3 次元仮想ゲノム空間の CAVE による可視化

佐々木 直文, 佐藤 直樹

東京大学大学院 総合文化研究科 広域科学専攻 生命環境科学系

1. 目的

生物情報科学（Bioinformatics）は生物由来の情報を計算機科学的な手法で処理することにより生物学的な知見を発見することを目的とした分野である。その中でもゲノム情報は生物のセントラルドグマの源泉となる情報であり、ヒトゲノムをはじめとして、多くの生物の塩基配列が解読されている。近年のゲノムデータの拡充により、生物をゲノムによって比較する試みが盛んになりつつあるが、遺伝子の分子進化に関しては既に確立された理論が有るのに対し、ゲノムの構造進化については未知であることが多い上に解析技術もなお発展の余地がある。これまでの我々の研究では、類縁の生物群のゲノムを使用し、これらのゲノム間で共通した構造を可視化する手法を開発した。本発表では、我々の開発した手法によって得られる仮想ゲノム空間を CAVE によって３次元的に表示し、これまでの２次元上では把握することが難しかった、仮想ゲノムの詳細な構造に関する知見を明らかにすることを目指す。

2. 方法

2.1 位置プロファイル法について

ゲノムの構造的進化の解析には遺伝子の並び方の共通部分（シンテニー）を検出することによって見つけることができる。シンテニーを探索するための計算機科学的な手法の多くは、遺伝子を頂点、それらの隣接関係を辺と見なすことで、シンテニー探索をグラフ構造の探索問題に帰着させている。しかし、このアプローチは近傍の情報のみが考慮されるため、互いに離れた位置にある遺伝子間の距離関係まで考慮することができない。我々はこの点に着目し、近傍だけでなくゲノム上で離れた位置も含め、遺伝子が等距離にある関係である'isoapostatic'な関係を検出することを目的として、新規手法である「位置プロファイル法」を開発した。

2.2 位置プロファイル法

本手法では、まず比較に使用するゲノム間に共通な遺伝子の集合である、ホモロジークラスターのデータが必要であり、これを本研究室で開発した GClust データベース(http://gclust.c.u-tokyo.ac.jp)から取得した。次にこのデータとゲノム上の遺伝子の位置情報からゲノム上に置ける遺伝子の距離関係を求め、遺伝子の位置関係のプロファイルを生成した。これらのプロファイル間の分散をとることによって'isoapostatic'な距離関係を求め、その情報を基にクラスタリングを行った。これらの処理によって得られたグループを'Virtual Linkage Group(VLG)'と定義した。また、位置プロファイルの可視化のため、多次元尺度法(Multidimensional Scaling)によって遺伝子の距離情報を 2 次元空間に縮約し、VLG の特徴空間上の位置を可視化した。

3. 結果

3.1 海洋性シアノバクテリアゲノムのモザイク構造

海洋性シアノバクテリアは酸素発生型の光合成を行うバクテリアであり、分子系統樹による系統関係から、*Prochlorococcus* 属と *Synechococcus* 属に分けられる。本解析手法をシアノバクテリアに適用した結果、8 個の VLG を得た。さらに VLG のゲノム上における位置関係を比較した結果、近縁のゲノム間では VLG のモザイクパターンがほとんど同じであり、進化的距離が離れるに従って、ゲノムの再編成が起きていることが分かった。(図.1) このことはゲノム構造と生息環境との間の相関を示唆している。

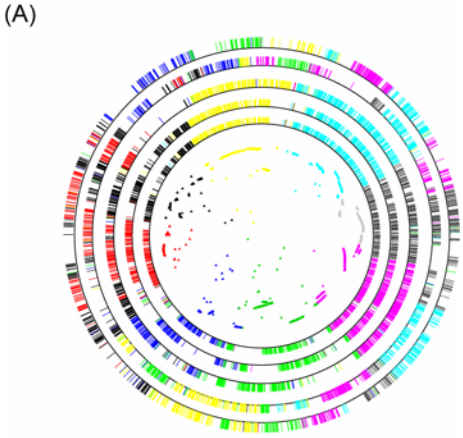

図.1: VLG のゲノム上での位置関係

3.2 位置プロファイル法の評価

MDS による次元縮約によって得られた遺伝子の 2 次元布置は、比較に使用した全ゲノム間で共通的である様な位置関係を示していることから、我々はこの 2 次元布置を「仮想ゲノム」と呼ぶことにした(図.2)。仮想ゲノム上に置ける VLG の位置は個々のクラスターにほぼ対応しており、クラスタリングと MDS による埋め込みがうまくいっていることを示している。次元縮約の効果の検証のため、埋め込む次元を変えて実験を行った結果、2~3 次元への縮約が効果的であることが明らかになった。また今回使用した K-means 法の改良法が今回使用したデータに適しているかどうかを検証した結果、距離行列から直接クラスタリングを行った場合と MDS によって次元縮約を行った上でクラスタリングした結果はほぼ同じ傾向を示した（表.1)。このことは、位置プロファイル法において、MDS による次元縮約が効果的であることを示唆している。

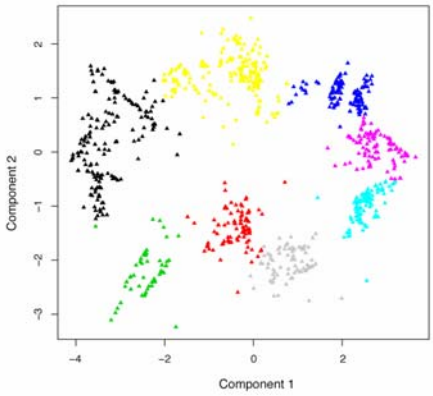

図.2: 仮想ゲノム上での VLG の位置関係

表.1: MDS による次元縮約の効果

Dimensions	Anabaena	Marine
1	43.5141	38.83796
2	23.95577	16.633578
3	8.445137	11.79412
4	5.000024	8.513441
5	5.017336	8.432034

4. 結論

本研究で開発された新規手法である「位置プロファイル法」の海洋性シアノバクテリアへの適用で得られた VLG のパターン領域の可視化により、これまで見過ごされてきた、ゲノムの骨格領域の再編成の様子が明らかになった。しかしながら本手法の拡張を考えたとき、クラスタリングの手法等についてはまだ改善の余地が大きい。本発表での CAVE における 3 次元空間での仮想ゲノムの可視化によって、仮想ゲノム空間の構造等について新たな知見が得られるものと期待している。