

平成 22 年 6 月 9 日現在

研究種目：基盤研究（B）

研究期間：2007～2009

課題番号：19300099

研究課題名（和文） ポストゲノム時代における分子系統樹推定法の研究

研究課題名（英文） Inferring Molecular Phylogenies in the Post-genomic Era

研究代表者

足立 淳（ADACHI JUN）

統計数理研究所・モデリング研究系・准教授

研究者番号：30370092

研究成果の概要（和文）：

ゲノム規模のデータから真獣類の系統樹を最尤推定すると、遺伝子ごとの違いを考慮に入れてモデルを改良していくにつれて正しいと思われる系統樹に収束していく。これは解析に用いるモデル選択の重要性を示しており、遺伝子ごとの進化速度や進化様式の不均一性が推定の偏りを生み出した原因だと考えられる。種子植物の系統解析でも、進化速度の変化や収斂的アミノ酸置換など、ゲノム系統学に内在する問題点を指摘し、より信頼性の高い系統樹推定法を提示した。

研究成果の概要（英文）：

The availability of large genomic sequence datasets allows us to perform the maximum likelihood analysis of the eutherian phylogenetic relationship. We demonstrate that the better model which takes into account the different tempos and modes of evolution among genes, should be used for phylogenetic inference in cases of large variation in evolutionary features among different genes, such as for genome-scale data.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2007 年度	5,200,000	1,560,000	6,760,000
2008 年度	4,500,000	1,350,000	5,850,000
2009 年度	4,500,000	1,350,000	5,850,000
年度			
年度			
総計	14,200,000	4,260,000	18,460,000

研究分野：分子進化学、分子系統学、情報生物学

科研費の分科・細目：統計科学

キーワード：ゲノム系統学、分子進化、分子系統樹、最尤法、モデル選択

## 1. 研究開始当初の背景

研究開始当初すでに、ポストゲノム時代といわれる時代に入っていた。それまでにゲノム・プロジェクトが急速に進展し、ヒトをはじめとしたいくつかの生物種で遺伝情報の全体であるゲノムの塩基配列が解読された。それに伴い、ゲノム・データの存在が生物学の多くの分野の研究の前提になってきた。ゲノムの全塩基配列を明らかにするということは、遺伝情報の全貌を知ることであり、それまで一般的であった限られた遺伝子の解析とは質的に異なる研究の展開が可能となってきたわけである。ゲノム・データを進化的な枠組みのなかで解析し、遺伝子機能の進化的な起源を明らかにするためには、まず系統関係を正しく捉える必要がある。ゲノム規模のデータから系統関係を推定する研究分野は、ゲノム系統学と呼ばれ、21世紀に入ってから世界中の研究室で活発な研究が行われていた。

## 2. 研究の目的

従来行われていた1個、あるいは数個の遺伝子の配列データの解析では、サンプリングの誤差が大きく、はっきりした結論が下せないことが多かった。ゲノム規模のデータではこの誤差がほとんどなくなり、はっきりした「結論」が下せるようになってきたように一般には思われていた。そのため、生物学者の間では、ゲノム・データさえ得られれば、系統学のあらゆる問題は簡単に解決するであろうという機運が生まれた。ところがゲノム・データが実際に得られてみると、問題はそのままに簡単なものではないことが次第に分かってきた。

系統樹推定の際に仮定される置換モデルは、実際の進化過程を近似するものに過ぎないから、必ず実際とのずれがある。このずれのことをモデル・ミススペシフィケーションという。このために系統樹推定に偏りが生じてしまうのである。データ量が多くなるために、サンプリングの誤差が実質上なくなる代わりに、強く支持される系統樹が間違っている危険性が増すのである。本研究の目的は、ゲノム系統学に内在するこのような問題点を明らかにして、膨大なデータから系統関係に関する情報を正しく抽出するためのデータ解析法の進展をはかることにあった。

