

分子系統進化学と分子系統樹作成法†

齋藤 成也

(国立遺伝学研究所)

1. 多種多様な分子系統樹作成法

生物進化の根本は、遺伝子が自己増殖していった軌跡である系統樹の上に突然変異が蓄積してゆく過程である。したがって、遺伝子の実体である塩基配列をさまざまな生物や遺伝子の種類と比較して、分子系統樹を作成することは、進化学研究において必須である^(1,2)。

現在までに多数の分子系統樹作成法が提唱されているが、用いるデータの種類によって、大きく2つに分けることができる。1つは、アラインメント(整列)を行った配列データをそのまま用いるものであり、形質状態法と呼ぶ。最大節約法⁽³⁾や最尤法⁽⁴⁾がこれに属する。一方、配列間の距離をなんらかの方法で計算し、これら距離行列を用いる方法が距離行列法である(塩基置換数の推定については文献(1)を参照されたい)。以下に述べる近隣結合法^(5,6)や、Fitch-Margoliash法⁽⁷⁾、進化速度一定を仮定するUPGMA⁽⁸⁾、距離ワグナー法⁽⁹⁾、最小進化法⁽¹⁰⁾などがこちらに含まれる。

また、系統樹の発見方法の違いによって、系統樹作成法を2つに分けることができる⁽¹¹⁾。1つは網羅的探索法であり、系統樹の可能な樹形(分岐パターン)を可能ならばすべて調べて、データと最もよく適合する樹形を選ぶものである。最大節約法、最尤法、Fitch-Margoliash法、最小進化法などがこちらに属する。しかし、可能な有根系統樹(共通祖先を示す根を持つ系統樹)の樹形の数は、 n 本の配列を比較すると $(2n-3)!/2^{n-2}(n-2)!$ となり、 n が増加するにつれて、爆発的に増大する⁽¹⁰⁾。そこで、もう1つの方法として、最終的な系統樹の樹形を段階的に決定してゆく方法がある。近隣結合法、UPGMA、距離ワグナー法などが含まれる。以下では、筆者らが開発した近隣結合法を中心に、系統樹作

成法の実際を紹介する。

2. 近隣結合法

近隣結合法(neighbor-joining method)は、距離行列データから簡単なアルゴリズムを用いて、段階的に近隣(無根系統樹において、1つの結節だけでつながった2つの配列のこと)を見出し、最終的に1個の無根系統樹を得る方法である⁽⁵⁾。この方法を用いると、配列の数が100を越えても、短い計算時間で系統樹を得ることができる。しかも進化速度一定を仮定していないので、系統によって進化速度に違いがあっても、比較的正確に真の系統樹を復元することができる。この方法は、系統樹の枝の長さの総和を最小にする樹形を選ぶので、最小進化の原理を用いているが、そのアルゴリズムは、同じく最小進化の原理を形質状態データに用いる最大節約法とはかなり異なっている。

以下に、近隣結合法のアルゴリズムを簡単に説明する(くわしくは文献(6)を参照されたい)。図1は、特定の1対の配列(i と j)を取り出し、それらが近隣となると仮定したものである。その全枝長の和 S_{ij} は、 D_{ij} を配列 i と j 間の距離とし、 $Q = \sum_{i < j} D_{ij}$ 、 $R_i = \sum_{j=1}^n D_{ij}$ とすると、

$$S_{ij} = D_{ij}/2 + [2Q - R_i - R_j]/2(n-2)$$

で与えられる。このように簡単な式なので、枝長の総和 S_{ij} の計算は、配列の数が比較的少なければ、コンピュータプログラムの助けを借りなくとも、電卓を用いて容易に行うことができる。

実際には、どの配列の対が近隣であるかわからないので、 $n(n-1)/2$ 個のすべての配列の対について、図1のような樹形を考え、それに対応する枝長の総和 S_{ij} を計算する。そうして、最小の S_{ij} を与える配列の対を近隣とみなすのである(最小進化の原理)。もちろん、実際には最小の S_{ij} を示す配列の対が近隣ではない可能性がある。しかし、用いられた距離が相加的である(系統樹において、任意の2個の配列を結ぶすべての枝の長さ

† Molecular Phylogenetics and Methods for Constructing Molecular Phylogenies.

