

分子系統樹推定におけるモデルの ミススペシフィケーション ——脊椎動物の系統進化を例として——

曹 纓¹・長谷川 政美²

(受付 2001 年 12 月 25 日 ; 改訂 2002 年 2 月 15 日)

要 旨

分子系統樹の推定は、仮定したモデルに依存する。このモデルは、進化過程における DNA の塩基置換や蛋白質のアミノ酸置換に関するものである。しかしモデルが現実的なものでない場合、偏った系統樹推定が行なわれる可能性がある。本稿では、脊椎動物の分子系統学的研究でみられる、このような置換モデルのミススペシフィケーションに関わる問題を、実例を通じて紹介する。

キーワード：分子系統樹推定，置換モデル，最尤法，モデルのミススペシフィケーション，脊椎動物進化。

1. はじめに

生物学のほとんどの問題は、本来進化の枠組みのなかで捉えられなければならないものである。なぜならば、それぞれの生物のもついろいろな特徴、生体器官や生体分子のもつそれぞれの機能はすべて長い進化の歴史の産物だからである。また、系統樹を推定することは、生物進化に関わるほとんどの問題を解くための出発点である。従来この分野の研究は、比較形態学による系統樹推定が主流であったが、近年は DNA の塩基配列や蛋白質のアミノ酸配列のデータを統計的に解析する分子系統学による方法が盛んになってきた。比較形態学では研究者の主観的な判断が入り込む余地が多く、サンプルできる形態的形質の数が限られているのに対して、分子系統学ではバクテリアからヒトに至るまでのあらゆる生物を共通の客観的な尺度で比較でき、しかも研究者がその気になりさえすれば（研究費が十分にあるという条件は必要であるが）、ほとんどいくらかでも多くの形質（実際にはゲノム全体という制限はあるが）をサンプルすることが可能であり、これによって地球上で実際に起った進化の系統樹を再構築することができると考えられている。このように分子系統学は非常に有望な方法であるが、実際にこの分野で使われている分子系統樹推定法はまだ未熟なものであり、さまざまな問題点を抱えている。分子系統樹を推定するにあたっては、進化における塩基置換やアミノ酸置換に関するモデルを仮定しなければならないが、このモデルが現実的なものでない場合には、偏った系統樹推定が行なわれる可能性がある。本稿では、脊椎動物の分子系統学的研究でみられる、このようなモデルのミススペシフィケーションに関わる問題を紹介することにする。

¹統計数理研究所：〒106-8569 東京都港区南麻布 4-6-7; cao@ism.ac.jp

²統計数理研究所：〒106-8569 東京都港区南麻布 4-6-7

2. 両生類の系統進化

今から 3 億 5 千万年から 4 億年前の古生代デボン紀の間に、硬骨魚類のうちで肺魚などを含む肉鱗類というグループの一部が水中生活から離れて陸上に進出し、陸上脊椎動物（四肢類）が進化したと考えられている。最初の陸上脊椎動物が両生類であるが、現生の両生類（分類学では両生綱という）は、カエルなどの無尾目（Anura）、サンショウウオ、イモリなどの有尾目（Urodela）、アシナシイモリなどの無足目（Gymnophiona）の 3 つの目（もく）に分類されている。ところが、これら 3 つの目の間の系統関係がどのようになっているかについては、さまざまな議論がある。最近、分子系統学から 2 つの対立する仮説が提案されている（図 1）。1 つは、Feller and Hedges（1998）の説で、有尾目と無足目が近縁であって、両生類のなかではカエルなどの無尾目が最初に他から分かれたというものである（Tree-1）。彼らは、ミトコンドリアの 12S リボソーム RNA、Val-tRNA、それに 18S リボソーム RNA の 5' 端の部分配列（合計 2040 塩基座位）を近隣結合法、最節約法、最尤法などの方法で解析したところ、いずれの方法でも有尾目/無足目のグルーピングが強く支持された。彼らの最尤法による解析では、有尾目/無足目のグルーピングのブートストラップ確率が 96% に達している。一方、Zardoya and Meyer（2001）は、ミトコンドリアの全塩基配列データを用いて、同じような方法で解析したところ、無尾目/有尾目のグルーピングが強く支持された（Tree-2）。最尤法の簡便法である Quartet-puzzling（Strimmer and von Haeseler（1996）、曹・長谷川（1998））では、無尾目/有尾目のグルーピングに対する支持が 100% に達し、両生類のなかではアシナシイモリなどの無足目が最初に他から分かれたという結論になった。このように、用いるデータによって、互いに矛盾する結論が導かれ、しかもそれぞれが非常に強く支持されたのである。この 2 つの研究グループは、どちらもミトコンドリアの塩基配列データを用いているから、正しい系統樹は一致

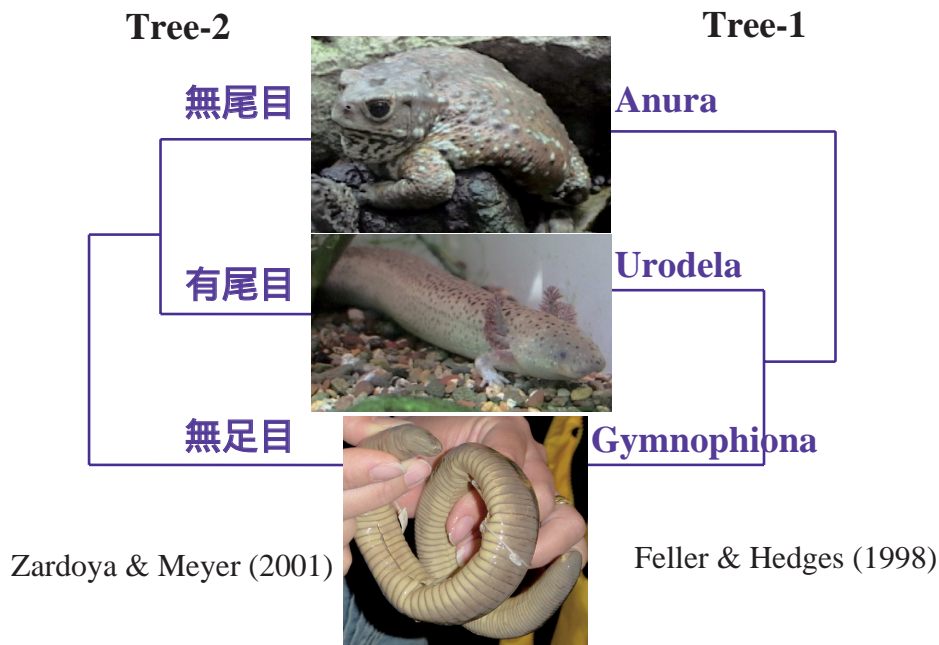


図 1. 両生類の 3 つの主要なグループの間の系統関係に関して対立する 2 つの仮説。

表 1. 図 1 に示された両生類の系統に関する 2 つの仮説 (Tree-1, 2) と第 3 の仮説 (Tree-3) の間の最尤法による分子系統学的な比較.

	a) 均質モデル			b) 不均質モデル (Γ 分布)		
	mt-RNAs	mt-proteins	Total	mt-RNAs	mt-proteins	Total
Tree-1 有尾/無足	<-17132.8>	<-39271.1>	<-56403.9>	-3.4 \pm 7.8	-2.5 \pm 9.4	-5.9 \pm 12.2
Tree-2 無尾/有尾	-2.5 \pm 11.8	-8.6 \pm 13.8	-11.2 \pm 18.1	<-16119.6>	<-37376.6>	<-53496.2>
Tree-3 無尾/無足	-16.6 \pm 8.9	-3.5 \pm 14.2	-20.1 \pm 16.8	-8.5 \pm 6.3	-9.6 \pm 8.0	-18.1 \pm 10.2

最尤系統樹の対数尤度はかぎ括弧内に、そのほかの系統樹については、最尤系統樹との対数尤度の差とその標準誤差 (\pm) が示してある。座位間の進化速度の違いを考慮しない均質モデルについては、塩基配列とアミノ酸配列データは、それぞれ MOLPHY の NucML, ProtML プログラム、 Γ 分布で座位間の進化速度の違いを考慮した不均質モデルについては、PAML の BaseML, CodeML プログラムを用いた。mt-RNAs は、Feller and Hedges (1998) のデータであり、HKY モデル (Hasegawa et al. (1985)) による解析を行なった。均質モデルの場合、塩基置換に関する HKY モデルにおける transition/transversion 比 $\kappa=2.8$ 、不均質モデルの場合には、 $\kappa=4.2$ 、 Γ 分布の shape parameter $\alpha=0.43$ 。mt-proteins は、次の 11 種の脊椎動物のミトコンドリア DNA の H 鎖にコードされた 12 個の蛋白質のアミノ酸配列データであり、mtREV-F モデルで解析した; *Didelphis virginiana* (オポッサム; データベースのアクセッション番号 Z29573), *Ornithorhynchus anatinus* (カモノハシ; X83427), *Gallus gallus* (ニワトリ; X52392), *Alligator mississippiensis* (ワニ; AF069428), *Eumeces egregius lividus* (トカゲ; AB016606), *Chrysemys picta* (カメ; AF069423), *Xenopus laevis* (カエル; M10217), *Typhlonectes natans* (アンナシイモリ; AF154051), *Mertensiella luschani* (イモリ; AF154053), *Protopterus dolloi* (ハイギョ; L42813), *Latimeria chalumnae* (シーラカンス; U82228)。

