

令和 5 年 6 月 22 日現在

機関番号：10106

研究種目：挑戦的研究（萌芽）

研究期間：2021～2022

課題番号：21K19058

研究課題名（和文）培地成分分析×人工知能によるスマート微生物ハンティングへの挑戦

研究課題名（英文）Smart hunting for novel microbes by medium components analysis and artificial intelligence

研究代表者

小西 正朗（Konishi, Masaaki）

北見工業大学・工学部・教授

研究者番号：90533860

交付決定額（研究期間全体）：（直接経費） 4,900,000円

研究成果の概要（和文）：AI活用基盤技術開発として、31種類の成分を含む完全合成培地を用いて、培地成分が大腸菌の増殖およびタンパク質生産へ与える影響を評価するシステムを構築し、有効性を検証した。ガスクロマトグラフィータンデム質量分析、液体クロマトグラフ質量分析、アミノ酸自動分析、イオンクロマトグラフィー、誘導結合プラズマ質量分析を用いた培地の網羅的定量分析法を確立した。複数の天然培地を分析し酵母エキスロット差を検出可能なシステムであることを確認した。導入した次世代シーケンサーiSeq100システムを用いて、下水汚泥の菌叢解析を実施した。構築したシステムを用いて、培地組成と菌叢の関係の解析を試みている。

研究成果の学術的意義や社会的意義

深層学習モデルおよびベイズ最適化により微生物培地中の31成分の濃度を同時に最適化できる手法は他になくオリジナリティーが高いものである。要素技術のひとつである天然成分の組成プロファイリングでは、5種の異なる分析機器を駆使し、7種の分析手法を組み合わせることで天然物の成分をほぼ網羅的に定量プロファイリングできることを示した。網羅性および定量性においてこれまでに報告がある培地分析手法と比べて優れたものとなっている。天然物のロットによる微生物挙動の変動要因を把握する技術としても発展が見込まれる。これらの手法を複合微生物（菌叢）に適用するという挑戦的な研究に取り組んだ。難培養微生物の挙動解析の高度化に資する。

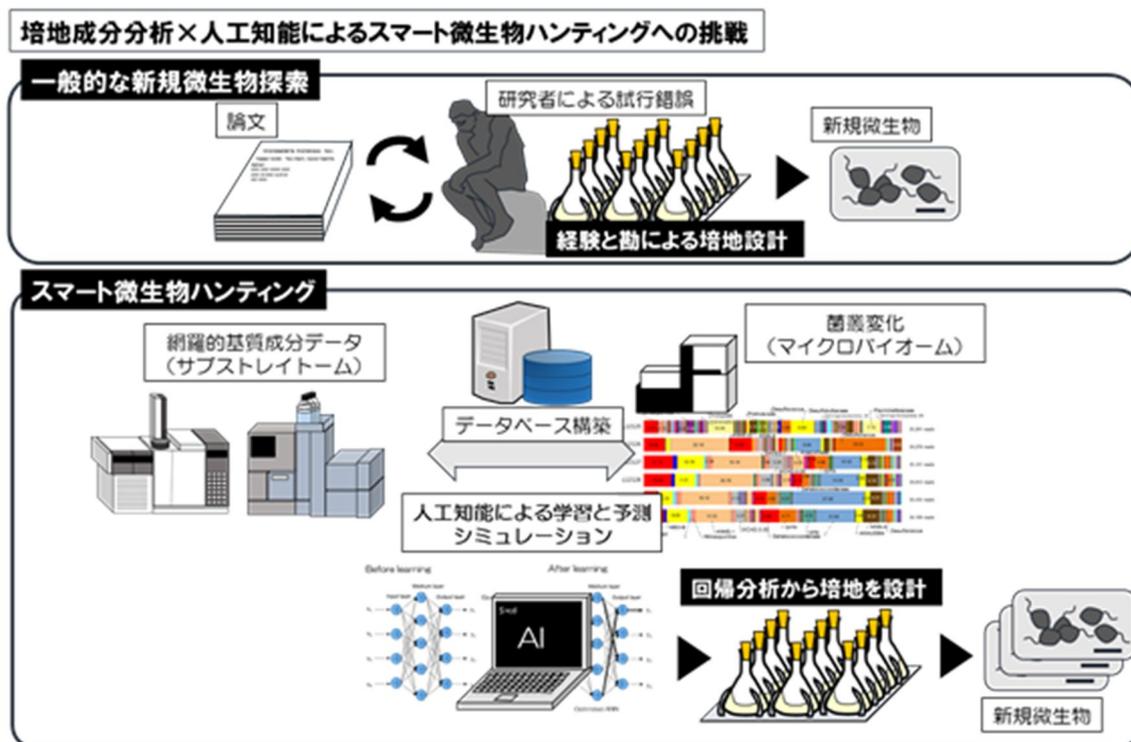
研究成果の概要（英文）：To confirm the availability of AI-based medium analysis, we constructed and validated a system for evaluating the effects of culture medium components on the growth and protein production of *Escherichia coli* using a chemically defined medium containing 31 different components. A comprehensive quantitative analysis method for culture media was established using gas chromatography tandem mass spectrometry, liquid chromatography mass spectrometry, automated amino acid analysis, ion chromatography, and inductively coupled plasma mass spectrometry. The system was confirmed to be capable of analyzing various natural culture media and detecting lot-to-lot differences in yeast extracts. Using the introduced iSeq100 next-generation sequencing system, we conducted a bacterial flora analysis of sewage sludge. Using the constructed system, we are attempting to analyze the relationship between the composition of culture media and the bacterial flora.

