へ伸びる枝の長さは0.0118/2=0.0059であり，CP, H, G の分岐点から H と G へ伸びる枝の長さは (0.0427+0.0382+0.0416+0.0327+0.0371)/10=0.0192 であり，CPHG と O の分岐点（すなわち根）から O へ伸びる枝の長さは (0.0953+0.0916+0.0965+0.0928)/8=0.0470 である（図23・4).

この方法は，平均距離法と同様に，進化速度が系統によらず一定であると仮定しているので，進化速度が一定でない場合には使えない.

23・3 最大節約法*

最大節約法は，系統樹の“枝の長さ”の合計を最小化するという最大節約原理のもとに系統推定を行う方法である．この最大節約原理は1970～80年代にかけて生物系統学の領域で大論争をまき起こした分岐分類学（cladistics）の方法論的基礎であり，この原理の妥当性をめぐっては現在でも論議がたえない[1]，

分子進化学における最大節約法は，最初 Eck と Dayhoff[2] によってアミノ酸配列データに適用された．しかし，現在では DNA や RNA の塩基配列データや制限酵素切断地図データなどさまざまな離散的形質に幅広く適用されるようになってきた．いわゆる最大節約法とよばれるカテゴリーにはいくつかの方法論が含まれている[3),4)]が，本節では，分子データの解析において最近広く用いられているタイプの最大節約法のみを論じることにする.

分子進化学では，対象群の共通祖先（根）の位置を指定する通常の意味での系統樹ではなく，根を指定しない無根系統樹としてつくられるのが普通である．そこで本節では，最大節約無根系統樹を作成する方法について論じる．以下では，まずはじめに最大節約法において最小化されるべき量（枝の全長）を定義する．ついで，最大節約法に基づく系統推定での二つの問題，すなわち無根系統樹の樹形の推定および各枝の長さの推定という二つの問題を論じる.

---

* 執筆担当：三中信宏，斎藤成也（§23・3）

1) E. Sober, “Reconstructing the Past: Parsimony, Evolution, and Inference”, MIT Press, Massachusetts (1988).
2) R. V. Eck, M. O. Dayhoff, “Atlas of Protein Sequence and Structure 1966”, National Biomedical Research Foundation, Silver Spring (1966).
3) E. O. Wiley, “Phylogenetics: The Theory and Practice of Phylogenetic Systematics”, John Wiley & Sons, New York (1981).
4) D. L. Swofford, G. J. Olsen, “Molecular Phylogenetics”, ed. by D. M. Hillis, C. Moritz, p. 411, Sinauer Associates, Sunderland (1990).

新生化学実験講座１６「分子進化実験法」，pp. 380-395
（五條堀・口野・西村・大沢共編）東京化学同人，1993年

23・3・1 最大節約規準

はじめに，無根系統樹 T の全長 $L(T)$ を定義する．塩基配列のような分類対象として n 個の操作的分類単位（operational taxonomic unit, OTU）があるとき，完全二分岐的な無根系統樹（図 23・5）の内部結節に対応する仮想的分類単位（hypothetical taxonomic unit, HTU）は必ず n-2 個存在する．分子データとして形質情報（各部位

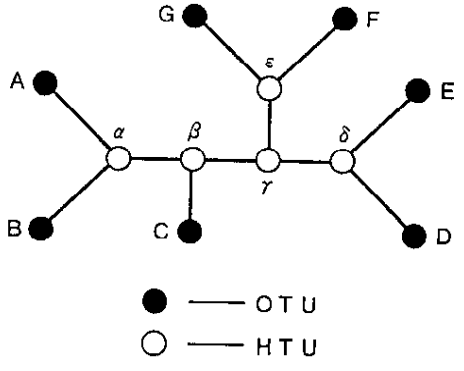

図 23・5 無根系統樹の構造．A～G の 7 個の OTU と α～ε の 5 個の OTU から成る完全二分岐的な無根系統樹．

ごとの塩基の種類）が与えられているのは実在する OTU だけであって，HTU に関する形質情報はない．HTU は無根系統樹の全体的最大節約性を達成するために構築または仮定される．ある無根系統樹 T を仮定したときに，OTU の形質状態の分布を説明するために最小限必要となる形質状態の変化の総数 $L(T)$ をその無根系統樹 T の“全長”という．いま，形質の総数（たとえば塩基配列の長さ）を N とし，与えられた無根系統樹 T のある枝の両端を A, B という二つの OTU（または HTU）とする．このとき，その枝の長さすなわち形質状態の変化数 $D[A, B]$ は，つぎの式で表される．

$$D[A, B] = \sum_{i=1}^{N} d_i(A, B) \tag{1}$$

式 (1) の $d_i(A, B)$ は，i 番目の形質において A, B の間で必要となる形質状態の変化数である．この無根系統樹 T の全長 $L(T)$ はすべての枝の長さの合計に等しいから，

$$L(T) = \sum_{\text{すべての枝}} \left(\sum_{i=1}^{N} d_i(A, B) \right) \tag{2}$$

式 (1) で定義される $D[A, B]$ は Manhattan 距離とよばれ，最大節約法では最もよく用いられている．この Manhattan 距離を用いると，式 (2) は

$$L(T)=\sum_{i=1}^{N}\left(\sum_{\text{すべての枝}} d_i(A,B)\right) \tag{3}$$

のように変形でき，各形質ごとの枝の長さの合計を別々に計算できる．以下の議論では，Manhattan 距離の利用を前提とする．

最大節約法とは，与えられた形質データのもとで式 (2) または式 (3) に示された全長 $L(T)$ の値を最小にするという最大節約規準を満たす無根系統樹 T を作成することである．データによっては，複数個の最大節約無根系統樹が存在することもあるが，その場合には最大節約規準を満足するすべての樹形を網羅する必要がある．

