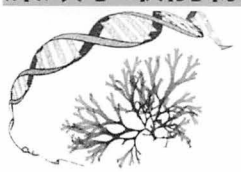
藻類 Jpn. J. Phycol. (Sôrui) 55: 117–120, July 10, 2007



坂口美亜子：複数遺伝子を用いた連結分子系統解析 －有孔虫及び有中心粒太陽虫を例として－

ある生物の単一遺伝子配列を用いて分子系統解析を行った場合，その遺伝子配列中の情報量が不十分なため，系統解析から推定された生物間の近縁関係がはっきりしないことがある。その場合，複数の遺伝子配列を連結することにより情報量を増やし，より解像度の高い分子系統解析を目指すことが重要となる。しかし，複数遺伝子を含む連結データを使った解析では，解析時に使用した配列進化モデルによって推定結果が異なることがある。また，連結データに含まれる特定の遺伝子が持つシグナルが，連結解析の全体の推定結果につよい偏りをもたらすことも考えられる。本稿では，3つの遺伝子 actin，α-tubulin，β-tubulin 配列データを用いて有孔虫類の系統的位置を検討した解析（Takishita *et al.* 2005）について解説する。この連結解析では，3遺伝子の配列進化的特性を無視したモデル（Concatenate モデル）に基づく解析と，3遺伝子の配列進化的特性を考慮したモデル（Separate モデル）に基づく解析を行った。興味深いことに，2つのモデルから推定された有孔虫類の系統的位置は著しく異なっていた。そこで Concatenate モデルと Separate モデルの違い，モデルの違いに起因する推定結果の違いについて解説し，複数遺伝子を用いた連結データ系統解析における問題点について議論する。

細胞骨格関連3遺伝子データによる有孔虫類の系統的位置の探索

今回の解析対象である有孔虫類は，主に石灰質からなる殻を持ち，細い仮足を網目状に伸ばすアメーバ状原生生物である。我々の研究以前に行われた単一分子系統解析の結果では，有孔虫類とケルコゾアとの近縁関係が示唆されていた（Keeling 2001; Berney & Pawlowski 2003; Longet *et al.* 2003）。また，これまで調べられたすべての真核生物のうち，有孔虫類とケルコゾアだけがポリユビキチンモノマー間に 1~2 アミノ酸残基のインサーションをもつことが分かっている。従って，分子系統解析とポリユビキチンモノマー間のインサーションは，共通して2つの生物群が進化的姉妹群であることを示唆する。さらに，小サブユニット（SSU）rRNA 遺伝子解析と actin 遺伝子解析とのコンセンサス系統樹から，有孔虫，ケルコゾア，そして太陽虫類の一部（Desmothoracida, Taxopodida），放散虫類は，単系統グループ（Rhizaria）を形成すると提唱された（Nikolaev *et al.* 2004）。我々は有孔虫類 *Planoglabratella opecularis* から actin，α-tubulin，β-tubulin 遺伝子を決定し，GenBank データベース中の別の有孔虫類 *Reticulomyxa filosa* の配列データと合わせて連結分子系統解析を行い，真核生物における有孔虫類の系統的位置について検討した。

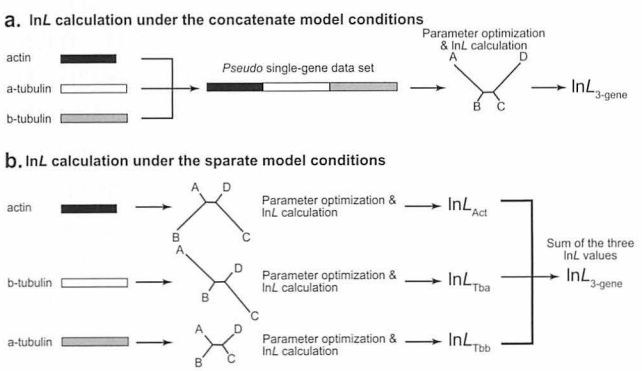


図1　2つのモデルによる連結分子系統解析　(a) Concatenate モデルでは，複数の各遺伝子アラインメントを連結させ，仮想的単一遺伝子データに対し，1セットのパラメーターの推定と対数尤度（$\ln L_{3\text{-gene}}$）計算を行う。(b) Separate モデルでは，actin，α-tubulin，β-tubulin 配列毎に1セットのパラメーターの推定と対数尤度（$\ln L_{Act}$, $\ln L_{Tba}$, $\ln L_{Tbb}$）の計算を行い，それぞれの遺伝子の対数尤度の総和（$\ln L_{3\text{-gene}}$）が最大となる系統樹を求める。actin，α-tubulin，β-tubulin 系統樹は，タクサ A，B，C，D の関係は同じだが，各遺伝子配列に基づき最適化された枝長が異なるはずである。