Naruya SAITOU (Laboratory of Evolutionary Genetics, National Institute of Genetics, Mishima 411)

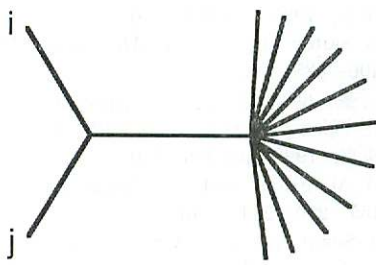


図 1 配列 i と配列 j が近隣と仮定された無根系統樹

を加えると必ず配列間距離に等しい場合) ときには、この方法は必ず近隣を見出すことがわかっている^(6,12)。また、距離が必ずしも相加的ではなくても、正しい系統樹では S_{ij} の期待値が最小になることが証明されている⁽¹³⁾。

いずれにせよ、最小の S_{ij} を与える配列の対が見つかったら、次にそれらを合体して1つの配列にし、また上の操作を繰り返すのである。配列 i と配列 j が合体した合体配列 (ij) と他の配列 k との進化距離 $D_{(ij)k}$ は単純平均を用いて計算する。これらの進化距離を含めて、新しい $(n-1)$ 個の配列間の距離行列ができる。この行列について上と同じ計算を行えば、次の近隣を決めることができる。この繰返し操作により、次々に近隣が見出されるので、それによって最終的な系統樹を作成することができる。また、それぞれの枝長も段階的に決まってゆく。進化距離がすべて相加的である場合には、近隣結合法は、最終的に得られる系統樹についても、すべての枝について正しい枝長を与える。

実際のデータは相加的でないのが普通なので、得られた系統樹が必ずしも正しいとは限らない。しかし計算機シミュレーションによると、近隣結合法は、他の方法に比べて真の樹形を復元する効率の高いことが知られている^(5,10,14,15)。

図 2 は、世界のあちこちからとられたヒト成人白血病ウイルス (HTLV-I) の塩基配列をもとに、近隣結合法で作成したものである⁽¹⁶⁾。近隣結合法は無根系統樹を作成するので、HTLV-II ウイルスの配列を外群として加えて、根の位置を与えている。各枝の下に数字は、それらの枝で生じた塩基置換数を示す。塩基配列データを近隣結合法で扱う場合、一般には塩基サイト当りの塩基置換数を距離として与えるので、各枝の長さは整数ではない。しかし、われわれが観察する特定の塩基配列におい

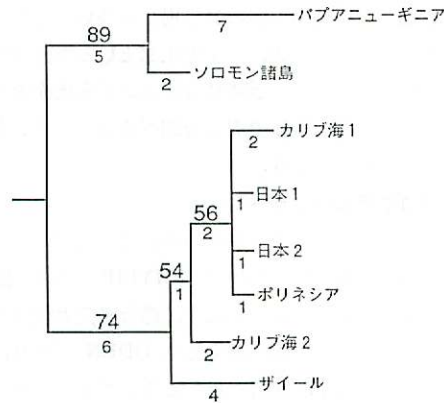


図 2 近隣結合法を応用した例 (文献 (16) より)

て生じる塩基置換は整数なので、この塩基サイト当りの塩基置換数に比較した塩基の長さをかけて、それをまとめて整数値に直している。このとき、枝の長さがゼロになったら、その枝は取り除く。このため、近隣結合法は基本的に二分岐型の系統樹を生成するが、図 2 の例では、日本で見つかったウイルスの 2 本の配列を含むクラスターは、1 点から 4 本の配列が分岐する形になっている。

各配列にはつながらない 4 本の枝の上の数字は、ブーツトラップ法⁽¹⁷⁾を用いて、樹形の各部分の分岐パターンの統計的信頼性を示したものである。1000 個のブーツトラップ系統樹を作り、そのなかの何パーセントが図 2 の系統樹の樹形を復元するかを与えている。95% 以上の頻度でもとの系統樹の分岐パターンが得られていれば、その枝はかなり信頼できると考えられる。逆に、それよりずっと小さい数字の場合 (図 2 でいうと、54% や 56% の枝)、このデータからははっきりしたことはいえないことになる。一般に、枝の長さが長くなると、ブーツトラップ確率も増加する。

