統計数理（1998）
第46巻 第2号 401-409

特集「計算と最適化」
［研究詳解］

最尤系統樹の探索法

統計数理研究所　曹　　纓・長谷川政美

（受付 1998年4月20日；改訂 1998年7月31日）

要　　旨

最尤法による分子系統樹推定法の問題点を議論した．ここでは特に，種の数が増えた場合に可能な系統樹の数が爆発的に増えるという問題を解決するために提案されている，局所再配置法とQuartet-Puzzling法という2つの方法について，実際のデータへの応用を通じて，それぞれの問題点を議論した．

キーワード：分子系統樹，最尤法，系統樹探索，局所再配置法，Quartet-Puzzling法．

1. はじめに

現在地球上には，細菌類からわれわれヒトに至るまで実に多種多様な生物種が生息しており，その数は数千万種にものぼるといわれている．これらの生物種の見かけ上の多様性にもかかわらず，DNAを基本とした遺伝的な仕組みが，調べられた範囲ではあらゆる生物で共通であることから，これらすべての生物は，もとをたどれば一つの共通祖先から由来したものであると考えられる．従って，地球上のあらゆる生物は，一本の巨大な系統樹のなかに位置づけられるはずである．このような作業をおこなう学問分野が，生物系統学である．

生物系統学は，単に地球上における生命の歴史を記載したり，われわれ自身の属する種であるヒトの歴史を明らかにするだけのものではなく，多種多様な生物種がどのような進化機構に基づいて生じてきたかを明らかにするためにも必要なものである．従来，生物系統学は主として現存生物やすでに絶滅した生物の化石などの形態を比較することによって行われてきた．ところがこのような比較形態学には研究者の主観的な判断の入る余地が多く，どのような形態的な特徴を重要であると考えるかによって，研究者の間で結論が分かれてしまうことが多かった．しかも，そのような論争を決着させるための客観的な基準が得られにくいということもある．分子系統学は，生物系統学に客観的な基準を持ち込むものととらえることができる（長谷川・岸野（1996））．これにより，形態学からは予想もできなかったような系統関係が真実である可能性が示唆されたり（Hasegawa et al. (1997), Penny and Hasegawa (1997)），形態学からは十分な情報が得られないような初期の生物進化における系統学的な問題についても，ある程度信頼できる答えを与えることができる（Iwabe et al. (1989), Hashimoto and Hasegawa (1996), Martin et al. (1998)）．しかしながら，既存の分子系統樹推定法はまだまだ未熟な段階にあり，かなり多くのデータに基づいても，いつも正しい結論を与えるとは限らない（Cao et al. (1997, 1998b, 1998c)）．

分子系統樹の推定は，現存生物から得られるDNAデータをもとにして過去の歴史を推定するということである．そもそも系統関係をしっかりと把握しない限り，ほとんどの進化的な議論は意味を持たないのである．しかしながら，直接確かめることのできない過去の進化の歴史を，現存生物のDNAデータから推測しようとするのだから，問題は簡単ではない．分子系統樹

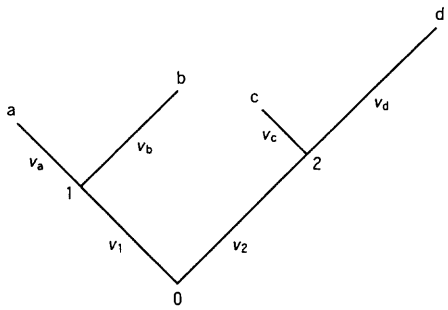


図1. 尤度の計算に用いた系統樹の例.