研究分野：応用微生物

キーワード：人工知能 深層学習 培地成分分析 次世代シーケンサー

### 1. 研究開始当初の背景

微生物の多くはいまだに分離培養できていないものが99%以上存在することが知られており、新たな微生物を分離し、純粋培養系でその性状を確認することは微生物の理解や応用に欠かすことができない基盤研究と言える。近年 DNA シークエンス技術の発展に伴い、シングルゲノムクスやメタマルチオミクス分析による環境微生物のゲノム構造や代謝機能、微生物相互作用が推定できるようになっている。しかしながら、依然として生理機能や代謝機能を予測数際に指標となる微生物データは単離培養できる微生物のものを指標としている。そのため、DNA を指標とした環境微生物の解釈には大きなバイアスが内包されている可能性がある。成功した微生物の単離培養は、多くの場合、培地や培養方法の工夫により達成されている。個々の微生物の分離条件は、新種記載論文等に記載されているが、体系化された情報がほとんどないのが現状である微生物の系統と生育環境の間には一定の相関があり、環境条件と複合微生物 (microbiome) の挙動解析を体系化できれば、微生物分離培養研究の強力なツールとなる。一方、深層学習 (DNN, Deep Neural Networks) は計算機能力の向上により様々な分野で活用されるようになっている。例えば、巨大地震後の余震の予測、タンパク質相互作用の予測、タンパク質の機能予測、バイオ医薬品の開発など幅広い分野で利活用されている。深層学習を含む機械学習は因果関係のある一定以上の多様性を持つデータセットから判別分析や回帰分析により、特定の因子を抽出することに長けており、データマイニングの手法の一つとして、有効な手段である。したがって、機械学習を用いて、培地組成のデータと複合微生物の微生物叢の関係を明らかにできれば、未培養微生物の挙動を解析することでスマートな微生物探索ができるか概念実証することを発案した。



### 2. 研究の目的

複合培養系における培養条件や培地条件と難培養微生物の挙動の相関を人工知能で予測モデル化し、モデルを用いて重要因子を抽出することで、分離培養手法を設計する新しい手法の実現可能性について検証する。(1) 培地組成から微生物の増殖挙動を予測する技術、(2) 培地組成の網羅的な分析手法、(3) 微生物群衆解析 (4) 培地組成と微生物群衆解析の相関モデルの構築の可能性について検討した。

