>0.1381</b>	0.0058
aro80	0.4853	0.5312	0.4794	0.4686	<b>0.4911</b>	0.0276	0.1442	0.1319	<b>0.1381</b>	0.0087
put3	0.4517	0.5197	0.4882	0.4974	<b>0.4892</b>	0.0283	0.1455	0.1291	<b>0.1373</b>	0.0116
oaf1	0.4883	0.4860	0.4708	0.4929	<b>0.4845</b>	0.0096	0.1436	0.1209	<b>0.1323</b>	0.0160
dal80	0.4446	0.5436	0.5483	0.4933	<b>0.5075</b>	0.0487	0.1256	0.1269	<b>0.1263</b>	0.0009
stp2	0.4846	0.5243	0.5007	0.4386	<b>0.4871</b>	0.0362	0.1212	0.1149	<b>0.1181</b>	0.0045
gcn4	0.4911	0.5307	0.4993	0.4425	<b>0.4909</b>	0.0365	0.1266	0.1037	<b>0.1152</b>	0.0162
rsf2	0.4384	0.5482	0.4984	0.5128	<b>0.4994</b>	0.0458	0.1127	0.1058	<b>0.1092</b>	0.0049
mot3	0.3891	0.3954	0.3995	0.4156	<b>0.3999</b>	0.0113	0.0714	0.0907	<b>0.0810</b>	0.0136
BY4742	0.4136	0.4934	0.4334	0.4068	<b>0.4368</b>	0.0394	0.1198	0.1215	<b>0.1207</b>	0.0012

**Table A4.** メタボローム解析に際して用いたデータ処理ソフトウェアのパラメータ

<u>MetAlign</u>		
Mass Mode	Nominal	
Mass Bin	0.60	
<u>Baseline and Noise Peak Elimination</u>		
<u>Params</u>	1	
Retention Begin (Scan nr)	24000	
Retention End (Scan nr)	7000000	
Maximum Amplitude	1	
Peak Slope Factor (x Noise)	4	
Peak Threshold Factor (x Noise)	200	
Peak Threshold (Abs. Value)	25	
Average Peak Width at Half Height (Scans)	no	
Keep Peak Shape (no alignment)		
Pre-Synchronize Scans	No Scaling	
<u>Scaling Options</u>	<u>Scan Nr.</u>	<u>Max. Shift</u>
	0	20
<u>Initial Peak Search Criteria</u>	24000	30
Begin of 1 <sup>st</sup> Region	0	0
End of 1 <sup>st</sup> Region	0	0
Begin of 2 <sup>nd</sup> Region		
End of 2 <sup>nd</sup> Region		
	Pre-align processing (Iterative)	
<u>Tuning Alignment Options and Criteria</u>	35	
Alignment Type	<u>1<sup>st</sup> Iteration</u>	<u>Last Iteration</u>
Maximum Shift per 100 Scans	7	7
	10	5

Mass Peak Selection

Min. Factor (x Noise)

Min. Nr. of Masses

**MSClust**

Effective Peaks	200
Peak Width, scans or time	10
P.W. Margin Softness	2
Correlation Threshold	0.8
C.T. Margin Softness	0.02
PD Reduction	0.8
PD Reduction Softness	0.01
Stop Criterion	2

**AIoutput2**

Peak Table Making

Height Threshold	50
RT Binning	2

Peak Identification and Annotation

Available Index	Retention Index
Analysis Type	Non targeted
RI Tolerance	5
Match Threshold (0.7-0.9)	0.8

Filtering

Type	Accurate
Height Filter	1000
RSD(CV) Filter	20

Table A5. GC/MS により検出および定量された化合物ピーク

Compound Name	RT	RI	RI lib.	RI deviation	QuantMS	Mass similarity
Unknown	247.75	No RI Inf.	-	-	134	-
Unknown	248	No RI Inf.	-	-	130	-
Unknown	257.55	No RI Inf.	-	-	221	-
n-Propylamine	282.1	1028	1028.23	0.649	174	0.993
Oxalacetic acid+Pyruvate	297.35	1050	1049.08	0.803	174	0.992
Unknown_2	302.15	1057	-	-	130	-
Lactic acid	306.75	1064	1062.36	1.270	117	0.995
Unknown_4	312.65	1072	-	-	221	-
Unknown_5	333.55	1103	-	-	188	-
Alanine_2TMS	336	1107	1105.68	1.023	116	0.988
Unknown_6_Organic acid like	336.05	1107	-	-	147	-
Unknown_8	354.5	1135	-	-	130	-
Unknown_10	366.3	1153	-	-	191	-
Leucine_1TMS	369.5	1158	1157.68	0.294	86	0.999
Unknown_13	381.95	1177	-	-	144	-
Unknown_14	409.2	1220	-	-	130	-
Valine_2TMS_major	411	1223	1221.7	1.445	144	0.994
Urea	419.05	1236	1236.49	0.074	147	0.996
Unknown_16_Amine like	422.3	1242	-	-	174	-
Serine_2TMS_minor	435.05	1263	1262.64	0.153	116	0.994
Leucine_2TMS	445.1	1279	1278.66	0.700	158	0.994
Unknown_23	446.35	1281	-	-	211	-