もちろん，最大節約法で得られた系統樹が正しいという保証はない．しかし，分岐の程度が比較的小さい場合には，他の系統樹作成法と同様に，正しい系統樹を復元する確率の高いことがコンピューターシミュレーションによってわかっている[1),2)]．一方，分岐の程度が大きく，しかも枝の長さがまちまちな系統樹を復元する際，条件によっては最大節約法で得られた系統樹が高い確率で誤りとなる場合があることが，4個の OTU の場合に理論的に[3)]，またそれ以上の個数の場合にはコンピューターシミュレーションによって明らかにされている[2),4),5)]．したがって，最大節約法で得られた系統樹の解釈には注意が肝要である．

23・3・2 最大節約系統樹探索のためのアルゴリズム

さて，最大節約規準を満たす無根系統樹を構築するアルゴリズムに話を進めることにしよう．とりわけ大量の分子データを処理するためには，系統分析用に開発されたコンピュータープログラムの利用は事実上不可欠である（詳しくは§23・3・9を参照）．以下では，最大節約法において最も広く用いられている PAUP および Hennig86 での無根系統樹構築アルゴリズムについて説明する．

最大節約規準を満たす無根系統樹の構築方法には，大きくわけて完全探索法と発見的構築法の二つの種類がある．

a. 完全探索法　完全探索法はあらゆる可能な無根系統樹を探索したうえで，各無根系統樹ごとに与えられたデータのもとでの全長を計算し，最大節約無根系統樹をみつける．したがって，この方法を用いて得られた無根系統樹は，真の最大節約無根系統

---

1) J. Sourdis, M. Nei, *Mol. Biol. Evol.*, 5, 298 (1988).
2) N. Saitou, T. Imanishi, *Mol. Biol. Evol.*, 6, 514 (1989).
3) J. Felsenstein, *Syst. Zool.*, 27, 401 (1978).
4) M. Hasegawa, T. Yano, *Bull. Biometric. Soc. Jpn.*, 5, 1 (1984).
5) N. Saitou, M. Nei, *J. Mol. Evol.*, 24, 189 (1986).

樹であることが保証される．一方，この方法の最大の問題点は，OTU 数が増えたときに評価しなければならない無根系統樹の数が爆発的に増えてしまうという点である．

この最大節約無根系統樹の完全探索の問題は，計算機科学でいう“NP 完全”という問題群に属していることが証明されており[1]，効率の良い完全探索アルゴリズムは存在しない．しかし，この完全探索に要する計算時間を大幅に節約できるアルゴリズムとして，“分枝限定法”（branch-and-bound method）が広く用いられるようになってきた．以下では Hendy と Penny[2] に従って，OTU を逐次的に付加する分枝限定法を説明するが，OTU ではなく形質を逐次付加する分枝限定法も提唱されている[3]．

i） 分枝限定法

分枝限定法の原理は，解くべき問題を部分問題に分割したうえで，ある限定条件のもとでいくつかの部分問題だけを解くことによって，もとの問題全体を効率的に解決するというアプローチをとる．分枝限定法が無根系統樹の完全探索をするときには，はじめに無根系統樹の全長の初期値 L を設定する（この初期値そのものが最小値である必要はない）．ついで，“深さ優先探索”という探索を開始する．たとえば，A, B, C, D, E から成る最大節約無根系統樹の探索を考えよう（図23・6）．最初，A, B, C から成る無根系統樹（*0*）を初期系統樹として D を付加する．最初，*0* に D を付加した無根系統樹 *1* を調べる．ついで，系統樹 *1* に E を付加した無根系統樹（*2* から *6*）を調べる．調べ終わった時点で，無根系統樹 *7* まで後戻りする．つぎに，その *7* に E を付加した無根系統樹（*8* から *12*）を調べる．その後 *13* に後戻りして，E を付加した無根系統樹（*14* から *18*）を調べる．この探索図に沿った深さ優先探索の途中で初期値 L を超えたならば，それよりさきの領域で最大節約無根系統樹を発見できる可能性はないので探索を中止して後戻りする．一方，先端まで探索したときに L よりも小さな値が得られたならば，その値を新たな初期値として設定し直し，さらに探索を続ける．その結果，すべての無根系統樹の樹形を逐一探索しなくても真の最大節約無根系統樹を発見できる．

この分枝限定法を用いれば，OTU 数が 20 程度ならば比較的短い計算時間での完全探索が可能である．もっと OTU 数が多い場合には，計算時間を節約するために，分枝限定法で真の最大節約無根系統樹を一つだけ発見し，あとで説明する分枝交換法を用いてそれ以外の真の最大節約無根系統樹を探索するという折衷法もある．たとえば，Hennig 86 では，“implicit enumeration”（ie-）コマンドを用いると，分枝限定法に

1) R.L. Graham, L.R. Foulds, *Math. Biosci.*, 60, 133 (1982).
2) M.D. Hendy, D. Penny, *Math. Biosci.*, 59, 277 (1982).
3) D. Penny, M.D. Hendy, *Comp. Appl. Biosci.* (CABIOS), 3, 183 (1987).

