

分岐分類学に基づく形質進化の最節約復元

三 中 信 宏

(農林水産省農業環境技術研究所)

系統推定における形質復元とは？：

分岐分類学・最節約性・アルゴリズム

系統推定とは、生物の形質データから進化史を復元することである。進化史の復元とは、系統樹の樹形（トポロジー）の復元と系統樹上の仮想的共通祖先に配置する形質状態の復元という2つの問題を含む。

樹形推定問題に対しては、距離法や形質状態法に属するいくつかの方法が利用でき（Swofford & Olsen, 1990）、とりわけ分子データに基づく系統推定では対立する系統推定法の相対的有効性をめぐる論議が今も活発である。一方、形質復元問題に関しては、その解決法を与えているのは、分岐分類学すなわち最節約法だけである。しかし、現在の進化生物学では、系統樹の樹形すなわち生物間の系統関係だけではなく、その系統樹の上での個々の形質進化の復元が要求される状況が増えてきている（Brooks & McLennan, 1991; Harvey & Pagel, 1991）。

しかし、形質復元ができる唯一の方法である最節約法においてさえ、形質復元にともなう問題点がこれまで十分に議論されてこなかったのは皮肉なことである。本論に入る前に、その歴史的理由について考えてみよう。第一の理由は、従来の分岐分類学における形質復元が、Willi Hennig (1966) のいう系統解析の論証法 *scheme of argumentation* という古いアルゴリズムの中に埋没してしまって、形質復元の論理が十分に鍛えられてこなかったことである。

分岐分類学の本来の目的は、30年前の論証法をかたくなに遵守することではなく、あくまでも派生形質の共有に基づく単系統群の発見である。Hennig 自身、「立証された固有派生形質の数が多いほど、その派生形質を共有する群が単系統群である確実性が増す」（Hennig, 1966）および「複数種に分布する、派生的であると確実に解釈できる形質（ただの形質ではなく！）の数が多いほど、これらの種が単系統群を作るという仮定を支持する裏付けは強くなる」（Hennig, 1982）と主張している。Hennig のこの一連の発言は、分岐分類学に基づく系統推定の論理的基礎が、ある目的関数の最適化であることを示唆している。その目的関数が最節約性 *parsi-*

mony にほかならないことは、Farris *et al.* (1970) がすでに指摘している。

系統樹の全長を最小化する（すなわち系統樹全体にわたってホモプラシーを最小化する）という最節約規準の達成が分岐分類学（最節約法）の目的であるとしたら、それを実行するアルゴリズムとしては何も30年前の論証法だけにこだわる必要はない。分岐分類学において論証法」にとって代わる改良されたアルゴリズムが1970年代以降議論されるようになったのは、当然のことである。

その結果として、形質進化方向性の判定や分岐図間の選択そして形質復元など分岐分類学に基づく系統推定プロセスの諸要素が最節約原理に照らして次々と再検討されていった。ただ、これらの再定式化の意義が十分に理解されているかどうかははなはだ心もとない。最節約原理を中心に据えるという点で、初期の Hennig 理論とパターン分岐分類学の理論とは対立するという見解を聞くことがある（三中 1993 a,b ; 直海 1993）。しかし、最節約原理は無から生じたものではなく、ほかならない Hennig の設定した目標（彼が考案したアルゴリズムではなく！）を完全に達成するための強力な道具としてこれまで利用されてきたことは明らかである（三中 1989）。

分岐分類学の中で形質復元問題がこれまで注目を集めなかった第二の理由は、その問題を解析的に解くための思考枠が成熟していなかったという点である。Farris (1970) は、最節約法のもとで系統樹を構築するにあたって、仮想的祖先の形質状態を最節約的に復元するためのアルゴリズム（Farris 最適化と呼ばれる）を最初に与えた。ちょうどその頃、Robinson (1972) は、数学的なグラフ理論の立場から、与えられたグラフのもとで、ある目的関数を最適化するように内部分岐点の値を決定する方法を論じた。これら2つの研究は、系統分類学とグラフ理論というまったく別の領域に属していたために、両者が実は同一の問題を異なる言語で論じていただけだということが当時は気づかれなかった。実際、最節約法のもとで形質復元問題を効率的に解くためには（さらに最節約分岐図を探索するためには）、後述するように、グラフ理論や数値計画法の諸手法が必要であ

る。しかし、グラフ理論などの離散数学が最も関心を持っていたのは、最短グラフ樹形の探索（スタイナー問題）であって、その部分問題である祖先形質復元問題にはほとんど注意を払ってこなかった。系統分類学者が解こうとしていた形質復元問題は、スタイナー問題の影に隠れてしまったのである（三中 1993c）。

スタイナー問題と MPR 問題：

最節約法における二つの最小化

形質状態法に属する系統推定法は、ある目的関数を最適化（最大化または最小化）するような樹形を探索する（Swofford & Olsen, 1990）。最節約法であれば、全長（あるいは全コスト）という目的関数を最小化し、最尤法ならば尤度という目的関数を最大化する。

用語の定義から始めよう。系統樹上の隣接 2 頂点を結ぶ線を枝と呼ぶ。頂点には、末端点と内部分岐点の 2 種類がある。末端点は実際にその形質状態が観察できる操作的分類単位（OTU：operational taxonomic unit）に、そして内部分岐点はその形質状態が未知の仮想的分類単位（HTU：hypothetical taxonomic unit）に、それぞれ対応する。

2 つの形質状態 a から b への状態遷移にともなうコスト（進化回数が直感的に最もわかりやすいコストの定義だが、それ以外にも形質状態遷移確率のある関数値としてコストを定義することがある：後述）を $\text{cost}(a; b)$ と表記する。

形質状態 a から b へのコストを表わす具体的な関数形としては、以下のものが利用されている。

1) マンハッタン距離： $\text{cost}(a; b) = |a - b|$

（Farris, 1970; Hanazawa *et al.*, 1995; Maddison & Slatkin, 1990; Robinson, 1972; Sankoff & Rousseau, 1975; Swofford & Maddison, 1987, 1992）,

2) 平方ユークリッド距離： $\text{cost}(a; b) = (a - b)^2$

（W. Maddison, 1991; McARDLE & Rodrigo, 1994; Robinson, 1972）,

3) 配列間距離： $\text{cost}(a; b) = \begin{cases} = 0 & (a = b \text{ のとき}) \\ = 1 & (a \neq b \text{ のとき}) \end{cases}$

（Fitch, 1971; Hartigan 1973）。

もっと一般的なコスト関数を用いることも可能である。その際は、形質状態の順序対ごとのコスト値は形質×形質のステップ行列 *stepmatrix* として表現できる（D. Maddison, 1994; Maddison & Maddison, 1992; Swofford & Begle, 1993; Swofford & Olsen, 1990）。