するはずである。従って、両方の結論が矛盾しているとしたら、いずれか一方、あるいは両方の系統樹推定法が適切でないために、間違っているということになる。

Feller and Hedges (1998) の最尤法の解析では、有尾目/無足目のグルーピングのブートストラップ確率が 96% に達したが、HKY モデル (Hasegawa et al. (1985)) のトランジション/トランスバージョン比 (κ) が 100 に固定されている。これは、彼らの考えではトランジションが飽和に達しているために、トランスバージョンの情報だけを解析に取り入れるためである。確かに、 κ を 100 に固定して、彼らのデータを MOLPHY (Adachi and Hasegawa (1996)) の NucML プログラムで解析してみると、Tree-1 (有尾目/無足目) が最尤系統樹 (対数尤度 = -20494.7) になり、Tree-2, 3 はそれよりも対数尤度がそれぞれ 33.7 \pm 23.2, 54.6 \pm 20.0 だけ低くなる。Tree-1 のブートストラップ確率は、93% ということになり、Feller and Hedges の解析結果とほぼ一致する。ただし、HKY モデルの κ パラメータを最尤法で最適化すると、表 1a のように、Tree-1 が最尤系統樹であることには変わらないが、対数尤度が -17132.8 と大きく改善され、しかも Tree-2 との対数尤度の差が 2.5 \pm 11.8 というふうになり、極くわずかなものになり、系統樹間に有意な差がみられなくなる。図 2 は、Feller-Hedges データを NucML で解析して得られた最尤系統樹であり、各節に添えられた数字は、局所的ブートストラップ確率である。局所的ブートストラップ確率は、系統樹内の当該の節に関わる 3 つの subtree (それにアウトグループ) の内部の関係は固定した上で、subtree の間で可能な 3 通りの系統樹の間でブートストラップ確率が 1 になるようにしたものである。subtree の内部の関係が間違っている場合は、この局所的ブートストラップ確率は意味をもたないが、この例の場合には無尾目、有尾目、無足目、それにアウトグループのなかの関係はかなり信頼できると思われるので、局所的ブートストラップ確率は、ここで問題にしている目間の関係に関してはある程度信頼できるであろう。ところが実際には、有尾目と無足目がグループを組む確率は、わずか 60% しかなく、ほかの可能性を棄却できない。さらに、PAML の BaseML プログラム (Yang (1997)) により、座位間の進化速度の違いを Γ 分布 (実際には離散化したものを用いる) で近似するモデルを採用すると、表 1b のように、Tree-2 の方がわずかながら Tree-1 よりも高い対数尤度 (-16119.6) をもつようになる。これらの結果は、Feller and Hedges の解析において、異なる系統樹の間でみかけ上有意な差がでたのは、モデルのミスマスペシフィケーションによる artifact である可能性を示唆するものである。

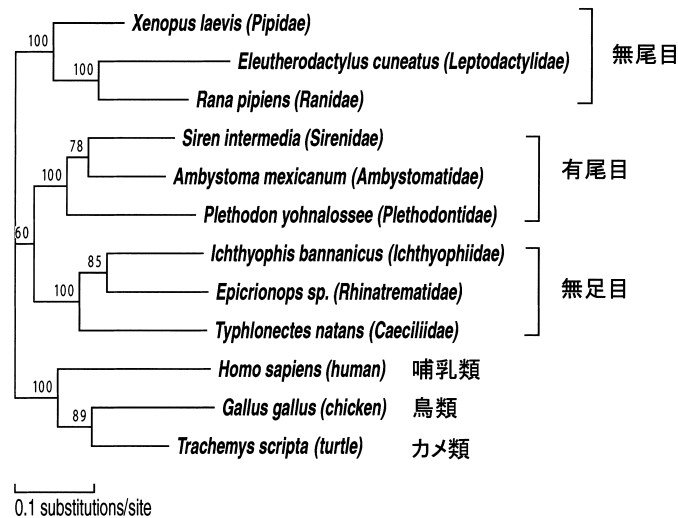


図 2. Feller and Hedges (1998) のデータを HKY モデル (Hasegawa et al.(1985)) により NucML で解析して得られた最尤系統樹．各節の数字は，局所ブートストラップ確率．

一方，Zardoya and Meyer (2001) の解析にも問題がある．彼らは，ミトコンドリアのゲノムのうち，蛋白質をコードしている遺伝子と tRNA, rRNA などの遺伝子が一つにつながった塩基配列について解析を行なったが，蛋白質遺伝子のコドン座位ごとの進化速度の違いを考慮していない．ミトコンドリアの進化速度は，核遺伝子にくらべて非常に高く，このような古い分岐を扱う際には，蛋白質遺伝子の同義置換（アミノ酸を変えないような塩基置換）は飽和状態に達しており，系統関係に関する情報はほとんどもたないであろう（例えば，Hasegawa et al. (1985)）．しかも，動物の系統によって塩基組成の偏りが違うために，すべての塩基座位を含めた解析は，塩基組成の偏りの違いを明示的に取り入れたモデルを使わない限り，系統樹推定を間違える可能性がある．このような場合，塩基配列データを扱うよりも，アミノ酸配列を解析する方が，より現実的なモデル化が可能である．アミノ酸組成の偏りの系統による違いは，一般には塩基配列の同義置換座位（主にコドンの 3 番目の座位）ほどではない．ミトコンドリアの rRNA の塩基配列は，Feller-Hedges データに含まれるから，蛋白質のアミノ酸配列だけを座位間の進化速度を考慮せずに MOLPHY の ProtML を用いて最尤法で解析すると（均質モデル），表 1a のように，Zardoya and Meyer (2001) と違って，Tree-1 が最尤系統樹になる．ただし，Tree-2 あるいは 3 の対数尤度には，Tree-1 とくらべて有意な差は認められない．ProtML による最尤系統樹が図 3 に示してあるが，この解析でも有尾目と無足目がグループを組むブートストラップ確率は，わずか 49% しかなく，ほかの可能性は棄却できない．また，PAML (Yang (1997)) の CodeML プログラムにより，アミノ酸座位の間の進化速度の違いを Γ 分布で取り入れて解析すると（不均質モデル；表 1b），今度は Tree-2 が最尤系統樹になる．この場合でも，他の系統樹との間に，対数尤度に有意な差は見られない．

最尤法の利点の一つは，これら 2 組のデータセットの解析結果を総合して評価することができることであろう．総合評価は，それぞれのデータセットの対数尤度を単に足し合わせることによって可能になる (Adachi and Hasegawa (1996), Hasegawa et al. (1997))．表 1a と b の Total の欄が，そのようにして得られた総合評価の結果を示している． Γ 分布を用いない均質モデルの場合は，Tree-1 が最尤系統樹になるが，Tree-2 の対数尤度の差は，その標準誤差よりも小さく，有意なものではない． Γ 分布を用いた不均質モデルでは，Tree-2 の方が最尤系統樹

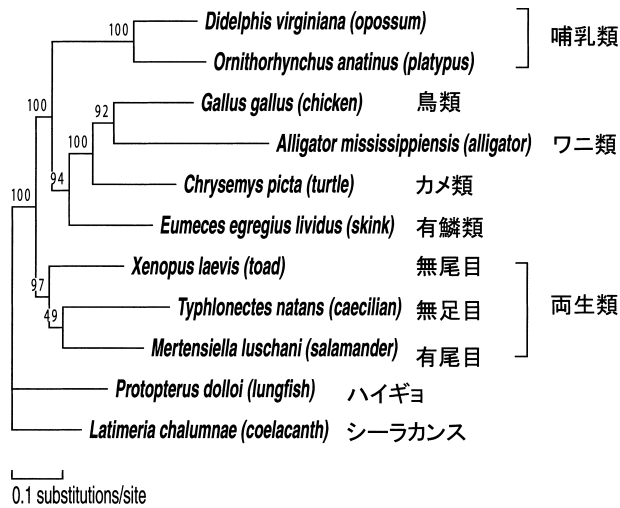


図 3. ミトコンドリアの蛋白質データを mtREV-F モデル (Adachi and Hasegawa (1996)) により ProtML で解析して得られた最尤系統樹. 各節の数字は, 局所ブートストラップ確率.