よって真の最大節約無根系統樹の一つをまず発見する．ついで，その無根系統樹に対して，“branch-breaker”（bb*）コマンドによる分枝交換を施して，異なる樹形をもつ真の最大節約無根系統樹がほかにあるかどうかを調べる．

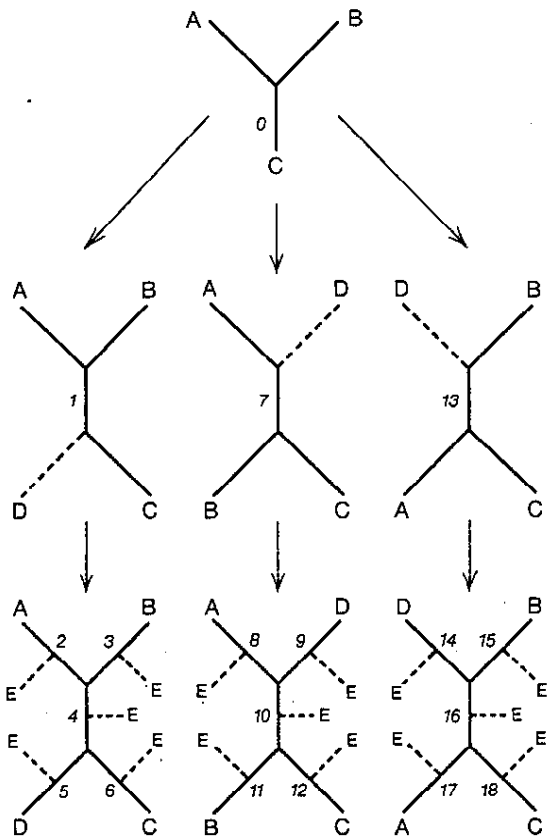

図 23・6 分枝限定法． A, B, C, D, E は OTU である．初期系統樹 *0* に D を付加し，さらに E を付加することにより，探索すべき無根系統樹を構築する．無根系統樹の番号は探索の順序に対応する．

ii） 網羅的探索法

最大節約解を速く発見するという目的からいえば，分枝限定法が最適である．しかし，すべての無根系統樹の樹形を探索しなければならない場合には，この網羅的探索法が用いられる．うえで指摘した OTU 数 n の増大に伴う無根系統樹の組合わせ論的爆発のため，計算時間を考えれば，n が 10 または 11 を超えるときには網羅的探索法を用いるべきではない．上述の分枝限定法が普及した現在，この網羅的探索法の唯一の存在意義は，無根系統樹の“樹長分布”が作成できることである．樹長分布とは，無根系統

樹の全長に関する頻度分布である．樹長分布から得られる第1の知見は，無根樹の全体集合のなかで最大節約的な無根系統樹の占める位置である．得られた最大節約無根系統樹が他の無根系統樹と比較してどの程度最大節約的であるのか，そして最大節約無根系統樹が準最大節約無根系統樹の近くでどのように分布しているのかは，樹長分布に照らして考察できる．第2の知見は，樹長分布の非対称性の尺度である歪度である．ある樹長分布が n 個の無根系統樹を含むとき，その分布の歪度 g_1 は式 (4) によって与えられる[1]．

$$g_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(L_i - \bar{L})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(L_i - \bar{L})^2\right)^{\frac{3}{2}}} \tag{4}$$

式 (4) において，L_i は i 番目の無根系統樹の全長であり，$\bar{L}$ は n 個の無根系統樹の全長の平均である．Hillis[1] は，データにノイズが少ないほど歪度 g_1 は絶対値の大きな負の値をとる，すなわち樹長分布は左裾の長い分布をすることを示した．いいかえれば，樹長分布の歪度はデータに含まれる系統学的情報量の尺度として利用できる．

b. 発見的構築法　発見的構築法では，はじめに逐次的に OTU を部分無根系統樹に付加して初期系統樹をつくる．ついで，その初期樹の枝を位置交換することによって全長がより短くなるような無根系統樹を発見的に構築する．ただし，発見的構築法を用いて得られた無根系統樹が真の最大節約解（大域的最大節約解）であるという保証はどこにもない．得られた解は，発見的構築の過程でたまたまおちいった局所的な解（大域的には最大節約的ではない）にすぎないかもしれない．

発見的構築法における大域解と局所解の存在は，無根系統樹集合での“族 (family)”あるいは“島”(island) の現象と深く関係している．ここでいう族または島とは，構造的に互いに類似する無根系統樹の集合である．ここで，二つの無根系統樹の間の構造的差異は，一方の無根系統樹を他方の樹形と一致させるのに要する枝のつけ替えの回数によって測られる[2]．与えられたデータによっては，無根系統樹の全体集合のなかにはいくつかの島（族）が存在することがある．同じ島に含まれる無根系統樹は全長の値も互いに類似する傾向があるが，島によってその頂上の標高（最大節約性の程度）には差がある．したがって，発見的構築の過程で局所解を頂点とする島に入込んでしまった場合，いかにしてその島から脱出して，大域的最大節約解を頂点とする島にたどり着くか

1) D.M. Hillis, “Phylogenetic Analysis of DNA Sequences”, ed. by M.M. Miyamoto, J. Cracraft, p.278, Oxford University Press, New York (1991).
2) D.F. Robinson, L.R. Foulds, *Math. Biosci.*, 53, 131 (1981).

が課題になる.