いま、ある系統樹の第 j 枝 e_j の両端点である頂点を v_j, v'_j とし、第 i 形質に関する v_j, v'_j の形質状態をそれぞれ $x(v_j; i), x(v'_j; i)$ と書く。この系統樹に含まれる

m 個の HTU の形質状態をベクトルとして $X = (x_1, x_2, \dots, x_m)$ と表わし、このベクトルをこの系統樹における HTU 形質状態の復元と呼ぶ。このとき、枝 e_j の長さは、各形質のその枝での状態遷移にともなうコストの合計、すなわち

$$\sum_i \text{cost}[x(v_j; i); x(v'_j; i)]$$

と定義できる。系統樹全体にわたって枝の長さを合計した値をこの系統樹の全長と呼ぶ。式で書けば、

$$\sum_j \left[\sum_i \text{cost}[x(v_j; i); x(v'_j; i)] \right]$$

となる。この式は、全長を個々の形質ごとに分割して別々に復元を考えればよいことを意味している。

復元の値（ベクトルの成分）を動かしたときに得られる全長の最小値をその系統樹の最小全長と定義し、最小全長を与える復元を最節約的復元 *most parsimonious reconstruction*（略して MPR）と呼ぶ（Swofford & Maddison, 1987）。OTU 形質状態行列と系統樹の樹形を与えたときに、HTU 仮想形質状態の最節約的復元を決定する問題を MPR 問題（Hanazawa *et al.*, 1995）と呼ぶ。

系統樹ごとに MPR 問題を解いて最小全長を決定できたならば、引き続いて可能な系統樹の樹形全体にわたって最小全長を比較し、最小全長を最小化する系統樹（スタイナー系統樹）を発見することが原理的には可能である。この最小全長の最小化はスタイナー問題と呼ばれる最短グラフ探索問題のカテゴリーに属しており、多項式計算時間では解決できないことが証明されている。スタイナー問題の観点から見れば、MPR 問題は樹形が固定されたときの部分問題とみなすことができるが、計算量の上では格段のちがいがあ。MPR 問題の計算量は OTU 数の多項式オーダーであるから（Hanazawa *et al.*, 1995）、スタイナー問題よりはるかに少ない計算量で解決できる。

最節約的形質復元のアルゴリズム

以下の説明は、三中（1995）に従う。ある系統樹の任意の HTU（たとえば R とする）を基準点に指定すると、系統樹全体は R を根とする複数の有根部分木に分割できる。ある頂点 q を直接祖先とする分岐点 p の 2 つの子孫分岐点を r, s とし、各頂点に配置する形質状態値をそれぞれ m, i, j, k とする（図 1）。形質状態遷移 $x \rightarrow y$ のコストを $\text{cost}(x; y)$ と表記する。 p, r, s を根とする部分木（クレード）をそれぞれ T_p, T_r, T_s と表わし、各分岐点での形質状態に対応するそれらの部分木の全長をそれぞれ $L(p; i), L(r; j), L(s; k)$ と表わす。

とき、 p の MPR 形質状態が決定できる。以下、末端点に向かって逐次的に最適化を反復すれば、基準点以外のすべての HTU の MPR 形質状態が決定できる。

上のアルゴリズムは、形質の種類には依存しない。しかし、上の定式化を踏まえて、個々の形質の種類ごとに、簡便な方法が開発されている。順序型形質の場合、形質状態変化のコストをマンハッタン距離で定義すると、ある HTU の特性区間は、それよりも末端の部分木にある複数 HTU（または OTU）の特性区間のメジアン演算（複数の閉区間の上端と下端の数値を大小順に並べ替えた数列の中央の 2 つの値を端点とする閉区間を求める演算）によって計算される（Farris, 1970; Hanazawa *et al.*, 1995; 三中 1995; 成嶋ら 1993; Robinson, 1972）。また、無順序型形質では、形質状態の区間ではなく集合によって MPR が求まる。このとき、メジアン演算の代わりにモード演算（複数の形質状態集合での最大出現頻度を持つ形質状態からなる集合を求める演算）が用いられる（Fitch, 1971; Hartigan,

$$L(p ; i) = \min_j [\text{cost}(i ; j) + L(r ; j)] + \min_k [\text{cost}(i ; k) + L(s ; k)].$$
$$L(p ; i) = \text{cost}(i ; j) + \text{cost}(i ; k)$$
$$\min_{\rho} L(R ; \rho)$$

2) 昇路ステップ

$$\min_i [\text{cost}(u ; i) + L(p ; i)]$$

(1)

A diagram showing a quantum network with four input nodes (A, B, C, D) and two output nodes (E, F). Nodes A and B are on the left, C and D on the right. A and B are connected to a central node X, while C and D are connected to a central node Y. X and Y are connected to each other. Arrows indicate the flow of information from inputs to outputs. The diagram is labeled (1) at the top.

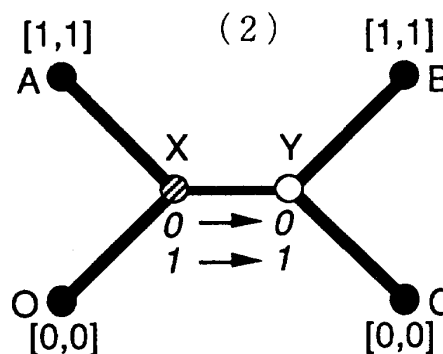


図 2. 最節約復元の 2 段階ステップ (三中 1995).

●:OTU, ○:HTU

1973; Steel, 1993). 一方、連続的実数値をとる形質の場合は、上のマンハッタン距離または平方ユークリッド距離 (W. Maddison, 1991; Robinson, 1972) を用いれば、同様の復元が可能である。

MPR 多重解と HTU 復元の歪み

前節では、与えられた系統樹の樹形のもとで最小全長を与える最節約復元の求め方について論じた。しかし、それは最節約復元が一意的に決定されることを意味するわけではない。ホモプラシーが含まれる現実のデータでは、仮想的共通祖先への最節約復元状態配置が複数解得られることがしばしばある。これを MPR 多重解問題