を推定するにあたって，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の法則に関する何らかの仮定が必要である．分子進化過程には，確率的な側面が重要である．なぜならば，DNA上の突然変異は多かれ少なかれランダムに起こるものであり，突然変異遺伝子が集団に広まって固定するのも，機会的な浮動による場合が多いと考えられるからである（木村(1986)）．このように，確率的な進化過程の産物である現存生物のDNAや蛋白質の配列データから，進化の歴史を推定することは，まさに統計的な推測になる．このような問題に対する統計科学における標準的なやり方が，確率モデルに基づく最尤法による解析である（長谷川・岸野（1996））．

分子系統樹の推定に最尤法を用いることを最初に提案したのはNeyman (1971)であったが，これをDNAの塩基配列データの実際の解析に適用されるかたちに定式化したのは，Felsenstein (1981)であった．

例えば，DNAの各塩基座位が互いに独立に同じ確率で置換する(i.i.d.)と仮定できる場合には，全体の尤度は各座位の尤度の積で表されるから，一つの座位について尤度が書き下せれば充分である．この場合，図1の系統樹について，尤度Lは次のように与えられる．

$$L=\sum_{s_0}\sum_{s_1}\sum_{s_2}\pi_{s_0}P_{s_0s_1}(v_1)P_{s_1s_a}(v_a)P_{s_1s_b}(v_b)P_{s_0s_2}(v_2)P_{s_2s_c}(v_c)P_{s_2s_d}(v_d) \tag{1.1}$$

ただし，S_iは点i（分岐点あるいは現存DNA）における塩基座位の状態である．ここで例えば$P_{s_0s_1}(v_1)$は，系統樹の根元で塩基s_0(T, C, A, Gのいずれか)であった座位が長さv_1の枝0-1を経て節1（aとbの共通祖先）でs_1になる確率を表す．したがって，上の式でPを6コ掛け合わせた部分は，根元でs_0であったとして，それが節1でs_1になり，さらに現存生物のDNAであるaでs_a, bでs_bになり，また節2ではs_2になり，さらに現存のcでs_c, dでs_dになる確率を表す．分岐した後は系統ごとに進化は独立に起こると考えられるので，枝ごとの確率を単に掛け合わせるわけである．ここで$\sum$はT, C, A, Gについてたし合わすことを示す．祖先の塩基座位の状態s_0, s_1, s_2は不明だから，可能な組合せについてたし合わす．π_{s_0}は系統樹の根元の点0で塩基s_0を見いだす確率であり，定常的なモデルを考えれば，現存DNAの塩基組成値になる．このように(1.1)式は，データとして与えられている現存生物のDNAの状態s_a, s_b, s_c, s_dが，進化の結果として実現する確率を表している．

Kishino et al. (1990)は，それまでDNA塩基のデータにしか適用されなかった最尤法を，蛋白質のアミノ酸配列データにも適用できるようにした．Adachi and Hasegawa (1992, 1996a)は，最尤法による分子系統樹推定プログラム・パッケージMOLPHYを開発し，1992年以来一般に公開してきた．MOLPHYには，塩基配列解析プログラムNucMLとアミノ酸配列解析プログラムProtMLが含まれている．MOLPHYは現在世界各地の研究者によって利用さ

れており，これを使った論文も最近急速に増えてきた．しかしながら，MOLPHY にはまだまだ改良を加えなければならない余地が多く残されている．

問題は大きく分けて2つある．一つは，最尤法で系統樹を推定するには，可能な系統樹のトポロジーのうちで尤度が最大になるものを選び出さなければならないが，扱う生物種の数 n が増えると，可能な2分岐無根系統樹の数 N が

$$N=\prod_{i=3}^{n}(2i-5)=\frac{(2n-5)!}{2^{n-3}(n-3)!} \tag{1.2}$$

のように，爆発的に増えるということがある．このような困難に対処するために，Adachi らは，近似尤度法や星型系統樹分解法，局所的再配置法などを開発し，MOLPHY に組み込んだ (Adachi and Hasegawa (1992, 1996a), Adachi (1995))．しかし，急速にデータが増えている現状で実用的な方法であるためには，これだけでは不十分である．また，Strimmer and von Haeseler (1996) は，n 種のうちから4種ずつを選ぶあらゆる組合せについてそれぞれ3通りずつの系統樹を最尤法で調べ上げ，その解析をもとにして n 種の系統樹を作り上げていく Quartet-Puzzling 法という方法を提案した．しかしながら，この方法も種の数が増えると必ずしも尤度最大の系統樹を選び出すとは限らず，最尤系統樹探索法には今後解決していかなければならない問題が多い (Cao et al. (1998a), Cao (1998))．