発見的構築法が局所解をしばしば導くというこの本質的欠点を補うために，以下に示すさまざまな方策が考案されている.

i) 逐次 OTU 付加

初期無根樹をつくるための逐次 OTU 付加には，単純にデータ行列に入力されている OTU の順番に OTU を部分無根系統樹に付加する方法，疑似乱数によって OTU 付加順を決める方法，あるいは三つの OTU の可能なすべての組合わせを考え，それぞれの組がつくる部分無根系統樹を出発点として全長が短くなるよう他の OTU を付加する方法などいくつかのやり方がある. さらに，最後の方法のように最大節約的に OTU 付

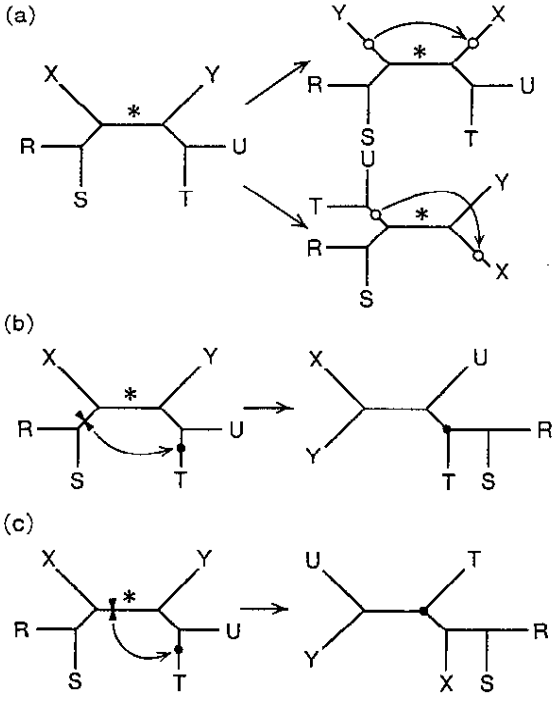

図 23・7　分枝交換法. 分枝交換の規準となる枝を＊で標識した. (a) 最近隣枝の交換. 枝＊をはさんで隣接する X と Y の交換あるいは X と U+T の交換をさす. (b) せん定と接ぎ木. 枝 R+S を枝＊に関して最近隣ではない枝 T につけ替える. (c) 切断と再接続. 枝＊そのものを切断して X+R+S を枝 T につけ替える.

加をする場合，各段階での最大節約部分無根樹だけでなくそれよりも少しだけ長い準最大節約部分無根樹をも保持させるようにすれば，大域的な最大節約解を導く初期無根樹

をOTU付加過程で棄却してしまう危険性が減るだろう．

どの付加方法がよいかは一概にはいえない．いくつかの方法を組合わせて反復するべきだろう．なぜなら，OTU付加法が異なればできる初期無根樹も異なり，結局は発見的構築の結果に影響するからである．

ii） 分枝交換

すべてのOTUが付加されると初期無根樹が完成する．もちろん，この初期無根樹は局所解ではあっても大域的最大節約解である保証はない．しかし，枝の位置を交換して樹形を変更すればより最大節約的な無根系統樹が得られるかもしれない．この操作を分枝交換とよぶ（図23・7参照）．分枝交換の規模が大きいほど，局所解におちいる危険性は小さくなるが，計算時間は増大する．分枝交換の過程では，その時点での中間解として生じうるいくつかの最大節約的な無根系統樹のみならず，複数個の準最大節約的な無根系統樹も同時に保持する必要がある．これもまた，最終的に大域的最大節約解となるかもしれない樹形を構築途中で捨ててしまわないための予防策である．

OTU数がかなり大きい場合や，個々の形質状態変化に対して異なる重みづけをする場合（後述）に，合理的な計算時間内で系統推定を行うためには，どうしても発見的構築法を用いざるを得ない．そのときは，局所解におちいらないようにOTU付加や分枝交換の方法をいろいろ変えながら繰返し試行することが肝要である．

23・3・3 最大節約無根系統樹の枝の長さ

これまで述べてきた方法により最大節約無根系統樹の樹形が計算されたとしても，もう一つの問題が残される．それは，この無根系統樹の各枝の長さの推定である．最大節約法のもとで，無根系統樹のそれぞれの枝の長さを求めるには，HTUの仮想的形質状態を決定しなければならない．一般に，ある最大節約無根系統樹の樹形のもとでは，その全長を変化させない，複数個の相異なるHTU形質状態配置が生じることがある[1]．HTU形質状態配置パターンが異なれば，それに対応して各枝の長さも変わる．

いま，アウトグループを用いてある最大節約無根系統樹を根づけることにする．この有根樹のHTU（仮想的共通祖先）に関する形質状態の再構築法としては，つぎの二つが提案されている[2]．1）形質変換遅延最適化（DELTRAN最適化）：形質状態の変化ができるだけ末端近くで生じるようにHTU状態を設定する（図23・8a）．形質の収斂回数は最大化され，逆に形質進化の逆転回数は最小化される．2）形質変換促進最適

1） W.M. Fitch, *Syst. Zool.*, 20, 406 (1971).
2） D.L. Swofford, W.P. Maddison, *Math. Biosci.*, 87, 199 (1987).

