



自然科学系 教授
阿部 貴志 ABE Takashi



http://bioinfo.ie.niigata-u.ac.jp

ライフサイエンス分野への機械学習の活用 ～ ゲノムビッグデータからの効率的な知識発見手法の開発 ～

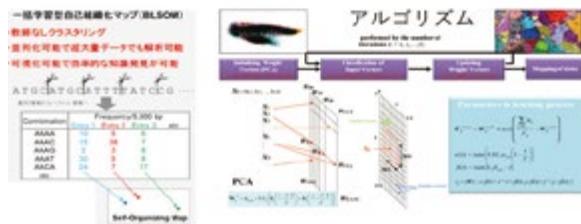
●キーワード● 一括学習型自己組織化マップ、連続塩基組成、微生物生態、メタゲノム、生物系統推定

研究の目的、概要、期待される効果

ゲノム解読技術の飛躍的進歩により、ゲノム配列データは指数関数的に増加しています。その中で最も容量を占めているのは微生物ゲノムです。新規遺伝子を豊富に保有すると考えられ、産業的・医学的にも注目を集めています。しかし、自然環境で生息する微生物の99%以上は難培養性で通常の実験的研究が行えず、またその新規性の高さから、配列相同性検索といった従来の情報学的手法を用いても、生物系統や遺伝子機能を推定する事はほぼ不可能です。

我々は、ゲノム配列自体はATGCの4塩基から成る単純な文字列であることから、連続塩基出現頻度を変数として、超大量ゲノム配列から生物種固有の配列特徴を俯瞰的に把握可能とする教師なし機械学習である一括学習型自己組織化マップ(Batch-Learning Self-Organizing Map、BLSOM)を開発しました。本手法は、大量ゲノム情報の特徴を網羅的、かつ、俯瞰的に可視化可能で、視覚的にも理解し易く把握できます。また、メタゲノム解析由来の各配列の生物系統や新規性を推定するための手法を開発し、より多くの研究者に利用できるソフトウェアを公開しています。

世界に先駆けて開発した技術を用いて、医学や産業的に有用な新規微生物や有用遺伝子を探索するための基盤情報の構築・提供を目指しています。



全既知微生物を対象にした断片化サイズ3kb、縮退4連続頻度でのBLSOM解析結果
19,341,836件、136次元データを対象に、地球シミュレータ (2048コア) 使用

BLSOMのアルゴリズム(上)とBLSOM解析結果の一例(下)



メタゲノム配列に対する系統推定ソフトウェアPEEMSの概要

関連する知的財産
論文 等

塩基配列の分類システムおよびオリゴヌクレオチド出現頻度の解析システム(特開2005-092786)
Abe et al., *Frontiers in Microbiology*, 11: 1486, 2020.
Okunaga et al., *npj Clean Water*, 5: 65, 2022.

アピールポイント

機械学習による、高精度なクラスタリングとその解析結果の俯瞰的可視化が可能。

ゲノム解析は、ウイルス・細菌・原虫・植物など幅広い生物種に対する実績を有します。

つながりたい分野(産業界、自治体等)

- ・ゲノム情報の活用に興味がある企業
- ・環境分析会社(水質、土壌、下水、etc)
- ・腸内細菌叢に興味のある食品会社や製薬会社
- ・新規微生物の活用を目指すバイオ系会社 等