

ミトコンドリアのタンパク質を用いた水棲生物の分子系統解析 Informatics of Protein Coded in Mitochondria of Aquatic Animal

坂楨 健伍[†] 蓬萊 尚幸[†]
Kengo Sakamaki Hisayuki Horai

1. はじめに

1.1 背景

我々の住むこの地球には様々な種の生物が生息している。このような生物がどのような進化を遂げて、現在の生物として分化してきたのか(系統)は系統解析により推定されている。推定には試料に DNA の塩基配列やタンパク質のアミノ酸配列を用いてクラスタリング^[1]することで行われる。推定された系統は系統樹として表され、生物を分類している。しかし、全ての試料で同じ結果になるとは限らず、作成された樹形図が現在知られている系統樹と異なる結果となる場合もある。それまでの系統では距離の離れた種が近い系統のものであることを示すような結果が得られることもある。

1.2 課題・目的

本研究では、エラーを持たず水中で生活をする生物の特徴や共通点、陸上の生物との違いを分子の視点から調査するため、水棲哺乳類を対象に分子系統解析を行う。その中でクラスタリングを行い、樹形図を作成し、各クラスター内に遠い関係にあると知られているはずの種同士が近い関係になっているなどの特異な存在の有無を検証する。

2. 手法

本研究では、分子系統解析の手法として配列データのクラスタリングを用いて解析を行う。

2.1 使用データ

本研究では解析対象としてイルカ、クジラ、アザラシを中心とする水棲の哺乳類 70 種を用いた。NCBI によって提供されている GenBank に登録されている配列データのうち、ミトコンドリアのゲノムにコーディングされているタンパク質のアミノ酸配列を実験データとした。使用した配列は 7 種の NADH dehydrogenase、4 種の cytochrome、2 種の ATP synthase の計 13 種である。

本研究では、水棲哺乳類という広い範囲を対象としているため、生物のエネルギー系に深く関与しており、あらゆる生物内に存在しているミトコンドリアは近縁の種のみでなく広範囲の種を対象とする解析を行うのに適していると考え、ミトコンドリアのタンパク質のアミノ酸配列を用いている。

2.2 距離行列の作成

距離行列の計算には EBI により提供されている多重配列プログラムの ClustalW^{[2][3][4]}を用いる。実験データであるアミノ酸配列にマルチプルアライメントを行い、類似度を計算することで距離行列を作成する。この際、反復改善法を用いてマルチプルアライメントを行っている。

2.3 樹形図の作成

マルチプルアライメントにより導かれた距離行列から樹形図の作成を行う。樹形図の作成も距離行列の計算と同様に ClustalW を用いて行う。この際、近隣結合法を用いている。

近隣結合法で作成される樹形図は無根のものであるため、有根のものにするための方法として解析対象から最も遠い関係にあると知られている生物種の配列を 1 つ以上含める、または最も離れた関係にある 2 つの配列を結ぶ枝の midpoint を内部ノードとするという 2 つがあり、本研究では後者の方法を用いて有根の樹形図としている。図 1 は近隣結合法を用いた際の距離行列と樹形図の例である。

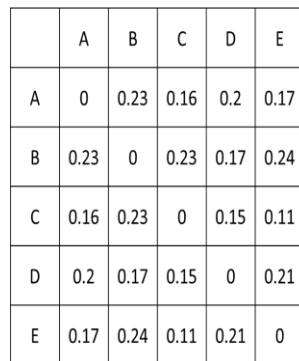


図 1 距離行列と樹形図の例

3. 結果

本研究では、GenBank に登録されている 13 種類のミトコンドリアのタンパク質のアミノ酸配列データを用いて、解析を行った。ここでは、例として、NADH dehydrogenase subunit 1 を用いてクラスタリングを行い、作成された樹形図と解析結果を示す。樹形図を図 2 破線の位置で分割し、14 個のクラスターに設定した。