もう一つの問題は，モデリングである．(1.1)式における P_{ij} をどのように与えるかということである．分子系統樹推定に際しては，DNA の塩基置換や蛋白質のアミノ酸置換に関する確率モデルが必要であるが，これは必然的に実際の過程の単純化になる．しかしながら，仮定したモデルが実際の過程と大きく違った場合には，偏った推定を行う可能性があり，なるべく現実の過程をうまく反映したモデルを用いることが望ましい．現実の進化過程は多様であり，それらの多様な状況をカバーするためには，パラメータ数の多い複雑なモデルが必要になる．従って，なるべく多くの事例を扱うことによって，どんな状況下で得られたデータに対しても対処できるようなモデル（モデルのセット）を追求していかなければならない．

これまでに開発されたモデルの例としては，DNA 塩基置換に関しては，HKY モデル (Hasegawa, Kishino and Yano (1985))，より一般的な可逆マルコフモデル (Adachi and Hasegawa (1996c))，アミノ酸置換に関しては，ミトコンドリア DNA にコードされた蛋白質の一般的可逆マルコフモデル (Adachi and Hasegawa (1996b))，葉緑体 DNA にコードされた蛋白質の一般的可逆マルコフモデル (Hasegawa et al., in preparation) などがある．今後は，座位ごとの進化速度の不均一性 (Yang (1994, 1996))，座位間の相関など様々な要素をモデルに取り入れていかなければならない．また，系統によって塩基組成が異なる場合があり，このことをモデルのなかにどのように取り入れていくかということもある (Hasegawa and Hashimoto (1993), Yang and Roberts (1995))．

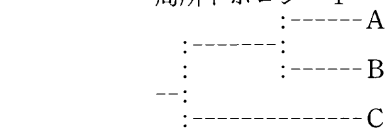
ここでは，これら2つの問題のうち，1番目の最尤系統樹探索法について，現在ひろく使われている2つの方法を紹介する．

2. 局所再配置法

これは，MOLPHY (Adachi and Hasegawa (1996a)) に組み込まれている最尤系統樹探索法の1つである．まず，距離行列法の一つである近隣結合法 (Saitou and Nei (1987)) やその他の簡便な系統樹推定法によって，最尤系統樹のトポロジー探索のための初期系統樹（近似系統樹）を求めておく．近隣結合法は，計算が簡単なわりに，最尤系統樹のよい近似を与えることが経験的に知られている (Hasegawa and Fujiwara (1993))．この方法で得られた系統樹が，

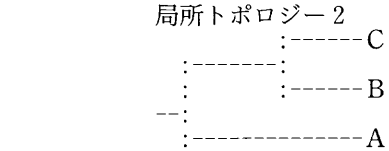
最尤系統樹のよい近似になっているとすれば，この系統樹に近いトポロジーだけを探索すれば十分であろう．初期系統樹の枝のうち，直接に現存生物につながらない内部の枝のまわりは，どれをとっても次のような構造になっている．

局所トポロジー1

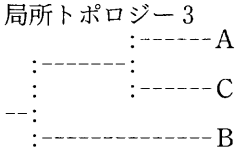


ここで，A，B，Cは部分系統樹である．この部分の局所再配置によって，次のような2つの局所トポロジーが可能である．

局所トポロジー2



局所トポロジー3



これら3つのトポロジーについて，尤度を比較して，より尤度の高いトポロジーに変えていく．初期系統樹のすべての内部枝について，このような再配置によって，より尤度の高いトポロジーを求めるわけである．この手続きを一通り行ってみると，以前は最適であった局所トポロジーが，まわりの再配置によって，さらに再配置を要するようになることもあるだろう．したがって，この手続きは，すべての内部枝が最適の状態，つまり局所再配置によっては尤度が改善されなくなるまで続けなければならない．

さらに，隣り合った2つ，3つ，あるいは4つの枝で最適なトポロジーの対数尤度と別のトポロジーの対数尤度の差が，その標準誤差（Kishino and Hasegawa (1989)）に比べて小さくて，分岐の順番がはっきりしない場合には，関連する15，105，あるいは945通りの局所トポロジーについて，最尤法による検討を行う(拡張局所再配置法)．これによって，純粋の局所再配置だけでは陥りやすい局所最適の可能性を少しでも減らすことができるであろう（Adachi (1995)）．

