

研究終了報告書

「有用物質生産を志向した機械学習支援ゲノムデザイン」

研究期間：2020年11月～2024年3月

研究者：山田 亮祐

1. 研究のねらい

石油資源の枯渇や、種々の環境問題を背景に、持続可能な社会を形成するため、再生可能な資源である植物バイオマス由来のグルコースなどの糖から、様々な有用物質を生産することが期待されている。遺伝子組換え技術を用いて生物の細胞内代謝を改変する代謝工学により、糖から、様々な微生物に様々な有用物質を生産させることが可能である。特に古くから食品の生産に利用され、広く安全であると認められているため、産業利用が容易な真核生物である酵母 *Saccharomyces cerevisiae* の代謝工学が盛んに研究されている。

グルコースをピルビン酸にまで分解する代謝経路は一般的に解糖系と呼ばれ、多くの目的有用物質はこのピルビン酸を経由して生産される。従って、グルコースをピルビン酸にまで代謝する速度を向上できれば、様々な有用物質を高い速度で生産することに役立つと考えられる。しかし、*S. cerevisiae* のような真核生物ではほとんど報告例がなく、真核生物においては実現困難な課題であった。研究代表者は既往の研究において、*S. cerevisiae* の解糖系に関連する13種類の酵素の発現量を同時に改変することで、これまで極めて困難であった酵母の解糖系によるグルコース代謝の高速化に成功した。しかし、主要代謝経路酵素遺伝子の発現量と、細胞増殖能や有用物質生産能との関係は未だ不明のままである。もしこの関係が明らかにできれば、種々の有用物質生産において、各遺伝子の最適発現量が分かり、合成生物学による有用物質生産細胞構築における、最適なゲノムデザインが明らかとなる。また、既存の遺伝子組換え酵母による物質生産系において、遺伝子発現量を再調整し、物質生産性を向上させるための指針を得ることが期待できる。

本研究では、主要代謝経路(解糖系およびメバロン酸経路)の遺伝子発現量を多様に変化させた酵母ライブラリーを作製する。次にそれらの遺伝子発現量と細胞増殖能および物質生産能との関係性を評価し、それにより得られたデータにより機械学習モデルを作製する。最後に作製した機械学習モデルを用いて、D-乳酸(解糖系経路生成物)およびβ-カロテン(解糖系およびメバロン酸経路経路生成物)の生産に最適な遺伝子発現量を予測することを目指す。

2. 研究成果

(1) 概要

機械学習による解糖系のデザインにおいては、まず様々な解糖系酵素遺伝子および *D-LDH* の発現量を有する酵母ライブラリーを作製し、遺伝子発現量と D-乳酸生産性を測定した。次に、遺伝子発現量を入力値、D-乳酸生産性を出力値とした機械学習モデルを構築した。構築したモデルの決定係数は、0.6628 となり、高い精度で遺伝子発現量から D-乳酸生産性を予測することに成功した。この機械学習モデルを用いて SHAP 値を算出したところ、特に *PFK1* と *ENO2* の発現量が D-乳酸生産性に大きく影響すると示された。作製した機械学習モ

デルを用い、高効率な D-乳酸生産性を実現するための遺伝子発現量を予測したところ、特に *FBA1*、*PFK1*、*PYK2* 等の発現量が高いことが重要であることが明らかとなった。D-乳酸生産性の予測値は 7.96 g/L/OD₆₀₀ であり、実際に作製した株 (5.84 g/L/OD₆₀₀) の 1.4 倍であった。

機械学習によるメバロン酸経路のデザインにおいては、まず様々なメバロン酸経路遺伝子の発現量を有する酵母ライブラリーを作製し、遺伝子発現量と β-カロテン生産性を測定した。次に、遺伝子発現量を入力値、β-カロテン生産性を出力値とした機械学習モデルを構築した。構築した機械学習モデルの決定係数は 0.6292 となり、高い精度で遺伝子発現量から β-カロテン生産性を予測することに成功した。この機械学習モデルを用いて SHAP 値を算出したところ、特に *tHMG1* と *ERG8* の発現量が β-カロテン生産性に大きく影響すると示された。また、作製したモデルを用いて、β-カロテン生産性が最大となる遺伝子発現量を予測した。予測された最大 β-カロテン生産性 (13.54 mg/L/OD₆₀₀) は宿主の約 4.3 倍、構築した株の約 1.4 倍であった。