になるが, この場合でも Tree-1 との差は有意ではない. このように, Feller and Hedges (1998) と Zardoya and Meyer (2001) では, 互いに矛盾した結論がそれぞれ統計的な有意さをもって主張されているが, より現実的なモデルによって解析すると, それぞれの有意さは消えてしまい, 2つのデータセットを総合的に評価しても, どちらかの系統樹が有意に支持されるということはない. 従って, 現在までに得られているデータからは, この問題に決着をつけることはできず, 両生類の主要なグループの間の系統関係を明らかにするには, さらに多くの遺伝子に関するデータが必要だということになる. 分子系統樹解析にあたっては, 非現実的なモデルを用いると, それが正しいかどうかは別として, そのデータが実際にもっている情報以上に特定の系統樹が強く支持される傾向があることに留意しなければならない (Hasegawa and Adachi, (1996)).

3. 羊膜類の系統進化

両生類は陸上に進出したものの, 完全には水中生活を捨てることができなかった. 彼らは現在でも卵からオタマジャクシの時代は水中で過ごしているが, それは卵のなかの胚を陸上での乾燥から防ぐための手段をもたないからである. 両生類から進化した爬虫類, 鳥類, 哺乳類は陸上でのそのような困難を克服するための手段を進化させた. それは, 羊膜と呼ばれるものであり, それによって胚を乾燥から防ぐことができるようになり, 一生を水中から離れて過ごすことが可能になった. これらの陸上脊椎動物は, 羊膜類と呼ばれている.

羊膜類のなかで最初に他から分かれたのは, 哺乳類であり, このことは形態学からも分子系統学からも支持される (例えば, Mindell et al. (1999)). ところが, 残りの爬虫類と鳥類の間の系統関係に関しては多くの議論がある. 爬虫類はトカゲ, ヘビなどの有鱗類, ワニ類, カメ類の3つの大きなグループに分類されるが, 実際には鳥類に対してこれら3つのグループは単系統の関係にはない. つまり鳥類が系統的には爬虫類のなかに入ってしまい, 系統関係を重視する限り, 爬虫類という分類単位は存在しないと考えられているのである. 形態学からは長

い間、カメ類がまず最初に有鱗類、ワニ類、鳥類などから分かれ、その後、有鱗類がワニ類と鳥類から分かれたと考えられてきた (Benton (1990)). つまり、鳥類はいわゆる爬虫類のなかではワニ類と最も近縁であるというものである。カメ類が最初に分かれたというこの伝統的な考えは、最近の分子系統学から否定されている (Zardoya and Meyer (1998) Kumazawa and Nishida (1999) Hedges and Poling (1999)). 多くの分子系統学的解析からは、カメ類よりもむしろトカゲ、ヘビなどの有鱗類の方が先に分かれたと考えられる。

ところが、カメ類、ワニ類、鳥類の間の関係は、依然としてまだよくわからない。カメ類は形態的に極めて特殊化したグループであり、比較形態学からその系統的な位置を決めるのは難しく、分子系統学の活躍が望まれるテーマであるが (Wilkinson et al. (1997)), カメ類、ワニ類、鳥類の間の関係は、分子系統学からも依然として不明なのである。Kumazawa and Nishida (1999) は、ミトコンドリアのゲノムデータから、鳥類とワニ類が最も近縁な関係にあると結論した (図4)。鳥類がワニ類に近縁だということは、形態学の方からも支持されていた。これに対して、Hedges and Poling (1999) は核遺伝子の配列データの解析からカメ類とワニ類が近縁であると主張している (図4)。Cao et al. (2000) は、Kumazawa and Nishida (1999) のデータにその後新たに得られたミトコンドリアのゲノムデータを付け加え、さらに Hedges and Poling (1999) の核遺伝子データも含めた総合評価を行ない、これらのデータからは2つの仮説のいずれかが正しいと結論することはできないとした。2つのデータセットからは、それぞれの仮説が支持されるが、他方の仮説を棄却するほどには強い証拠になっていないのである。

Hedges and Poling (1999) と同じような結論が、Mannen and Li (1999) からも得られている (図4)。彼らは、エノラーゼ、LDHa、それに LDHb などの蛋白質の遺伝子データを用いているが、これは Poling and Hedges のデータにも含まれる遺伝子である (種のサンプリングは増えている。Hedges and Poling のデータには、この他に21個の核遺伝子と9個のミトコンドリア遺伝子が含まれる)。これらのデータを解析した結果、カメ/ワニのグルーピングが、近隣結合法、最節約法、最尤法でそれぞれ100、100、99%の高いブートストラップ確率で支持された。ただし、彼らの解析では、コドン座位による進化速度の違いが考慮されていない。彼らと同じように、これら3つの遺伝子の塩基配列をつなげたものをそのまま座位間の進化速度の不均質性やコドン構造など考慮しない単純なモデル (均質モデル) に基づいて最尤法で解析すると、カメ/ワニが単系統のグループになる系統樹 (Tree-1, 2, 3) 以外は、表2aで示したいかなる検定を用いても、5%の有意水準ではいずれも棄却されてしまう。表2では、鳥類、ワニ類、カメ類、それに有鱗類との間で可能な15通りの系統樹について、最尤系統樹との対数尤度の差とその標準誤差、それに近似的に不偏な検定 (AU test; Shimodaira and Hasegawa (2001) 下平 (2002)), RELL ブートストラップ確率 (BP; Kishino et al. (1990)), Kishino-Hasegawa 検定 (KH test; Kishino and Hasegawa (1989)), Kishino-Hasegawa 検定に多重比較の補正を加えた Shimodaira-Hasegawa 検定 (SH test; Shimodaira and Hasegawa (1999)) などの P 値を示した。これらの P 値は、CONSEL プログラム (Shimodaira and Hasegawa (2001)) を用いて計算された。カメ/ワニが単系統のグループになる系統樹 Tree-1, 2, 3 のブートストラップ確率の合計は、99.3%になる。コドン構造を明示的には考慮せず、座位間の進化速度の不均質性を Γ 分布で取り入れた解析では (不均質モデル; 表2a), Tree-1, 2, 3 以外では Tree-5 が SH 検定では5%水準ぎりぎり棄却されなくなるが、ミトコンドリアの蛋白質データが支持する Tree-4 は、いずれの検定でも棄却されてしまう。この場合、系統樹 Tree-1, 2, 3 のブートストラップ確率の合計は、97.8%になる。この問題のように数億年という時間スケールの分岐についての解析では、コドンの3番目の座位は明らかに塩基置換が飽和に達しており、そのような座位を解析に含めても有効な情報は得られないと考えられる。むしろ、塩基組成、あるいはコドンの使用頻度、が系統によって異なるということから、misleading な情報しか得られないかもしれない。

そこでわれわれは、コドンの3番目の座位を取り除いて、1番目と2番目の座位だけで再び塩基配列データを解析してみた(表2b)。全ての座位を含めた解析と同様に、均質モデルでも不均質モデルでも、カメ/ワニのグルーピングが強く支持され、Tree-1, 2, 3のブートストラップ確率の合計は、均質モデルと不均質モデルとで、それぞれ98.5%, 99.3%もの高い値になり、いずれのモデルでもカメ/ワニでない系統樹は全て5%の有意水準で棄却されてしまう(均質モデルでのTree-7以外)。従って、コドンの3番目の座位を含めても含めなくとも、結果にはあまり違いが見られない。このように、塩基配列データをどのような方法で解析しても、カメ/ワニのグルーピングが強く支持されるということは、このような仮説が真実だということの意味するものであろうか?これまでのところ、比較形態学からは、このような仮説を支持するデータは得られていない。分子系統学は有効な方法であるが、多くの問題を抱えており、比較形態学と大きく矛盾した結論が得られた場合には、その結論を安易に受け入れられないような慎重な態度が必要である(長谷川・岸野(1996))。上の解析では、コドンの3番目の座位を取り入れた場合も、除いた場合も、それぞれの塩基座位がお互いに独立に進化することを仮定している。実際には、3つの連なった塩基が1つのアミノ酸を指定するようになっており(それをコドンと呼ぶ)、3つの塩基の置換は強い相関をもって起こっているものである。そのような相関を取り入れて、より現実的なモデルに基づいた解析を行なうにはどのようにすればよいであろうか?そのためには、コドン置換モデル(Yang et al.(1998))を用いることもできるが、同義置換が飽和状態に達しているような時間スケールの問題に対しては、塩基配列よりもアミノ酸配列のレベルで解析するほうがよいだろう。