こうして得られた系統樹は，尤度に関しては初期系統樹を改良したものであることは確かであるが，必ずしもこれが(1.1)式で与えられる数の可能な系統樹のうちで尤度が最大であるという保証はない．いずれにしても，初期系統樹に依存した系統樹なのである．そのため，いろいろな初期系統樹を用いて局所再配置を行い，得られた系統樹のなかで一番尤度の高いものを選ぶということが必要である．いろいろな初期系統樹を得るためには，ブートストラップ(Felsenstein (1985))で得られたサンプルデータについて近隣結合系統樹を推定したり，次に紹介するQuartet-Puzzling法のpuzzlingステップで得られる多数の系統樹を用いることなどが考えられる．

3. Quartet-Puzzling法（QP法）

この方法は，Strimmer and von Haeseler (1996) によって開発されたものであり，3つのステップから成る．まず，n種のなかから4種，i，j，k，l，を選び出す$\binom{n}{4}$通りの組合せすべてについて，それぞれの4つ組に可能な3つの無根系統樹，$((i,j),k,l),((i,k),j,l),((i,l),j,k)$，を最尤法で解析する．次に，得られた$\binom{n}{4}$個の4つ組最尤系統樹を組み合わせて$n$種の系統樹を作り上げる．これがpuzzlingステップと呼ばれるものである．このステップでは，すでに作られた部分木に新たな種をランダムな順番で付け加える．加える種の順番が，A，B，C，D，E，

... とする．4つ組 (A, B, C, D) の最尤系統樹を核として，n 種の系統樹を作っていくわけである．4つ組 (A, B, C, D) の最尤系統樹が図2 (a) のようであったとすると，隣接関係 (neighbor relation) が，$AB\|_{ml}CD$ であると定義する．

新たな種 E を付け加える際には，次のようにする (図2)．E を含むすべての4つ組の隣接関係が最尤法によって，$ij\|_{ml}kE$ のようにすでに決められているが，この場合，もちろん E は i と j とを結ぶパスの上に置くべきではない．すべての4つ組 (i, j, k, E) について，E を置くべきでない枝に印を付けておく．こうして部分木のすべての枝にペナルティの点数が割り振られる．すべての4つ組 (i, j, k, E) について調べ上げて，最もこの点数が低くなるような場所に E を置く．このように一種づつ順次付け加えていって，n 種から成る系統樹が得られる．一般には，結果は種を付け加えていく順番に依存するので，puzzling ステップは，ランダムな順番でなるべく多数回繰り返されなければならない．

最終ステップでは，puzzling ステップで得られた多数の n 種系統樹から，多数決による consensus の系統樹として Quartet-Puzzling (QP) 系統樹が得られる (Margush and McMorris

