

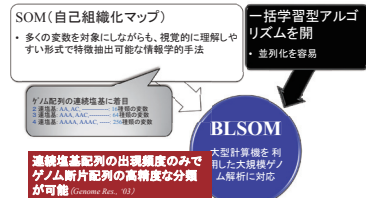
地球環境変動・保全に係わる全地球レベルでの微生物群集構造把握のためのゲノム情報基盤整備

阿部貴志¹, 和田健之介², 池村淑道²

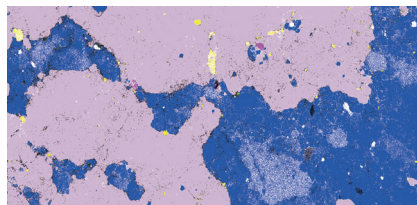
¹, 新潟大学, ², 長浜バイオ大学

概要 地球環境の変動が生物生態系に多大な影響を与えと共に、生物生態系側も地球環境へ多大な影響を与えて来た。海洋に生息する植物プランクトン・光合成微生物類の様に、炭素循環に重要な役割を果たしているが、肉眼的な観察が困難な場合、関与している微生物類の生態系の実態、特に地球レベルでの生態については未知に残されて来た。更に重要な点は、光合成微生物類を例に取っても、地球環境への影響を知るには、個々の微生物種ではなく、微生物群集構造の把握が必須である。しかしながら、自然環境の微生物類の99%以上が実験室では培養が困難であり、関与する微生物類の特定を困難にして来た。最近のDNA配列の読取技術の飛躍的な高速化は、「メタゲノム解析法」と呼ばれる革新的な手法を確立させた。海洋を代表例として全地球レベルでの大規模なメタゲノム解析が進行し、超大量なゲノム配列が集積している。我々が開発して来た、一括学習型自己組織化マップ法(BLSOM)は、断片ゲノム配列を生物系統ごとに高精度に分離(自己組織化)する能力を持つ。解読された全ゲノム配列を対象にESで大規模BLSOMを作成・更新して行けば、メタゲノム解析で得られる大量配列をマップすることで、各環境中で生息する生物群集の全体像の把握が可能になり、加えて、環境保全・浄化に役立つ新規微生物類やそれらの保持する有用遺伝子類を発掘できる。

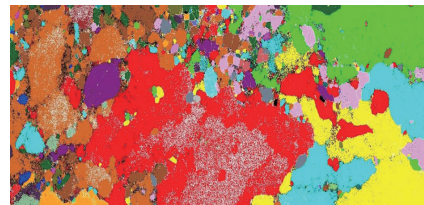
一括学習型SOM (BLSOM)について



高等動植物ゲノムを含む全ゲノム配列を対象にしたBLSOM



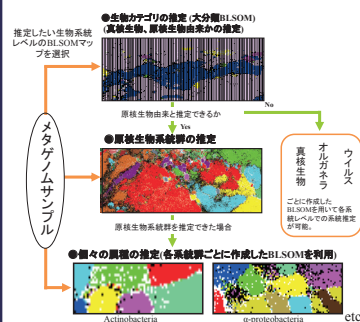
既知原核生物3157種を対象にした縮退4連続BLSOM



メタゲノム配列に対する系統推定ソフトウェア～PEMS (Phylogenetic Estimation of Metagenomic sequece using BLSOM)～