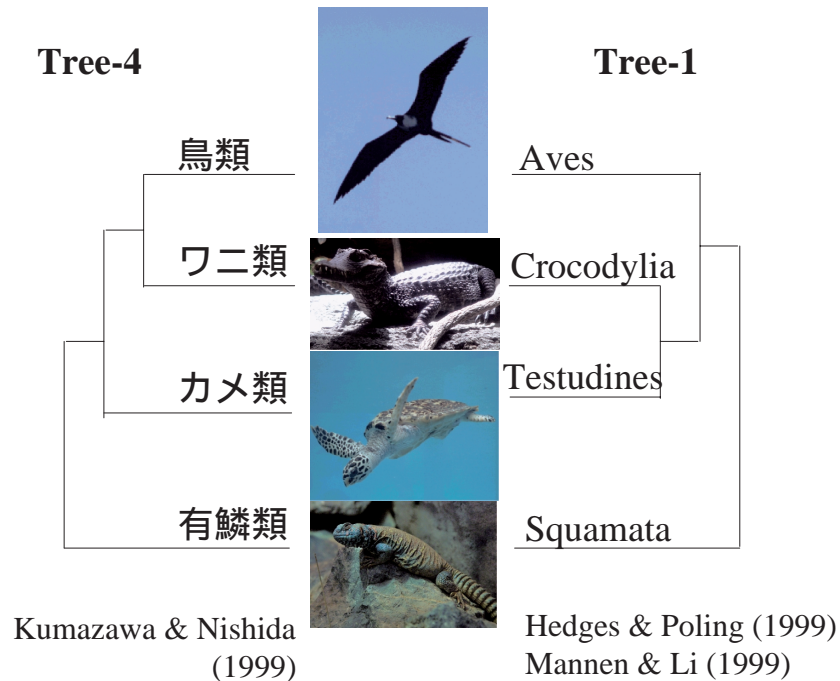


図 4. 羊膜類の5つの主要なグループの間の系統関係に関して対立する2つの仮説。

表 2. 図 4 に示された羊膜類の系統に関する 2 つの仮説 (Tree-1, 4) のほか, 鳥類 (Bi), ワニ類 (Al), カメ類 (Tu) それに有鱗類 (Li) の間で可能な合計 15 通りの系統樹についての Mannen and Li (1999) のデータの最尤法による解析.

a)

系統樹トポロジー	塩基配列, 均質モデル (HKY モデル)					塩基配列, 不均質モデル (HKY+ Γ モデル)				
	$\Delta \ell \pm SE$	AU	BP	KH	SH	$\Delta \ell \pm SE$	AU	BP	KH	SH
1 ((Bi,(Tu,Al)),Li)	<-15531.0>	0.968	0.918	0.944	0.999	<-14984.7>	0.837	0.698	0.778	0.972
2 (Bi,((Tu,Al),Li))	-21.8 \pm 13.8	0.081	0.046	0.056	0.377	-4.2 \pm 5.6	0.356	0.172	0.222	0.669
3 ((Bi,Li),(Tu,Al))	-23.6 \pm 13.4	0.064	0.029	0.043	0.333	-5.0 \pm 5.3	0.248	0.108	0.169	0.630
4 (((Bi,Al),Tu),Li)	-54.7 \pm 16.4	0.001	0.000	0.000	0.022	-23.0 \pm 9.1	0.009	0.002	0.009	0.036
5 (Bi,((Tu,Li),Al))	-57.2 \pm 23.5	0.015	0.005	0.008	0.015	-21.2 \pm 11.4	0.036	0.012	0.032	0.051
6 (((Bi,Tu),Al),Li)	-61.5 \pm 15.3	0.002	0.000	0.000	0.007	-24.1 \pm 8.7	0.014	0.001	0.005	0.024
7 (((Bi,Li),Tu),Al)	-65.2 \pm 21.9	0.002	0.000	0.002	0.004	-24.5 \pm 11.4	0.021	0.005	0.018	0.028
8 (Bi,(Tu,(Al,Li)))	-74.0 \pm 21.6	0.001	0.000	0.000	0.001	-24.7 \pm 10.5	0.005	0.000	0.010	0.017
9 (((Bi,Li),Al),Tu)	-75.0 \pm 20.4	0.000	0.000	0.000	0.001	-27.3 \pm 10.6	0.004	0.000	0.007	0.011
10 ((Bi,(Tu,Li)),Al)	-87.0 \pm 24.6	0.000	0.000	0.001	0.001	-30.0 \pm 12.3	0.016	0.002	0.009	0.012
11 ((Bi,Al),(Tu,Li))	-90.3 \pm 23.8	0.000	0.000	0.000	0.000	-31.9 \pm 11.8	0.012	0.000	0.005	0.007
12 (((Bi,Tu),Li),Al)	-106.4 \pm 22.4	0.001	0.000	0.000	0.000	-33.3 \pm 11.6	0.022	0.000	0.003	0.004
13 (((Bi,Al),Li),Tu)	-108.1 \pm 21.6	0.000	0.000	0.000	0.000	-35.1 \pm 11.1	0.015	0.000	0.002	0.002
14 ((Bi,Tu),(Al,Li))	-114.4 \pm 21.2	0.000	0.000	0.000	0.000	-36.6 \pm 10.8	0.000	0.000	0.001	0.001
15 ((Bi,(Al,Li)),Tu)	-114.9 \pm 21.3	0.000	0.000	0.000	0.000	-36.6 \pm 10.8	0.000	0.000	0.001	0.001

b)

系統樹トポロジー	塩基配列, コドン 1 番号+2 番号, 均質モデル (HKY モデル)					塩基配列, コドン 1 番号+2 番号, 不均質モデル (HKY+ Γ モデル)				
	$\Delta \ell \pm SE$	AU	BP	KH	SH	$\Delta \ell \pm SE$	AU	BP	KH	SH
1 ((Bi,(Tu,Al)),Li)	<-6258.6>	0.602	0.419	0.502	0.877	-0.7 \pm 2.6	0.493	0.337	0.381	0.816
2 (Bi,((Tu,Al),Li))	-3.7 \pm 7.7	0.303	0.154	0.303	0.794	-1.5 \pm 2.2	0.199	0.099	0.239	0.759
3 ((Bi,Li),(Tu,Al))	-0.1 \pm 8.7	0.625	0.412	0.498	0.895	<-6096.0>	0.705	0.557	0.619	0.923
4 (((Bi,Al),Tu),Li)	-40.9 \pm 16.8	0.011	0.003	0.009	0.030	-24.0 \pm 10.2	0.016	0.002	0.013	0.017
5 (Bi,((Tu,Li),Al))	-50.0 \pm 17.4	0.003	0.000	0.003	0.005	-25.0 \pm 9.2	0.006	0.000	0.005	0.008
6 (((Bi,Tu),Al),Li)	-45.5 \pm 15.9	0.011	0.000	0.003	0.013	-35.7 \pm 10.4	0.006	0.000	0.001	0.001
7 (((Bi,Li),Tu),Al)	-33.0 \pm 17.9	0.022	0.005	0.031	0.071	-30.6 \pm 9.7	0.001	0.000	0.002	0.004
8 (Bi,(Tu,(Al,Li)))	-42.7 \pm 18.7	0.022	0.003	0.013	0.023	-24.6 \pm 9.5	0.004	0.001	0.007	0.011
9 (((Bi,Li),Al),Tu)	-33.4 \pm 18.0	0.022	0.004	0.031	0.064	-31.2 \pm 9.6	0.004	0.001	0.002	0.003
10 ((Bi,(Tu,Li)),Al)	-55.4 \pm 18.2	0.011	0.001	0.002	0.003	-36.2 \pm 10.7	0.000	0.000	0.001	0.001
11 ((Bi,Al),(Tu,Li))	-53.6 \pm 18.3	0.008	0.000	0.003	0.004	-25.8 \pm 10.1	0.020	0.005	0.008	0.011
12 (((Bi,Tu),Li),Al)	-53.4 \pm 18.5	0.010	0.000	0.003	0.005	-36.2 \pm 10.7	0.000	0.000	0.001	0.001
13 (((Bi,Al),Li),Tu)	-48.9 \pm 19.2	0.010	0.001	0.007	0.011	-25.1 \pm 10.3	0.015	0.003	0.011	0.015
14 ((Bi,Tu),(Al,Li))	-53.2 \pm 18.5	0.009	0.000	0.003	0.005	-37.3 \pm 10.5	0.000	0.000	0.001	0.001
15 ((Bi,(Al,Li)),Tu)	-49.0 \pm 19.3	0.024	0.001	0.007	0.011	-26.0 \pm 10.1	0.008	0.001	0.007	0.011

表 2c は, アミノ酸配列データの最尤法による解析結果を示している. ここでも, アミノ酸座位間の塩基速度の違いを考慮しないモデル (均質モデル) とそれを Γ 分布で考慮したモデル (不均質モデル) の両方で解析した. どちらのモデルでも, 塩基配列の解析と同様にカメ/ワニのグルーピングが支持されるが, 支持の度合いは低くなり (ブートストラップ確率で, それぞれ 88.4%, 89.5%), Tree-7, 8, 9 という奇妙な系統樹が, どの検定を用いてもかなり大きな P 値をとって, 棄却されなくなる (ただし, Tree-8 だけは, 不均質モデルを用いると Kishino-Hasegawa 検定 (Kishino and Hasegawa (1989)) では 5%水準ぎりぎり棄却されてしまうが).