化（ACCTRAN 最適化）：DELTRAN とは逆に，形質状態の変化ができるだけ根の近くで生じるように HTU 状態を設定する（図 23・8c）．形質進化の逆転回数は最大化され，逆に形質の収斂回数は最小化される．

これら二つの方法は，HTU 形質状態配置としては両極端であり，実際にはそれらの中間に位置する配置がいくつか存在することもある（図 23・8b 参照）．どの HTU 形

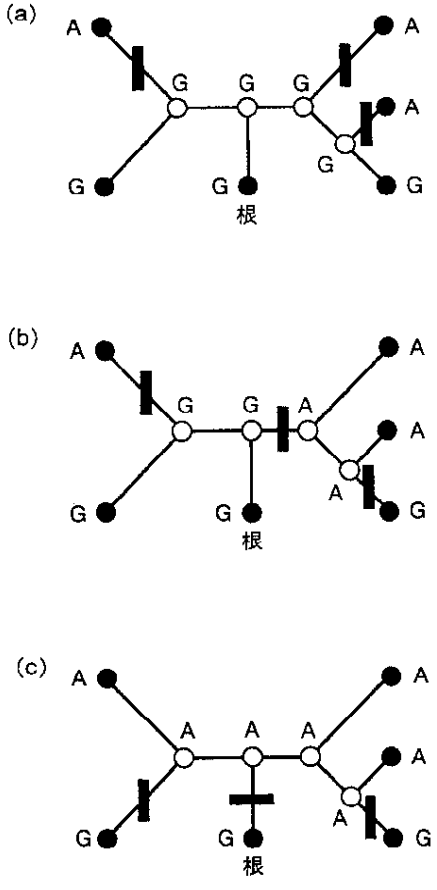

図 23・8 HTU 形質状態の再構築．●で表示した 6 個の OTU に関して，ある部位の塩基が A または G であるとする．無根系統樹の樹形はすでに与えられているものとする．G をもつアウトグループ OTU（根）によってこの無根系統樹を根づけた．(a) DELTRAN 最適化．根からできるだけ離れた枝で塩基置換（■で表示）が生じたという仮定のもとで，HTU（○で表示）への形質状態を最大節約的に配置する．(b) DELTRAN 最適化と ACCTRAN 最適化の中間に位置する，HTU への形質状態配置．(c) ACCTRAN 最適化．塩基置換の収斂回数が最小になるように，HTU への形質状態を最大節約的に配置する．

質状態配置を選択するかは，収斂と逆転の相対的頻度に関しておかれる仮定によって異なる．

最大節約法を用いて枝の長さを推定する場合に注意すべき点がある．分子データの場合，比較する配列間の分岐の程度が大きいと，同一の部位で複数回の置換が生ずることがある．最大節約法は，並行置換や復帰置換をみつけるにはある程度まで有効だが，複数の置換が継続して生じる場合は検出できない，したがって，置換数の過小推定を与えることになる．Saitou[1] は塩基置換の場合を分析して，配列間の塩基置換数が 0.2 未満の場合には最大節約法による過小推定の程度が小さいが，それ以上では枝の長さが大幅に過小推定されることを示している．したがって，大きく異なっている配列を比較した場合，枝の長さはあまりあてにならない．

23・3・4 形質や形質状態変化の重みづけ

すでに述べた最大節約規準は，各形質に等しい重みを与え，また一形質のなかでの形質状態変化 a→b と b→a とを等価に扱っている．しかし，以下にあげるいくつかの状況では，これらの仮定は合理的ではないことがある．1) 各 DNA 部位での転位と転換の差：DNA に含まれる 4 塩基 A, G, C, T については，トランジションの方がトランスバージョンよりも多く生じている．この場合，各 DNA 部位では形質状態の変化として，トランジションよりもトランスバージョンに大きく重みをつける方が合理的である．最も極端な場合，トランジションをまったく無視することがある．2) 各コドンにおける第 3 塩基の進化速度の違い：一般に，コドンの第 3 塩基は置換されてもアミノ酸の変化をひき起こすことは少ない（同義置換）ので，進化速度が速い．このとき形質としてのコドンの第 1 塩基，第 2 塩基を第 3 塩基よりも大きく重みづけるべきである．3) 制限酵素切断地図における制限部位の生成と消失：ある制限酵素で切断されるためには特定の認識配列（数塩基）が存在しなければならないが，その制限部位がなくなるためにはその認識配列のうちどれか一つの塩基が置換されればよい．したがって，制限部位は生成する方が消失するよりも生じにくいと考えられるので，ある制限部位においては消失よりも生成に対してより大きな重みをつける方がよい．

うえの 2) では，特定の形質に重みをつける必要があるのに対し，1) と 3) ではある形質のなかの形質状態変化ごとに別々の重みをつけなければならない．しかし，重みの値を実際にどのようにして決めるのかという問題はまだ解決されていない．

形質ごとの重みづけをしたうえで全長を最小化させる無根系統樹を求めるのは，たいていの最大節約法のソフトウエアで可能である．一方，形質状態変化ごとに重みづけし

1) N. Saitou, *Syst. Zool.*, 38, 1 (1989).

