4F-05

ミトコンドリアゲノムのグラフ表示と生物分類への応用

水田 智史† 齋藤 達‡ 小田桐 健§
弘前大学大学院理工学研究科† 弘前大学理工学部‡ 弘前大学理工学部§

1 はじめに

DNA の塩基配列やタンパク質のアミノ酸配列などの生物学的配列の類似性の評価は、分子生物学における重要な手続きの一つであり、その方法としては一般的にアライメントが用いられている。しかし、ゲノム配列を対象とした場合においては、時間計算量の問題や大規模なゲノム再編成に対処する困難さなどから、様々なアライメントに依らない配列比較の方法が提案されている。

その一つとして、配列を 2 次元平面あるいは 3 次元空間中のグラフに表示し、その形状を比較するという手法がある [1]。この手法においては、何らかの測度によって定量的に求めたグラフ間の距離により配列間の類似性を数値化するというのが標準的な流れであるが、本研究ではこの手法を異なる観点から利用することを試みる。すなわち、目、科、属、種など、各分類階級毎に抽出したグラフの特徴に基づいた生物分類が可能であるかどうかについて検討を行う。

本稿では、グラフ表示の方法と、魚類および鳥類のミトコンドリアゲノムを対象とした解析結果の一部について述べる。

2 配列のグラフ表示

2.1 開始点の整列

ミトコンドリアゲノムは環状であるため、データベース中の配列の先頭がすべての生物種で統一されているとは限らない。そこで、グラフ表示をするに当たって、魚類、鳥類それぞれにおいて、対象となるゲノム配列の Clustal Omega を用いた多重アライメントを行い、保存の度合いが大きい領域の先頭をグラフ表示の開始点とした（図 1）。なお、一部のゲノムでは DNA の二重鎖の別の側の鎖が配列データとして格納されているものがあったため、それらについては他の生物種と同じ側の鎖のデータを用いて以後の処理を行った。

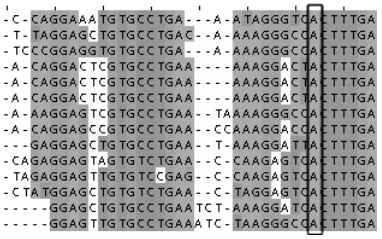

図 1 開始点の整列。保存の度合いが大きい領域の先頭（図の枠で囲った塩基）をグラフ表示の開始点とする。

生物種を新たに加える場合は、多重アライメントを実行した段階で作成しておいたコンセンサス配列との間でペアワイズ・アライメントを実行することにより開始点を決定することが可能である。

2.2 ベクトルの割り当て

配列の各塩基を 2 次元平面上の点に対応付けるため、4 種の各塩基に対して図 2 に示すベクトルを割り当てる。割り当ての仕方は、90 度の回転、および x 軸もしくは y 軸に関する鏡像変換によって重なるものを除けば、全部で 3 通りある。ここでは、その中でグラフの特徴が最も明確に現れた割り当てを採用している。

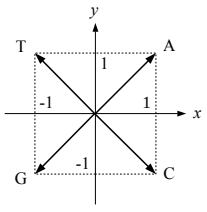

図 2 4 種の塩基へのベクトルの割り当て

Graphical representation of mitochondrial genomes and application to biological classification

† Satoshi Mizuta, Graduate School of Science and Technology, Hirosaki University
‡ Tohru Saitoh, Faculty of Science and Technology, Hirosaki University
§ Ken Odagiri, Faculty of Science and Technology, Hirosaki University

2.3 重み付け

グラフの特徴がより顕著に現れるように、各塩基に割り当てたベクトルに塩基の出現確率に基づいて算出した重み $W(z|xy) = -\log_2 P(z|xy)$ を乗じてグラフ表示をする [2]。ここで x, y, z はそれぞれ塩基 A, T, G, C のいずれかを表し、$P(z|xy)$ は塩基列 xy の次に塩基 z が出現する条件付き確率で、解析対象とする生物種全体のゲノム配列中に出現する塩基列の出現数 N_{xyz} から $P(z|xy) = N_{xyz} / \sum_{s\in\{\mathrm{A,\ T,\ G,\ C}\}} N_{xys}$ によって算出した。