本研究では、実測した 13 種類の解糖系遺伝子から D-乳酸生産性を、8 種類のメバロン酸経路遺伝子の発現量から β-カロテン生産性を予測する機械学習モデルの構築に成功した。また、構築した機械学習モデルを用い、D-乳酸や β-カロテンの生産に最適な遺伝子発現量を明らかにすることに成功した。

(2) 詳細

研究テーマ A「機械学習支援解糖系デザイン」

代謝工学の発展に伴い、微生物により種々の有用物質を生産する研究が盛んに行われている。特に、酵母 *S. cerevisiae* は安全な微生物であり産業規模での培養技術が確立されており、解糖系を経由して生成する D-乳酸や、解糖系およびメバロン酸経路を経由して生成する β-カロテンなどの様々な有用物質の生産への応用が期待されている。本研究では、様々な解糖系酵素遺伝子および *D-LDH* の発現量を有する酵母ライブラリーを作製し、遺伝子発現量と D-乳酸生産性を測定した。また、取得したデータを基に、高効率な D-乳酸生産を実現するための遺伝子発現量を予測する機械学習モデルの構築を目指した。

酵母 YPH499/dPdA/DLDH/A5 株を *D-LDH* 発現・解糖系改変プラスミドで形質転換して、形質転換体ライブラリーを作製し、形質転換体 200 株の D-乳酸生産性を測定した。200 株中 44 株が親株よりも高い D-乳酸生産性を示した。その中で、YPH499/dPdA/DLDH/A5/C69 株が最も高い D-乳酸生産性 (5.84 g/L/OD₆₀₀) を示した。

次に、D-乳酸生産性を算出した 200 株の遺伝子発現量を測定した。200 株の形質転換体の中では、*HXT7* と *PYK2* の発現量が高い形質転換体の割合が高いことが確認された。遺伝子発現量を入力値、D-乳酸生産性を出力値とした機械学習モデルを構築した。モデルの決定係数は、訓練データを使用した際に 0.6932、テストデータを使用した際に 0.6628 となった。

次に、この機械学習モデルを用いて SHAP 値を算出した (図 1)。*PFK1*、*HXK2*、*FBA1*、*GPM1*、*PYK2*、*PGK1*、*PFK2* の発現量は D-乳酸生産性と正の相関があり、*ENO2*、*TDH3*、*TPI1*、*HXT7*、*D-LDH*、*PGI1* の発現量は負の相関があると判明した。特に *PFK1* と *ENO2* の発現量が D-乳酸生産性に大きく影響すると示された。

作製した機械学習モデルを用い、高効率な D-乳酸生産性を実現するための遺伝子発現量を予測した (図 2)。最大の D-乳酸生産性を示すと予測された遺伝子発現量では、特に

FBA1、*PFK1*、*PYK2* 等の発現量が高かった。D-乳酸生産性の予測値は 7.96 g/L/OD₆₀₀ であり、実際に作製した株のうち D-乳酸生産性が最大の YPH499/dPdA/DLDH/A5/C69 (5.84 g/L/OD₆₀₀) の 1.4 倍であった。

本研究では、*D-LDH* 発現・解糖系改変酵母ライブラリーを作製し、細胞濃度、D-乳酸生産量、および 12 種類の解糖系遺伝子と *D-LDH* の遺伝子発現量を測定し、機械学習に用いるデータを取得した。また、取得したデータから、D-乳酸生産性を予測する機械学習モデルの構築に成功した。構築した機械学習モデルを用いて D-乳酸を高効率に生産するための遺伝子発現量を予測し、実際に作製した株の 1.4 倍の D-乳酸生産性を示すと予測される遺伝子発現量を見出した。また、SHAP 値の計算により、特に *PFK1* と *ENO2* の発現量が D-乳酸生産性に大きな影響を与えることを明らかにした。