Glycerol	446.7	1282	1280.81	1.188	147	0.966
Threonine_2TMS_minor	458.15	1301	1299.31	1.635	117	0.967
Isoleucine_2TMS_major	458.5	1302	1300.64	0.930	158	0.990
Proline_2TMS	461	1306	1306.38	0.351	142	0.999
Glycine_3TMS	466.95	1317	1316.4	0.243	174	0.996
Uracil	483.4	1346	1344.43	1.556	99	0.991
Serine_3TMS_major	496.75	1370	1370.14	0.340	204	0.998
Threonine_3TMS_major	512.45	1398	1396.08	1.726	117	0.988
Malic acid	564.45	1497	1496.36	0.455	147	0.986
Aspartic acid_3TMS	580.95	1530	1529.21	0.981	232	0.995
5-Oxoproline	582.4	1533	1532.89	0.253	156	0.981
Glutamic acid_3TMS	628.65	1629	1628.11	0.803	246	0.993
Phenylalanine_2TMS	634.8	1642	1642.75	0.577	218	0.996
Asparagine_3TMS	653.55	1683	1682.41	0.190	116	0.994
Phthalic acid	668.25	1715	1707.28	7.816	147	0.823
2-Aminoadipic acid	673	1726	1725.27	0.641	217	0.966
Quinolinic acid	680	1742	1742.32	0.471	147	0.922
Ribitol	684.5	1752	1753.64	1.545	217	0.978
Orotic acid	686.6	1757	1757.76	0.884	254	0.964
Ornithine_3TMS	691.3	1768	1768.56	0.983	174	0.959
Glutamine_3TMS	698.9	1785	1785.52	0.638	156	0.994
Isocitric acid+Citric acid	721.75	1839	1837.41	1.370	147	0.968
4-Aminobenzoic acid	723.4	1843	1843.24	0.513	192	0.970
Cadaverine	729.3	1857	1856.9	0.058	174	0.988
Glucose_1	760.85	1934	1932.24	1.722	205	0.956
Lysine_4TMS	763.35	1940	1939.94	0.306	156	0.968
Galactose_2+Glucose_2	768.75	1954	1950.45	3.371	147	0.931
Tyrosine	770.45	1958	1956.8	1.295	218	0.993



Sorbitol	777.7	1976	1977.61	1.290	205	0.944
Glucarate_1_same	781.1	1985	1980.9	3.967	217	0.874
Plamitic acid(16:0)	804.85	2047	2047.65	1.041	117	0.966
Inositol	837.25	2133	2132.25	1.019	217	0.961
Tryptophan_1TMS	867.5	2217	2217.35	0.148	218	0.997
Tryptamine_1	873.3	2234	2233.05	0.780	128	0.984
Stearic acid(17:0)	876.9	2244	2244.1	0.051	117	0.950
Tryptophan_3TMS	878.85	2250	2249.15	0.592	202	0.997
Spermidine	891.65	2286	2285.18	1.259	144	0.941
Cystine	904.05	2323	2324.78	1.843	218	0.912
Fructose 6-phosphate_2	915.2	2356	2355.34	0.940	315	0.895
Fructose 6-phosphate_3	920.15	2371	2370.11	0.973	387	0.928
Unknown_98	942.65	2440	-	-	150	-
Sucrose	1024.4	2706	2706.2	0.432	361	0.920
Trehalose	1056.3	2817	2816.81	0.552	191	0.949
Unknown_102	1108.3	3008	-	-	105	-

本表では AIoutput2 の出力 (PeakTableUpdate および IdentificationTable ワークシート) より得られたピーク情報をまとめた。