Concatenate モデルと Separate モデル：　今回の解析ではアミノ酸配列に基づくベイズ法の解析を行った。ベイズ・最尤法により連結データを解析する際，Concatenate（または Linked）モデルと Separate（または Unlinked）モデルによる解析方法が考えられるが，これらのモデルでは系統解析において，パラメーターを最適化する方法が異なる。ここでの「パラメーター」とは，配列データから最適化する系統樹の枝長と置換モデルパラメーター（例えばアライメント座位間の置換速度差をガンマ分布によりモデル化する際のパラメーター）を指す。ここでは，タクサ A，B，C，D から構成される4 taxon tree を考え，actin，α-tubulin，β-tubulin 遺伝子が連結データを構成すると仮定する（図1）。

Concatenate モデルによる解析（Concatenate 解析）では，3種類の単一遺伝子アライメントを連結し，その仮想的単一遺伝子データをもとに1セットのパラメーターの最適化と系統樹の対数尤度の計算を行う（図1a）。最終的に，探索した系統樹の中から最大の尤度をもつものを最尤系統樹として選択する。Concatenate 解析では，既存の分子系統解析プログラム（例えば PHYLIP，PAUP*）を用いて，最尤系統樹の自発的推定（heuristic tree search）が可能である。

Separate モデルによる解析（Separate 解析）では遺伝子配列を連結せず，系統樹のパラメーターを単一遺伝子配列毎に最適化し対数尤度を計算する（図1b）。図の例では，それぞれの遺伝子に1セットのパラメーターを設定することに

