

氏名 足 立 淳

学位（専攻分野） 博士（学術）

学位記番号 総研大甲第110号

学位授与の日付 平成7年3月23日

学位授与の要件 数物科学研究科 統計科学専攻

学位規則第4条第1項該当

学位論文題目 Modeling of molecular evolution and maximum
likelihood inference of molecular phylogeny

論文審査委員 主査 助教授 佐藤 俊哉

教授 田 辺 國 士

教授 長谷川 政 美

教授 伊 藤 栄 明

助教授 岸 野 洋 久（東京大学）

生物の核酸 (DNA, RNA) の塩基配列や、タンパク質のアミノ酸配列を比較することによって、生物の系統樹 (系統関係) を推定する「分子系統学」の研究をおこなった。これまでの研究から、分子データから系統樹を推定する方法の中で、最尤法は推定精度が高いことが知られていた。しかし、最尤法は他の方法に比べて圧倒的に計算量が多く、またデータ (taxon) の数が増えると系統樹の組合せ爆発が起こり、最尤系統樹を探索することが困難になるため、現実にはあまり使われていなかった。本研究は、最尤法による系統樹推定を実用的なものにするための新しい方法と、その方法を実装したソフトウェアの開発、そしてそれらを用いて現実のデータを解析して得られた知見からなる。

第1章で分子系統樹推定の導入をした後、第2章では分子進化のモデリングを行った。最尤法によって分子系統樹を推定するには、塩基やアミノ酸の置換モデルをマルコフモデルとして記述する。最尤法でこれまで扱えたモデルは固定されていたが、本研究では様々な置換モデルを扱えるように、相対的な置換速度の比をあらわす相対置換頻度行列を導入し、組成率と分けて置換モデルを記述した。このモデルの特徴は、置換する配列の組成率が無限時間後にその組成率に収束することである。また、系統関係が既知のデータから相対置換頻度行列を推定することが可能なので、置換モデルを定義する相対置換頻度行列を最尤推定する方法を開発した。

塩基配列データの場合では、ミトコンドリアのタンパク質をコードしている領域の4重縮退のコドンの3番目の座位だけを取り出し、置換モデルを最尤推定した。これらの座位は、置換が起こってもコードしているアミノ酸は変わらないので、置換に対する制約が一番緩く、現実には起きた突然変異を反映したものになっている。このデータから HKY85, TN93, General Stationary Markov の各モデルをそれぞれ推定した。AIC によると General Stationary Markov モデルが一番当てはまりがよい。

次にアミノ酸の置換モデルを最尤推定した。ミトコンドリアのゲノムは核のゲノム比べて進化速度が速いことから、比較的近縁な生物種間の系統樹推定のためによく使われているが、核のゲノムとはコドン表が異なるなどの違いがあるので、ミトコンドリア独自のアミノ酸置換確率行列の確立が望まれてきた。そこでミトコンドリアのタンパクをコードしている領域のアミノ酸配列から、アミノ酸の置換確率行列も推定した。この置換確率行列は、ミトコンドリアのデータによりフィットするため、タンパクの分子系統樹推定の精度が高まることが期待できる。

第3章は、最尤法による分子系統樹の推定法的高速化と実用化の研究である。

これまでの方法と比べ、部分系統樹の部分尤度を保存するデータ構造を変え、各パラメータをニュートン法で更新するアルゴリズムを非再帰的ループによる新しいアルゴリズムにすることによって、最尤なパラメータを推定する高速なアルゴリズムを開発した。

OTU の数が増加すると、系統樹の2分木のトポロジーの組み合わせ数は爆発的に増大するので、すべてトポロジーの尤度を計算することは不可能になる。最尤法は最尤なパラメータを推定するために Newton 法を使用しているが、これはパラメータが収束するまで計算量の多い尤度関数を頻繁に呼び出さなければならない。そこで、パラメータ推定を計算量の少ない最小二乗法でおこない、その値を使って尤度を1度だけ計算し、それを近似

尤度とすることを考えた。この近似尤度を系統樹選択の規準とする方法を「近似尤度法」と呼ぶ。近似尤度の値が上位のトポロジィだけ取り出し、その中から最尤法により最尤系統樹とそれに準ずる系統樹を得ればよい。また、最小二乗法の残差二乗和を足切りの規準として使うこともできる。

総当たり探索ができない領域では、何らかのヒューリスティックな最尤系統樹探索法が必要になる。そこで *Star Decomposition* と名付けた最尤系統樹探索法を考案した。始めに、全 OTU が1つの仮想的な root から同時に分岐した、star 型の系統樹を仮定する。この tree 型の系統樹から任意の OTU のペアを近縁であると仮定しクラスターを組ませる。ペアを選ぶ全ての組合せについてその系統樹の尤度を計算し、その中であるペアの組がそのペアの一方を含んだ別の組よりも有意であれば、そのペアはクラスターを組み独立できる。

第4章では、2、3章の研究成果を基に開発した、最尤法による分子系統樹推定ソフトウェア MOLPHY について述べてある。MOLPHY は Internet 上の Anonymous FTP site にポストし、フリープログラムとして一般に公開している。現在、海外からだけで200人以上の研究者がダウンロードするくらい需要がある。MOLPHYは、世界で唯一のタンパク質系統樹の最尤推定プログラム "ProtML"、核酸系統樹の最尤推定プログラム "NucML"、近隣結合法のプログラム NJdist、そしてアミノ酸（塩基）配列データの基本的統計量出力プログラム ProtST (NucST) 等、によって構成される。また各種のユーティリティ等も開発した。

第5章は、これまで研究開発してきた方法を用い、実際のデータを解析した応用である。「霊長類の分岐年代推定」、「ミトコンドリアのアミノ酸置換の傾向」、「クジラ目の系統解析」、「4 Taxa サンプリング解析」、「Cytochrome b による哺乳類と鳥類の系統解析」、そして、「Cytochrome Oxidase Subunit II による哺乳類の系統解析」等の解析を行い、新たな生物学的知見を得ている。

論文審査の結果の要旨

分子系統樹の最尤推定にはこれまで塩基配列、アミノ酸配列の置換に関して固定された経験的な置換確率しか用いられていなかった。そのため解析する配列がミトコンドリアのDNAにコードされたアミノ酸配列でも核のDNAにコードされたアミノ酸配列でも同一の置換確率を使わざるを得なかったが、本論文では置換確率そのものを推定する方法を与えることにより、それぞれのアミノ酸の置換確率を用いて分子系統樹の推定を行うことを可能にした。また系統樹推定のアルゴリズムの高速化、総当たり探索のための最大尤度の近似、系統樹分解のヒューリスティックな方法の提案、またこれらの結果から分子系統樹推定ソフトウェア MOLPHY を構築し公開したことは、提案した統計的方法が理論的にだけでなく実際に適用可能であることを広く一般に示しており高く評価できる。ソフトウェア MOLPHY によるデータ解析結果も興味のある新しい知見を与えており、本論文の一部はすでに *Japanese Journal of Genetics*, *Journal of Molecular Evolution* に発表され、また *Journal of Molecular Evolution*, *Molecular Biology and Evolution* に掲載が決定している。

審査委員会は数物科学研究科統計科学専攻足立淳氏の論文を審査した結果、博士論文として十分な内容を備えていると判断した。