新潟大学
ビッグデータアクティベーション研究センター

ライフサイエンス分野への機械学習の応用

【キーワード】

自己組織化マップ	一括学習型	塩基組成比較	メタゲノム	生物系統推定

■概要

大規模ゲノム情報解析のための高速データマイニングシステム

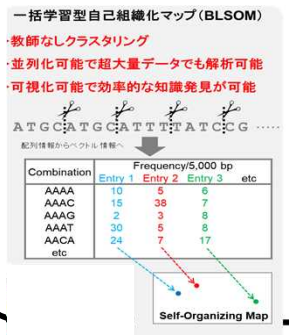
地球環境の保全や修復において，微生物を含む生物生態系の役割が大きい．ゲノム解読技術の発展は，「メタゲノム解析」と呼ばれる新実験分野を生み，全地球レベルでの生物生態系の把握を目標にした大規模解析が行われている．

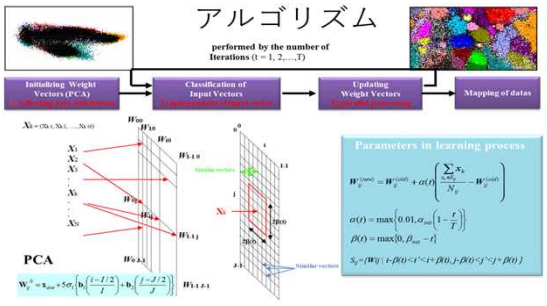
我々が開発した一括学習型自己組織化マップ（Batch-Learning Self-Organizing Map, BLSOM）は，大量ゲノム情報からの知識発見において，当初予想を遥かに超える優れた能力を持つことが明らかになった．大量ゲノム情報の特徴を網羅的，かつ，俯瞰的に可視化可能で，視覚的にも理解し易く把握できる．大規模なBLSOM解析結果を用いて，メタゲノム解析由来の各配列の生物系統や新規性を推定するための手法を開発し，より多くの研究者に利用できるソフトウェアを公開している．

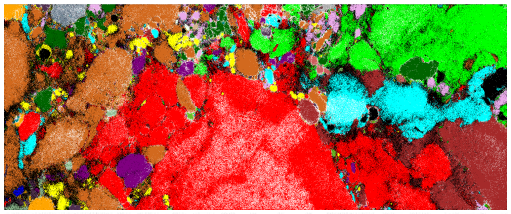
世界に先駆けて開発した技術を用いて，「全地球レベルでの微生物群集の多様性」を俯瞰し，そこから効率的に医学や産業的に有用な新規微生物や有用遺伝子を探索するための基盤情報の構築・提供を目指している．

■詳細

一括学習型自己組織化マップ（BLSOM）



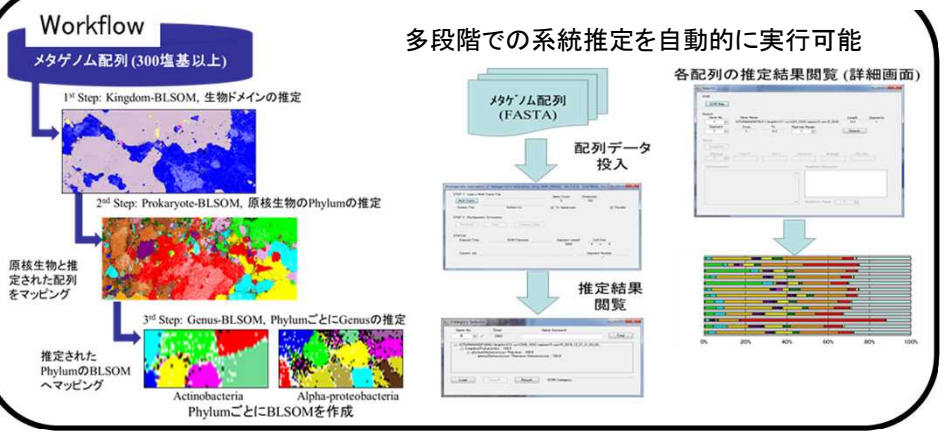




全既知微生物を対象にした断片化サイズ3kb，縮退4連続頻度でのBLSOM解析結果

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

大規模BLSOM解析結果の活用メタゲノム配列に対する系統推定ソフトウェアPEMS



○競合研究に対する優位性
- 配列相同性とは異なるアプローチのため、高感度，ならびに，高精度
- 配列情報のみで，推定可能
- 段階的な予測により，新規性の高い微生物種の系統も検出可能

○想定される実施例，応用例
- 環境中の微生物叢の理解
- 新規有用微生物や遺伝子の探索
- 環境中の有害微生物検出システム

○今後の課題，展望
- 真核生物やウイルス由来メタゲノム配列の検出
- 水平伝播候補遺伝子の検出
- 超大規模データ解析手法の確立
- 様々なビックデータからの効率的なデータマイニングシステムへの適応

■連携提案先（産業界・行政等）

➢ 環境分析会社（水質，土壌 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しているバイオ系会社

本技術の問い合わせ先　新潟大学　地域創生推進機構

TEL：025-262-7554　FAX：025-262-7513　E-mail：onestop@adm.niigata-u.ac.jp