Table A6. 第2章で得られた新株のストレスフリーおよび1.5% (v/v) 1-ブタノール条件下比増殖速度 (各 N=4)

Strain	n1	n2	n3	n4	Average	Standard deviation	p-value
<b><math>\mu_{\text{nonstress}}</math></b>							
<i>his3</i>	0.4131	0.4331	0.3967	0.4020	<b>0.4112</b>	0.0161	-
<i>cha1</i>	0.4335	0.4602	0.4508	0.4439	<b>0.4471</b>	0.0112	0.0525
<i>met2</i>	0.4344	0.4156	0.4502	0.4281	<b>0.4321</b>	0.0144	0.4215
<i>aat1</i>	0.4392	0.4659	0.4494	0.4352	<b>0.4474</b>	0.0137	0.0554
<i>aat2</i>	0.4471	0.4680	0.3957	0.4537	<b>0.4411</b>	0.0315	0.2377
<i>cit1</i>	0.4604	0.4333	0.4056	0.4557	<b>0.4387</b>	0.0251	0.2526
<i>cit2</i>	0.4735	0.4297	0.4439	0.4210	<b>0.4420</b>	0.0230	0.1642
<i>cit3</i>	0.3814	0.4302	0.3776	0.4172	<b>0.4016</b>	0.0261	0.1652
<b><math>\mu_{\text{stress}}</math></b>							
<i>his3</i>	0.1329	0.1280	0.1565	0.1378	<b>0.1388</b>	0.0124	-
<i>cha1</i>	0.1558	0.1503	0.1526		<b>0.1529</b>	0.0028	0.1178
<i>met2</i>	0.1696	0.1635	0.1646		<b>0.1659</b>	0.0033	0.0156
<i>aat1</i>	0.1456	0.1327	0.1463	0.1527	<b>0.1443</b>	0.0084	0.4927
<i>aat2</i>	0.1523	0.1640	0.1451	0.1421	<b>0.1509</b>	0.0098	0.1785
<i>cit1</i>	0.1378	0.1553	0.1508	0.1433	<b>0.1468</b>	0.0078	0.3175
<i>cit2</i>	0.1720	0.1873	0.1787	0.1816	<b>0.1799</b>	0.0064	0.0011
<i>cit3</i>	0.1332	0.1420	0.1398	0.1519	<b>0.1417</b>	0.0077	0.7064

各株について、異なるシングルコロニーから生物学的反復のサンプルを用意した。p-valueはStudent's *t*検定(two-tailed 両側, homoscedastic 等分散)より計算した。

Table A7. 各 OSC 成分により除かれたデータ分散の割合

NO. OF OSC COMPONENTS	OSCFRAC. VARIANCE REMAINING	TOTAL FRAC. VARIANCE EXPLAINED BY OSC COMPONENT	TOTAL FRAC. VARIANCE EXPLAINED (CUMULATIVE)
0	1	-	0
1	0.8268988	0.1731012	0.1731012
2	0.699962	0.1269368	0.300038
3	0.6463233	0.0536387	0.3536767
4	0.5974177	0.04890561	0.4025823

本研究では4つまでの OSC 成分を構築してみた。各成分により説明された変分をデータセットから引き算した際の、分散の減少を表で示した。第1、第2の OSC 成分はそれぞれデータ分散の17.31%と12.69%を説明したので、有意な OSC 成分として考えた。一方、第3の OSC 成分はデータ分散のわずか5.36%しか説明せず、それ以降の OSC 成分は更に小さな分散の割合を説明したので、適用する OSC 成分の数を2とした。

Table A8. 第3章のデータマイニング法の各パラメータ

**Data preprocessing**

Scaling method: autoscale (mean-centering + scaling by standard deviation)

No. of OSC components: 2

**RANSAC-PLS**

Initial sample size: 8

No. of PLS components: 1

Inlier  $y$ -error threshold (initial value): 0.5\*

\*adjusted with  $y$ -error threshold =  $0.5 - 0.25 * \text{stringency}$

Min. score (initial value) = 1

\*\*adjusted with min. score =  $1 * \text{stringency}$

Stringency (initial value) = 1

\*\*\*adjusted with stringency =  $\text{stringency} * 0.9$

PLS cross validation method: 7-fold jackknifing cross validation (CV7)

Threshold num. models for termination: 1000

**HCA**

Distance measure: Pearson correlation

Linkage method: Ward

Table A9. 第3章で得られた新株の 1.5% (v/v) 1 - ブタノール条件下比増殖速度 (N=4)

