



分子系統学の基本的方法論

DNAの塩基配列情報が容易に得られるようになったこと、使い勝手のよい解析ソフトが容易に入手できるようになったことにより、分子系統樹の構築はルーチンワークとなった。しかし、基本的原理の理解なしに得られた系統樹を評価することは出来ない。

分子系統樹を構築する方法論には3種類ある。距離行列法、最節約法、最尤法である。これらの方法論を網羅的に眺めるのであれば「分子進化実験法」がよい。また概念と成果を知るには宮田(1994)が手頃である。なお、いまだに誤解があるようであるが、分子進化時計は厳密には成り立たない。瀬戸口烈司が「講座進化」の中で、分子系統学(分子進化学)は「自ら一定ではないことを認めている分子進化速度を一定と仮定して系統、進化を論じている」と断じ、このような学問は第二科学であると非難しているが、まったくの誤解である。系統樹構築において進化速度一定性はまったく仮定していない。

距離行列法は対象の種類(OTU: Operational Taxonomical Unit)間の遺伝的距離を基に系統樹を構築する。根井(1990)やLi and Graur(1994 訳本)に詳しい。

2OTU間の塩基配列の違いをみるには比較している配列全体中の何%が異なっているか(何%同じか)を見ればよい。例えば、ヒトエグサとアオサとでは95%の相同性がある、というような言い方をする。しかし、この値をただちに遺伝的距離として扱うことは出来ない。transition/transversionにおける置換速度の違い、コドンの1, 2, 3番目における同義置換、非同義置換の頻度の違いによる速度差など、さまざまな進化速度のばらつきを考慮に入れなければならないからである。これらを考慮した様々な補正式が考案されている。こうして得られた遺伝的距離を行列にして表す。ここまですが1段階目である。この距離行列に見られる値を出発点とし、それがどのような系統関係から生じたのかを数学的に解くのが第二段階である。これには多くの方法が開発されているが、現在もっとも普及しているのが近隣結合法Neighbor Joining (NJ) 法である。

距離行列法の最大の欠点は、遺伝的距離を計算するときに位置情報を完全に無視することである。置換している塩基が配列中のどこにありうと同一遺伝的距離しか答えとして得られない。図1のAとBとはCに対して同距離になるわけである。これでは情報の大きな欠落である。また、距離行列から統計学的にもっとも優れた系

分子系統を構築するとき用いる方法には、幾つかのものがあるようですが(最大節約法、近隣結合法、距離行列法、最尤法)それらの簡単な原理と特徴は何でしょうか?またブートストラップについてもわかりやすく教えてください。(学芸大院生T.O.)

```
A  C T T A G G C T T G A C C
B  C T A A G C C A T G A C G
C  C T T A C G G A T G A C G
```

図1. Cに対し、AとBとは同じ遺伝的距離を示すが(Cからどちらも4塩基異なっている)、塩基置換が生じている場所は異なっている。なお、すべての座位間において差はないものとする。

統樹を選択する方法は結局のところ探索法(後述)的であり、得られた答えが最適解であるという保証が、その考え方の範疇の中においてすらも、存在しない。

一方、最大の利点は計算が速いことである。他の2方法とはまったく比較にならない。また、系統樹が1つしか出ないので、よくも悪くも悩む必要がない(最尤法では最節約法と異なり1つしか答えがないように思われているが、実は統計学的に排除できない樹形がいくらでも出る可能性がある)。

ソフトとしてはClustalVおよびClustalWがある(フリーソフトウェア:FSW)。特徴は、Multiple Alignmentが容易;アミノ酸、DNA両方が対象(後述するソフトではすべて別個のソフトを用いる)、ブートストラップ確率が容易にNJ系統樹に対して求められる、等である。また、分子系統学最大のプログラムパッケージ、PHYLIP(この膨大なソフト群がFSWである)に含まれるDNADIST/PROTDISTとNEIGHBORを用いる事ができ、さらにSEQBOOT、CONSENSEによりブートストラップ確率を求めることが可能である(後述するように注意が必要)。

これに対し、最尤法と最節約法はどちらも位置情報をそのまま用いる。一方、可能なすべての樹形を総当たりして比較した後でしか最適解が決定できないために、とんでもなく計算時間がかかる。3OTUの無根系統樹はただ1通りしか存在しない。しかし、4OTUでは3通り、以下、15, 105, 945, 10,395, 135,135,そしてわずか100TUで2,027,025通りとなる。現時点で最高速のパソコンをやっと購入して演算速度が従来機の10倍になったとしても、90TUを計算した時間で100TUは計算できない。スパコンを独り占めして100倍の速度を得たとしても、同じ時間内に110TUの計算は出来ない。すなわち、これらの方法で真の解を求めることは、ほんのちょっと種類数が増えれば不可能なのである。そこで、すべてを比較せずすむ技術が開発されており、それを探索法と言う。しかし、探索法で得られた解が最適解である保証はまったくない。