て全長を最小化させる無根系統樹を求めるのは，たとえばPAUPにおいてstepmatrix（形質状態間の変化に伴う負荷量の行列）を各形質に対して指定すれば可能である．もっとも，その分析には多大な計算時間を要する．

23・3・5 データと系統樹の一致性の尺度

$L(T)$の値に関与する部位の形質状態の分布がある無根系統樹とどの程度一致するかは，“一致指数”（consistency index）[1]によって表される．ある部位iの塩基の分布を説明するのに必要な塩基置換数の最小値と最大値をそれぞれm_i，g_iとし，その無根系統樹のもとで要求される塩基置換数をs_iとする．このとき，部位iの一致指数c_iは，

$$c_i = \frac{m_i}{s_i} \tag{5}$$

と定義される．その部位iの塩基分布と無根系統樹が完全に整合的で1回だけの形質状態変化で分布が説明できるならば，$c_i=1$となる．一方，一致の程度が低く，余分な塩基置換（これをhomoplasyという）を必要とするときには，c_iの値は小さくなる．c_iの最小値はm_i/g_iであるが，この値は部位ごとに異なる．そこで，c_iを修正して，その値域が区間$[0, 1]$になるように標準化する．この修正値c'_iは次式で表される．

$$c'_i = \frac{c_i - \dfrac{m_i}{g_i}}{1 - \dfrac{m_i}{g_i}} = \frac{g_i - s_i}{g_i - m_i} \times \frac{m_i}{s_i} \tag{6}$$

式(6)において，

$$r_i = \frac{g_i - s_i}{g_i - m_i} \tag{7}$$

とおく．この係数r_iは“保持指数”（retention index）[2]とよばれる．式(6)は結局，

$$c'_i = r_i \times c_i \tag{8}$$

となるが，この修正値c'_iを“修正一致指数”（rescaled consistency index）[2]とよぶ．これらの量を，形質データ全体と無根系統樹との一致性に拡張する．ある無根系統樹のもとでの形質データ全体の一致指数CI，保持指数RI，修正一致指数RCはそれぞれ

$$CI = \frac{M}{S} \tag{9}$$

$$RI = \frac{G - S}{G - M} \tag{10}$$

---

1) A.G. Kluge, J.S. Farris, *Syst. Zool.*, 18, 1 (1969).
2) J.S. Farris, *Cladistrics*, 5, 417 (1989).

$$RC = RI \times CI \tag{11}$$

と定義される．式(9)，式(10) 右辺の S, G, M はそれぞれすべての形質（部位）についての総和であり，$S=\sum s_i$, $G=\sum g_i$, $M=\sum m_i$ と定義される．特定の OTU にだけ異なる塩基が存在する部位の数が増加するとともに，一致指数も増加する．しかし，保持指数の値はそれらの部位の個数の影響を受けないので，みかけ上の一致性の影響を除去することができる．

23・3・6 最大節約無根系統樹の信頼性の検討

最大節約無根系統樹の信頼性を調べるために，もとのデータ行列からの形質の無作為再抽出に基づくブーツストラップ法[1]（§24・1参照）が広く用いられている．最大節約法でのブーツストラップ法は，もとのデータから重複を許して同数の形質を無作為抽出し，得られたデータに対して最大節約無限系統樹（ブートストラップ系統樹）を作成する．このプロセスを何度も（数千回程度）反復することによって，無根系統樹の枝の出現率が計算できる．この出現率がきわめて高い（たとえば95％以上）ならば，その枝は十分に有意であって，信頼度が高いと判定される．

このブーツストラップ法を用いて単一の枝の信頼性を評価するのは問題ない．しかし無根系統樹の樹形全体の信頼性をブーツストラップ法によって評価するに当たってはいくつかの点で注意が必要である．たとえば，無根系統樹に含まれる複数の枝の信頼性を同時に評価することは，有意でない枝を有意であると判定する第1種過誤の確率を高めてしまう（多重検定問題）ので，そのための補正が必要になる[1]．また，複数のブーツストラップ系統樹に共通する出現率の高い枝を組合わせてつくった合意系統樹（consensus tree）を信頼度の高い樹形であると解釈するのは誤りである[2]．ブーツストラップ法によって信頼度が評価されたのは個々の枝であって，同意系統樹の樹形そのものではないからである．

最大節約系統樹の各枝の統計学的信頼性を調べるには，ブーツストラップ法のほかにも，各部位がどの系統樹を支持するかをカウントし，二項分布を用いてその観察値が生じる確率を計算する方法もある[3]．

---

1) J. Felsenstein, *Evolution*, 39, 783 (1985).
2) M. J. Sanderson, *Cladistics*, 5, 113 (1989).
3) S. A. Williams, M. Goodman, *Mol. Biol. Evol.*, 6, 325 (1989).

23・3・7 霊長類のミトコンドリアDNAの塩基配列データの例

Horai ら[1]による霊長類のミトコンドリアDNAの塩基配列データの一部分（5個のtRNA遺伝子と1個のタンパク質遺伝子を含む，350～1797 bpの1448塩基）を用いて，うえで述べた最大節約法の基本原理を説明する．このデータは，OTU数（塩基配列数）が6である．

最大節約法の重要な特徴は，すべてのDNAの部位が系統推定に貢献するわけではないという点である．たとえば，この例で6個のOTUが同一の塩基を共有している部位（図23・9の無印の部位）は，塩基置換をまったく想定する必要がないから，式

```
 351                                                                420
      +     !  +      !       ! !!   ! !!  ++    !  ! ! +  !  !  !+!!!
Hu  TTAATCCCCT GGCCCAACCC GTCATCTACT CTACCATCTT TGCAGGCACA CTCATCACAG CGCTAAGCTC
Ch            A          A                 C   A       G        T
Py            A          A                     C           T    T      T
Go      C                A                     C        T     T  T      A
Or      C     A  T       A         C TC   G      CA           T  T      A  G
Si      C     A  T       A           A     T          T          T  C   ATCT

 421                                                                490
   +        + ! !+     !  !                 +  + ! !  !  !   !   ! !!     !
Hu  GCACTGATTT TTTACCTGAG TAGGCCTAGA AATAAACATG CTAGCTTTTA TTCCAGTTCT AACCAAAAAA
Ch  A           C                          T  A            C   A C
Py  A           C                          T  A            C     C
Go  C            G                            A            C     C
Or  C       C    G     C  G                   A   C    C   C     C
Si  A       C C CG                              T    C      C     CT     A
```