	<i>n1</i>	<i>n2</i>	<i>n3</i>	<i>n4</i>	<b>averag e</b>	<i>s.d.</i>	<i>p-value</i>
<b>non-stress</b>							
<i>his3</i>	0.417	0.417	0.437	0.421	<b>0.423</b>	0.010	<i>1.0000</i>
<i>oxp1</i>	0.435	0.437	0.426	0.423	<b>0.430</b>	0.007	<i>0.2434</i>
<i>bat2</i>	0.443	0.446	0.440	0.444	<b>0.443</b>	0.003	<i>0.0060</i>
<i>ath1</i>	0.424	0.446	0.433	0.437	<b>0.435</b>	0.009	<i>0.1124</i>
<i>nth1</i>	0.444	0.435	0.452	0.437	<b>0.442</b>	0.008	<i>0.0220</i>
<b>stress</b>							
<i>his3</i>	0.090	0.115	0.083	0.101	<b>0.097</b>	0.014	<i>1.0000</i>
<i>oxp1</i>	0.131	0.114	0.147	0.163	<b>0.139</b>	0.021	<i>0.0164</i>
<i>bat2</i>	0.144	0.135	0.136	0.157	<b>0.143</b>	0.010	<i>0.0019</i>
<i>ath1</i>	0.106	0.124	0.096	0.136	<b>0.115</b>	0.018	<i>0.1636</i>
<i>nth1</i>	0.142	0.117	0.153	0.118	<b>0.133</b>	0.018	<i>0.0209</i>

各株について、異なるシングルコロニーから生物学的反復のサンプルを用意した。ストレス、ストレスフリー条件において同じ番号の繰り返し (“n1” など) には、同じシングルコロニーからのものを用いた。p-value は Student’s *t* 検定 (two-tailed 両側, homoscedastic 等分散) より計算した。

## 発表論文

### 第一著者論文

Shao Thing Teoh, Sastia Putri, Yukio Mukai, Takeshi Bamba, and Eiichiro Fukusaki, “A metabolomics-based strategy for identification of gene targets for phenotype improvement and its application to 1-butanol tolerance in *Saccharomyces cerevisiae*,” *Biotechnology for Biofuels*, vol. 8, no. 1, p. 144, Sep. 2015.

### 共著者論文

Zanariah Hashim, Shao Thing Teoh, Takeshi Bamba, and Eiichiro Fukusaki, “Construction of a metabolome library for transcription factor-related single gene mutants of *Saccharomyces cerevisiae*,” *Journal of Chromatography B*, vol. 966, pp. 83–92, Sep. 2014.

## 学会発表

1) Shao Thing Teoh, Sastia Putri, Yukio Mukai, Takeshi Bamba and Eiichiro Fukusaki. “Metabolomics-based prediction of 1-butanol tolerance in *Saccharomyces cerevisiae*” (Oral). *Society for Biotechnology Japan (SBJ) 64th Annual Meeting*, Kobe, Japan. 23-26 October 2012.

2) Shao Thing Teoh, Sastia Putri, Yukio Mukai, Takeshi Bamba and Eiichiro Fukusaki. “Metabolome analysis and construction of a prediction model for 1-butanol tolerance in *Saccharomyces cerevisiae*” (Poster). *International Joint Symposium on Single-Cell Analysis (The 6th International Workshop on Approaches to Single-Cell Analysis & The 8th International Forum on Post-Genome Technologies)*, Kyoto, Japan. 27-28 November 2012.

3) Shao Thing Teoh, Sastia Putri, Yukio Mukai, Takeshi Bamba and Eiichiro Fukusaki. “Metabolomics-based semi-rational identification of gene targets conferring 1-butanol tolerance in *Saccharomyces cerevisiae* using gas chromatography-mass spectrometry and OPLS modeling” (Poster). *Society for Biotechnology Japan (SBJ) 65th Annual Meeting*, Hiroshima, Japan. 18-20 September 2013.

4) Shao Thing Teoh, Sastia Putri, Yukio Mukai, Takeshi Bamba and Eiichiro Fukusaki. “Metabolomics-based semi-rational identification of gene targets conferring 1-butanol tolerance in *Saccharomyces cerevisiae*” (Oral; selected for 2014 Metabolomics Society Student Travel Award, 2014 Metabolomics Society Student Prize). *10th International Conference of the Metabolomics Society*, Tsuruoka, Japan. 23-26 June 2014.