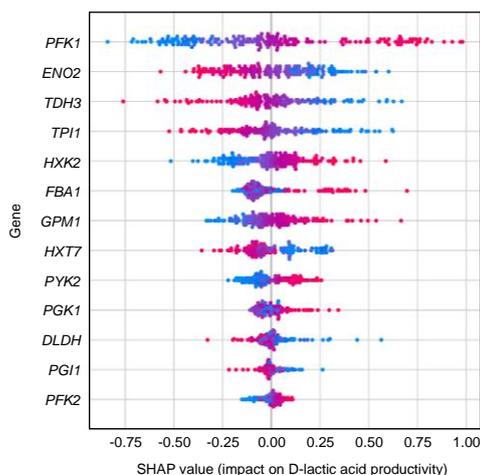


図 1 解糖系モデルの SHAP 値

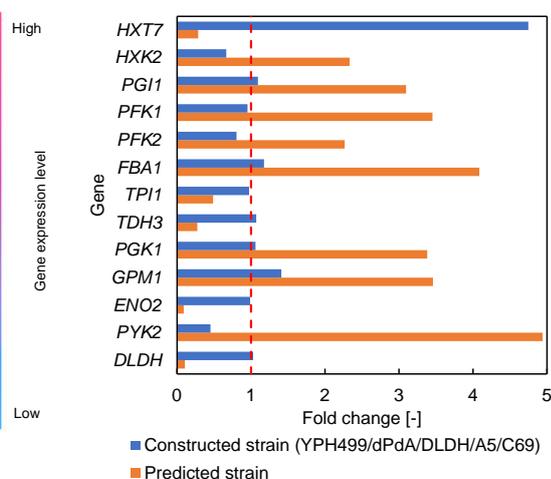


図 2 D-乳酸生産に最適な遺伝子発現量

研究テーマ B「機械学習支援メバロン酸経路デザイン」

β -カロテンは高い抗酸化作用を有することから、食品や化粧品を始めとした幅広い産業で使用される有用な化合物である。既往の研究で、多数の遺伝子の発現量を同時に調節可能なカクテル δ -インテグレーション法が開発され、この手法を用いて 3 種類の外来遺伝子を酵母 *S. cerevisiae* に発現させることで、グルコースから、解糖系およびメバロン酸経路を經由して β -カロテンを生産する酵母が作製された (Yamada et al., *Bioresour Technol*, 268, 616, 2018)。カクテル δ -インテグレーション法を用いて β -カロテン生産に関連する 8 種類のメバロン酸経路遺伝子の発現量を改変することで、 β -カロテン生産性の向上が期待できる。しかし、8 種類の酵素遺伝子の発現量比は多様であり、それを網羅的に検証し最適化することは非常に困難である。本研究では、8 種類のメバロン酸経路遺伝子の発現量と β -カロテン生産性に関するデータを取得し、各遺伝子の発現量から β -カロテン生産性を予測する機械学習モデルの構築を目指した。また、作製したモデルを用いて、 β -カロテン生産性が最大となる遺伝子発現量の予測を試みた。

β -カロテン生産酵母 YPH499/Mo2Crt105 を宿主とし、8 種類のメバロン酸経路遺伝子の発現量を改変するプラスミドを用いて形質転換を行った。作製した形質転換体の β -カロテン生産性を測定した。 β -カロテン生産性を測定した 193 株のうち、155 株が宿主よりも高い β -カロテン生産

性を示した。最も高い β -カロテン生産性を示した形質転換体の β -カロテン生産性 (9.49 mg/L/OD₆₀₀) は宿主 (3.18 mg/L/OD₆₀₀) の約3.0倍であった。