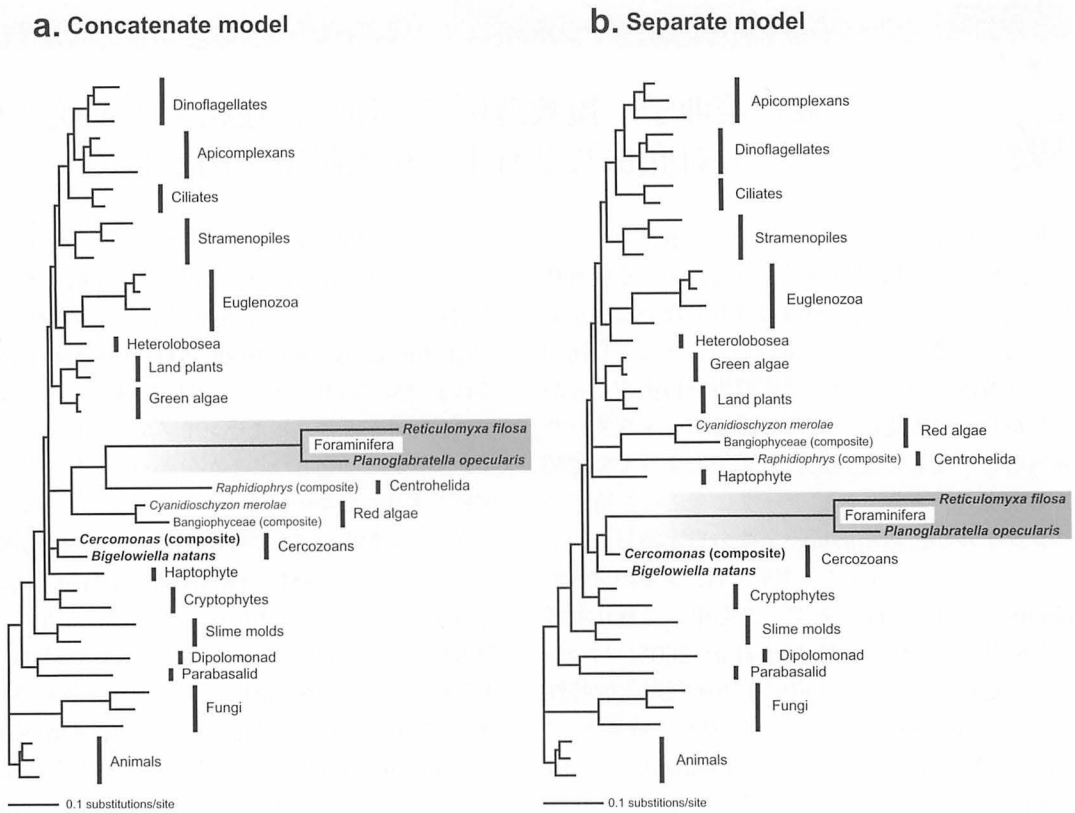


図2　ベイズ法による actin ＋α -tubulin ＋β -tubulin 連結データ解析　(a) Concatenate モデルによる解析結果：ロングブランチである有孔虫類配列と有中心粒太陽虫類配列が単系統となったが，ロングブランチアトラクションアーティファクトだと考えられる。(b) Separate モデルによる解析結果：有孔虫類配列は有中心粒太陽虫類配列でなく，ケルコゾア配列と単系統となった。

なる。その系統樹の連結データにもとづく対数尤度は，3つの単一遺伝子から計算された対数尤度の総和とし，その「総和」対数尤度が最大となる系統樹を最尤系統樹として選ぶ。つまり，Separate 解析からの最尤系統樹は，連結データに含まれる複数の単一遺伝子配列から「平均的に」もっとも支持される樹形となり，各単一遺伝子配列から推定される最尤系統樹と同一である必要はない。また，同一連結データを解析しても，Separate モデルからの最尤系統樹と Concatenate モデルからの最尤系統樹が必ずしも同じにはならない。現在の最尤法系統解析プログラムには Separate モデルをもちいた自発的樹形探索は実装されていないが，ベイズ法プログラム MRBAYES ではそれが実装されている。本研究での Separate モデルに基づく自発的推定には MRBAYES v.3.0 (Ronquist & Huelsenbeck 2003) を使用した。

2つのモデルを用いた解析：　actin，α -tubulin，β -tubulin 遺伝子から構成される「Act + Tba + Tbb」データセット（合計914アミノ酸ポジション）を Concatenate 解析したところ，有孔虫類は有中心粒太陽虫類 *Raphidiophrys contractilis* と姉妹群となった（図2a）。この推定結果は，これまでの単一遺伝子配列にもとづく系統解析や，2つのグループに特異的なポリユビキチン配列中のアミノ酸残基のインサーションから推定された，有孔虫類－ケルコゾア間の近縁性と整合性がない。我々は *Raphidiophrys* ポリユビキチン遺伝子を解析したが，モノマー間にアミノ酸インサーションは存在しなかった（Takishita *et al.* 2005）。また，有中心粒太陽虫類の系統的位置はいまだに不明であり（Sakaguchi *et al.* 2005），有孔虫類との近縁性はこれまで積極的にサポートされていない。注目すべき点は，有孔虫類配列と有中心粒太陽虫類の枝長である。有孔虫類配列は，解析した配列のうち最もロングブランチとなった。また有孔虫類配列と比べると短いが，有中心粒太陽虫類配列もロングブランチである。従って，この有孔虫類–有中心粒太陽虫類クレードは LBA アーティファクトである可能性がある。シミュレーション解析からの知見では，最尤法（およびベイズ法）で著しい LBA アーティファクトが観察される場合，モデル不整合が起こっている可能性が高い。もしこの仮説が正しいとすると，今回の Concatenate モデルは，Act + Tba + Tbb データセット中の重要な配列進化特性を十分に記述できなかった可能性が高い。反対に，Concatenate 解析でミスモデルされた配列進化特性を考慮する解析では，LBA アーティファクトである有孔虫類–有中心粒太陽虫類クレードは復元されず，（正しいであろう）有孔虫類とケルコゾアとの単系統性が復元されるはずである。