## 3. 研究の方法

それまでのわれわれのグループを含めた世界のいくつかのグループの研究により、真獣類（哺乳類のなかで雌が胎盤をもったもの）が進化的には異なる地理的な起源をもつ3つのグループに分類されることが明らかになった。1つは中生代の北半球の大陸であったローラシア大陸起源のボレオ真獣類、2つめはアフリカ大陸が孤立していた時代にそこで進化したアフリカ獣類、3つめは南米大陸起源の異節類である。このことはいろいろなデータから支持されており、ほぼ確立した考えになっているが、この3者の関係に関しては論争が続いている。この3者間を含む系統樹の根元がどこにあるかという問題は、初期の真獣類がいつ頃どこで進化したかを明らかにする上で非常に重要である。

Nishihara et al. (2007) では、ゲノム規模のデータが蓄積している哺乳類10種について、共通にアラインメントできる2,789個の蛋白質遺伝子(1,011,870塩基)のデータセットを用意して、この問題に関する系統樹推定を行った。その結果、広く一般に用いられている単純なモデルでは、推定の偏りが大きく、明らかに間違った系統樹が非常に強く支持されること、さらにモデルを改良していくにつれて正しいと思われる系統樹に収束していくことを示した。このことは、分子系統樹解析におけるモデルの重要性と統計学的に基礎づけられた方法の必要性を強く示している。

本研究では、この問題をはじめとしたゲノム系統学の諸問題においてモデルのミススペシフィケーションが系統樹推定にどのような偏りをもたらすのか、またそのような偏りをなくするにはどのようにモデルを改良していけばよいのかについて検討した。

## 4. 研究成果

真獣類の系統進化に関しては、われわれのグループをはじめとした世界のいくつかの研究グループの成果として、ボレオ真獣類、アフリカ獣類、異節類の3大グループに分類できることが明らかになった。ところが多くのデータが蓄積しているにもかかわらず、これら3者の関係については依然として不明であり、論争が絶えない。Nishihara et al. (2007) は、ゲノム規模のデータが蓄積している哺乳類10種について、共通にアラインメ

ントできる 2,789 個の蛋白質遺伝子 (1,011,870 塩基)のデータセットを用意して、3者の関係について系統樹推定を行った。その結果、データを1つながりの均一な配列とみなす単純なモデルでは推定の偏りが大きく、明らかに間違った系統樹が非常に強く支持されること、さらに遺伝子ごとの違いを考慮に入れてモデルを改良していくにつれて正しいと思われる系統樹に収束していくことが分かった。このことは解析に用いるモデル選択の重要性を示しており、遺伝子ごとの進化速度や進化様式の不均一性が推定の偏りを生み出した原因だと考えられる。

ゲノム系統学のモデルケースとして、葉緑体ゲノム・データによる種子植物の系統解析を行った。種子植物のうちグネツム科、サバクオモト科、マオウ科の3科は、グネツム目 Gnetales という分類単位を構成する。それぞれの科の構成メンバーは1属ずつしか現存せず、進化的には衰退傾向にあるグループと考えられている。グネツム目の系統的な位置づけは、植物系統学において長い間論争の的であった。最近になって葉緑体ゲノム・データが多くの子植物から得られるようになり、大量データをもとにしたこの問題の解析が可能になってきた。このような古い分岐を扱うには同義置換はほとんど飽和状態に達しているために、あまり系統に関する情報をもたない。そのために、葉緑体ゲノムにコードされたたんぱく質のアミノ酸配列を扱う。アミノ酸の置換モデルとしては、われわれが開発した cpREV model (Adachi et al. 2000) があるが、これを開発した当時、葉緑体ゲノム・データはわずか 10 種についてしか得られていなかった。今回、その後得られたデータを加えて 64 種について解析し、新たなアミノ酸置換モデルを構築し、Gnetales(グネツム目)の系統的な位置づけに関する解析を行った (Zhong et al. 投稿準備中) (独立性を確保するために、アミノ酸置換モデルの構築に際しては、Gnetales のデータは入っていない)。

グネツム目の解析を通じて、ゲノム系統学の様々な問題点が浮かび上がってきた。まず通常の方法で解析すると、Gnetales が球果類のうちでマツ科 Pinaceae を除くグループである Cupressophyta (Non-Pinaceae conifer、スギなどを含む)に近縁であるという系統樹 (Gnecup 仮説)が 100%のブートストラップ確率で支持された。ところが、詳細な解析により、Gnetales と Cupressophyta を結び付けるこの系統樹は、2つの枝の間の Long-branch attraction と、いくつかのたんぱく質においてこの2つの枝の間で起こった収斂的アミノ酸置換による artifact であることが明らか