Mannen and Li (1999) のデータに関する上記の解析では, より現実的なモデルのもとでは, Tree-7, 8, 9 が棄却されなくなることを見た. 従って, 単純なモデルのもとでそれらが棄却されたのは, データが系統に関する十分な情報をもっていたからというよりは, むしろ方法論上の artifact であると考えざるを得ない. ところが, そのような解析でも, ミトコンドリアの蛋

表 2. (つづき)

系統樹トポロジー	アミノ酸配列, 均質モデル (JTT-F モデル)					アミノ酸配列, 不均質モデル (JTT-F+ Γ モデル)				
	$\Delta \ell \pm SE$	AU	BP	KH	SH	$\Delta \ell \pm SE$	AU	BP	KH	SH
1 ((Bi,(Tu,Al)),Li)	-5.1 \pm 7.7	0.370	0.165	0.244	0.718	-3.6 \pm 4.1	0.265	0.117	0.186	0.635
2 (Bi,((Tu,Al),Li))	-5.2 \pm 7.7	0.381	0.159	0.242	0.714	-3.7 \pm 4.1	0.281	0.109	0.183	0.630
3 ((Bi,Li),(Tu,Al))	<-5890.4>	0.881	0.560	0.756	0.976	<-5789.3>	0.928	0.669	0.814	0.980
4 (((Bi,Al),Tu),Li)	-25.6 \pm 14.8	0.053	0.010	0.041	0.083	-18.5 \pm 9.3	0.038	0.004	0.025	0.037
5 (Bi,((Tu,Li),Al))	-26.1 \pm 12.7	0.003	0.000	0.026	0.060	-16.2 \pm 8.0	0.011	0.000	0.026	0.049
6 (((Bi,Tu),Al),Li)	-29.8 \pm 13.8	0.007	0.000	0.018	0.034	-19.3 \pm 9.0	0.014	0.000	0.019	0.027
7 (((Bi,Li),Tu),Al)	-15.9 \pm 10.4	0.105	0.029	0.069	0.311	-10.1 \pm 7.7	0.148	0.063	0.093	0.269
8 (Bi,(Tu,(Al,Li)))	-20.9 \pm 13.8	0.087	0.024	0.064	0.161	-15.0 \pm 8.5	0.076	0.013	0.042	0.080
9 (((Bi,Li),Al),Tu)	-14.6 \pm 10.8	0.153	0.052	0.091	0.353	-11.4 \pm 7.3	0.074	0.019	0.064	0.203
10 ((Bi,(Tu,Li)),Al)	-35.5 \pm 14.1	0.000	0.000	0.008	0.014	-18.9 \pm 9.7	0.030	0.002	0.028	0.040
11 ((Bi,Al),(Tu,Li))	-33.6 \pm 14.7	0.055	0.001	0.017	0.026	-20.4 \pm 9.5	0.004	0.000	0.018	0.027
12 (((Bi,Tu),Li),Al)	-35.0 \pm 14.3	0.001	0.000	0.010	0.016	-18.8 \pm 9.7	0.030	0.004	0.029	0.041
13 (((Bi,Al),Li),Tu)	-30.2 \pm 15.2	0.039	0.002	0.030	0.047	-19.5 \pm 9.7	0.043	0.004	0.023	0.035
14 ((Bi,Tu),(Al,Li))	-33.3 \pm 14.7	0.009	0.000	0.017	0.025	-20.0 \pm 9.6	0.016	0.000	0.022	0.032
15 ((Bi,(Al,Li)),Tu)	-30.0 \pm 15.3	0.028	0.002	0.028	0.047	-19.2 \pm 9.8	0.049	0.003	0.027	0.039

哺乳類(ヒト, ラット, マウス)とカエル (*Xenopus*) をアウトグループとして用いた。

$\Delta \ell \pm SE$: 最尤系統樹との対数尤度の差 \pm 標準誤差 (Kishino and Hasegawa (1989)), ただしかぎ括弧内は最尤系統樹の対数尤度, AU: Approximate unbiased 検定(Shimodaira and Hasegawa (2001)), BP: RELI ブートストラップ確率 (Kishino et al. (1990)), KH: Kishino-Hasegawa 検定 (Kishino and Hasegawa (1989)), SH: Shimodaira-Hasegawa 検定(Shimodaira and Hasegawa (1999)).

白質データが示唆している Tree-4 は, 不均質モデルではどの検定でも 5%水準で棄却され, 均質モデルでも AU 検定, SH 検定で P 値がそれぞれ 5.3%, 8.3%でぎりぎり棄却されないに過ぎず, このデータでみる限り Tree-4 が正しい可能性はあまり高くないように思われる。ミトコンドリアの蛋白質データでは, Tree-4 が支持されるものの, Tree-1 の可能性は棄却できないし, Hedges and Poling (1999) のデータと併せた総合評価では Tree-1 が支持されるから, Tree-1 が真の系統樹と考えるとよいであろうか?

ここで解析した遺伝子のうちで LDHa と LDHb は脊椎動物の進化の過程で遺伝子重複を起こしていることが知られており (Stock et al. (1997)), 種の分岐以前に重複した遺伝子を混ぜあわせて比較している可能性がある。これをパラローガスな比較というが, そのような比較を行なうと生物学的な種の系統樹は得られないことになる。従って羊膜類の系統進化を明らかにするためには, LDH 遺伝子ファミリーの詳細な解析とともに, 遺伝子重複を起こしていない別の遺伝子についても解析を進めることが必要であろう。これまでで最も多くの遺伝子データを用いた Cao et al. (2000) の解析では, Tree-1 が支持されたものの, Tree-4 が正しい可能性もかなり残っていたことから, 鳥類, ワニ類, カメ類の間の分岐は, 地質学的には短期間の間に続いて起こったようであり, この分岐の順番を明らかにするためには, さらに多くの遺伝子データが必要であろう。その際, あまりにも単純な非現実的なモデルに基づいた解析では, データが実際にもっている情報以上に, 見かけ上明確な結論が得られることがあることに注意することが必要であろう。

4. 食虫類の系統的位

モグラ, トガリネズミ, ハリネズミなどは, 真獣類(哺乳類のなかで雌が胎盤をもつもの)のなかで食虫目 (Insectivora) に分類されている。食虫目には, このほかにアフリカ大陸に生息するキンモグラ, マダガスカル島のテンレックなどが入ると考えられてきたが, 最近の分子

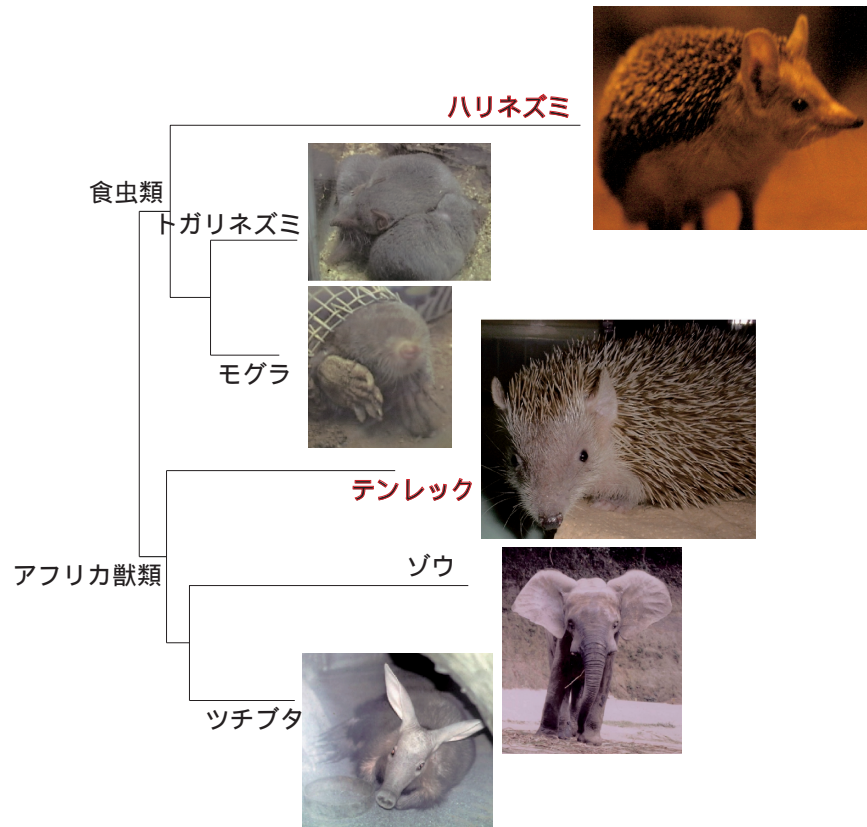


図 5. ミトコンドリア蛋白質データから推定されたハリネズミとテンレックの系統的な関係。テンレックの写真は、寺井洋平博士が撮影したもの。

系統学から、これらはモグラなどのいわゆる食虫類よりは、ゾウ、マナティー、ハイラックス、ツチブタなどアフリカ起源の一群の哺乳類に近縁である可能性が高いことが指摘されている (Springer et al(1997) 長谷川・曹(1999) Mouchaty et al(2000) Madsen et al(2001) Murphy et al.(2001a, 2001b))。そのため最近ではこれらを総称してアフリカ獣類 (Afrotheria)、それに対してこれまで食虫類と呼ばれてきたもののうちでアフリカ獣類に属するものを除いたモグラ、トガリネズミ、ハリネズミなどは、真食虫類 (Eulipotyphla) と呼ばれるようになってきた (Waddell et al.(1999a))。アフリカ獣類は、およそ 1 億年前にアフリカ大陸が南米大陸から分かれた後、1800 万年前にユーラシア大陸とつながるまでの孤立した大陸であった間、そこで進化したものと考えられている。かつてアフリカ、南米、マダガスカル、インド、南極、オーストラリアなどは一つにまとまっていて、ゴンドワナ超大陸を形成していたが、次第に小さな大陸に分かれていった。アフリカと南米との分離は、その最後の段階であったが、マダガスカルはそれ以前にアフリカから分かれていた。従って、テンレックなどのアフリカ獣類は、その祖先が孤立したアフリカ大陸で進化した後に、なんらかの機会に海を越えてマダガスカルに渡ったものと考えられている。マダガスカルの原猿類であるキツネザルの仲間も、同じようにその祖先がアフリカ大陸から海を越えて渡ったものであろう。

テンレックのうちでヒメハリテンレックは体毛が針になっていて、図 5 から明らかなよう

表 3. ハリネズミ, テンレックの系統的位に関するミトコンドリア蛋白質の最尤法による解析.