2.4 グラフ表示

グラフ表示は、前述の重みを乗じたベクトルを配列の各塩基の順番に接続していくことにより行う。

例として、重み W(A|AC)、W(T|CA)、W(A|AT)、W(T|TA)、W(G|AT) の値がそれぞれ 2.3, 0.60, 1.3, 1.7, 1.2 であった場合の配列 “ACATATG” のグラフ表示の結果を図 3 に示す。なお、配列の先頭 2 塩基目までは重みが定義されないため、実際のグラフ表示は第 3 塩基から行った。

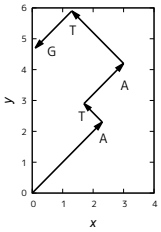

図 3　配列 “ACATATG” のグラフ表示。先頭の 2 塩基は重みが定義されないため、グラフ表示は第 3 塩基から行う。

3 結果

ここでは、魚類、鳥類のグラフ表示の結果の内、それぞれにおいて特徴的なものを選別して示す。

3.1 魚類ミトコンドリアゲノムのグラフ表示

図 4 はスズキ目カワスズメ科の *Maylandia zebra*（メイランディア・ゼブラ）、*Neolamprologus brichardi*（ネオランプロローグス・ブリシャディ）、*Oreochromis niloticus*（ナイルティラピア）のグラフ表示の結果である。配列の先頭近くと末尾近くに、それぞれ、**突出した棚**状および**スパイク**状の形状をもつことがわかる。

3.2 鳥類ミトコンドリアゲノムのグラフ表示

図 5 はタカ目タカ科の *Spilornis cheela*（カンムリワシ）、*Nisaetus nipalensis*（クマタカ）、*Accipiter gentilis*

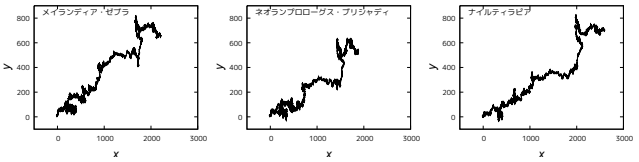

図 4　スズキ目カワスズメ科の 3 生物種のグラフ

（オオタカ）、*Buteo buteo*（ヨーロッパノスリ）のグラフ表示の結果である。配列の末尾から 3 分の 1 ほどの場所に、**S 字**状の形状をもつことがわかる。また、これら 4 種の生物種はそれぞれタカ科の異なる属 (genus) に属しているが、上列の 2 種（カンムリワシ、クマタカ）と下列の 2 種（オオタカ、ヨーロッパノスリ）の 2 グループに何らかのレベルで分類できるのではないか、ということがこの図から推測することができる。

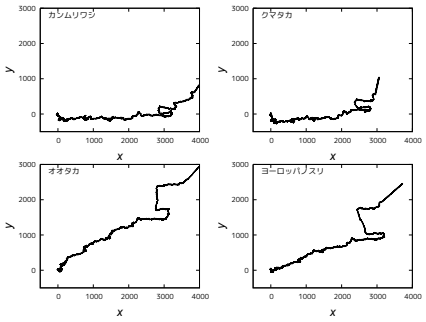

図 5　タカ目タカ科の 4 生物種のグラフ

4 まとめ

アライメントに依らない生物学的配列比較の一手法である、配列のグラフ表示について、生物分類への応用の可能性を検討した。その結果、魚類、および鳥類において、一部の生物種のグラフに特徴的な形状が現れることが確認できた。このような形状の組み合わせやそれらが現れる配列上の位置などを系統的に整理することにより、生物分類へ応用することが可能であると期待される。

参考文献

[1] Zhu-Jin Zhang, “DV-Curve: a novel intuitive tool for visualizing and analyzing DNA sequences,” *Bioinformatics* **25** (2009), pp.1112–1117.

[2] Yusei Kobori and Satoshi Mizuta, “Similarity Estimation Between DNA Sequences Based on Local Pattern Histograms of Binary Images,” *Genomics Proteomics Bioinformatics* **14** (2016), pp.103–112.