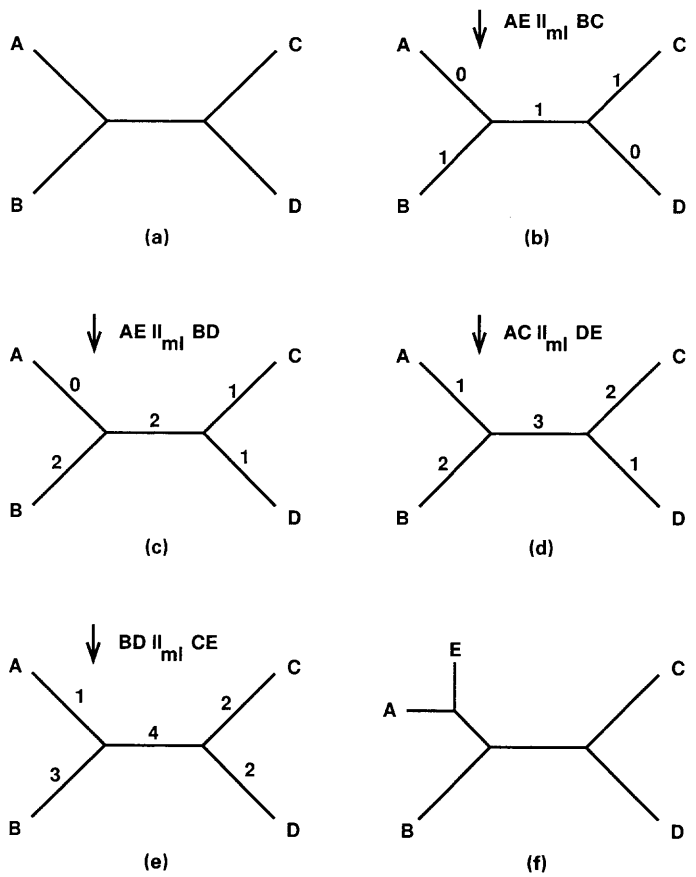


図2．新たな種 E を4つ組系統樹 ($(A, B), C, D$) (a) に付け加える方法 (Strimmer and von Haeseler (1996) の Fig. 2 を改変)．E を含む4つ組の隣接関係が，(b) $AE\|_{ml}BC$, (c) $AE\|_{ml}BD$, (d) $AC\|_{ml}DE$, (e) $BD\|_{ml}CE$ のようになっているとする．それぞれの隣接関係を考慮すると，E を置くことに対するペナルティの点数が図のように加算されていく．E を含むすべての4つ組の隣接関係を考慮に入れて得られた点数 (e) が一番少ない枝 A に E を置く (f)．

(1981)).

puzzling ステップで特定のグルーピングが得られた回数は，ブートストラップ確率と同様に，データがもつ系統関係に関する情報の量を反映していると考えられる．Strimmer and von Haeseler (1996) は，これを QP 系統樹の“信頼度”と呼んでいる．彼らのシミュレーションによれば，この信頼度はブートストラップ確率と高い相関をもっているという．

4. 具体例についての方法の比較

Strimmer and von Haeseler (1996) は，様々な状況でのシミュレーションを行って，近隣結合法，QP 法，exhaustive な最尤法でそれぞれどのような割合で真の系統樹が再現されるかを調べた．一般に QP 法は近隣結合法よりもよい成績をあげるが，進化速度が系統ごとに大きく異なり，しかも多重置換の効果が大きな状況では，exhaustive な最尤法よりも成績が大きく劣ることが分かった．彼らは，応用例としてこの QP 法を Hedges (1994) の用いた脊椎動物ミトコンドリアの 12S rRNA, 16S rRNA, それに $tRNA^{Val}$ の遺伝子 DNA 配列データに，Janke et al. (1994) によるオポッサム *Didelphis virginiana* のデータと Janke et al. (1996) によるカモノハシ *Ornithorhynchus anatinus* のデータを付け加えたものに適用した．

われわれは，彼らのデータに対してまず，あらゆる 2 種の組み合わせについて最尤法で距離を推定し，それをもとに近隣結合法 (Saitou and Nei (1987)) により系統樹を推定した (Cao et al. (1998a))．得られた系統樹は図 3 の左と同じトポロジーのものであり，これは Strimmer and von Haeseler の QP 法によって得られたものと一致した．彼らによると，この系統樹では，*Sceloporus* (トカゲ) が鳥類/ワニ類/ムカシトカゲ (*Sphenodon*) の作るグループの姉妹群になっており，この関係の QP 法による信頼度は，94%もの高いものになる (HKY モデルによる)．ところが，この系統樹で他の関係を固定した場合に，ムカシトカゲ，トカゲ，鳥類/ワニ類，そ