系統樹トポロジー	a) 均質モデル					b) 不均質モデル				
	$\Delta \ell \pm SE$	AU	BP	KH	SH	$\Delta \ell \pm SE$	AU	BP	KH	SH
1 ((ShrMol,Hed),(EleAar,Ten))	-23.2±17.8	0.152	0.082	0.092	0.402	<-28553.9>	0.838	0.676	0.732	0.988
2 ((ShrMol,(EleAar,Ten)),Hed)	<-29813.4>	0.903	0.734	0.819	0.995	-6.4±10.9	0.390	0.243	0.268	0.650
3 (((ShrMol,Hed),Ten),EleAar)	-53.9±24.1	0.015	0.001	0.013	0.038	-13.4± 7.4	0.069	0.021	0.039	0.355
4 (((ShrMol,Hed),EleAar),Ten)	-46.3±23.8	0.031	0.005	0.026	0.068	-13.6± 7.3	0.049	0.011	0.034	0.350
5 (ShrMol,((EleAar,Ten),Hed))	-47.3±14.5	0.001	0.000	0.001	0.062	-18.5± 8.4	0.005	0.001	0.017	0.181
6 (((ShrMol,EleAar),Ten),Hed)	-15.7±17.3	0.254	0.165	0.181	0.572	-18.9±15.0	0.130	0.045	0.102	0.183
7 (((ShrMol,Ten),EleAar),Hed)	-37.5±14.1	0.002	0.000	0.004	0.154	-26.0±13.7	0.002	0.000	0.029	0.049
8 ((ShrMol,EleAar),(Ten,Hed))	-44.1±24.8	0.042	0.009	0.039	0.092	-28.6±14.0	0.014	0.002	0.020	0.038
9 (ShrMol,(EleAar,(Ten,Hed)))	-54.6±22.4	0.013	0.002	0.009	0.029	-28.6±11.6	0.008	0.001	0.006	0.028
10 (((ShrMol,EleAar),Hed),Ten)	-60.3±23.4	0.000	0.000	0.005	0.014	-30.1±13.9	0.009	0.000	0.014	0.027
11 ((ShrMol,(EleAar,Hed),Ten))	-82.1±19.9	0.000	0.000	0.000	0.000	-30.9±11.2	0.011	0.000	0.004	0.016
12 ((ShrMol,(Ten,Hed)),EleAar)	-64.3±25.0	0.007	0.000	0.005	0.008	-35.1±12.2	0.004	0.000	0.002	0.007
13 (((ShrMol,Ten),Hed),EleAar)	-86.9±22.7	0.013	0.000	0.000	0.000	-37.7±11.8	0.015	0.000	0.001	0.003
14 ((ShrMol,(EleAar,Hed)),Ten)	-89.0±22.4	0.000	0.000	0.000	0.000	-40.0±11.8	0.013	0.000	0.001	0.001
15 ((ShrMol,Ten),(EleAar,Hed))	-96.0±21.2	0.023	0.000	0.000	0.000	-41.1±11.8	0.025	0.000	0.001	0.001

次の9種の哺乳類のミトコンドリアDNAのH鎖にコードされた12個の蛋白質のアミノ酸配列データをmtREV-Fモデルで解析した; *Sorex unguiculatus* (オオアシトガリネズミ; データベースのアクセッション番号 AB061527), *Talpa europaea* (ヨーロッパモグラ; Y19192), *Loxodonta africana* (アフリカゾウ, AJ224821), *Orycteropus afer* (ツチブタ; Y18475), *Echinops telfairi* (ヒメハリテンレック; AJ400734), *Erinaceus europaeus* (ナミハリネズミ; X88898), *Macropus robustus* (ケナガワラルー; Y10524), *Didelphis virginiana* (キタオポッサム; Z29573), *Ornithorhynchus anatinus* (カモノハシ; X83427). 肉食虫類のトガリネズミ(Shrew; 略号 Shr)とモグラ(Mole; Mol), アフリカ獣類のゾウ(Elephant; Ele)とツチブタ(Aardvark; Aar)がそれぞれ単系統のグループを作ることを仮定し, これらのグループ(ShrMolとEleAar), それにハリネズミ(Hedgehog; Hed), テンレック(Tenrec; Ten)の間で可能な15通りの系統樹について最尤法による解析を行った.

に形態的にはハリネズミとそっくりである. 従って, この動物が食虫類ではなくてアフリカ獣類に属するという主張は, 形態学者にとっては大きな驚きであった. この問題に関連して, ミトコンドリアのゲノムデータが得られている哺乳類について解析した結果が, 表3に示してある. 肉食虫類としては, ハリネズミのほか, モグラ, トガリネズミ, アフリカ獣類としてはテンレックのほか, ゾウ, ツチブタのデータを解析した. アウトグループとしては, 有袋類のワラルーとオポッサム, 単孔類のカモノハシを用いた. ここでは, ミトコンドリアDNAのH鎖にコードされている12種類の蛋白質のアミノ酸配列をつなげたものを解析した. アミノ酸置換モデルとしては, ミトコンドリアの蛋白質用に開発されたmtREV-Fモデル(Adachi and Hasegawa (1996a, 1996b))を用いた.

表3aは, 全てのアミノ酸座位が同じ進化速度をもつという均質モデルの結果であるが, Tree-2が最尤系統樹になる. この系統樹では, テンレックがゾウ, ツチブタなどとともにアフリカ獣類に入るが, ハリネズミはほかの肉食虫類とは別の系統になり, 真獣類の系統樹の根元付近から分岐している. Krettek (1995) は, 最初にハリネズミのミトコンドリアのゲノムデータを発表した論文で, ハリネズミが真獣類のなかで最初にほかのグループから分かれたと主張している. Tree-2が最尤系統樹になるということは, これと一致するものであるが, ハリネズミがモグラ, トガリネズミなどの肉食虫類と一緒にグループに入らないということは, 主に核遺伝子のデータによるTree-1と一致する最近の分子系統樹(Madsen et al.(2001), Murphy et al.(2001a, 2001b))と矛盾するものである.

ハリネズミのミトコンドリアの分子進化速度は, ほかの真獣類に比べて速く, DNAの塩基組成や蛋白質のアミノ酸組成の偏り方が独特であるため, ミトコンドリアの分子系統樹が間違っている可能性がある(Sullivan and Swofford (1997), Waddell et al. (1999b)). 進化速度の座位間の違いを取り入れた不均質モデルでは, Tree-1が最尤系統樹になる(表3b). 不均質モデルは, アミノ酸組成の系統による違いを明示的に取り入れているわけではないが, 均質モデルに比べるとはるかにデータとの適合性が改善されているため(自由パラメータの数は1つ増え

ているだけなのに、対数尤度は 1000 以上も高くなっている), より正しい系統樹を与えていると考えられる. 均質モデルでも Tree-1 が棄却されるわけではないし, 逆に不均質モデルでも Tree-2 が棄却されるわけではないので, ミトコンドリアの蛋白質データだけからこの系統学的問題に結論を下すことはできないが, 置換モデルがより現実的なものになるにしたがって, 系統樹推定もより正しいものに近づいていく傾向が明らかである (別の系統学的問題で似たよう