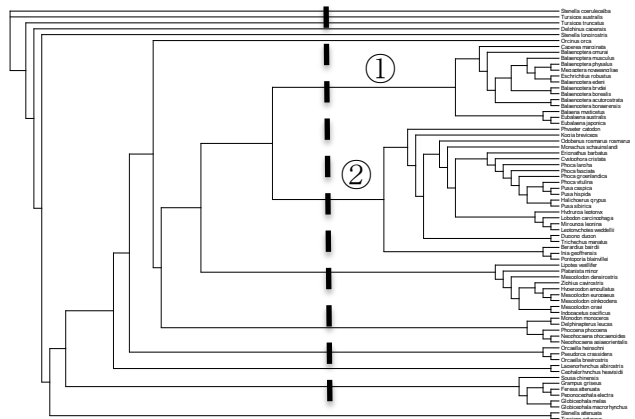


図 2 NADH dehydrogenase subunit 1

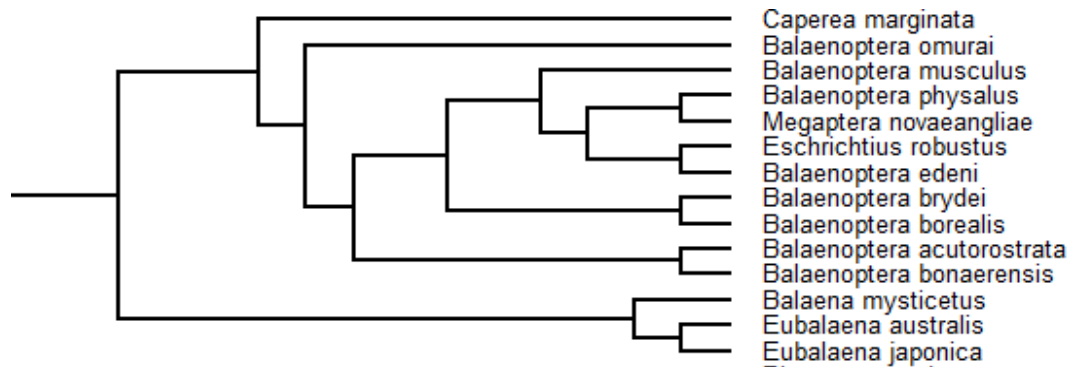


図 3 NADH dehydrogenase subunit 1 クラスタ①

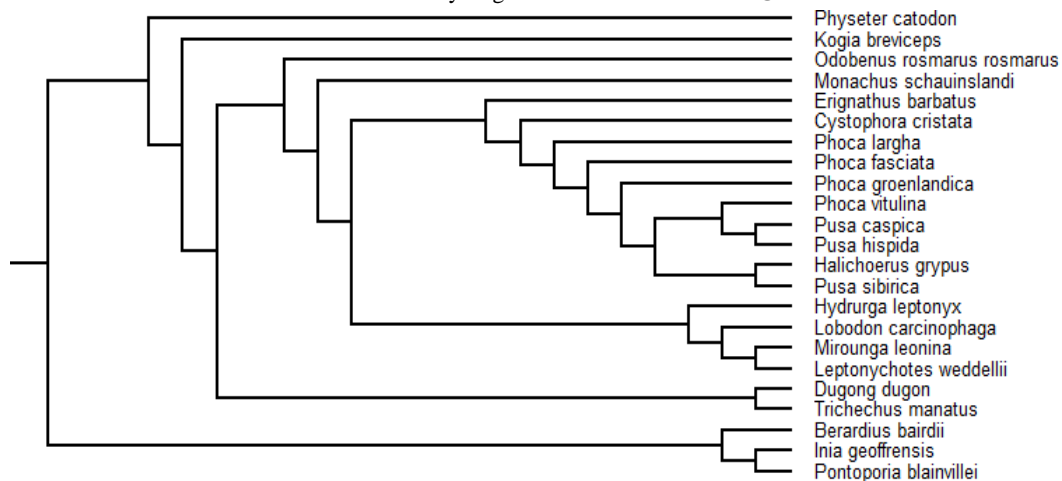


図 4 NADH dehydrogenase subunit 1 クラスタ②

図 3 は図 2 の①のクラスタを抜粋したものである。このクラスタはクジラのみで形成されており、イルカなどの他の動物は含まれていなかった。

図 4 は図 2 の②のクラスタを抜粋したものである。このクラスタの大半がアザラシで構成されているが、カワイルカやジュゴン、マナティといったアザラシ以外の種が混在している。このことからカワイルカは海洋に生息するイルカよりアザラシに近い種であるということが確認できる。また、解析対象の中の 4 種のカワイルカのうち 3 種がこのクラスタに分類されている。

4. 考察

図 4 のクラスタに存在していたカワイルカやジュゴン、マナティは本来、アザラシとは遠い種であるとされていることから、これらは 1.2 節で述べた特異な存在となっている可能性が高いと考えられる。NADH dehydrogenase subunit 1 を用いた樹形図と NADH dehydrogenase subunit 2 を用いた樹形図を比較すると、両者の類似度は高かったが、樹形図を有根のものとする際の内部ノードとする枝を作る 2 つの配列が異なっていた。内部ノードの位置は ClustalW で樹形図を作成する際に自動で設定されるため、最も遠い関係にある配列が複数存在する場合によるものであるとも考えられるが、タンパク質ごとの特徴であると捉えることもできるため、さらなる分析が必要であると考えられる。

2 つの樹形図を比較したとき、階層下部ではなく階層上部でつながっている部分に、差異が大きく現れており、ク

ラスタ内の特異な存在ではなく、階層上部に着目して今後の解析を行うことも重要であると考えられる。

5. おわりに

13 種のタンパク質から樹形図を作成したが、そのうちのすべてを分析するまでには至らなかったため、他の分析を行う必要がある。今回、NADH dehydrogenase の subunit 1 と subunit 2 の 1 対比較を行ったので他のタンパク質とも 1 対比較を行っていく。

遠い関係にあると知られているはずの種同士が近い関係になっていると考えられるものが確認できたので、これについても深く調査していく必要がある。

また、樹形図を比較したとき、樹形図同士の差異や類似を数値として表すことが可能であれば、より高度な分析と評価が可能になると考える。そのための手法の提案なども今後の研究の課題として検討していく。

参考文献

- [1] B. S. Everitt et al. "Cluster Analysis", Wiley, (2011)
- [2] Clustal W and Clustal X version 2.0. (2007 September 10) Bioinformatics (Oxford, England) 23 (21) :2947-2948
- [3] The EMBL-EBI bioinformatics web and programmatic tools framework. (2015 April 06) Nucleic acids research 43 (W1) :W580-4
- [4] Analysis Tool Web Services from the EMBL-EBI. (2013 May 13) Nucleic acids research 41 (Web Server issue) :W597-600