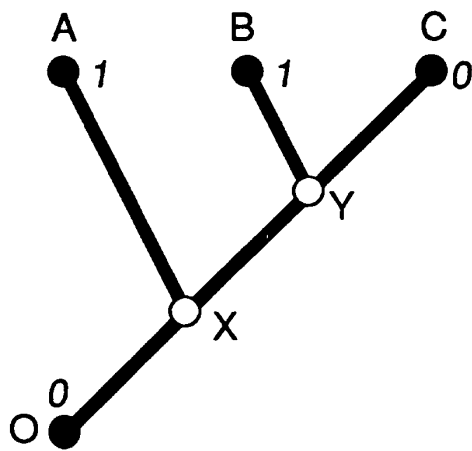


図3 HTU 形質状態復元の仮想例 (三中, 1995).
●:OTU, ○:HTU

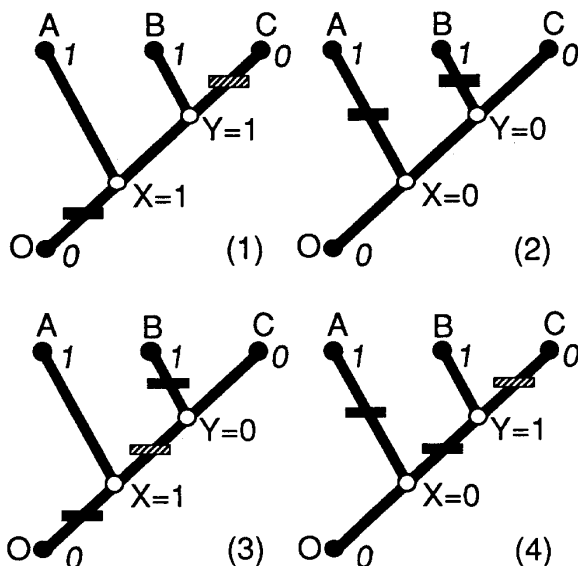


図4. 図3に対する可能なすべての HTU 復元配置 (三中 1995).

- (1) ACCTRAN 配置;
(2) DELTRAN 配置;
(3), (4) 非最節約復元配置.
黒短線は0→1遷移を, 縞短線は1→0遷移を表す。

と呼ぶことにする。図3の例では、ある二値的形質 (0/1) に関して、外群Oを含む OTU 集合 {A, B, C, O} のうち、A, Bが1, C, Oが0を持つとする。外群Oで根付けられた有根系統樹 $T = ((O, (A, (B, C))))$ にはXとYの2つの HTU がある (図3)。この2つの HTU にどのような形質状態を配置するかによって、2つの MPR が決まる (図4-1, 2)。どちらの MPR も最小全長2を与える。図4の中の太短線は形質状態の変化を表す。

図4-1の MPR は、根にすぐ続く枝で0→1が、そして末端枝の一つで逆転1→0が生じている。一方、図4-2は形質変化 (0→1) を根からできるだけ遠い末端部分で生じさせる MPR であり、前者の型の MPR は形質変換促進最適化配置 (ACCTRAN 最適化配置)、後者は形質変換遅延最適化配置 (DELTRAN 最適化配置) と呼ばれる。もっと大きなデータでは ACCTRAN, DELTRAN 以外の MPR も存在し得る (Swofford & Maddison, 1987)。たとえば、図5aの場合は、上述のアルゴリズムにしたがって形質状態を最節約復元すると、図5bのようになり、全部で8通りの最節約復元が導かれる。ACCTRAN と DELTRAN を含む MPR

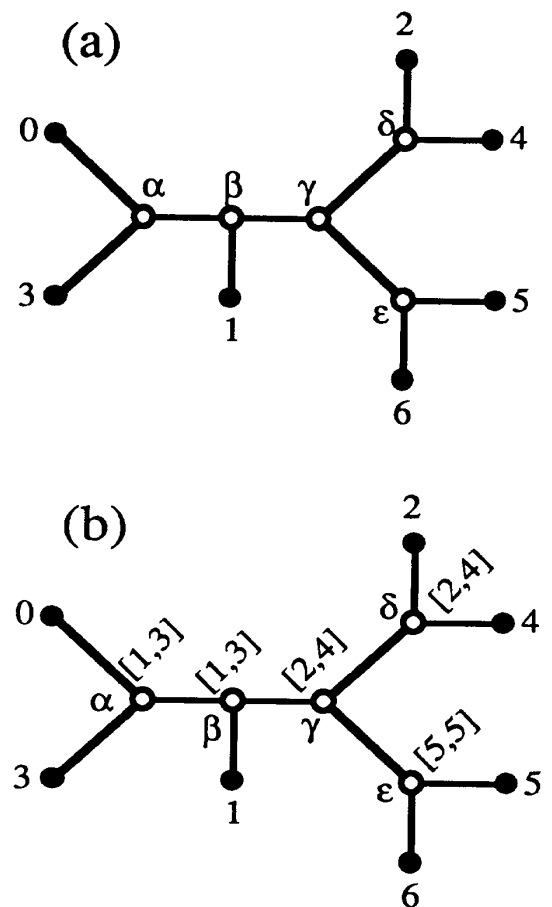


図5. 整数値を形質状態とするある仮想的順序型形質の最節約復元。

HTU の MPR 集合は閉区間で表わされる。

多重解は、どれも全長に関しては定義によりまったく差がない。そのため、全長ではない別の観点に立って、MPR 多重解間の比較をする必要がある。

図3では、 $T_1 = (A, (B, C))$ および $T_2 = (B, C)$ という2つの部分木がある。ACCTRANでの長さは $L(T_1) = 1, L(T_2) = 1$ であるのに対し、DELTRANでは $L(T_1) = 2, L(T_2) = 1$ である(図4-1, 図4-2)。順序型形質では、ACCTRANは任意の部分木の長さを最小化する唯一のMPRであることがすでに証明されている(Misheva & Narushima, 1994)。したがって、DELTRANを含む非ACCTRANはどれかの部分木がACCTRANのそれよりも長くなる。あるHTU復元Rにおける各部分木(全体系統樹も含めて)の長さ $L(T; R)$ とACCTRANでの対応する長さとの偏差の合計値を歪み指数 distortion index (Minaka, 1993) と呼ぶことにする。すなわち、歪み指数

$$I_D = \sum_T \{ L(T; R) - L(T; \text{ACCTRAN}) \}$$

と定義される(図6)。上の例でのACCTRANの歪み指数は $I_D(R_1) = 0$ (定義により)、DELTRANのそれは $I_D(R_2) = (2-2) + (2-1) + (1-1) = 1$ である。非MPRには、 R_3, R_4 の2つが存在するが(図4-3, 図4-4)、歪み指数はそれぞれ $I_D(R_3) = 2$ および $I_D(R_4) = 3$ である。

HTU復元の歪みは、ゴム紐のアナロジーによって視覚的に説明できる。図7では、図5の例でのACCTRANとDELTRANを3次元的に表示した。縦軸は形質値である。HTU(○)の最節約復元形質値のMPR集合(図5b)を太点線で表した。形質状態が既知で