図 23・9 **霊長類6種のミトコンドリアDNA塩基配列データ**（部分）．Horai ら[1]のミトコンドリアDNA塩基配列データのうち，351番から490番までの140部位を示した．各部位は，変異のない部位（無印），変異はあるが情報のない部位（！印）および情報をもつ部位（＋印）に分類される．Hu: ヒト，Ch: チンパンジー，Py: ピグミーチンパンジー，Go: ゴリラ，Or: オランウータン，Si: フクロテナガザル．

(2) に示した全長 $L(T)$ には寄与しない．このような部位は，“変異のない部位”（invariant site）とよばれる．この例では，1448塩基部位中1023個が変異のない部位（図23・9の無印の部位）である．残る425部位はそれぞれ少なくとも1回は塩基置換がなければならないので，$L(T)$ の値の大小に関与する．しかし，そのうち特定のOTUにだけ異なる塩基が存在する部位（図23・9の！印の部位）は，任意の無根系統樹において等しく枝（その種を末端とする枝）の長さを1だけ増加させるにすぎないから，全長 $L(T)$ に差異を生じさせる原因とはならない．このような部位は，上述の変異のない部位とともに“情報をもたない部位”（uninformative site）[2]とよばれる．残る

1) S. Horai, Y. Satta, K. Hayasaka, T. Inoue, T. Ishida, S. Hayashi, N. Takahata, *J. Mol. Evol.*, **35**, 32 (1992).
2) W.M. Fitch, *Am. Nat.*, 111, 223 (1977).

425 部位中，262 部位がこの種類に属する．最終的に残った部位（図 23・9 の＋印の部位）は，少なくとも 2 種類の塩基のそれぞれが少なくとも二つの OTU に分布している部位であり，“情報をもつ部位”（informative site）とよばれる．これらの部位だけが最大節約無根系統樹の構築にとって意味のある情報をもたらす．この実例では，163 個の情報をもつ部位が存在する．

OTU 数が 6 のとき完全二分岐的な無根系統樹の樹形は 105 通り存在する．網羅的探索法により，それぞれの樹形について，図 23・9 にその一部を示した塩基配列データのもとでの全長 $L(T)$ を計算した結果，一つの最大節約無根系統樹（全長 564）が発見された（図 23・10 a）．これらの無根系統樹はテナガザルをアウトグループとして根づけると有根系統樹になる（アウトグループによる根づけについては文献[1]を参照）．この有根系統樹に関して，DELTRAN 最適化と ACCTRAN 最適化の 2 通りの方法で枝の長さを決定した（図 23・10 b）．

また，この最大節約無根系統樹のもとで，$L(T)$ の値に関与する部位すべてをこみに

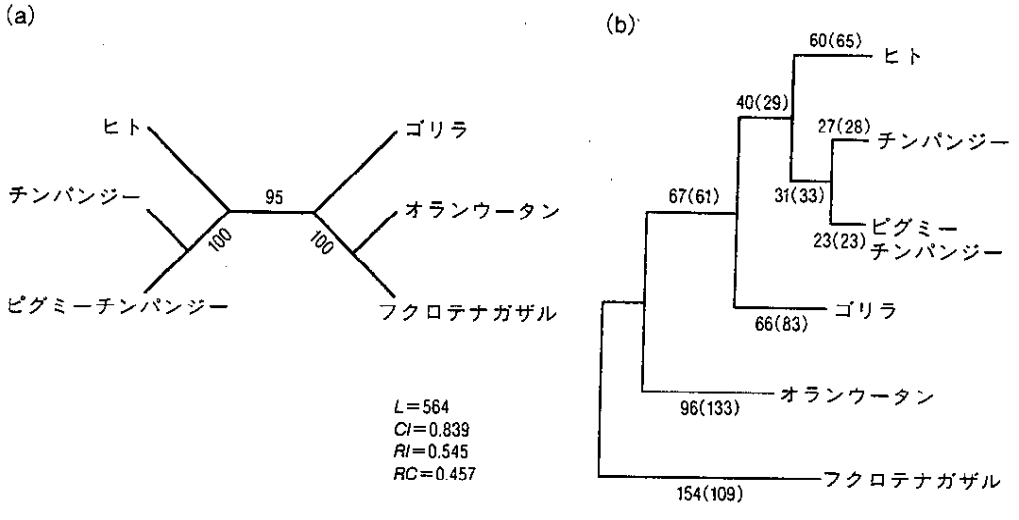

図 23・10 ミトコンドリア DNA 塩基配列データに基づく最大節約系統樹． Horai ら[2]の部分データに基づく最大節約分析．(a) 最大節約無根系統樹について，全長 (L)，一致指数 (CI)，保持指数 (RI)，修正一致指数 (RC) を示した．また，各枝の数値はブーツストラップ法（10,000 回反復）による信頼度（%）である．(b) フクロテナガザルをアウトグループとして根づけた有根系統樹．枝上の数字は，ACCTRAN（かっこ内は DELTRAN）による枝長を示す．図の枝長は，ACCTRAN で得られたものに比例してかいてある．

1) W. P. Maddison, M. J. Donoghue, D. R. Maddison, *Syst. Zool.*, 33, 83 (1984).
2) S. Horai, Y. Satta, K. Hayasaka, T. Inoue, T. Ishida, S. Hayashi, N. Takahata, *J. Mol. Evol.*, 35, 32 (1992).

