

氏名（本籍）	石川 奏太		
学位の種類	博 士（ 理 学 ）		
学位記番号	博 甲 第	7317	号
学位授与年月日	平成 27 年 3 月 25 日		
学位授与の要件	学位規則第4条第1項該当		
審査研究科	生命環境科学研究科		
学位論文題目	How to Reconstruct Accurate Phylogenetic Trees from Nucleotide Sequence Data with Extraordinary Compositional Bias: Assessment of the Performance of Data-Recoding Methods and Non-Homogeneous Models (塩基組成の不均一な遺伝子配列データに基づく適切な進化系統樹推測法の探索：データリコーディング法および非一様モデルの有用性の検証)		
主査	筑波大学准教授	博士（理学）	稲垣 祐司
副査	筑波大学教授	学術博士	橋本 哲男
副査	筑波大学教授	理学博士	佐藤 三久
副査	筑波大学教授	博士(理学)	石田 健一郎

## 論 文 の 要 旨

分子系統解析では複数の生物種より得られたアライメントに基づき、配列間における塩基やアミノ酸の形質の違いを比較することでそれぞれの配列の進化を推測する。このとき「遺伝子配列の進化では塩基・アミノ酸組成はすべての系統で保存される」と仮定される。しかし、実際の遺伝子配列では塩基・アミノ酸組成には系統間で不均一性が生じ、程度の違いはあるが上記の仮定が崩れる。特に、塩基配列におけるアデニンとチミンの合計含量（AT含量）には極端な偏りが生じることが知られ、配列間でのAT含量の均一性を仮定する従来の解析手法では誤った系統関係（アーティファクト）が再構築される場合がある。この問題を解決するために、これまでに二つの手法が提案され、既に実データ解析に適用されている。一つは、アライメント中の4つの塩基をプリンとピリミジンに変換しAT含量の偏りを平均化する方法（RY-coding法）、もう一方はAT含量が系統樹の各枝で変化することを許容する置換モデル（非一様モデル）である。上記2つの手法が、一般的な解析手法に比べ系統解析の精度をどの程度向上させるのかについては、詳細な検証が不可欠である。そこで、本研究ではシミュレーションデータおよび実配列データを用い、RY-coding法および非一様モデルの有用性に量的評価を与えることを目的とした。

AT含量に系統間で偏りのある4配列からなる塩基配列シミュレーションデータを準備した。このデータを、（1）系統間で均一な塩基組成を仮定するHKYモデルに基づく最尤法、（2）系統間でAT含量の不均一性を許容する非一様モデル（GG98モデル）に基づく最尤法により解析した。また（3）RY-codingしたシミュレーションデータを、系統間で均一な塩基組成を仮定するモデルを用いた最尤法により解析した。その結果、アライメント中のAT含量の不均一性に対し、HKYモデルに比べRY-codingおよびGG98モデルは頑健な系統樹推測を可能とすることが示された。ただし、GG98モデルがモデル化しきれない、あるいはRY-codingがAT含量を平均化できない複雑

な塩基組成の不均一性を含むシミュレーションデータの解析では、2手法による解析の頑健性は低下することが判った。またRY-coding法および非一様モデルを実データの解析に用い、AT含量の不均一性によって誘導されるアーティファクトを克服できることを確認した。

上記のシミュレーションおよび実配列データの解析結果から、データリコーディング法および非一様モデルの頑健性は示された。ただし、複雑な非一様モデルに基づく系統樹推測では膨大な計算時間が必要となる可能性がある。そこで、非一様モデルを用いた分子系統解析に要する計算時間を短縮するため、非一様モデルを実装した系統解析用プログラムのハイブリッド並列化を行い、大規模並列計算機上で性能を評価した。その結果、計算時間の大幅な短縮を実現した。

## 審 査 の 要 旨

これまで塩基配列に基づく分子系統解析では、アライメント中の配列の塩基組成（あるいはA/T含量）は均一性を仮定する置換モデルが主に用いられてきた。しかし、現実には生物間でAT含量が大きく異なりうる。この配列間でのAT含量の不均一性に起因する系統樹の誤推測を避けるため、(1) 非一様置換モデルと (2) RY-coding法が提案され、実データ解析に用いられてきた。著者はシミュレーションデータと実データの解析から、上記2つの手法を用いた系統解析の有効性を世界に先駆けて検証した。以上の結果から、今後の実データ系統解析を行う際のガイドラインを提供することができたことは系統解析の分野における大きな貢献である。

パラメーターリッチな非一様モデルに基づく系統解析の計算コストは高く、実データ解析には大規模並列計算機が必要である。しかし、非一様モデルを実装し、かつ並列計算可能な系統解析プログラムは存在しなかった。著者は、非一様モデルを実装したNHMLプログラムをMPI/OpenMPによりハイブリッド並列化し、その性能評価を行った。並列化により計算時間の短縮を実現したことで、大規模計算機上での非一様モデルに基づく実配列データ解析が現実的になったことは高く評価できる。

平成27年2月3日、学位論文審査委員会において、審査委員全員出席のものに論文の審査及び最終試験を行い、本論文について著者に説明を求め、関連事項について質疑応答を行った。その結果、審査員全員によって合格と判定された。

よって、著者は博士（理学）の学位を受けるのに十分な資格を有するものとして認める。