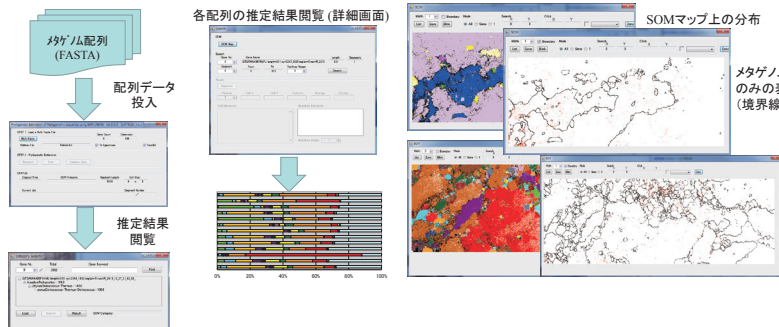
●BLSOMによる系統推定(原核生物)の流れ

既知生物由来配列を対象にしたBLSOMを用いて、各系統での推定を実施することによってメタゲノム配列の系統の絞り込みが可能



公開サイト: http://bioinfo.ie.niigata-u.ac.jp/?PEMS_Soft

Kingdom ⇒ Phylum ⇒ Genusと多段階での系統推定を自動的に実行



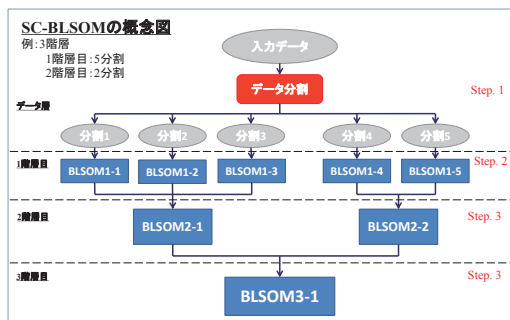
自己圧縮BLSOM (SC-BLSOM)

SC-BLSOM: BLSOMを階層的に構築することで高速化する手法

BLSOMは学習過程で入力データの特徴を、より数の少ないリファレンスベクトルへと集約している。SC-BLSOMでは入力データをデータの属する分類に沿って分割しBLSOMを行い、分割ごとの連続塩基組成の特徴をリファレンスベクトルへ抽出する。そのリファレンスベクトルを元に再度BLSOMを行う。BLSOMは入力データ数の3乗オーダーで計算量が増加するが、データの分割とリファレンスベクトルへの圧縮で各BLSOMの入力データが減少するため、トータルでの計算時間は通常のBLSOMよりも短くなる。

SC-BLSOMの作成手順

- ① 入力データの分割
- ② 分割したデータごとにBLSOM解析
- ③ 解析結果(リファレンスベクトル)を入力データとしてBLSOM解析(繰り返し)



基本性能評価

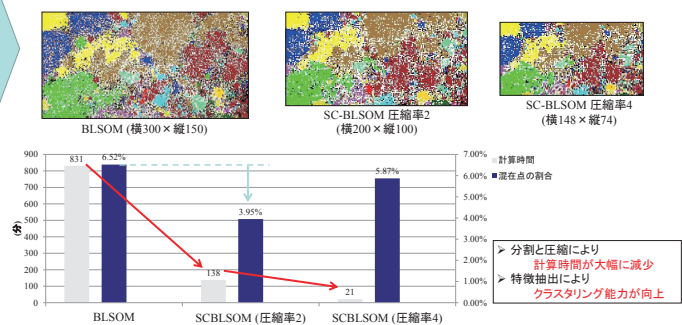
データセット

配列データ: 原核生物完全ゲノム817種よりランダムに10kb抜き出し元の配列の1/10になるまで併合した配列断片化サイズ: 5kb, 連続塩基: 縮退4連続塩基頻度 (454,990,000 → 90998個の136次元ベクトルデータ)

実験内容

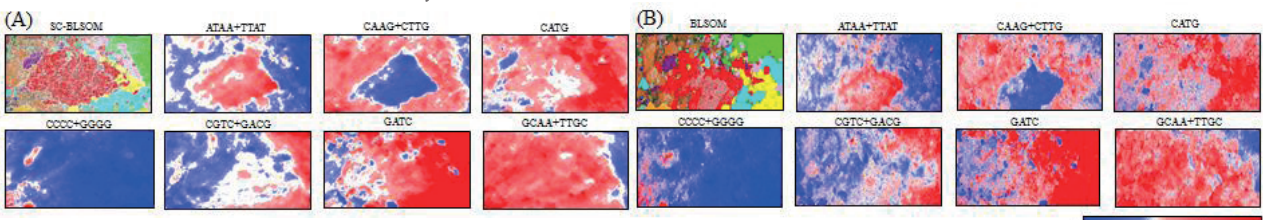
- 上記データでBLSOMとSC-BLSOMを作成して計算時間とクラスティング能力^{a)}を測定した。SC-BLSOMは圧縮率^{b)}の場合と圧縮率4の場合の2通りを作成した。
- クラスティング能力として、データが分離できず複数種のデータが混在している格子点の割合を調べた
 - 圧縮率=1階層目のBLSOMの1格子点あたりのベクトル数

結果



比較ゲノム解析への応用

原核生物3,157種によるSC-BLSOM(A)と従来のBLSOM(B)との頻度マップ両方共、各Phylumで同様の特徴的な縮退4連続塩基が検出されている。



より大規模なゲノム配列情報を対象にしたゲノム情報基盤の提供が可能