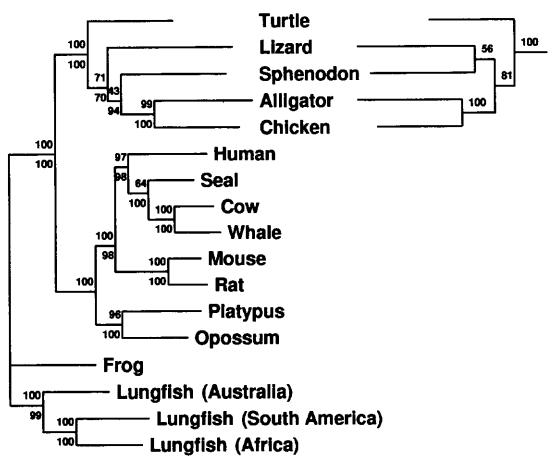


図 3. 左側：ミトコンドリアの 12S リボソーム RNA+16S リボソーム RNA+$tRNA^{Val}$ の塩基配列データから近隣結合法によって推定された系統樹トポロジーについて MOLPHY の NucML プログラムによって枝の長さ (塩基置換数) を推定したもの (HKY モデル，$\alpha/\beta=2.32$)．これは，Strimmer and von Haeseler (1996) によって QP 法で得られたトポロジーと一致している．右側：近隣結合系統樹から出発して局所再配置法によって得られた最尤系統樹．近隣結合系統樹と異なる部分だけを示した．枝の上に示した数字は局所ブートストラップ確率(%)，下の数字は QP 法による信頼度．Cao et al. (1998a) より．

れにアウトグループの間で可能な3通りの系統樹について，最尤法でブートストラップ確率(系統樹の他の部分の関係を固定しているので，局所的ブートストラップ確率と呼ばれる；Adachi and Hasegawa (1996a)) を求めると，43% (図3左) にしかならない．Strimmer and von Haeseler (1996) は，QP法の信頼度はブートストラップ確率と高い相関があると主張しているが，彼らが例題として取り上げたこの問題では，彼らの主張は成り立っていない．

図3の左のトポロジーを初期系統樹として，局所再配置法でより尤度の高い系統樹を探索してみたところ，図3の右の系統樹が得られた．この図で省略してある部分は，図3の左と同じ系統関係になっているということである．実は，尤度最大の系統樹は，QP法で得られた図3の左とは違っていて，ムカシトカゲがトカゲ類と系統的に一つのグループを作っているというものであった．ただし，この系統樹の対数尤度は，図3の左に比べて1.6±7.8 (±は1標準誤差であり，Kishino and Hasegawa (1989) の式によって推定された) 低いだけであり，尤度の規準ではほとんど区別できない．このことは，QP法で得られたこの系統関係に関する94%という信頼度は，実際よりも誇張されたものであることを示している．

この例では，QP法が尤度最大に近い系統樹を選び出すことはできるが，計算量がはるかに少ない，近隣結合系統樹から出発して局所再配置法によってより尤度の高い系統樹を探索するという方法のほうが，効率よく尤度最大の系統樹を探しあてることができるのである．局所再配置法が有効に働くためには，初期系統樹として最尤系統樹に近いものがあらかじめ得られていなければならないが，そのための一つの方法として，QP法は有望であろう．

5. おわりに

最尤系統樹探索法にはいまだにここで述べたようなheuristicなものしかなく，今後はもっと理論的なバックグラウンドをもった方法の開発が必要である．最適化の一つの問題として，多くの研究者がこの方面の研究に参加されることが望まれる．

参考文献

Adachi, J. (1995). Modeling of molecular evolution and maximum likelihood inference of molecular phylogeny, Ph. D. Dissertation, The Graduate University for Advanced Studies, Tokyo.

Adachi, J. and Hasegawa, M. (1992). MOLPHY: Programs for molecular phylogenetics, I; PROTML: Maximum likelihood inference of protein phylogeny, *Comput. Sci. Monographs*, No. 27, The Institute of Statistical Mathematics, Tokyo.

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.

Adachi, J. and Hasegawa, M. (1996c). Tempo and mode of synonymous substitutions in mitochondrial DNA of primates, *Molecular Biology and Evolution*, **13**, 200-208.

Cao, Y. (1998). Molecular phylogeny and evolution of vertebrates, Ph. D. Dissertation, Tokyo Institute of Technology.

Cao, Y., Okada, N. and Hasegawa, M. (1997). Phylogenetic position of guinea-pigs revisited, *Molecular Biology and Evolution*, **14**, 461-464.

Cao, Y., Adachi, J. and Hasegawa, M. (1998a). Comment on the quartet puzzling method for finding maximum-likelihood tree topologies, *Molecular Biology and Evolution*, **15**, 87-89.