あるOTU(●)はこの3次元空間内のある定点とみなせるのに対し、形質状態が未知のHTUは空間内の浮動点と解釈できる。いま、定点であるOTUと浮動点であるHTUを、与えられた樹形にしたがって、仮想ゴムひもで結ぶ。このとき手を放すと、浮動点は仮想ゴムひもの張力エネルギーがすべての部分木で最小になる(歪みが最小ということ)座標を占める。これがACCTRAN復元に相当する(図7の濃い実線)。一方、非ACCTRAN、とくにその中でも最大の歪み指数をもつDELTRANは、逆にどこか特定の部分木での仮想ゴムひもの張力エネルギーが最小ではなくなっている(図7の淡い実線)。

歪み指数は、各部分木(クレード)における形質進化の集中度を反映していると考えられる。したがって、形質進化が集中しているという証拠がとくにないならば、あるいは形質進化のクレード間均等分布の帰無仮説が何らかの規準のもとで棄却されないうちは、少なくとも最初に考察されるべきMPRとしては、歪みゼロのACCTRANによるHTU形質状態の復元が最も妥当であるといえるだろう。別の言い方をすれば、最初から理由もなく非ACCTRAN復元を採用する必然性はないということである。

すべての部分木に歪みの生じないACCTRANがほかのMPRよりも優先されるという上の主張は、グラフ理論的な考察を踏まえたものだった。しかし、別の観点からACCTRANを支持すべきであるという意見が

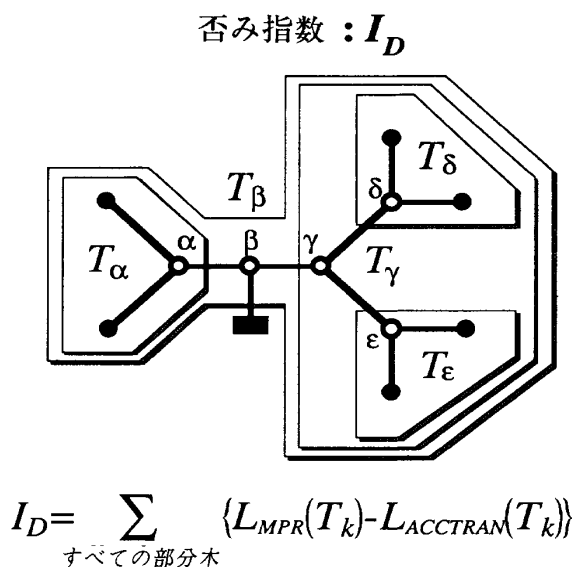


図6. 歪み指数の定義 (Minaka, (1993)を改変)。

ある根に対して規定される部分木の長さを最節約復元ごとに集計し、ACCTRAN配置の集計値との差を取ることで、ある最節約復元の歪み指数が計算できる。

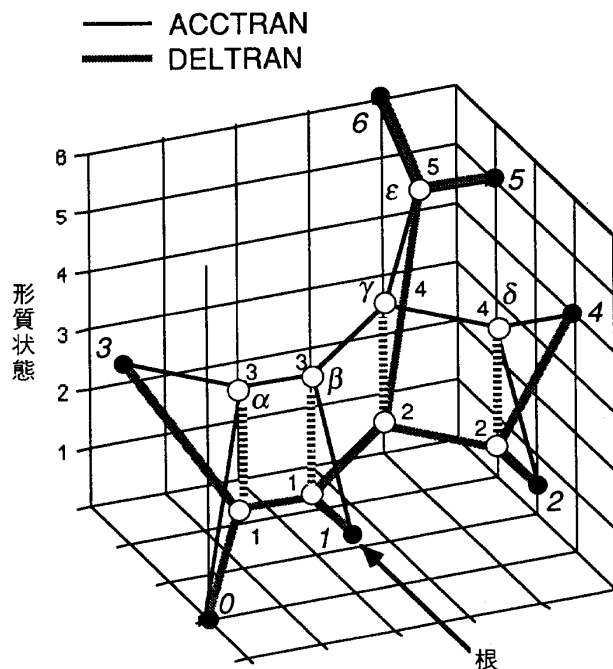


図7. 最節約復元の3次元表示。

図5から導かれる8通りの最節約復元のうち、ACCTRANとDELTRANを立体的に表示した。OTUのx,y座標値は恣意的にである。z軸の形質状態値だけに着目されたい。

最近強まってきた。たとえば、de Pinna (1991)は、ACCTRAN は派生的形質状態の共有を相同形質としてそのまま保持するが、DELTRAN は最初からホモブラスシーとして解釈してしまうから、相同性（共通祖先性）を重視する立場からは ACCTRAN の方が支持されると主張した。また、Allard (1994) は、最節約法が統計学的な一致性 consistency を持つかどうかの議論と関連づけて、DELTRAN の場合は不一致の危険性 (Huelsenbeck & Hillis (1993) のいう Felsenstein 結界に踏み込む可能性) が増大することを示した。

ベイズ推定に基づく最節約復元： 最節約性と確率モデルの接点

進化プロセスに関する仮定を系統推定にどのように組み込むかという問題は、分岐分類学に基づく系統推定と形質復元を進めていく上で、ますます重要になってきた。ここでいう形質進化モデルとは、たとえば形質状態の変化順序・遷移確率・重み付け・形質相関などの仮説である。とりわけ、形質進化モデルが構築しやすい分子レベルの配列データについては、これらの仮説のいくつかは経験的に検証可能であり、十分に裏付けられた形質進化モデルは積極的に系統推定に利用すべきだろう。

形質進化モデルを組み込む効用は次の2つである。第一に、形質進化モデルの組み込みにより、分岐分析の精度を上げることができる (Hillis *et al.*, 1994; Huelsenbeck & Hillis, 1993)。たとえば、コード領域の DNA 塩基配列では、第3コドンはほかのコドン位置と比べて置換確率が数倍大きいとされている。このとき、推定された塩基置換数の（たとえば）逆数によって各コドン位置を重みづけすれば、第3コドンの影響を小さくし、系統関係を示すシグナルの検出が容易になるだろう。

十分に裏付けられたプロセス理論をも排除してしまうと、かえって説明の最節約性が損なわれるおそれがある。たとえば、コドンごとの置換確率の差に関する十分に裏付けられた仮説を無視して各コドンをあえて等しく重みづけするためには、かえってそういう等価な重み付けを正当化するための弁明が必要だろう (Maddison & Maddison, 1992)。

モデルを組み込む第2の効用は、確率に基づく系統仮説（および系統推定法）の定量的比較が可能になるという点である。ある確率モデルを前提としたとき、与えられた OTU 形質状態のもとでその系統樹が生じる確率（事後確率）またはその系統樹のもとで観察された OTU 形質状態が生じる確率（尤度）は、系統仮説の相対的妥当性を評価する規準として有効である (Farris, 1973, 1983, 1986; Sober, 1988)。

