

## 論文の内容の要旨

論文題目: 位置プロファイル法に基づくゲノム構造進化の解析  
(Analysis of genomic structure evolution using positional profile method, )

氏名: 佐々木 直文

### はじめに

ある一つの共通祖先から系統分岐した近縁の 2 つの生物種のゲノムは、遺伝子のセットも互いに類似していると考えられる。これらの互いに共通祖先からの派生関係にある遺伝子はオルソログと呼ばれている。これらオルソログのそれぞれのゲノム上における位置の対応関係を比較することで、ゲノム間の構造的進化を見ることができる。近年のシーケンス技術の発展により解読済みのゲノム数は急激に増加しており、それに伴い、ゲノム比較解析によるゲノムの構造的進化の様子が明らかになりつつある。しかし、遺伝子の配列進化に関しては既に確立された理論があるのに対し、ゲノムの構造的進化については未知であることが多い上に解析技術も発展の余地がある。本研究において、著者は近縁種のゲノムを網羅的に比較することにより、これらのゲノム構造を可視化する手法の開発を行い、ゲノム構造の進化的側面を明らかにすることを目的とした。また、解析の結果を組み込んだデータベースを開発した。

ゲノムの構造的進化はゲノムの一部の領域が別の領域と入れ替わるゲノム再編成 (genomic rearrangement) が深く関係しており、その様子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、遺伝子の隣接関係と見なすことで、シンテニー探索を頂点と辺の集合から構成された「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に重点が置かれているため、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。グラフ探索法にはこのような欠点があるものの、これまでの類縁関係の遠いゲノム間の比較の結果から一般にゲノム間の進化的距離が離れると急速にシンテニーが崩れ、その断片がゲノム上にランダムに分布することが知られていたため、この問題点は重要視されていなかった。

本研究ではこの点に着目し、近傍だけでなくゲノム上で離れた位置も含め、遺伝子が同期的に移動する現象を検出することを目的とした。このため、従来のグラフ探索のアプローチではなく、多変量解析の考え方を応用することで、ゲノム間での遺伝子の位置関係について同期的挙動を示す遺伝子のグループ(Virtual Linkage Group, VLG)を抽出する手法について検討し、新規手法である「位置プロファイル法」を開発した。さらに、開発された手法を海洋性シアノバクテリア 14 種のゲノムに適用した結果、これらのゲノム間に存在する VLG を検出することができた。さらにそれぞれのゲノム上における VLG の位置を見ることにより、ゲノムの構造的進化を可視化することに成功した。

#### 方法

本研究で開発した手法は、比較に使用するゲノム間に共通なオルソログを含んだ、ホモロジークラスターのデータとゲノム上の遺伝子の位置情報を基に、オルソログ間の距離関係のプロファイルを作成し、これを多変量解析を用いて統計的性質から分類する手法である。入力となるデータセットは本研究室において開発された Gclust データベースを利用した。それぞれのゲノムに含まれる全オルソログ遺伝子間の距離の分散を計算して得られた非類似度行列を基にクラスタリングを行ったところ、8 個のクラスターが得られた。多次元尺度法 (Multi Dimensional Scaling, MDS) による可視化の結果、これらのクラスターが比較に使用した全ゲノムを重ね合わせた「仮想ゲノム」における遺伝子の位置関係を反映していることが明らかになったため、このクラスターを「Virtual Linkage Group (VLG)」と呼ぶことにした。本手法を海洋性シアノバクテリアに適用した結果、それぞれのゲノム上で、同じ VLG に属する遺伝子が断続的に隣接していることが分かったため、この隣接関係が連続している範囲を推定する手法を、HMM (隠れマルコフモデル) をベースとして開発した。

#### 海洋性シアノバクテリアのゲノムのモザイク構造

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る。そのなかで海洋性のシアノバクテリアである *Prochlorococcus* 属と *Synechococcus* 属は、光条件や栄養条件といった生息環境の違いによっていくつかの生態型(ecotype)に分類することができ、これらについて研究が行われている。これらの生態型は 16S rRNA と 23S rRNA 間領域の DNA 配列(internal transcribed spacer sequence)を用いた分子系統解析の結果から、互いに近縁であることが知られている。海洋性シアノバクテリアを含め、幾つかの近縁種や変異株などの非常に近縁なゲノムの網羅的比較解析の結果、これらのゲノム間には、同属の種間においても驚くほどの多様性があることが明らかになっている。これらのゲノムにおいては近縁種間で保存されている安定な領域 (ゲノム・コア) と、不安定な領域 (ゲノム・アイランド) の 2 つの性質の異なったゲノム領域が存在していることが知られているが、本研究で開発した位置プロファ