### 3. 研究の方法

#### (1) 培地組成から微生物の増殖挙動を予測する技術

機械学習を用いて培地組成と増殖挙動の相関モデルを構築することで、培地組成から微生物の増殖挙動が予測できることを確認するため、大腸菌および合成培地を用いてモデルし、タンパク質生産に適切な培地を設計できるか検証した。大腸菌の M9 培地をベースにアミノ酸やビタミン

ンを加えた 31 種類の成分を含む合成培地を設計し、各成分の濃度とプラスミドにて GFP 生産性を付与した大腸菌(BL21(DE3)pLysS/pRSETemGFP)による GFP 生産の関係を深層学習によりモデル化した。ディープウェルスケールの培養で学習データを収集し、蛍光プレートリーダーによる緑色蛍光測定により GFP の発現量を相対的に評価した。学習モデルを用いてベイズ最適化(DNN-BO)により適正培地を探索した。DNN-BO が推定した適性培地の性能を確かめるため、実際に培地を調合し、ディープウェル培養にて、GFP 発現量を評価した(Yoshida et al. 2023)。

## (2) 培地組成の網羅的な分析手法

網羅的な培地成分の分析手法として、ガスクロマトグラフ質量分析計、液体クロマトグラフ質量分析計、イオンクロマトクロマトグラフィー、高周波誘導結合プラズマ(ICP)発光分光分析法(AES)を用いた天然成分の分析手法を検討した。

ガスクロマトグラフィー質量分析は分析する培養基材を 10 g/L になるように水に溶解し、分析用サンプルを調製した。希釈した 1 g/L サンプル, 100  $\mu$ L に 20 mg/mL リピトール 60  $\mu$ L, 抽出溶媒(Methanol:Chloroform:水=5:2:2)900  $\mu$ L を加えて、抽出した。抽出サンプルはトリメチルシリル化による前処理をした。分析には Varian 社製 GC450-300MS システムを使用した。キャリアガスにはヘリウムを使用し、流速は 2.25 mL/min とした。試料の注入量は 1  $\mu$ L、スプリット比は 25:1 とした。オープン温度は、初期温度 60°C で 1 min 保持したのち、10 °C/min で 210°C まで上昇させた。210°C から 5 °C/min で 230°C まで上昇させたのち、15 °C/min で 325°C まで上昇させ、10min 保持し、分析時間は 36.3 min とした。注入口、イオン源、MS トランスファーラインの温度はそれぞれ、250°C、280°C、280°C とした。定量分析対象項目は定量成分の内訳は、アデニン、アデノシン、クエン酸、シトルリン、フルクトース、ガラクトース、マンノース、フマル酸、ガンマアミノ酪酸、グルコース、グリセロール、イソクエン酸、乳酸、オルニチン、リンゴ酸、マルトース、myo-イノシトール、リン酸、ソルビトール、コハク酸、スクロース、トレハロース、トリブトファン、ウラシル、尿素、キサンチンとした。

ガスクロマトグラフィーと同様のサンプル抽出操作により抽出されたサンプルを分析した。標準物質には 20  $\mu$ g/mL Piperazine-1,4-bis(2-ethanesulfonic Acid) Monosodium Salt を使用した。Waters™ e2695 Separation Module と Waters ACQUITY QDa 検出機を使用し、カラムには Shodex HILICpak VG-50 2D (2.0 mm I.D.×150 mm) を使用した。試料の注入量は 5  $\mu$ L とした。溶離液 A には 0.5% アンモニア水溶液を使用し、溶離液 B にはアセトニトリルを使用し、流速は 0.2 mL/min でグラジエントモードにて分析した。グラジエントは、溶離液 B の初期割合を 80% として 2min 保持し、10 min かけて 10% まで下降させた。10% で 3 min 保持し、5 min かけて 80% まで上昇させ、分析時間は 20 min とした。試料のイオン化にはエレクトロイオンスプレー化 (Electrospray ionization) を使用し、プローブ温度は 300°C とした。定量分析対象項目はピオチン、ピリドキシン、パントテン酸、チアミン、リボフラビンとした。