Concatenate 解析で考慮されていない Act + Tba + Tbb データセット中の配列進化特性のなかでも，最も重要であ

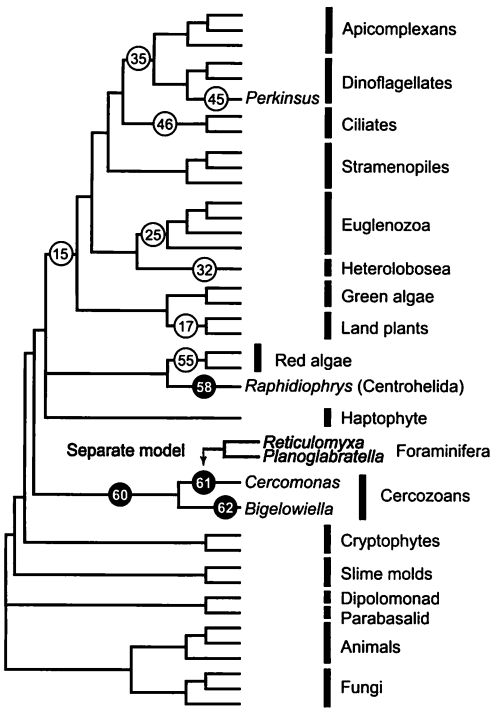


図3　AUテストの結果　Separate解析で得られた樹形をもとに，可能性のある全ての系統的位置に有孔虫類配列を付け替えた系統樹（71個）を作成し，AUテストを行った。対立仮説のうち，P値が0.01以上のもの白丸，P値が0.05以上のものは黒丸で系統樹上に示した。対数尤度はSeparateモデルを用いて計算した。

ると考えられるのはactin，α-tubulin，β-tubulin特異的な配列進化パターンである。これら3つの遺伝子配列が受ける機能的・構造的制約は本質的に互いに異なる。しかし，Concatenate解析では3遺伝子配列全体から1セットのパラメーターを推定しており，その連結データ全体に平均化されたパラメーターは，遺伝子特異的な進化パターンには十分フィットしていない。ここに強烈なモデル不整合が発生している可能性があると考え，遺伝子配列毎にパラメーターを最適化するSeparate解析をおこなった。

Separate解析では，有孔虫類は有中心粒太陽虫類とクレードを形成せず，その代わりにケルコゾアと姉妹群となった（図2b）。ここで復元された有孔虫類–ケルコゾアクレードは，これまで蓄積されてきた真核生物系統に関する知見とも矛盾しない。Separate解析でも有孔虫類・有中心粒太陽虫類配列は，ともにロングブランチであるが，Concatenateモデルよりも Separateモデルと連結データの配列進化パターンとの適合性が良くなったため，著しいモデル不整合が起こらず，LBAアーティファクトの影響を排除できたと考えられる。

Act + Tba + Tbbデータセットの2つの解析からは，有孔虫類–ケルコゾア単系統がもっともらしいと考えられるが，より客観的にSeparateモデルとCocatenateモデルとの評価ができないだろうか。もし，何らかの基準によりどちらか一方のモデルが他方よりもAct + Tba + Tbbデータセットの解析にふさわしいと判断されれば，より「ふさわしい」モデルに基づく推定結果が，より尤もらしいと考えられる。

パラメーター数の異なる統計モデルを使用したとき，そのモデルを比較する基準として赤池情報量基準（Akaike information criterion or AIC; Akaike 1974）が使用される。AIC値は以下の通りに計算できる：AIC = -2（最大対数尤度）＋2（モデルで推定されるパラメーター数）。前式に従い比較すべき複数のモデルに対するAICを計算し，最小のAIC値をもつモデルを「もっとも適切なモデル」として選ぶことが出来る。今回Act + Tba + Tbbデータセットの解析に用いたConcatenate及びSeparateモデルのAICの値を計算した結果，前者ではAIC = 38964.75，後者ではAIC = 37989.09となり，Concatenateモデルに対してSeparateモデルの優位性が示された。従ってAICに基づくモデル選択の観点からも，Separate解析で復元された有孔虫類–ケルコゾア単系統が，Concatenate解析からの有孔虫類–有中心粒太陽虫類単系統性よりも尤もらしいと考えられる。