上に挙げた形質進化モデルの多くは、後述のステップ

行列を用いて分岐分類学に組み込むことができる。ステップ行列のコストは形質状態の遷移確率から計算できる。たとえば、DNA 塩基配列データを考えよう。いま、DNA 塩基配列での塩基置換確率を任意の枝 e で塩基 i から j に置換する確率 $p(i; j)$ と表わし、置換に伴うコスト $\text{cost}(i; j)$ をその自然対数で定義することにする：

$$\text{cost}(i; j) = -\ln [p(i; j)]$$

(Albert *et al.*, 1993; Felsenstein, 1981)。コスト $\text{cost}(i; j)$ を並べたステップ行列（ 4×4 型）は、対角要素が塩基の非置換確率に、非対角要素が置換確率に相当する。制限酵素切断サイトデータでも、同様の対数変換によって形質状態遷移のコストを計算できる (Albert *et al.*, 1992)。

ステップ行列にしたがって系統樹全体にわたるコストの総和を求め、それを最小にする樹形を選ぶ系統推定法を一般化最節約法 generalized parsimony method と呼ぶ (Swofford & Maddison, 1992)。

一般化最節約法は、形質復元と樹形選択の双方に関わってくる。上のように形質状態の遷移確率の対数（の負値）を状態遷移に伴うコストと定義するならば、コストの総和を最小化する系統樹は、必ずベイズ事後確率を最大化する。同様に、MPR 問題においても、ベイズ事後確率を最大化する HTU 復元は、この一般化最節約規準のもとでの MPR にほかならない。以下では、この点について解説する。

与えられた有根樹 T のもとで、観察された OTU 形質状態データを O 、第 i 番目の HTU 復元（MPR も含む）を R_i と表わす。異なる HTU 復元は互いに排反事象であるから、 T のもとで OTU データ O が生じる確率 $P(O)$ は、

$$P(O) = \sum_i P(O \& R_i)$$

と分割できる。右辺の各項はさらに

$$P(O \& R_i) = P(R_i) P(O | R_i)$$

と変形できる。 $P(R_i)$ は HTU を結ぶ内部枝で形質変化が生じる確率、 $P(O | R_i)$ は R_i という HTU 復元のもとで末端 OTU にいたる枝で形質変化が生じる確率である。したがって、形質進化は枝の間でたがいに独立であると仮定すると、確率積 $P(O \& R_i)$ は、HTU 復元 R_i のもとでの有根樹 T 全体にわたる形質進化確率の総積である。

図4の形質復元を例にとろう。0/1という二値的状态をとる一つの形質を考え、任意の枝の両端での形質状態遷移確率を $P(0 \rightarrow 1) = a$, $P(0 \rightarrow 0) = 1 - a$, $P(1 \rightarrow 0) = b$, $P(1 \rightarrow 1) = 1 - b$ で規定する単純な形質進化モデルを

仮定する ($0 \leq a, b \leq 1$ とする). この例での可能な 4 通りの HTU 復元の確率積は次のようになる.

$$\begin{aligned} \text{ACCTRAN} (R_1) &: ab(1-b)^3; \\ \text{DELTRAN} (R_2) &: a^2(1-a)^3; \\ \text{非 MPR} (R_3, R_4) &: a^2(1-a)b(1-b). \end{aligned}$$

各 HTU 復元の生起確率曲面の 3 次元グラフ (図 8) からわかるように, この確率積の相対的な大小関係は形質状態の遷移確率 a, b の値によって変わる. OTU データ O のもとで, ある HTU 復元 R_i が生じる事後確率は, ベイズの定理により

$$P(R \mid O) = P(R_i)P(O \mid R_i) / P(O)$$

となる. 分母 $P(O)$ は定数だから, 分子の確率 $P(R_i)P(O \mid R_i)$ すなわち確率積 $P(O \& R_i)$ の値は, HTU 復元のベイズ事後確率 $P(R_i \mid O)$ の値に比例する.

MPR のベイズ的定式化は, 次の二つのことを意味する. 一つは, 与えられた樹形のもとである形質の MPR 多重解を導かれたとき, それらの MPR の妥当性をベイズ事後確率の大小によって相互比較できるという点である. これは, ステップ行列のコストと遷移確率とを連動させない場合 (たとえばコスト=進化回数と解釈するとき) の話である. もう一つは, コストを遷移確率のある関数と定義する一般化最節約法の場合 (たとえば遷移確率の対数値をコストとみなすとき), MPR 自体がベイズ事後確率を最大化する復元と同一になるという点である.

なぜ最節約復元なのか?

なぜ最節約的に復元するのか? 最も自明な理由は, 形態データ・分子データの別を問わず, 最節約法以外に形質状態の復元法が開発されていないからである. UPGMA や近隣結合法のような距離法では, 形質状態データを距離に変換した時点で HTU 形質状態を復元することは原理的に不可能になる. 一方, 現在用いられている最尤法は, 確率和 (積分) を取ることにより特定の HTU 形質状態配置への依存を回避しているから, これまた形質復元には適さない方法論である.

樹形 T と HTU 形質状態 R を与えたときに OTU 形質状態データ O が生じる確率を $P(O \mid T \& R)$ と表わすと, 最尤法では

$$\text{尤度 } L(T) = \sum_R P(R \mid T) P(O \mid T \& R)$$

と定義することにより攪乱母数の一つである R の影響を除去し, 最大尤度を与える樹形 T を探索する. それに対して, 上述の一般化最節約法では $Pr(O \mid T \& R)$ そのものを最大化する T と R を同時に推定しようとする. 最尤法と一般化最節約法との近縁性はこれまでも指摘されてきたが (Goldman, 1990; Swofford & Olsen, 1990), 樹形の選択基準のちがいは決定的である. 最尤法はすべての HTU 復元に関する和を取るが, 最節約法は最節約復元だけを考察の対象とするからである.

形質状態の最節約復元は, はたして生物学的に妥当な祖先復元をしているのだろうか? この方面の研究はあまり進んでいないが, 最近の重要な研究成果を 2 つだけ挙げておく.