イオンクロマトグラフィーでは 10 g/L になるように溶解したサンプルを 0.22 $\mu$ m のフィルターでろ過後、100 倍から 1000 倍希釈したサンプルを分析した。東ソー イオンクロマトグラフィー・システム IC-2010 を使用し、カチオン分析のカラムには TSKgel IC-Cation I/II HR(4.6 mm I.D.×10 cm)、アニオン分析には TSKgel SuperIC-AZ(4.6 mm I.D.×15 cm)を使用した。両分析において、注入量は 30  $\mu$ L、カラム温度は 40°C、流量は 0.8 mL/min とした。使用した溶離液について、カチオン分析では 0.2 M 硝酸を使用し、アニオン分析では 1.9 mM 炭酸水素ナトリウムと 3.2 mM 炭酸ナトリウムを使用した。またアニオン分析ではサプレッサーを用いた。定量分析対象項目は塩化物イオン、亜硝酸イオン、硝酸イオン、リン酸イオン、硫酸イオン、ナトリウムイオン、アンモニウムイオン、カリウムイオン、マグネシウムイオン、カルシウムイオンとした。

アミノ酸量の分析では、10 g/L サンプル 10  $\mu$ L をガラス試料管 S に入れ、デシケーターで乾固した。乾固したサンプルに低沸点塩酸 (20% w/v HCl) 200  $\mu$ L 加え、減圧封管した。サンプルを乾熱機で 110°C、20 時間反応させたのち、遠心エバポレーターで乾固した。乾固したサンプルに超純水 200  $\mu$ L 加えて溶解し、分析に供した。SHIMADZU Prominence UFLC と Shimadzu RF-10AXL Fluorescence Detector を使用してポストカラム法にて分析した。カラムには Shim-pack Amino-Na (6.0 mm I.D. × 100 mm) を使用し、アンモニアトラップカラムには Shim-pack ISC-30/S0504 Ns (4.0 mm I.D. × 50 mm) を使用した。注入量は 10  $\mu$ L とし、カラム温度および試料の反応温度は 60°C とした。移動相には Amino Acid Mobile Phase Kit (Na Type) を使用し、流速は 0.5 mL/min とした。反応液 Amino Acid Reaction Reagent Kit (Solutions A and B) を使用し、流速は 0.2 mL/min で使用した。蛍光検出機の励起光は 350 nm、蛍光は 450 nm とした。定量分析対象項目はアスパラギン酸、スレオニン、セリン、グルタミン酸、プロリン、グリシン、アラニン、バリン、メチオニン、イソロイシン、ロイシン、チロシン、フェニルアラニン、ヒスチジン、リジン、アルギニン、システインとした。

誘導結合プラズマ発光分析では、10 g/L になるように溶解したサンプルを 0.1 M HNO<sub>3</sub> で 1 g/L、0.1 g/L に希釈し、分析に供した。シーケンシャル型 ICP 発光分光分析装置 SPS3100HV UV にて分析した。分析対象項目はアルミニウム(396.152)、ホウ素(249.773)、カルシウム(393.366)、カドミウム(214.506)、クロム(205.618)、銅(324.754)、鉄(238.277)、マンガン(257.610)、亜鉛(213.924)、コバルト(228.687)、ニッケル(221.716)、鉛(220.422)とした。括弧内は分光器の検出波長を示す。