次に、宿主よりも高い β -カロテン生産性を示した 155 株の遺伝子発現量を測定した。155 株の形質転換体の中では、特に *ERG13*、*tHMG1*、*ERG8* の発現量が高い形質転換体の割合が高いことが確認された。宿主より高い生産性を示した 155 株を機械学習用データとして用い、種々のアルゴリズムを検討した結果、最も高い予測精度を示したモデルの訓練データを使用した際の決定係数は 0.7386、テストデータ使用した際の決定係数は 0.6292 となった。

次に、この機械学習モデルを用いて SHAP 値を算出した (図 3)。*tHMG1*、*ERG20*、*ERG19*、*ERG13* の発現量は β -カロテン生産性と正の相関があり、*ERG8*、*ERG10*、*IDI1*、*ERG12* の発現量は負の相関があると判明した。特に *tHMG1* と *ERG8* の発現量が β -カロテン生産性に大きく影響すると示された。また、作製したモデルを用いて、 β -カロテン生産性が最大となる遺伝子発現量を予測した (Predicted strain) (図 4)。予測された最大 β -カロテン生産性 (13.54 mg/L/OD₆₀₀) は宿主の約 4.3 倍、Constructed strain の約 1.4 倍であった。

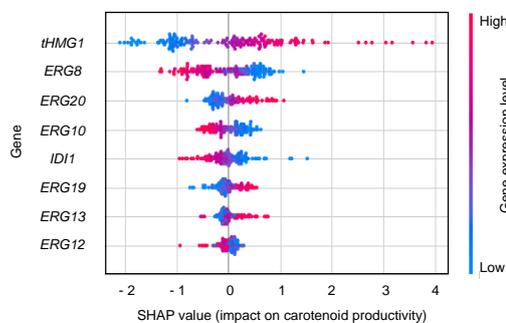


図 3 メバロン酸経路モデルの SHAP 値

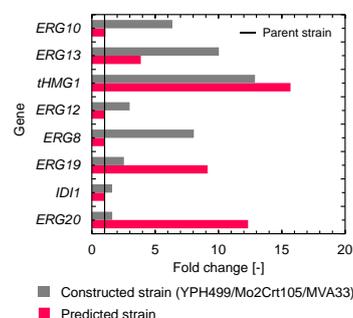


図 4 β -カロテン生産に最適な遺伝子発現量

本研究では、実測した 8 種類のメバロン酸経路遺伝子の発現量から β -カロテン生産性を予測する機械学習モデルの構築に成功した。また、構築したモデルを用いて、 β -カロテン高生産を実現する遺伝子発現量比の予測に成功した。

3. 今後の展開

本研究の成果により、酵母を用いた、解糖系やメバロン酸経路を経由する有用物質の高効率生産に資するゲノムデザイン的设计指針が得られた。今後の展開としては、1)他の主要代謝経路の設計に本研究で確立した技術を応用する 2)本研究により得られたゲノム設計指針をもとに有用物質生産酵母を開発する の 2 つの研究の展開が考えられる。いずれの研究についても、本研究において、既にそれらの基盤技術は確立済みである。従ってさらに 5 年程度研究を進めることでこれらの目標は達成できると考えられる。

本研究成果の社会実装に向けては、酵母などの微生物による有用物質の生産性を更に高めると共に、化学産業における、環境負荷の高い化学プロセスから環境負荷の低いバイオプロセスへのさらなる転換が求められる。バイオプロセスへの転換は欧米で先行しており、これに追従して日本においても今後 10 年程度でバイオプロセスへの転換が加速するものと考えられる。本研究の成果を今後さらに発展させることで、日本におけるバイオプロセスによるものづくりに貢献すること