になった。このような問題点を考慮すると、最も可能性の高い系統樹は Gnetales が Pinaceae と姉妹関係にある (Gnecup 仮説) というものであることが分かった (Zhong et al. 投稿準備中)。このように本研究は、ゲノム系統学に内在するいくつかの問題点を指摘し、より信頼性の高い系統樹推定法を提示した。

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計8件)

1. A. Matsui, F. Rakotondraparany, I. Munechika, M. Hasegawa, and S. Horai, Molecular phylogeny and evolution of prosimians based on complete sequences of mitochondrial DNAs. *Gene*, 査読有, 441, 2009, pp.53-66.
2. T. Yonezawa, N. Kohno, and M. Hasegawa, The monophyletic origin of sea lions and fur seals (Carnivora; Otariidae) in the Southern Hemisphere. *Gene*, 査読有, 441, 2009, pp.89-99.
3. Z. Ren, B. Zhu, E. Ma, J. Wen, T. Tu, Y. Cao, M. Hasegawa, and Y. Zhong, Complete nucleotide sequence and gene arrangement of the mitochondrial genome of the crab-eating frog *Fejervarya cancrivora* and evolutionary implications. *Gene*, 査読有, 441, 2009, pp.148-155.
4. H. Miwa, I.J. Odrzykoski, A. Matsui, M. Hasegawa, H. Akiyama, Y. Jia, R. Sabirov, H. Takahashi, D.E. Boufford, and N. Murakami, Adaptive evolution of *rbcl* in *Conocephalum* (Hepaticae, Bryophytes). *Gene*, 査読有, 441, 2009, pp.169-175.
5. Y. Murata, T. Yonezawa, I. Kihara, T. Kashiwamura, Y. Sugihara, M. Nikaido, N. Okada, H. Endo, and M. Hasegawa, Chronology of the extant African elephant species and case study of the species identification of the small African elephant with the molecular phylogenetic method. *Gene*, 査読有, 441, 2009, pp.176-186.
6. A. Kurabayashi, M. Sumida, H. Yonekawa, F. Glaw, M. Vences, and M. Hasegawa,

Phylogeny, recombination, and mechanisms of stepwise mitochondrial genome reorganization in mantellid frogs from Madagascar. *Mol. Biol. Evol.*, 査読有, 25, 2008, pp.874-891.

7. H. Nishihara, N. Okada, and M. Hasegawa, Rooting the eutherian tree: the power and pitfalls of phylogenomics. *Genome Biology*, 査読有, 8:R199, 2007, pp.1-10.

8. T. Yonezawa, M. Nikaido, N. Kohno, Y. Fukumoto, N. Okada, and M. Hasegawa, Molecular phylogenetic study on the origin and evolution of Mustelidae. *Gene*, 査読有, 396, 2007, pp.1-12.

〔学会発表〕(計2件)

1. Y. Cao,  
Molecular phylogeny and evolution of river dolphins: integration of mitochondrial DNA and SINE flanking sequences,  
Evolution 2007 Conference, 2007/6/18,  
Christchurch, New Zealand

2. J. Adachi,  
MolPhy: new software for phylogenetic analyses using maximum likelihood,  
Evolution 2007 Conference, 2007/6/17,  
Christchurch, New Zealand

## 6 . 研究組織

### (1)研究代表者

足立 淳 (ADACHI JUN)  
統計数理研究所・モデリング研究系  
・准教授  
研究者番号：30370092

### (2)研究分担者

曹 纓 (CAO YING)  
統計数理研究所・モデリング研究系・助教  
研究者番号：20370091

### (3)連携研究者

長谷川 政美 (HASEGAWA MASAMI)  
統計数理研究所・モデリング研究系  
・名誉教授  
研究者番号：60011657

西原 秀典 (NISHIHARA HIDENORI)  
東京工業大学大学院・生命理工学研究科  
・助教  
研究者番号：30370092

松井 淳 (MATSUI ATSUSHI)  
京都大学霊長類研究所・遺伝子情報分野  
・gCOE 研究員

米澤 隆弘 (YONEZAWA TAKAHIRO)  
復旦大学 (中国)・生命科学学院・講師