### (3) 微生物群衆解析

微生物群衆を網羅的に解析するため、iSeq100 システムを導入し、16S リボゾーマル RNA の DNA シーケンスの部分配列(V3-V4 領域)をターゲットとしたアンプリコンシーケンスを実施した。微生物細胞を遠心分離で回収し、ISOIL for Beads Beating kit を用いてゲノム DNA を回収し、イルミナシーケンス用アダプター配列を付加したプライマー(16S amplicon PCR forward primer: 5'- TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG - CCTACGGGNGGCWGCAG -3', 16S amplicon PCR reverse primer: 5'- GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG - GACTACHVGGGTATCTAATCC -3')を用いて、PCR 増幅した。増幅産物を 16S Metagenomic Sequencing Library Preparation マニュアルに従って、インデックス付加、ライブラリーを調製し、iSeq100 シークエンサーにより、1 検体あたり 200,000 リードのシーケンスを得た。得られたシーケンスデータは CutAapt を用いてアダプター配列を除去後、FastQC にてシーケンスクオリティーを確認した。Qiime2 による DADA2 アルゴリズムを用いて、denoising し、Silva 132 99%OTUs from 515F・806R region of sequences を用いて系統分類分析を行った。

## 4. 研究成果

### (1) 培地組成から微生物の増殖挙動を予測する技術

学習データを3水準81条件となるように直交計画(OA)により学習データ用の培地を設計した。ディープウェル培養による評価では、各培地の GFP の蛍光強度は  $8.0 \times 10^3 \sim 2.7 \times 10^4$  の範囲で観測された。培地組成が GFP 発現量に影響することが明らかになったため、データを DNN の学習に用いた。学習用データおよび評価用データに分割し、予測精度を検証したところ、平方和誤差が小さく、決定係数が 0.95 以上の高い予測精度の DNN モデルを構築できた。そこで学習済み DNN モデルを予測関数とした B0 により、GFP 蛍光強度を最大化するように B0 を実施した。B0 により提案された培地を実際に調合し GFP 蛍光強度を確認したところ、学習データより優れた培地組成が見出された。DNN による最適培地の GFP 強度の予測値と実際の GFP 蛍光強度が一致しなかった。そこで、検証で得られたデータを追加して DNN を再学習し、再度 B0 により適正な培地組成を提案させたところ、当初の学習データの 1.4 倍の GFP 蛍光強度が得られる培地組成を見出すことができた。以上により、培地組成と微生物の挙動をモデル化し、培地組成を最適化できることを示すことができた(Yoshida et al. 2023)。この結果は微生物群衆中の特定の微生物の挙動を培地により制御できる可能性を示唆する結果である。

### (2) 培地組成の網羅的な分析手法

つぎに微生物の培養に用いる天然物の組成を正確に把握することを目的とし、市販の酵母エキスやペプトンの成分データベースの構築に取り組んだ。酵母エキスおよびペプトン中の成分を各種機器分析で定量分析し、全体の 95%程度の組成を明らかにすることができた。ある酵母エキスの組成は遊離アミノ酸 42.5%、ペプチド性アミノ酸 15.4%、糖および糖アルコール 21.1%、無機塩類 7.3%、有機酸 2.8%、核酸 1.6%、不明 60%であった。酵母エキス 15 種類、ペプトン類 23 種類の組成データベースを構築した。

### (3) 微生物群衆解析

微生物解析手法を導入する必要があったため、北見市浄化センターから提供していただいた活性汚泥をモデルにアンプリコンシーケンスによる微生物群衆解析を実施した。サンプリングした時期により門レベルでの菌叢の変動を捉えることができた(図 1)。以上により培養条件に依存した菌叢の変動を捉えることが可能となった。

### (4) 培地組成と微生物群衆解析の相関モデルの構築の可能性

上述の(1)~(3)の要素技術を組合せることで、天然物を含む培地組成と複合微生物中の未培養微生物の挙動の相関モデルを構築し解析できる可能性を示すことができた。引き続き、具体的な未培養微生物の挙動解析を進めている。

#### <参考文献>

Yoshida K., Watanabe K., Chiou T-Y, Konishi M.\*: Escherichia coli protein expression using deep learning and Bayesian optimization. *J. Biosci. Bioeng.* **135**, 127-133 (2023)