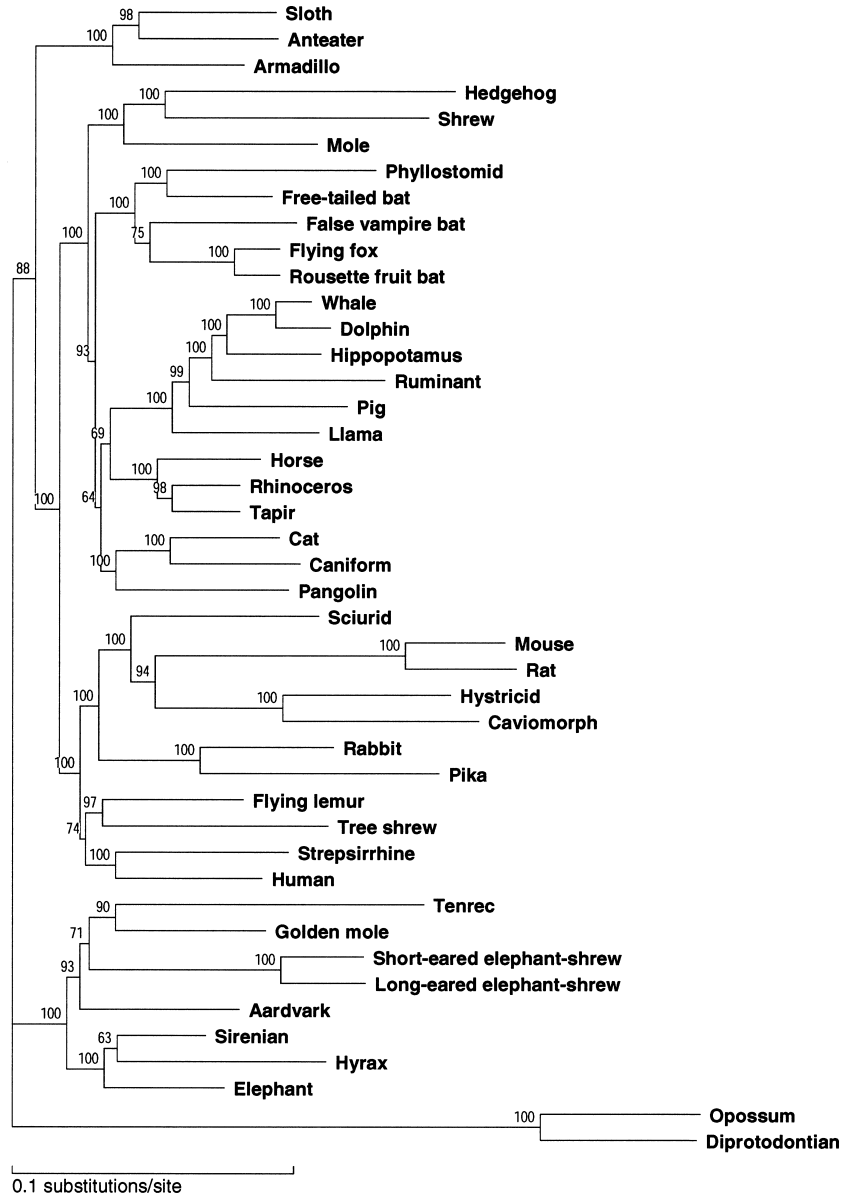


図 6. Murphy et al.(2001b) の哺乳類塩基配列データを HKY モデル (Hasegawa et al.(1985)) に より NucML で解析して得られた最尤系統樹. 各節の数字は, 局所ブートストラップ確率.

表 4. ハリネズミとテンレックの食虫類, アフリカ獣類におけるそれぞれの系統的位に関するミトコンドリア蛋白質の最尤法による解析.

系統樹トポロジー	a) 均質モデル					b) 不均質モデル				
	$\Delta \ell \pm SE$	AU	BP	KH	SH	$\Delta \ell \pm SE$	AU	BP	KH	SH
1 (((Shr,Mol),Hed),((Ele,Aar),Ten))	<-29836.6>	0.697	0.572	0.636	0.917	<-28553.9>	0.842	0.776	0.791	0.972
2 (((Shr,Mol),Hed),((Ele,Ten),Aar))	-4.7±14.2	0.435	0.328	0.364	0.750	-10.3± 7.0	0.073	0.028	0.075	0.351
3 (((Shr,Mol),Hed),(Ele,(Aar,Ten)))	-11.9±13.7	0.182	0.102	0.194	0.571	-6.6± 8.1	0.246	0.194	0.209	0.526
4 ((Shr,(Mol,Hed)),((Ele,Aar),Ten))	-88.6±19.3	0.001	0.000	0.000	0.000	-30.2±9.9	0.002	0.001	0.002	0.014
5 ((Shr,(Mol,Hed)),((Ele,Ten),Aar))	-89.5±23.6	0.001	0.000	0.000	0.000	-39.9±12.0	0.001	0.000	0.001	0.001
6 (((Shr,Hed),Mol),((Ele,Aar),Ten))	-90.3±19.0	0.000	0.000	0.000	0.000	-30.8± 9.7	0.001	0.000	0.001	0.012
7 (((Shr,Hed),Mol),((Ele,Ten),Aar))	-90.5±23.4	0.006	0.000	0.000	0.000	-40.2±11.9	0.001	0.000	0.000	0.001
8 ((Shr,(Mol,Hed)),(Ele,(Aar,Ten)))	-97.2±23.2	0.000	0.000	0.000	0.000	-35.9±12.7	0.002	0.000	0.003	0.003
9 (((Shr,Hed),Mol),(Ele,(Aar,Ten)))	-97.7±23.0	0.000	0.000	0.000	0.000	-36.2±12.7	0.003	0.000	0.002	0.003

な状況は、橋本 他 (2002) でも扱われている)。

形態的には非常によく似ているハリネズミとテンレックが一番近い関係にあるという系統樹は、Tree-7, 9, 12 であるが、これらが正しい可能性は 5% の有意水準でほとんど棄却される (ただし、Tree-7 だけは均質モデルのもとでは、AU および SH 検定で棄却されない)。特に古典的な分類学で考えられていたハリネズミとテンレックがともにモグラやトガリネズミと同じく食虫類に入るといふ Tree-12 は、いずれのモデルでも非常に強く棄却されてしまう。

図 6 は、Murphy et al. (2001b) の哺乳類塩基配列データを HKY モデル (Hasegawa et al. (1985)) により NucML で解析して得られた最尤系統樹である。各節に添えられた数字は、局所ブートストラップ確率である。図 6 でも Murphy et al. (2001b) でも、ハリネズミ (hedgehog) がモグラ (mole)、トガリネズミ (shrew) とともに真食虫類に入ることが強く支持されるが、ハリネズミはモグラよりもトガリネズミに近い関係になる。また、アフリカ獣類のなかでは、テンレックがゾウよりもツチブタ (aardvark) に近くなる。このことは、表 3 および図 5 で仮定したトガリネズミ/モグラ、ゾウ/ツチブタのグルーピングと矛盾しているので、表 4 では、真食虫類、アフリカ獣類のなかでそれぞれ可能な 3 通りの系統関係、あわせて 9 通りの系統樹について、ミトコンドリア蛋白質データの解析を行なった。表 4 の Tree-1 は、表 3 の Tree-1 (図 5) と同じものであるが、いずれのモデルでも最尤系統樹になる。ツチブタがゾウよりもテンレックに近くなるという系統樹のうちで、Tree-3 はどちらのモデルでもどの検定でも棄却されず、Murphy et al. (2001b) のデータによる解析とは矛盾がないが、Murphy データが示しているハリネズミがモグラよりもトガリネズミに近くなるという系統樹 (Tree-6, 7, 9) は完全に棄却されてしまう。真食虫類の内部系統に関するこの矛盾の原因としては、(1) ハリネズミのミトコンドリア蛋白質の速い進化速度とアミノ酸組成の独特の偏りがモデルに十分反映されていないためのミトコンドリア系統樹の誤りと、(2) 逆に Murphy データの解析に使われたモデルの方になんらかの欠陥があることが考えられるが、いずれが本当の原因であるかについては、今後の研究に待たなければならない。

5. おわりに

本稿では、生物系統学の一つの問題について、異なる遺伝子の解析結果が異なる結論を導いているいくつかの例を、脊椎動物の系統進化において取り上げた。遺伝子の系統樹が生物種の系統樹と食い違う原因としては、祖先集団における多型、遺伝子重複に伴うパラローガスな比較、遺伝子の水平伝播などさまざまなことが考えられるが (例えば、長谷川・岸野 (1996))、ここではモデルのミスマスペシフィケーションによる系統樹推定の偏りの問題を中心に議論した。分子系統樹推定における最尤法の利点の一つは、データが蓄積するにつれてより現実的なモデ

ルを構築することが可能になり, しかもその際 AIC (Akaike (1974)) のように, 考え得るモデルのうちでどれが一番よいモデルであるかを判定する客観的な規準があるということであろう。これまでのモデルは, 基本的にはどの系統においても同じような分子進化の様式に従った置換が起こることが仮定されていたが, 最近のコバリオン・モデルのように変異可能な座位が系統によって違ってくるというモデルや (Penny and Hasegawa (2001), Penny et al. (2001), Galtier (2001)), 塩基組成の偏りが系統によって違ってくる (Galtier et al. (1999)) などといった, より現実的なモデルが開発されつつある。このようなモデルの開発を進めることによって, より信頼性の高い分子系統樹推定法が得られるようになるであろう。

謝 辞

本稿は, 統計数理研究所共同研究 (13-共研-2041), 統計数理研究所・プロジェクト研究, 日本学術振興会科学研究費・基盤研究 (B) (#12554037) の研究成果の一部をまとめたものである。