1) 実験系統学から: Hillis *et al.*, (1992) は, 実験

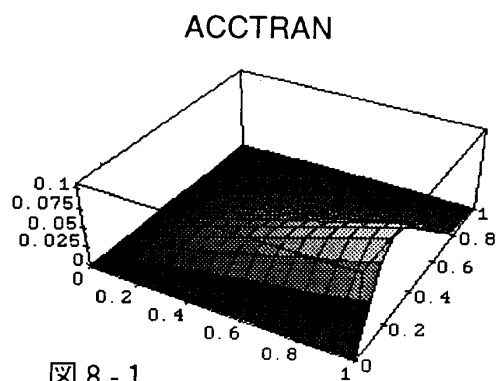


図 8-1

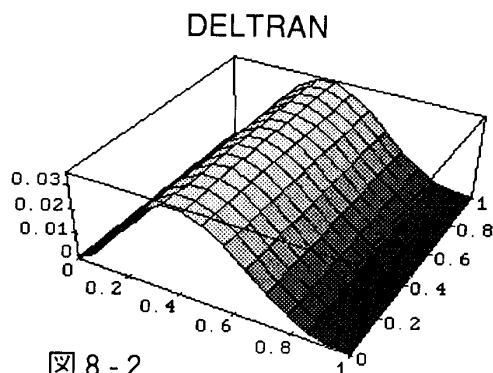


図 8-2

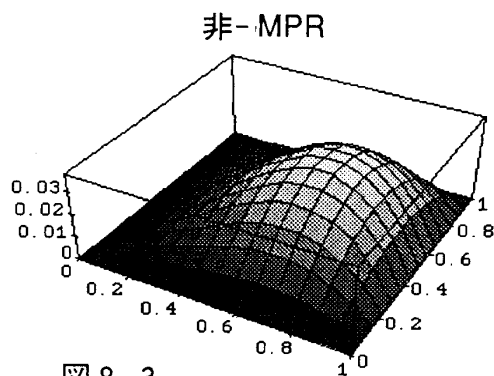


図 8-3

図 8. HTU 復元状態配置の確率曲面 (三中 1995).

本文中で説明した形質状態遷移モデルのもとで図 4 の 4 つの HTU 復元の生起確率を 3 次元的に表示した. x 軸は 0 → 1 遷移確率を, y 軸は 1 → 0 遷移確率を表わす.

室内での T₇ファージの既知系統樹と既知祖先塩基配列に照らして、最節約法に基づく HTU 復元の正答率を調べたところ 97% から 98% を越える高い値が得られた。

2) 実験生化学から: *Jermann et al.*, (1995) は偶蹄類の腓リボヌクレアーゼについて現存の DNA 塩基配列から祖先配列を最節約復元した (この過程については *Fitch* (1971) および 三中・斎藤 (1993) を参照)。さらに、その復元 DNA 塩基配列から翻訳されたポリペプチド鎖を実験生化学的に合成したところ、それらの復元リボヌクレアーゼは正常な酵素活性を有することが判明したばかりでなく、反芻動物類への進化過程で酵素活性に大きな変化が生じたことまで推定された。

これらの知見は、最節約復元が、単にアルゴリズム上の虚構ではなく、生物学的に意味のある推定を行なっていることを示唆していると私は解釈する。

では、仮想祖先形質状態の最節約復元の成果を求めているのは誰だろうか? 次に列挙するように進化生物学のさまざまな領域でそれは求められ、同時に未解決の問題も残されている。

1. 種間比較をしなければならない生態学者・行動学者は、系統樹の樹形だけではなく、HTU の最節約復元まで知らねばならないことが多い。

1-1. ある形質の系統樹上での進化回数のカウント (*Brooks & McLennan*, 1991)・形質相関の検出 (*W. Maddison*, 1989)・適応の検証 (*Coddington*, 1988; *Harvey and Pagel*, 1991) など比較生態学の諸問題の考察には系統学の知見を必要とするものが多い。ここでの最も大きな問題は、最節約復元間の比較評価である。最節約復元間の選択基準についてはさらに検討の余地が残されている。上述のいくつかの理論的根拠で ACCTRAN が支持されるが、場合によっては非 ACCTRAN がデータによって支持されることもあるだろう。たとえば、形質進化が特定のクレードに集中していることが判明したならば、非 ACCTRAN の方がより妥当だろう。

1-2. 形質復元の枚挙と形質進化仮説の検証とは明確に分けるべきだろう。この区別は、系統推定に用いた形質の進化を解析するとき問題となる。系統樹構築に関与した形質についてその HTU 形質状態配置を最節約復元すること自体は認識論的問題を生まないだろう。それは最節約的に可能な形質配置の枚挙だからである。一方、その形質の進化をそれによって推定された系統樹の上で検証することは、論理循環の問題を引き起こすはずである。したがって、ある形質進化仮説を系統樹を用いて検証したいので

あれば、その系統樹を導いた形質群は検証したい形質とは別の情報源から得るべきだろう。たとえば、ある形態形質の進化を分子系統樹の上で検証するのであれば、論理循環という認識論的問題は生じないだろう。

2. 分岐分類学者も次に挙げるさまざまな状況で、最節約復元が必要になる。

2-1. 外群規則

形質状態の方向性すなわち原始の状態と派生的状態を仮説化するための規準として分岐分類学でこれまで広く用いられてきた外群比較 outgroup comparison の基礎は、上述の最節約復元の降路ステップである。外群比較の基礎が最節約原理であることを理解すれば (*Farris*, 1982), その欠点がはっきり認識できるだけでなく、さらに方向性と樹長の同時推定という解決への道も開ける (*Maddison et al.*, 1984)。

2-2. 側系統・多系統の区別

単系統群ではない側系統群や多系統群をどのように定義し、判定するかは議論の的だった。*Farris* (1974) は群定義形質という仮想二値的形質 (群内の種は状態 1 を、群外は状態 0 を取る) の系統樹上での最節約復元に基づく側系統群と多系統群の判定規準を与え、現在では広く受け入れられている。しかし、最節約復元配置が複数生じる場合では判定が分かれることがある。*Farris* (1991) は、この多重解の問題を解決するための修正案を提出しているが、根本的な解決にはなっていない。

2-3. 共有派生形質の認識

ある単系統群を認識する規準は、共有派生形質の存在である。しかし、複数の最節約復元が存在する場合には、仮想共通祖先に最節約配置される形質状態によって共有派生形質と認識されたり、されなかったりする。これは、該当する枝の長さにそのまま反映される。最節約復元によって枝の長さが異なるということである。*Swofford & Maddison* (1987) は、ある枝の長さの最小値が非ゼロすなわちすべての最節約復元を考慮してもなおその枝において形質状態の遷移を想定しなければならないときに限り、共有派生形質状態と考えるべきであると主張する。一方、*de Pinna* (1991) は、DELTRAN 復元は共有派生形質の数を減らすだけだから、ACCTRAN 復元を選択すべきであると反論する。

2-4. 祖先地域の推定

分岐図の末端 OTU の形質としてその地理的分布を考えると、地理的分布の最節約復元による祖先地域 (HTU の持つ仮想分布域) の推定が可能にな

る。たとえば, D. Maddison (1991) と D. Maddison *et al.*, (1992) は, ヒトのミトコンドリア DNA のアフリカ起源説を検証するにあたって, 最節約復元に基づく祖先地域の推定を行なった。