が期待できる。

4. 自己評価

本研究の目的は、機械学習を用い、酵母による有用物質生産において特に重要な代謝経路である解糖系およびメバロン酸経路の代謝酵素遺伝子の発現量から有用物質の生産性を予測し、目的化合物の生産に最適な遺伝子発現量を明らかにすることであった。本研究により、実際に解糖系およびメバロン酸経路の多数の遺伝子の発現量から目的物質の生産性を予測し、最適な遺伝子発現量を予測することが可能であることを初めて明らかにし、本研究の目的は達成した。これまでに本研究のように、多数の遺伝子の発現量を説明変数として酵母の物質生産性を予測する機械学習モデルの報告例は知られておらず、顕著な研究成果が得られたと考えられる。得られた成果により4件(内投稿中1件)の国際誌での論文発表、7件の講演・学会発表、1件の総説発表を行い、1件の賞を受けた。

本研究では、解糖系やメバロン酸経路を経由する高効率有用物質生産に資するゲノムデザイン的设计指針が得られた。今後の展開としては、他の主要代謝経路の設計に本研究で確立した技術を応用することや、得られたゲノム設計指針をもとに有用物質生産酵母を開発することを考えている。本研究で得られた技術を今後さらに発展させることで、持続可能な社会の構築に貢献する、バイオものづくり技術の確立に大きな貢献ができると考えられる。

5. 主な研究成果リスト

(1) 代表的な論文(原著論文)発表

研究期間累積件数: 4件

1. Yoshiki Yamamoto, Ryosuke Yamada, Takuya Matsumoto, Hiroyasu Ogino, Construction of a machine-learning model to predict the optimal gene expression level for efficient production of D-lactic acid in yeast, World Journal of Microbiology and Biotechnology, 2023, 39, 69

本研究では、解糖系の13種類の遺伝子の発現量から、D-乳酸生産性を予測する機械学習モデルの構築を試みた。本研究で構築したモデルの決定係数は0.6628であり、遺伝子発現量から高い精度でD-乳酸生産性を予測することに成功した。また、作製した機械学習モデルを用いることでD-乳酸生産に適した13種類の遺伝子の発現量を同定することに成功した。本成果は、解糖系を経由して生産される様々な有用物質の生産に応用可能であり、今後の発展が期待される。

2. Shun Shimazaki, Ryosuke Yamada, Yoshiki Yamamoto, Takuya Matsumoto, Hiroyasu Ogino, Building a machine-learning model to predict optimal mevalonate pathway gene expression levels for efficient production of a carotenoid in yeast, Biotechnology Journal, 2024, 19, 2300285

本研究では、メバロン酸経路上の8種類の遺伝子の発現量から、 β -カロテン生産性を予測する機械学習モデルの構築を試みた。本研究で構築したモデルの決定係数は0.6292であり、遺伝子の発現量から高い精度で β -カロテン生産性を予測することに成功した。また、作製した機械学習モデルを用いることで β -カロテン生産に適した遺伝子の発現量を同定することに成功し

た。本成果は、メバロン酸経路を経由して生産されるテルペノイド類など、他の様々な有用物質の生産に応用可能であると考えられる。

(2) 特許出願

研究期間全出願件数： 0 件

(3) その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

1. 【招待講演】山田亮祐, 酵母によるものづくりとゲノムデザイン, 第 14 回 CHEMBIO ハイブリッドレクチャー (2021/09/28)
2. 【招待講演】山田亮祐, 酵母による有用物質生産を志向した遺伝子発現最適化技術の開発, JACI ライフサイエンス技術部会反応分科会講演会 (2023/07/26)
3. 【受賞】山田亮祐, 南部陽一郎記念若手奨励賞, 大阪公立大学 (2023/07/31)
4. 【学会発表】島崎舜, 山田亮祐, 山本祥輝, 松本拓也, 荻野博康, カロテノイド高生産酵母の作製を目指した機械学習モデルの構築, 化学工学会第 54 回秋季大会 (2023/09/11)
5. 【招待講演】Ryosuke Yamada, Machine-Learning-Assisted Gene Expression Optimization for Production of Useful Chemicals, International Symposium of the Graduate School of Engineering, Osaka Metropolitan University (2023/09/20)