参 考 文 献

- Adachi, J. and Hasegawa, M. (1996a) MOLPHY Version 2.3: Programs for molecular phylogenetics based on maximum likelihood, *Comput. Sci. Monographs*, No. 28, The Institute of Statistical Mathematics, Tokyo.
- Adachi, J. and Hasegawa, M. (1996b) Model of amino acid substitution in proteins encoded by mitochondrial DNA, *Journal of Molecular Evolution*, **42**, 459–468.
- Akaike, H. (1974) A new look at the statistical model identification, *IEEE Trans. Automat. Control*, **19**, 716–723.
- Benton, M. J. (1990) *Vertebrate Palaeontology*, Unwin Hyman, London.
- 曹 纓, 長谷川政美 (1998) 最尤系統樹の探索法, *統計数理*, **46**, 401–409.
- Cao, Y., Sorenson, M. D., Kumazawa, Y., Mindell, D. P. and Hasegawa, M. (2000) Phylogenetic position of turtles among amniotes: Evidence from mitochondrial and nuclear genes, *Gene*, **259**, 139–148.
- Feller, A. E. and Hedges, S. B. (1998) Molecular evidence for the early history of living amphibians, *Molecular Phylogenetics and Evolution*, **9**, 509–516.
- Galtier, N. (2001) Maximum-likelihood phylogenetic analysis under a covarion-like model, *Molecular Biology and Evolution*, **18**, 866–873.
- Galtier, N., Tourasse, N. and Gouy, M. (1999) The last common ancestor to extant life forms was not hyperthermophilic, *Science*, **283**, 220–221.
- Hasegawa, M. and Adachi, J. (1996) Phylogenetic position of cetaceans relative to artiodactyls: Reanalysis of mitochondrial and nuclear sequences, *Molecular Biology and Evolution*, **13**, 710–717.
- 長谷川政美, 曹 纓 (1999) 大陸の移動と哺乳類の進化, *科学*, **69**(5), 440–448.
- 長谷川政美, 岸野洋久 (1996) 『分子系統学』, 岩波書店, 東京。
- Hasegawa, M., Kishino, H. and Yano, T. (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA, *Journal of Molecular Evolution*, **22**, 160–174.
- Hasegawa, M., Adachi, J., and Milinkovitch, M. C. (1997) Novel phylogeny of whales supported by total molecular evidence, *Journal of Molecular Evolution*, **44**(Suppl. 1), 117–120.
- 橋本哲男, 有末伸子, 長谷川政美 (2002) 分子系統樹法の応用と現状の問題点——真核生物の初期進化の解析を例として——, *統計数理*, **50**, 45–68.
- Hedges, S. B. and Poling, L. L. (1999) A molecular phylogeny of reptiles, *Science*, **283**, 998–1001.

- Kishino, H. and Hasegawa, M. (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea, *Journal of Molecular Evolution*, **29**, 170–179.
- Kishino, H., Miyata, T. and Hasegawa, M. (1990) Maximum likelihood inference of protein phylogeny, and the origin of chloroplasts, *Journal of Molecular Evolution*, **31**, 151–160.
- Krettek, A., Gullberg, A. and Arnason, U. (1995) Sequence analysis of the complete mitochondrial DNA molecule of the hedgehog, *Erinaceus europaeus*, and the phylogenetic position of the Lipotyphla, *Journal of Molecular Evolution*, **41**, 952–957.
- Kumazawa, Y. and Nishida, M. (1999) Complete mitochondrial DNA sequences of the green turtle and blue-tailed mole skink: Statistical evidence for archosaurian affinity of turtles, *Molecular Biology and Evolution*, **16**, 784–792.
- Madsen, O., Scally, M., Douady, C. J., Kao, D. J., Debry, R. W., Adkins, R., Amrine, H. M., Stanhope, M. J., de Jong, W. W. and Springer, M. S. (2001) Parallel adaptive radiations in two major clades of placental mammals, *Nature*, **409**, 610–614.
- Mannen, H. and Li, S. S.-L. (1999) Molecular evidence for a clade of turtles, *Molecular Phylogenetics and Evolution*, **13**, 144–148.
- Mindell, D. P., Sorenson, M. D., Dimcheff, D. E., Hasegawa, M., Ast, J. C. and Yuri, T. (1999) Interordinal relationships of birds and other reptiles based on whole mitochondrial genomes, *Systematic Biology*, **48**, 138–152.
- Mouchaty, S. K., Gullberg, A., Janke, A. and Arnason, U. (2000) Phylogenetic position of the Tenrecs (Mammalia: Tenrecidae) of Madagascar based on analysis of the complete mitochondrial genome sequence of *Echinops telfairi*, *Zoologica Scripta*, **29**, 307–317.
- Murphy, W. J., Eizirik, E., Johnson, W. E., Zhang, Y. P., Ryder, O. A. and O'Brien, S. J. (2001a) Molecular phylogenetics and the origins of placental mammals, *Nature*, **409**, 614–618.
- Murphy, W. J., Eizirik, E., O'Brien, S. J., Madsen, O., Scally, M., Douady, C. J., Teeling, E., Ryder, O. A., Stanhope, M. J., de Jong, W. W. and Springer, M. S. (2001b) Resolution of the early placental mammalian radiation using Bayesian phylogenetics, *Science*, **294**, 2348–2351.
- Penny, D. and Hasegawa, M. (2001) Covarion model of molecular evolution, *Encyclopedia of Genetics* (eds S. Brenner and J. H. Miller), 473–477, Academic Press, San Diego.
- Penny, D., McComish, B. J., Charleston, M. A. and Hendy, M. D. (2001) Mathematical elegance with biochemical realism: The covarion model of molecular evolution, *Journal of Molecular Evolution*, **53**, 711–723.
- 下平英寿 (2002) ブートストラップ法によるクラスタ分析のパラッキ評価, *統計数理*, **50**, 33–44.
- Shimodaira, H. and Hasegawa, M. (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference, *Molecular Biology and Evolution*, **16**, 1114–1116.
- Shimodaira, H. and Hasegawa, M. (2001) CONSEL: A program for assessing the confidence of phylogenetic tree selection, *Bioinformatics*, **17**, 1246–1247.
- Springer, M. S., Cleven, G. C., Madsen, O., de Jong, W. W., Waddell, V. G., Amrine, H. M. and Stanhope, M. J. (1997) Endemic African mammals shake the phylogenetic tree, *Nature*, **388**, 61–64.
- Stock, D. W., Quattro, J. M., Whitt, G. S. and Powers, D. A. (1997) Lactate dehydrogenase (LDH) gene duplication during chordate evolution: The cDNA sequence of the LDH of the tunicate *Styela plicata*, *Molecular Biology and Evolution*, **14**, 1273–1284.
- Strimmer, K. and von Haeseler, A. (1996) Quartet puzzling: A quartet maximum-likelihood method for reconstructing tree topologies, *Molecular Biology and Evolution*, **13**, 964–969.
- Sullivan, J. and Swofford, D. L. (1997) Are guinea pigs rodents? The importance of adequate models in molecular phylogenetics, *Journal of Mammalian Evolution*, **4**, 77–86.
- Waddell, P. J., Cao, Y., Hauf, J. and Hasegawa, M. (1999a) Using novel phylogenetic methods to

- evaluate mammalian mtDNA, including amino acid-invariant sites-LogDet plus site stripping, to detect internal conflicts in the data, with special reference to the positions of hedgehog, armadillo, and elephant, *Systematic Biology*, **48**, 31–53.
- Waddell, P. J., Okada, N. and Hasegawa, M. (1999b) Towards resolving the interordinal relationships of placental mammals, *Systematic Biology*, **48**, 1–5.
- Wilkinson, M., Thorley, J. and Benton, M. J. (1997) Uncertain turtle relationships, *Nature*, **387**, p. 466.
- Yang, Z. (1997) PAML: A program package for phylogenetic analysis by maximum likelihood, *CABIOS*, **13**, 555–556.
- Yang, Z., Nielsen, R. and Hasegawa, M. (1998) Models of amino acid substitution and applications to mitochondrial protein evolution, *Molecular Biology and Evolution*, **15**, 1600–1611.
- Zardoya, R. and Meyer, A. (1998) Complete mitochondrial genome suggests diapsid affinities of turtles, *Proc. Nat. Acad. Sci. U.S.A.*, **95**, 14226–14231.
- Zardoya, R. and Meyer, A. (2001) On the origin of and phylogenetic relationships among living amphibians, *Proc. Nat. Acad. Sci. U.S.A.*, **98**, 7380–7383.

Model Misspecification in Molecular Phylogenetic Inference as Illustrated in Evolutionary Study of Vertebrates

Ying Cao and Masami Hasegawa

(The Institute of Statistical Mathematics)

Molecular phylogenetic inference depends on the assumed model that describes the substitution process of nucleotides or amino acids during evolution, and model misspecification can give a misleading estimate of molecular phylogeny.

During our study on the phylogenetic evolution of vertebrates, we have encountered several cases of putatively misleading trees probably due to misspecification of substitution models, and these examples are presented in this article.