すると CI=0.839 となるが，そのうち特定の OTU にだけ異なる塩基が存在する部位を除去するとその値は0.668にまで低下する．一方，RI はどちらの場合も 0.545 である．この RI の値は，情報をもつ部位の共有塩基全体の 50％強がその無根系統樹のもとでもやはり共有状態として保持され，残りの共有塩基は homoplasy として説明されることを意味している．

最後に，10,000回反復のブーツストラップを行い，この最大節約無根系統樹の枝の信頼性を評価した（図 23・10 a）．その結果，どの枝も 95％以上のきわめて高い信頼性をもつことが示された．

なお，うえの解析では塩基配列のみを形質として用いたが，部位の挿入，欠失の共有も形質として利用することができる．

23・3・8 最大節約法のための主要ソフトウエアリスト

1) PAUP: Phylogenetic Analysis Using Parsimony (Version 3.0): さまざまなタイプの最大節約法ならびにブーツストラップ，樹長分布解析を実行できる．現在はマッキントッシュ版だけがリリースされている．PHYLIP および Hennig86 とのデータファイルやツリーファイルの変換機能をもつ．MacClade とはデータファイル自身を共有できる．全般的に完成度のきわめて高いソフトである．価格は US $ 50.00 である．申込先: D. L. Swofford, Illinois Natural History Survey, 607 East Peabody Drive, Champaign, Illinois 61820, USA.

2) Hennig86 (Version 1.5): MS-DOS マシン専用のソフトで，かなり大きなデータ行列でも高速に最大節約系統樹を計算できるという特長がある．その反面，ユーザーインターフェイスはそれほどよくない．価格は US $ 50.00 である．申込先: J. S. Farris, 41 Admiral Street, Port Jefferson Station, New York 11776, USA. 日本語 MS-DOS 環境で Hennig86 を用いる場合にはいくつか問題があるが，その解決策については三中信宏（〒305 つくば市観音台 3-1-1 農業環境技術研究所計測情報科調査計画研究室）まで連絡されたい．

3) MacClade: Interactive Analysis of Phylogeny and Character Evolution (Version 3.0): マッキントッシュ専用のソフトである．ユーザーインターフェイスの良さは定評があり，分岐図の分枝交換や形質進化の復元をマウス操作で簡単に行うことができる．ファイル形式が完全互換である PAUP といっしょに用いれば，さらに精密な系統解析を行える．Sinauer Assosiates (Sunderland, Massachusetts, USA) より販売されている (W. P. Maddison, D. R. Maddison (1992) MacClade Version 3.0).

4) PHYLIP: Phylogeny Inference Package (Version 3.4): 最大節約法だけでなくさまざまな系統分析法のソフトを含んだ汎用のパッケージである. PASCAL でかかれたソースプログラムおよび MS-DOS マシン用の実行ファイルが無料で配布されている. しかし, コンパイラさえあれば, マッキントッシュ, UNIX, メインフレームなどさまざまな計算環境で用いることができる. 最大節約法のプログラムとしては, MIX, PENNY, DNAPARS, PROTPARS, DNAPENNY が含まれている. また, ブーツストラップを実行するプログラムとしては BOOT, DNABOOT, SEQBOOT がある. PAUP や Hennig86 と比較すれば計算速度は遅い. 申込先: J. Felsenstein, Dept. of Genetics SK-50, Univ. of Washington, Seattle, Washington 98195, USA. 電子メール (joe@genetics.washington.edu) から PHYLIP プログラムを入手することも可能である.

うえにあげた以外にも多くのソフトが開発されているが, それらについての最新の情報 (入手方法, 特長, 関連文献) を集約した下記のリストを筆者らは配布している. 関心のある読者は連絡されたい (三中信宏, 斎藤成也, 系統樹作成のためのソフトウエアリスト (第1版) (1992)).

23・4 Farris 法と改 Farris 法*

改 Farris 法 (modified Farris 法)[1] は, Farris 法[2] を改良する試みの結果生まれた方法である. また, Farris 法もそれ以前に提唱されていた Wagner 法 (文献[3] を参照) をもっと一般化した方法で, 種間の数値化された距離から系統樹を作成できるようにしたものである. このため, Farris 法は距離 Wagner 法ともよばれている. この意味で, 改 Farris 法は系統樹作成法に対する考えの一つの流れのなかに位置するといえる.

さて, 改 Farris 法による系統樹のつくり方であるが, うえに述べたように, この方法は Farris 法を基盤としているので, Farris 法を説明しながらその相違点について述べるのがよいと思われる. したがって, 表 23・2 の距離行列を用いて, まず Farris 法について説明する.

距離行列のなかで D_{12} が最小値と仮定すると (どの値を最小と仮定しても同じように説明される), 種1と種2が一つのグループ (1,2) をつくる. つぎに, (1,2) と他の

---

* 執筆担当: 舘野義男 (§23・4)

1) Y. Tateno, M. Nei, F. Tajima, *J. Mol. Evol.*, 18, 387 (1982).
2) J.S. Farris, *Am. Nat.*, 106, 645 (1972).
3) J.S. Farris, *Syst. Zool.*, 19, 83 (1970).