

ゲノムスケールの DNA 設計・合成による細胞制御技術の創出
2020 年度採択研究代表者

2021 年度 年次報告書

榊原 康文

慶應義塾大学工学部
教授

深層学習を用いたゲノムスタイル特徴抽出と DNA 配列 de novo 設計と合成

§ 1. 研究成果の概要

深層学習を用いて人工 DNA 配列を設計する DNA 配列変換器の汎化性能の向上を目的として、第一バージョンを開発した。DNA 配列変換器の過学習を回避するために事前学習方法を実装し、RefSeq から取得した 23,037 本の CDS 配列を用いて事前学習を行った。それにより、DNA 配列の文脈依存的な埋め込み表現を獲得した。次に、菌種間のオーソログ遺伝子配列を変換する深層学習モデルを構築した。他菌種間のオーソログ配列変換を補助情報として利用することにより、学習に用いていない未知の DNA 配列の変換に関して精度の向上が得られた。

PET 分解から TCA サイクルまでの代謝経路を人工的に枯草菌内に構築することを目的として、代謝経路上の全 14 酵素の DNA 配列の設計を行った。これらの DNA を合成して、枯草菌のプラスミドに集積し、遺伝子発現までを確認した。

ゲノムスタイルという新しいゲノム特徴量を用いたメタゲノム配列ビニング手法の開発を完了し、ベンチマークデータセット CAMI チャレンジを用いて既存の代表的ビニングツールより高精度であることを示した。その結果を国際学術誌に掲載し、プログラムを github 上で公開して、広く世界中の研究者に使用可能とした。

枯草菌を用いた DNA 合成プラットフォームを構築した。取得された DNA を接合伝達で放線菌に速やかに再導入できる新たな枯草菌システムを構築するために、枯草菌、放線菌、大腸菌で複製できる 3 者間プラスミドベクターを構築し、その安定性を調べた。

PETase (PET 分解酵素) の発現を確認するための抗体の作成を行った。PETase 遺伝子の発現用プラスミドが導入された大腸菌と枯草菌の活性の確認を行ったが、PETase のタンパク質ならびに PET 分解活性を確認することができなかった。発現系がうまく動かない原因について調査を行い、その結果を他グループにフィードバックした。

§ 2. 研究実施体制

(1) 榊原グループ

① 研究代表者: 榊原 康文 (慶應義塾大学理工学部 教授)

② 研究項目

- DNA 配列変換の深層学習アルゴリズムの開発と DNA 配列設計.
- 挿入欠失のある DNA 配列変換と機能分析器を加えた学習への拡張.
- ゲノムスタイルの分子生物学的解釈とメタゲノムビニングツールの開発.

(2) 片岡グループ

① 主たる共同研究者: 片岡 正和 (信州大学工学部 准教授)

② 研究項目

- 長鎖 DNA プラスミド全合成手法の開発.
- 放線菌プラスミドの再設計と全合成.
- 多様な GC 含量放線菌プラスミドの枯草菌での複製, 安定性, 遺伝子発現の解析.

(3) 宮本グループ

① 主たる共同研究者: 宮本 憲二 (慶應義塾大学理工学部 教授)

② 研究項目

- PET 分解酵素群の遺伝子配列再設計と機能検証.
- 人工 PET 分解菌メガプラスミドの設計と構築.

【代表的な原著論文情報】

- 1) Yoshimura Y, Hamada A, Augey Y, Akiyama M, Sakakibara Y. “Genomic style: yet another deep-learning approach to characterize bacterial genome sequences”, *Bioinformatics Advances* 1(1), vbab039 (2021).
- 2) Akiyama M and Sakakibara Y. “Informative RNA base embedding for RNA structural alignment and clustering by deep representation learning”, *NAR Genom Bioinform*, 4(1), lqac012 (2022).