AUテスト：　Separate解析によって有孔虫類–ケルコゾアの姉妹群が復元された。このSeparate解析から推定された系統樹と，有孔虫類–有中心粒太陽虫類単系統性をふくめて他の可能性を示唆する系統樹との間に有意差があるかどうかを確率値（P値）で評価するため，AUテストを行った（図3）。AUテストに使用するアライメントサイト毎の対数尤度データはTREE-PUZZLE v.5.2（Schmidt *et al.* 2002）を使用し計算した。この尤度データをもとにCONCEL v.0.1（Shimodaira & Hasegawa 2001）を用いてAUテストを行った。その結果では，対立仮説となる系統樹のP値が$P < 0.05$であれば5%の有意水準で棄却される。

AUテストの結果，Separate解析から推定された系統樹における，有孔虫類とケルコゾアの*Cercomonas*との姉妹群関係（樹形No. 61）に対し，同じケルコゾアの*Bigelowiella natans*と姉妹群となる可能性（樹形No. 62）やケルコゾアのクレードの根元に位置する可能性（樹形No. 60）は否定できなかった。また，Concatenate解析からの推定と同じ有孔虫類が有中心粒太陽虫類と姉妹群関係となる可能性（樹形No. 58）も同じく否定されなかった。従ってAct + Tba + Tbbデータセットは，有孔虫類と有中心粒太陽虫類との近縁性も否定しない。

まとめ：　Act + Tba + Tbbデータセットを Separate解析した場合，有孔虫類とケルコゾアとの近縁性が示された。一方，同じ連結データをConcatenate解析した場合，有孔虫類と有中心粒太陽虫類が単系統となった。これまでに蓄積された3つの生物群に関する知見，AICに基づく系統解析モデルの評価を総合すると，Separate解析による推定結果が尤もらしいと考えられる。一方，Concatenate解析では，機能・構造が異なる3遺伝子の特異的配列進化パターンを無視したために著しいモデル不整合が発生し，ケルコゾアと近縁であるはずの有孔虫類が有中心粒太陽虫類とLBAアーティファクトによりグルーピングしてしまったと考えられる。

連結データ解析の落とし穴

系統樹のより「深い」分岐の解像度を得るために，複数の遺伝子配列を連結した解析が数多く行われているが，多くの連結解析ではConcatenateモデルが使用されている。その原因の1つは，連結データのConcatenate解析には単一遺伝子系統解析の手法をそのまま当てはめることで推定結果を得ることができることであろう。ところが本稿で解説したように，Concatenate解析には必ずモデル不整合が発生し，LBAアーティファクトの下地を提供することになる。実際Kolaczkowski & Thornton (2004) のシミュレーション解析でも，Concatenate解析の推定は著しい偏りを示すことが実証されている（稲垣の稿参照)。特に，LBAアーティファクトの強度とデータサイズとは正の比例関係にあることに注意すべきである（稲垣のモデル不整合条件下でのシミュレーション解析参照)。従って，より頑健な系統推定を目指しデータサイズを増加させたにも関わらず，モデル不整合が存在するConcatenate解析ではアーティファクトがブートストラップ解析やAUテストできわめて強く支持されることになる。連結データ解析は「諸刃の刃」であり，きわめて慎重なモデル選択を行わない限りアーティファクトが高い支持を受ける可能性がある。

今回の連結データ解析では，有孔虫類と有中心粒太陽虫類配列が極端なロングブランチであることから，これらの配列が単系統となる樹形はLBAアーティファクトであることが分かりやすかった。しかし，系統解析におけるアーティファクトの原因はLBAだけではないのは自明である。例えば，配列間でのアミノ酸・塩基組成の偏りやコドン使用頻度の平行進化などから引き起こされるアーティファクトは，必ずしも2組のロングブランチを必要としない（例えばInagaki & Roger 2006)。従って，解析データ組成の慎重なチェックは偏りの少ない系統推定を行うためには必須である。