最節約Maximum Parsimony (MP) 法は最小ステップ数の系統樹を求め、それを最適解とする(図2)。言い換え

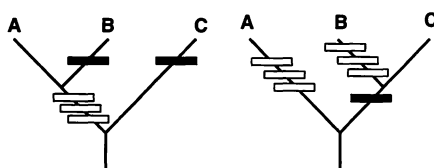


図2. 3 OTU間の同じ形質を用いて組み立てた有根系統樹。左では全部で5ステップ、1つのHomoplasyがあるのに対し、右の図では7ステップで3つのHomoplasyが存在することになる。したがって、左の図がより節約的である。

れば最も並行進化の少ない樹形を選択する。入門解説は「分子進化実験法」を見られたい。ソフトとしてはマッキントッシュ用のPAUP（市販）が事実上唯一のものである。PHYLIPにも多くのソフトが含まれているが使いにくい。

裸のままでは真の系統樹を求める能力を大きく欠くことが知られているが、そのための補正值を得る方法が最節約法にはない。したがって、最尤推定など別の方法論で推定して入れ込む必要がある。なおかつ、多重置換を有効に補正する方法は今のところ無い。これらについてはNewsletter for Plant Molecular Systematics (NPMS) 15号に詳しい入門解説がある。

最尤法 Maximum Likelihood (ML) は進化過程に仮定を置くことから始まる。例えばコドンの1, 2, 3番目の進化速度の比を1:0.3:6.4などと置く。こうした仮定に基づいて個々の樹形の尤度を計算し、それを比較して最大の尤度を示す樹形を最適解として選ぶ。仮定は最尤推定の中で自らより正しい値に近づけることが出来る。多重置換などあらゆる過程を組み込み、かつ検証可能である。長谷川 (1989) がよい入門書である。またNPMS16号に詳しい入門編が書かれている。DNAではPHYLIPに含まれるDNAMLが元祖である。計算速度を速くしたFastDNAMLや足立・長谷川によるMOLPHYというパッケージ中のNUCMLでは多少現実的な時間で計算できる。アミノ酸配列は後者のパッケージ中のPROTMLが唯一のソフトである。ML法は最も計算時間がかかるが、改良が進められている。

これら3つの方法のどれを選ぶべきか、ということは最大の関心事であろう。実は、十分な情報量を持ったデータさえあればどの方法でも同じ樹形が得られることがシミュレーションによって示されている。しかし実際には、そのような「優れた」データを用いることは無理である。対象の種類のばらつきや個々のデータの質により答えは異なることがある。3方法とも使うべきである、というのが筆者の考えである。

アメリカではMP法が大半のシェアを占めており、日本でもその影響が強いが、シェアと信頼性が一致するとは限らない。距離行列法には伝統があるがシェアは狭

い。最尤法は世界的に高い評価を受けてはいるが、計算時間がかかりすぎるため、シェアは低い。

さて、得られた系統樹には統計学的な信頼性が表示されなければならない。ところが、分子系統樹を構築するデータには母集団というものがない。このために登場したのがブートストラップ法である。利用できるものは元のデータしかない。そこで、コンピューターの力を使って強引に元データから擬似的に「母集団」を作り上げ、それがどの程度得られた樹形を支持しているかを表現させる。

方法は簡単で、元データから繰り返しを許して配列要素（塩基/アミノ酸残基の数）の数だけランダムに選びだし、あらたな配列を作り上げる。これを毎回違う乱数に基づいて繰り返し行っていく。そして、それぞれのデータセットにおいて系統樹を構築し、元の系統樹の内枝が何回現れたかを見る。ある内枝がブートストラップのデータセットによる系統樹群において1,000個の中に862回出現したとすると、この場合、86.2%の確率でこの内枝が支持されたということになる。これをブートストラップ確率と言う。

ところで、ブートストラップ確率を出すための値を目的の値と誤解した学会発表や論文が最近目に付く。すなわち、PAUPやCONSENSEではブートストラップ時の個々の樹形の合意系統樹をも示すのであるが（Clustalは正しい表示をする）、これは無論、決して採用してはならない。生物学的にも統計学的にも意味が無いからである。必ず、得られたNJ (ML, MP) 系統樹の内枝に対する値を読んで適用しなければならない。

参考文献

- 長谷川政美. 1989. DNAからみた人類の起源と進化, 増補. 海鳴社.
- 宮田 隆. 1994. 分子進化学への招待 (ブルーボックス). 講談社.
- 日本生化学会編. 1993. 分子進化実験法 (新生化学実験講座16). 東京化学同人.
- 根井正利. 1990. 分子進化遺伝学. 五條堀孝・斎藤成也共訳, 根井正利監訳・改訂. 培風館.
- 柴谷篤弘・長野敬・養老孟司編. 1991. 講座進化 (全7巻). 東京大学出版会.
- Wen-Hsiung Li and Dan Graur. 1994. 分子の進化, 遺伝子レベルでみる生物の進化 (化学と生物実験ライン32). 館野義男・山崎由紀子共訳. 廣川書店.
- 植物分類・分子系統研究グループ発行 (季刊). Newsletter for Plant Molecular Systematics, (既刊: 1~19号). (和文ニュースレター. 問い合わせ先: 筆者まで)

(植田邦彦 金沢大学理学部)