系統樹的信頼性を調べるには、その他に、得られた系統樹の内部にあって系統樹の樹形を決めている枝の長さの標準誤差を計算し、それらが有意かどうかを調べる方法がある^(18,19)。もっとも、配列が 5 個以上の場合には、これらの方法にしたがって標準誤差を計算するのはかなり面倒になってくる。最終的な系統樹の枝の長さを最小 2 乗法で推定し、各枝に標準誤差を与える方法も最近開発されている⁽²⁰⁾。

近隣結合法は、最終的に 1 つの系統樹が得られる。しかし、同一のデータを用いれば、どの系統樹作成法を用

いても似通った系統樹が得られる場合が多い。したがって、最大節約法、最尤法、最小進化法といった網羅的探索法を用いる場合でも、近隣結合法による系統樹をまず求めておいて、その樹形の周辺を調べるという方法をとることが合理的であろう。

3. プログラムパッケージ

多数の方法のプログラムが含まれているものとして、J. Felsenstein 氏の開発した PHYLIP がある。国立遺伝学研究所のコンピュータには、総合研究大学院大学遺伝学専攻の伊奈康夫氏が開発した ODEN があり、さまざまな分子進化的解析を行うことができる。また、最大節約法だけに関するものだが、使いやすいソフトとして、PAUP がある。近隣結合法としては、筆者の作成したプログラムパッケージ、東京都立大学理学部生物学科の田村浩一郎氏の作成したプログラムパッケージ、EMBL データライブラリーの D. Higgins 氏の作成したプログラムパッケージ CLUSTAL V など、数種類のものがある。

- (1) 根井正利：「分子進化遺伝学」，五條堀孝・齋藤成也共訳，培風館，東京，1990。
- (2) N. Saitou: in "Handbook of Statistics, vol. 8," eds. by R. Chakraborty and C. R. Rao, Elsevier Science Publishers B. V., Amsterdam, 1991, pp. 317—346.
- (3) 三中信宏，齋藤成也：新生化学実験講座第 16 卷「分子進化実験法」，五條堀孝ほか共編，東京化学同人，東京，1993，pp. 380—395.
- (4) 岸野洋久：新生化学実験講座第 16 卷「分子進化実験法」，五條堀孝ほか共編，東京化学同人，

東京，1993，pp. 410—416.

- (5) N. Saitou and M. Nei : *Mol. Biol. Evol.*, **4**, 406—425 (1987).
- (6) 齋藤成也：新生化学実験講座第 16 卷「分子進化実験法」，五條堀孝ほか共編，東京化学同人，東京，1993，pp. 400—410.
- (7) W. M. Fitch and E. Margoliash : *Science*, **155**, 279—284 (1967).
- (8) R. Sokal and P. H. A. Sneath : in "Principles of Numerical Taxonomy," W. H. Freeman, San Francisco, 1963.
- (9) J. S. Farris : *Am. Nat.*, **106**, 645—668 (1972).
- (10) L. L. Cavalli-Sforza and A. W. F. Edwards : *Am. J. Hum. Genet.*, **19**, 233—257 (1967).
- (11) N. Saitou and T. Imanishi : *Mol. Biol. Evol.*, **6**, 514—525 (1989).
- (12) J. A. Studier and K. J. Keppler : *Mol. Biol. Evol.*, **5**, 729—731 (1988).
- (13) A. Rzhetsky and M. Nei : *Mol. Biol. Evol.*, **10**, 1073—1095 (1993).
- (14) J. Sourdis and M. Nei : *Mol. Biol. Evol.*, **5**, 298—311 (1988).
- (15) J. Li and M. Nei : *Mol. Biol. Evol.*, **7**, 82—102 (1990).
- (16) V. R. Nerurkar, K. -J. Song, N. Saitou, R. R. Melland and R. Yanagihara : *Virology*, **196**, 506—513 (1993).
- (17) 長谷川政美：新生化学実験講座第 16 卷「分子進化実験法」，五條堀孝ほか共編，東京化学同人，東京，1993，pp. 417—422.
- (18) W. -H. Li : *Mol. Biol. Evol.*, **6**, 424—435 (1989).
- (19) F. Tajima : *Mol. Biol. Evol.*, **9**, 168 (1992).
- (20) A. Rzhetsky and M. Nei : *Mol. Biol. Evol.*, **9**, 945—967 (1992).