3. 形質進化プロセスモデルの構築においても, 最節約復元が必要となる状況がある。仮想祖先の形質状態復元は特定の系統樹の特定のクレードに限定された推定問題である。一方, その知見に基づいて形質の進化特性の推定を行なうのは, もっと一般的な形質進化モデルの構築につながる。

3-1. ステップ行列要素の推定

一般化最節約法でのステップ行列の各要素(コストの関数値)をどのようにして推定するかは, この方法を利用するに当たって最重要問題であるにもかかわらず, 私の知るかぎりまだ決定的方法は提案されていない。とくに, 手持ちの形質データだけからステップ行列を推定し, さらに形質復元や樹形推定を論理循環なしに行なえるのかどうかは, まだ未解決である。D. Maddison (1994) は, 無順序的な形質タイプ(ステップ行列の非対角要素がすべて1)を帰無モデルとして系統樹上で最節約復元された形質状態配置からステップ行列を推定し, それが帰無モデルからどの程度外れるかを調べるという方法を提案している。ただし, 最節約復元を行なう際, 樹形への依存性をいかにして除去するかという問題が残る。

- 3-2. 祖先形質状態をパラメーターとして含む進化モデルには, 末端種の形質状態からの HTU 復元を前提とするものがある。たとえば, Farris (1973) の形質進化モデルでは, 与えられた OTU の形質状態から最節約復元された HTU 形質状態配置に基づいて, 形質の進化経路 pathway (Felsenstein, 1973) の生じる確率を計算する。Sober (1985) のモデルでは, OTU 形質状態が得られる確率をある分岐図樹形のもとで可能なすべての HTU 復元(非最節約復元まで含む)について合計し, その分岐図の尤度と解釈した。

3-3. バイアスの問題

塩基組成の不均一性が仮想的祖先形質状態の復元にバイアスをかける危険性があることが, 最近指摘されている (Collins *et al.*, 1994)。また, 何らかの仮定あるいは方法に基づいて推定されたステップ行列が系統推定にどのようなバイアスを与えるのかについては, ほとんど研究されていない (D. Maddison, 1994)。Albert *et al.*, (1993, 1994) は, 葉緑体 *rbcL* 遺伝子の塩基配列の場合, ステップ行列に基づく重み付けをしたときと何も重み付けしない無

順序的形質として処理したときでは, 得られる系統樹にそれほど差がないと述べている。しかし, 彼らの主張は, 「種子植物の *rbcL* 遺伝子では, いろいろな重み付けしてみたけれどももしないときと大差がない」という主張であり, ほかの遺伝子や生物にまですぐさま外挿できる一般的主張ではないと私は考える。

祖先形質状態の復元問題は, それを必要としている生物学者が期待するほど, 系統学者の関心を集めてはいないように私は感じている。多くの系統学者の興味は生物間の系統関係(樹形)だけに向けられているからである。この認識のずれをなくしていくことがまず先決だろう。

引用文献

- Albert, V.A., A. Backlund & K. Bremer 1994. DNA characters and cladistics: the optimization of functional history. *In* : R.W. Scotland, D.J. Siebert & D.M. Williams (eds.), *Models in Phylogeny Reconstruction*, p. 249-272. Oxford Univ. Press, Oxford.
- Albert, V.A., M.W. Chase & B.D. Mishler 1993. Character-state weighting for cladistic analysis of protein-coding DNA sequences. *Ann. Missouri Bot. Gard.* 80 : 752-766.
- Albert, V.A., B.D. Mishler & M.W. Chase 1992. Character-state weighting for restriction site data in phylogenetic reconstruction, with an example from chloroplast DNA. *In* : P.S. Soltis, D.E. Soltis & J.J. Doyle (eds.), *Molecular Systematics of Plants*, p. 369-401. Chapman and Hall, New York.
- Allard, M.W. 1994. An empirical example of parsimony behaviour. *In* : R.W. Scotland, D.J. Siebert & D.M. Williams (eds.), *Models in Phylogeny Reconstruction*, p. 231-248. Oxford Univ. Press, Oxford.
- Brooks, D.R. & D.A. McLennan 1991. *Phylogeny, Ecology, and Behavior: A Research Program in Comparative Biology*. Univ. of Chicago Press, Chicago.
- Coddington, J.A. 1988. Cladistic tests of adaptational hypotheses. *Cladistics* 4 : 3-22.
- Collins, T.M., P.H. Wimberger & G.J.P. Naylor 1994. Compositional bias, character-state bias, and character-state reconstruction using