イル法を海洋性シアノバクテリアのゲノムに適用した結果、得られた VLG が特にゲノム・コアの再配置ブロックに対応していることが明らかとなった。さらに、得られた VLG のパターンを分子系統樹に対応させて比較検討したところ、近縁のゲノム間ではシンテニー内の VLG の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、VLG の連続領域を単位としてゲノム再編成が起きていることが分かった。また *Prochlorococcus* 属において、光適応条件が類似したゲノムのシンテニーマップは類似しており、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。

#### ゲノムの構造的な制約に関する解析

本研究において明らかとなった、海洋性シアノバクテリアゲノムの VLG のパターンの観察から、VLG が連続的に存在する領域はゲノムの骨格のようなものであり、付加的遺伝子はこの領域の間隙に寄せ集まるように存在していることが分かったが、これは、他の研究で報告されているように、ゲノム・アイランドが水平移動遺伝子を多く含んでいることとも一致しており、VLG コア領域のパターンによって、ゲノムの構造とその構造の進化について視覚的にもアプローチすることが可能になった。著者は、VLG のパターンの構造をよりはっきりさせるため、VLG が連続した領域を推測する手法を、HMM (隠れマルコフモデル) をベースとして開発した。HMM では内部的な情報として確率構造が必要であり、連続領域推定の最適化は確率構造の最適化に対応している。本手法ではこの確率構造を最適化するために反復 Viterbi 法を用いた。この手法を前述の海洋性シアノバクテリアの VLG パターンに適用した結果、ゲノム全体の VLG のモザイク構造をより明確に可視化することに成功した。

本研究ではさらに、これらの VLG パターンが安定である理由について、ゲノムの構造的側面からアプローチした。ゲノムの一定区間のグアニン-シトシン塩基対におけるコーディング鎖でのグアニンの出現頻度の偏りは GC skew と呼ばれており、バクテリアではこの値の最大値と最小値がそれぞれ染色体複製開始点 (*ori*) と、複製終点 (*ter*) の位置にあたることが多い。アデニン-チミン塩基対においても同様の結果が得られるため、両者はあわせて GC/AT skew と呼ばれている。本解析で得られた VLG 遺伝子と、GC/AT skew から予測された *ori*, *ter* からの距離を測定し、統計を取ったところ、VLG 遺伝子は VLG 毎に *ori*, *ter* から一定の距離に多く分布していることが明らかになった。また、それぞれの VLG に含まれる遺伝子の機能分類を調べたところ、VLG 毎に特徴的な機能遺伝子が含まれていることが明らかになった。

#### VLG 情報の CyanoClust データベースへの取り込み

シアノバクテリアは葉緑体の細胞内共生の起源と考えられている生物であり、葉緑体の進化を考える上で重要な生物である。しかし、葉緑体とシアノバクテリアを繋ぐような比較ゲノムに特化したデータベースは存在していなかった。そこで本研究では光合成生物の比較ゲノム解析のプラットフォームとして、Gclust データベースの派生である、CyanoClust データベースに本研究によって得られた VLG の情報を取り込み、インターネット上のウェブサービスとして公開した。CyanoClust に含まれるクラスターの内、917 個のオルソログに対して VLG 番号を割り当てることができた。本データベースは、他にもオルソログの系統解析機能等が実装されており、ゲノム比較解析における、オルソログの機能アノテーション付加等の解析プラットフォームとして利用できることが期待される。

## まとめ

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 コア領域のパターンの可視化により、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。このことは本手法の有用性を示している。また、VLG 遺伝子の *ori*, *ter* からの距離関係は、これまでランダムであると考えられてきたゲノム再編成が、ゲノム構造の制約を受けている可能性を示唆した。さらに、VLG と遺伝子機能との関係は、細胞の活動とゲノムの構造との関連を示唆しており、これらのことから、ゲノム比較によって明らかとなるゲノム構造とその進化が生物にとって重要な意味を持っていることが示唆された。(3994 字)