最後に，連結データ解析における解析対象とする遺伝子の選択についてコメントしたい。最近の解析では，巨大連結データから「進化速度の遅い」遺伝子やアライメント座位を選択的に選ぶ傾向がある（例えばHackett *et al.* 2007)。確かに進化速度の速い遺伝子・座位は，通常の置換モデルでは対応できない配列進化が蓄積している可能性があり，その影響を排除することはもっともらしい手法である。しかし，ほとんどの解析では，進化速度の速い遺伝子・座位の削除した解析に対する，対照解析（実験）が行われていない。例えば，高進化速度の遺伝子・座位のみを解析すれば，これらの遺伝子・座位からのアーティファクトが高いサポートで再現されるはずである。また進化速度に関係なく，ランダムに遺伝子・座位を削除し，高進化速度の遺伝子・座位の削除とは異なる推定結果になるかどうか，厳密にチェックするべきである。分子系統解析は本質的に実験科学であるはずだが，残念なことに基礎中の基礎である対照実験を行い，本実験の結果との比較・検討を行っている論文があまりにも少なすぎる印象がある。このままでは，特定の遺伝子・座位を排除するという連結データの解析手法は，都合の良い結果を得るための意図的な操作であるとの批判を招きかねない。

分子系統解析は，生物進化の道筋を解明することを目標にしている実験科学であり，唯一の真実を目指すための推定手法の妥当性・客観性は極めて重要である。従って，不適切な解析をして得た推定は，我々の目指す真実とはほど遠い。また，客観性に乏しい系統解析の前提・推定手法から得た結果は信用できない。ましてや結果が自分の仮説に適合しない，解析法が理解できない等の理由で，適切な手法を用いた解析を行わない，あるいは認めないという態度は科学者として問題がある。

引用文献

Akaike, H. 1974. A new look at the statistical model identification. IEEE Trans. Automat. Contr. 19: 716–723.

Berney, C. & Pawlowski, J. 2003. Revised small subunit rRNA analysis provides further evidence that Foraminifera are related to Cercozoa. J. Mol. Evol. 57: S120–S127.

Hackett, J. D., Yoon, H. S., Li, S., Adrian-Prieto, A., Rümmele, S. E. & Bhattacharya, D. 2007. Phylogenomic Analysis Supports the Monophyly of Cryptophytes and Haptophytes and the Association of 'Rhizaria' with Chromalveolates. Mol. Biol. Evol. (in press).

Inagaki, Y. & Roger, A. J. 2006. Phylogenetic estimation under codon models can be biased by codon usage heterogeneity. Mol. Phylogenet. Evol. 40: 428–434.

Keeling, P. J. 2001. Foraminifera and Cercozoa are related in actin phylogeny: two orphans find a home? Mol. Biol. Evol. 18: 1551–1557.

Kolaczkowski, B. & Thornton, J. W. 2004. Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. Nature 431: 980–984.

Longet, D., Archibald, J. M., Keeling, P. J. & Pawlowski, J. 2003. Foraminifera and Cercozoa share a common origin according to RNA polymerase II phylogenies. Int. J. Syst. Evol. Microbiol. 53: 1735–1739.

Nikolaev, S. I., Berney, C., Fahrni, J. F., Bolivar, I., Polet, S., Mylnikov, A. P., Aleshin, V. V., Petrov, N. B. & Pawlowski, J. 2004. The twilight of Heliozoa and rise of Rhizaria, an emerging supergroup of amoeboid eukaryotes. Proc. Natl. Acad. Sci. USA 101: 8066–8071.

Ronquist, F. & Huelsenbeck, J. P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572–1574.

Sakaguchi, M., Nakayama, T., Hashimoto, T. & Inouye, I. 2005. Phylogeny of the Centrohelida inferred from SSU rRNA, tubulins, and actin genes. J. Mol. Evol. 61: 765–775.

Schmidt, H. A., Strimmer, K., Vingron, M. & von Haeseler, A. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18: 502-504.

Shimodaira, H. and Hasegawa, M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17: 1246–1247.

Takishita, K., Inagaki, Y., Tsuchiya, M., Sakaguchi, M. & Maruyama, T. 2005. A close relationship between Cercozoa and Foraminifera supported by phylogenetic analyses based on combined amino acid sequences of three cytoskeletal proteins (actin, α-tubulin, β-tubulin). Gene 362: 153–160.

(筑波大学)