Cao, Y., Janke, A., Waddell, P., Westerman, M., Takenaka, O., Murata, S., Okada, N., Pääbo, S. and Hasegawa, M. (1998b). Conflict amongst individual mitochondrial proteins in resolving the

phylogeny of eutherian orders, *Journal of Molecular Evolution*, **47**, 307-322.
Cao, Y., Waddell, P., Okada, N. and Hasegawa, M. (1998c). The complete mitochondrial DNA sequence of the shark *Mustelus manazo*: evaluating rooting contradictions to living bony vertebrates, *Molecular Biology and Evolution* (in press).
Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach, *Journal of Molecular Evolution*, **17**, 368-376.
Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap, *Evolution*, **39**, 783-791.
Hasegawa, M. and Fujiwara, M. (1993). Relative efficiencies of the maximum likelihood, maximum parsimony, and neighbor-joining methods for estimating protein phylogeny, *Molecular Phylogenetics and Evolution*, **2**, 1-5.
Hasegawa, M. and Hashimoto, T. (1993). Ribosomal RNA trees misleading?, *Nature*, **361**, p. 23.
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. (1997). Novel phylogeny of whales supported by total molecular evidence, *Journal of Molecular Evolution*, **44** (Suppl. 1), 117-120.
Hashimoto, T. and Hasegawa, M. (1996). Origin and early evolution of eukaryotes inferred from the amino acid sequences of translation elongation factors 1α/Tu and 2/G, *Advances in Biophysics*, **32**, 73-120.
Hedges, S. (1994). Molecular evidence for the origin of birds, *Proc. Nat. Acad. Sci. U.S.A.*, **91**, 2621-2624.
Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S. and Miyata, T. (1989). Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes, *Proc. Nat. Acad. Sci. U.S.A.*, **86**, 9355-9359.
Janke, A., Feldmaier-Fuchs, G., Thomas, W., von Haeseler, A. and Pääbo, S. (1994). The marsupial mitochondrial genome and the evolution of placental mammals, *Genetics*, **137**, 243-256.
Janke, A., Gemmell, N., Feldmaier-Fuchs, G., von Haeseler, A. and Pääbo, S. (1996). The complete mitochondrial genome of a monotreme, the platypus (*Ornithorhynchus anatinus*), *Journal of Molecular Evolution*, **42**, 153-159.
木村資生（1986）．『分子進化の中立説』（木村資生，向井輝美，日下部真一 訳），紀伊國屋書店，東京．
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.
Margush, T. and McMorris, F. (1981). Consensus n-trees, *Bulletin of Mathematical Biology*, **43**, 239-244.
Martin, W., Stoebe, B., Goremykin, V., Hansmann, S., Hasegawa, M. and Kowallik, K. (1998). Gene transfer to the nucleus and the evolution of chloroplasts, *Nature*, **393**, 162-165.
Neyman, J. (1971). Molecular studies of evolution: a source of novel statistical problems, *Statistical Decision Theory and Related Topics* (eds. S. Gupta and J. Yackel), 1-27, Academic Press, New York.
Penny, D. and Hasegawa, M. (1997). Molecular systematics: the platypus put in its place, *Nature*, **387**, 549-550.
Saitou, N. and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees, *Molecular Biology and Evolution*, **4**, 406-425.
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.
Yang, Z. (1994). Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods, *Journal of Molecular Evolution*, **39**, 306-314.
Yang, Z. (1996). Among-site rate variation and its impact on phylogenetic analyses, *Trends in Ecology and Evolution*, **11**, 367-372.
Yang, Z. and Roberts, D. (1995). On the use of nucleic acid sequences to infer early branchings in the tree of life, *Molecular Biology and Evolution*, **12**, 451-458.

Methods for Searching the Maximum Likelihood Tree

Ying Cao and Masami Hasegawa
(The Institute of Statistical Mathematics)

Phylogenetic knowledge is indispensable in biology, and molecular phylogenetics has become an important tool in inferring evolutionary history of organisms. Among many methods used in inferring molecular phylogeny, the maximum likelihood method has a sound statistical ground, but its computational burden due to a huge number of possible trees for even a moderate number of species prevents its application to a large number of species. In this article methods for searching the maximum likelihood tree among a huge number of possible trees are reviewed, and their efficiency were compared by using a real data.

Key words: Molecular phylogeny, maximum likelihood, topology search, local-rearrangement method, Quartet-Puzzling method.