- parsimony. *Syst. Biol.* **43** : 482-496.
- Farris, J.S. 1970. Methods for computing Wagner trees. *Syst. Zool.* **19** : 83-92.
- Farris, J.S. 1973. A probability model for inferring evolutionary trees. *Syst. Zool.* **22** : 250-256.
- Farris, J.S. 1974. Formal definitions of paraphyly and polyphyly. *Syst. Zool.* **23** : 548-554.
- Farris, J.S. 1982. Outgroup and parsimony. *Syst. Zool.* **31** : 328-334.
- Farris, J.S. 1983. The logical basis of phylogenetic analysis. In: N.I. Platnick and V.A. Funk (eds.), *Advances in Cladistics*, volume 2, p.7-36. Columbia Univ. Press, New York.
- Farris, J.S. 1986. On the boundary of phylogenetic systematics. *Cladistics* **2** : 14-27.
- Farris, J.S. 1991. Hennig defined paraphyly. *Cladistics* **7** : 297-304.
- Farris, J.S., A.G. Kluge & M.J. Eckardt 1970. A numerical approach to phylogenetic systematics. *Syst. Zool.*, **19** : 172-189.
- Felsenstein, J. 1973. Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. *Syst. Zool.* **22** : 240-249.
- Felsenstein, J. 1981. A likelihood approach to character weighting and what it tells us about parsimony and compatibility. *Biol. J. Linn. Soc.* **16** : 183-196.
- Fitch, W.M. 1971. Toward defining the course of evolution: minimum change for a specific tree topology. *Syst. Zool.* **20** : 406-416.
- Goldman, N. 1990. Maximum likelihood inference of phylogenetic trees, with special reference to a Poisson process model of DNA substitution and to parsimony analysis. *Syst. Zool.* **39** : 345-361.
- Harvey, P.H. & M.D. Pagel 1991. *The Comparative Method in Evolutionary Biology*. Oxford University Press, Oxford.
- Hanazawa, M., H. Narushima & N. Minaka. 1995. Generating most parsimonious reconstructions on a tree: a generalization of the Farris-Swofford- Maddison method. *Discr. Appl. Math.* **56** : 245-265.
- Hartigan, J.A. 1973. Minimum mutation fits to a given tree. *Biometrics* **29** : 53-65.
- Hennig, W. 1966. *Phylogenetic Systematics*. University of Illinois Press, Urbana.
- Hennig, W. 1982. *Phylogenetische Systematik*. Verlag Paul Parey, Berlin.
- Hillis, D.M., J.J. Bull, M.F. White, M.R. Badgett & I.J. Molineux 1992. Experimental phylogenetics: generation of a known phylogeny. *Science* **255** : 589-592.
- Hillis, D.M., J.P. Huelsenbeck & C.W. Cunningham 1994. Application and accuracy of molecular phylogenies. *Science* **264** : 671-677.
- Huelsenbeck, J.P. & D.M. Hillis 1993. Success of phylogenetic methods in the four-taxon case. *Syst. Biol.* **42** : 247-264.
- Jermann, T.M., J.G. Opitz, J. Stackhouse & S.A. Benner 1995. Reconstructing the evolutionary history of the artiodactyl ribonuclease superfamily. *Nature* **374** : 57-59.
- Maddison, D.R. 1991. African origin of human mitochondrial DNA reexamined. *Syst. Zool.*, **40** : 355-363.
- Maddison, D.R. 1994. Phylogenetic methods for inferring the evolutionary history and processes of change in discretely valued characters. *Annu. Rev. Entomol.* **39** : 267-292.
- Maddison, D.R., M. Ruvullo & D.L. Swofford 1992. Geographic origins of human mitochondrial DNA: phylogenetic evidence from control region sequences. *Syst. Biol.* **41** : 111-124.
- Maddison, W.P. 1989. A method for testing the correlated evolution of two binary characters: are gains or losses concentrated on certain branches of a phylogenetic tree? *Evolution* **44** : 539-557.
- Maddison, W.P. 1991. Squared-change parsimony reconstructions of ancestral states for continuous valued characters on a phylogenetic tree. *Syst. Zool.* **40** : 304-314.
- Maddison, W.P., M.J. Donoghue & D.R. Maddison 1984. Outgroup analysis and parsimony. *Syst. Zool.* **33** : 83-103.
- Maddison, W.P. & D.R. Maddison 1992. *MacClade, Version 3: Analysis of Phylogeny and Character Evolution*. Computer Software and User's Manual. Sinauer Associates, Sunderland.
- Maddison, W.P. & M. Slatkin 1990. Parsimony reconstructions of ancestral states do not depend on the relative distances between linearly-ordered character states. *Syst. Zool.* **39** : 175-178.

- McArdle, B. & A.G. Rodrigo 1994. Estimating the ancestral states of a continuous-valued character using squared-change parsimony: an analytical solution. *Syst. Biol.* 43 : 573-578.
- 三中信宏 1989. 現代進化生物学における分岐分類学: Hennig 以後の理論展開とその積極的評価. 種生物学研究, 13 : 18-44.
- 三中信宏 1993 a. 穏やかに且つ執拗に: 分岐分類学の草の根布教のバイブルを読む重層的悦楽. *Shinka* 3 : 125-138.
- 三中信宏 1993 b. 形質極性の事前決定は分岐分析の根幹ではない: 外群規則に基づく Hennig の論証法は最節約法と論理的に等価である. *Shinka* 3 : 179-184.
- 三中信宏 1993 c. 組合せ論的視点から見た系統推定: 最節約法と離散数学の接点. 千葉県立中央博物館自然誌研究報告 2 : 83-98.
- Minaka, N. 1993. Algebraic properties of the most parsimonious reconstructions of the hypothetical ancestors on a given tree. *Forma* 8 : 277-296.
- 三中信宏 1995. 分岐分析にもとづく系統推定の論理とその応用: 系統樹推定と祖先形質復元. 馬渡峻輔 (編) 動物分類学の多様な展開, 北海道大学図書刊行会
- 三中信宏・斎藤成也 1993. 系統樹作成法: 最大節約法. 新生化学実験講座第16巻: 分子進化実験法 (日本生化学会編), p. 380-395. 東京化学同人
- Misheva, N. & H. Narushima 1994. On the positions of Acctran and Deltran in the MPR-poset. 日本数学会1994年度年会応用数学分科会講演アブストラクト 35-38.
- 直海俊一郎 1993. 形質の極性決定は分岐分類学の根幹ではなくなったのか?: 三中信宏氏の書評へのコメント. *Shinka* 3 : 170-178.
- 成嶋弘・花沢正純・三中信宏 1993. Generating most parsimonious reconstructions on a tree: 数理生物学からの問題. 日本数学会1993年度年会応用数学分科会講演アブストラクト 109-112.
- de Pinna, M.C.C. 1991. Concepts and tests of homology in the cladistic paradigm. *Cladistics* 7 : 367-391.
- Robinson, D.F. 1972. Extending a function on a graph. *Discr. Math.* 6 : 89-99.
- Sankoff, D. & P. Rousseau 1975. Locating the vertices of a Steiner tree in an arbitrary metric space. *Math. Program.* 9 : 240-246.
- Sober, E. 1985. A likelihood justification of parsimony. *Cladistics* 1 : 209-233.
- Sober, E. 1988. Reconstructing the past: parsimony, evolution, and inference. xv + 265 pp. The MIT Press, Massachusetts.
- Steel, M.A. 1993. Distributions on bicoloured binary trees arising from the principle of parsimony. *Discr. Appl. Math.* 41 : 245-261.
- Swofford, D.L. & D.P. Begle 1993. PAUP: Phylogenetic Analysis Using Parsimony, Version 3.1. User's Manual. Smithsonian Institution, Washington, D.C.
- Swofford, D.L. & W.P. Maddison 1987. Reconstructing ancestral character states under Wagner parsimony. *Math. Biosci.*, 87 : 199-229.
- Swofford, D.L. & W.P. Maddison 1992. Parsimony, character-state reconstructions and evolutionary inferences. In : R.L. Mayden (ed.), *Systematics, historical ecology, and North American freshwater fishes*, p. 186-223. Stanford Univ. Press, Stanford.
- Swofford, D.L. & G.J. Olsen 1990. Phylogeny reconstruction. In : D.M. Hillis & C. Moritz (eds.), *Molecular systematics*, p. 411-501. Sinauer Associates, Sunderland.
- Wiley, E.O., D. Siegel-Causey, D.R. Brooks & V.A. Funk 1991. *The Compleat Cladist: A Primer of Phylogenetic Procedures*. Univ. of Kansas Mus. of Nat. Hist., Lawrence.