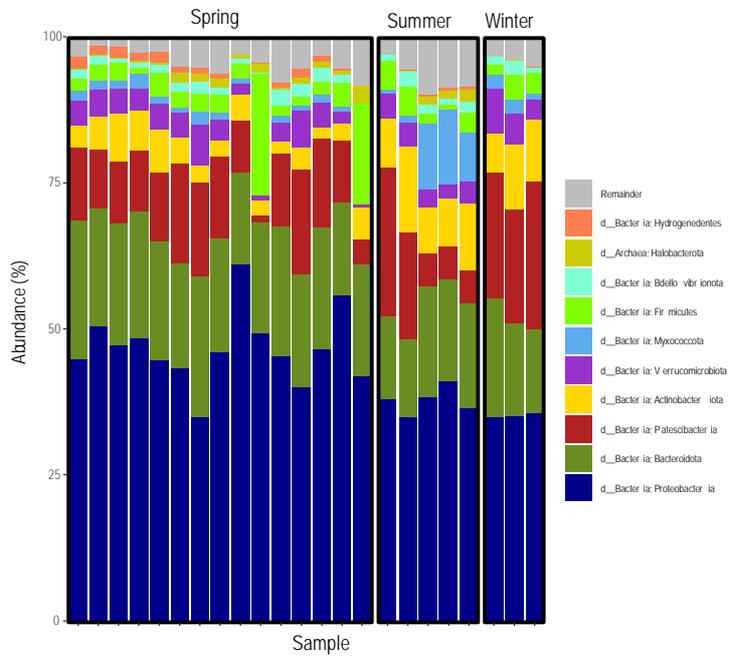


図 1. アンプリコンシーケンスによる菌叢解析結果

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件 / うち国際共著 0件 / うちオープンアクセス 0件）

1. 著者名 Yoshida K., Watanabe K., Chiou T-Y, Konishi M.	4. 巻 135
2. 論文標題 Escherichia coli protein expression using deep learning and Bayesian optimization	5. 発行年 2023年
3. 雑誌名 J. Biosci. Bioeng.	6. 最初と最後の頁 123-127
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.jbiosc.2022.12.004	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計4件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 吉田果菜子, 邱泰瑛, 小西正朗
2. 発表標題 深層学習-ベイズ最適化による効率的な培地探索
3. 学会等名 第31回化学工学・粉体工学研究発表会（化学工学会・北海道支部）
4. 発表年 2021年

1. 発表者名 中島 拓都, 邱 泰瑛, 小西 正朗
2. 発表標題 酵母エキスおよびペプトンの定量データを用いた成分プロファイリング
3. 学会等名 日本生物工学会第74回年会
4. 発表年 2022年

1. 発表者名 中島拓都, 邱泰瑛, 小西正朗
2. 発表標題 酵母エキス及びペプトンの網羅的プロファイリング
3. 学会等名 化学工学会 第87年会
4. 発表年 2022年

1. 発表者名 吉田果菜子, 邱泰瑛, 小西正朗
2. 発表標題 機械学習とハイスループットな評価系を組み合わせたシミュレーションによる培地の設計と最適化
3. 学会等名 化学工学会-第52回秋季大会
4. 発表年 2021年

〔図書〕 計1件

1. 著者名 小西 正朗	4. 発行年 2023年
2. 出版社 シーエムシー出版	5. 総ページ数 5
3. 書名 第9章 機械学習を活用した培地分析および探索の未培養生物への適用可能性 in 未培養微生物研究の最新動向	

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究協力者	邱 泰瑛  (Chiou Tai-Yang)  (60777978)	北見工業大学・工学部・准教授   (10106)	
研究協力者	吉田 果菜子  (Yoshida Kanako)	北見工業大学・工学研究科・大学院生   (10106)	
研究協力者	中島 拓都  (Nakajima Takuto)	北見工業大学・工学研究科・大学院生   (10106)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------