

水産資源管理談話会報

第 31 号

日本鯨類研究所 資源管理研究センター

2003年 6月

翻訳・公表希望者は以下の手続きとり、著者の許可を得た上で翻訳・公表する。

1. 翻訳・公表希望者は文章（FAX、手紙）で著者、表題および会報の号を明記し、資源管理談話会事務局を通じて要請し、著者の許可を得て翻訳・公表する。
2. 翻訳公表物を資源管理談話会事務局に送付する。

## 目次

お知らせ	.....	2
系統群あれこれ	中坊 徹次	..... 3
遺伝距離と系統推定	斉藤 憲治	..... 23
まぐろ・かじき類の系群判別	張 成年	..... 29

財団法人 日本鯨類研究所  
資源管理研究センター

〒104-0055 東京都中央区豊海町 4-5 豊海振興ビル

TEL 03-3536-6521

FAX 03-3536-6522

## 遺伝距離と系統推定

齊藤憲治（東北区水産研究所）

分子遺伝学的手法の水産への応用範囲は、養殖魚介類の育種や防疫といった飼育下／実験室内のものから、資源管理のための系群解析や繁殖生態などのフィールドにおける生態調査、加工流通段階での原材料の識別など、産地から食卓までの多岐にわたる。系群解析は、この手法の応用範囲として注目されるのは早い方で、それなりの研究蓄積がある。ここでは、具体的な系群識別手法というより、その基礎となる、DNA配列間の遺伝距離の算出方法と、DNA配列の系統類縁関係の推定について述べる。

### 遺伝距離とは何か

DNAは4種類の塩基（アデニン[A]、グアニン[G]、シトシン[C]、チミン[T]）が固有の配列で並んでいる巨大分子で、この塩基配列に生じたあらゆる変化が系群解析の手がかりになる。DNA配列間の遺伝距離の尺度は、調べた配列の長さに対する変化の数、あるいは1塩基座位あたりの変化の平均で表すことができる。

塩基配列の変化にはある塩基が他の塩基へ置き換わる塩基置換と、ギャップと呼ぶ、塩基と塩基の間に別の塩基が挿入されたり塩基が抜け落ちたりする場合、あるいはある一部の配列の向きが逆転する逆位がある。ギャップや逆位はその発生パターンのモデル化が困難で、遺伝距離など量的評価への利用は限られている。対して塩基置換についてはこれまで多くのモデルが考案され、量的な評価が広く行われている。また、系群という、種内の問題にかかわるような、近縁なDNA配列の間に見られる変異の多くは塩基置換である。そこで一般に、DNA配列間の遺伝距離は、先ほどの「変化」を「塩基置換」と読み替えたもの、調べた配列の長さに対する塩基置換数、あるいは1塩基座位あたりの平均塩基置換数で表す。

塩基置換とは、4種類あるうちのある塩基が他の3種類のどれかに変化することである。当然他の3種類の塩基についても、それぞれ別の3種類の塩基に変化する。どの塩基からどの塩基に変わるときにも、その変わりやすさに差がないとすると、話は簡単である。この場合、近縁な2つのDNA配列間の遺伝距離は、比較したDNA配列の長さに対して何カ所塩基が一致しなかったかで近似的に表すことができる。これを一般に、 $p$ 距離と呼ぶ。ただし、DNA配列間の距離が大きくなると、同じ場所で2回以上置換が起きている（多重置換）可能性が無視できなくなる。多重置換が起きていると $p$ 距離は過小評価となる。

### JCモデル

そこで、多重置換の可能性を織り込んだ距離の推定方法がJukes & Cantor (1969)により考案された。ここではそのベースになったモデルをJCモデルと呼ぶ。実際に計算してみると、JCモデルにより求めた距離は $p$ 距離よりも若干大きな値になる。これ以降、いくつかの遺伝距離推定法を列挙するが、それらはすべて多重置換の可能性を織り込んでいる。

実際には化学的な原因により、変化のしやすさは一様ではないので、塩基置換のパターンをリーグ戦の星取り表のようなもので整理してみよう（図1）。対角線上のます目は塩基に変化がない場合で、ここで問題にしているのはその右上と左下にある6個ずつのます目である。星取り表（マトリックス）は、上の行にある塩基が左側の列の塩基に置換することを示す。JCモデルでは、どの塩基からどの塩基に変わるときにも、その変わりやすさに差がないので、問題にしているすべてのます目を1色に塗っておく（図1A）。

### K2Pモデル

化学式の似ている、 $A \leftrightarrow G$ と $C \leftrightarrow T$ 間の変化（トランジションと呼ばれる）は、そうでない $A \leftrightarrow T$ 、 $A \leftrightarrow C$ 、 $G \leftrightarrow C$ 、 $G \leftrightarrow T$ 間の変化（トランスページョンと呼ばれる）よりも生じやすいことが知られている。

そこで、塩基置換率をトランジションとトランスページョンの2通りに分けて推定して遺伝距離を求める方法がKimura (1980)により考案された（図1B）。ここではK2Pモデルと呼ぶ。マトリックスでは置換率を2通りに考慮するので、それぞれのます目を2色で塗り分けて示す。K2Pモデルに基づく遺伝距離の計算は簡単で、K2Pモデルは遺伝距離の推定に多用されている。

### HKY85モデル

現在、系群解析や系統解析に多用されているmtDNAには塩基組成に偏りがある。たとえば、ヒラメのmtDNAではA、C、Tがそれぞれ27.4%、29.7%、26.1%に対してGは16.8%しかない（多くの遺伝子が乗っているL鎖側の場合で、当然その相補鎖ではGだけが頭抜けて多い）。仮にマトリックスのまず目に入る塩基置換の観察値が均等だとすると、組成の低い塩基の場合にはむしろ塩基置換率は高いことを意味する。すると、より単純なK2Pモデルで推定した遺伝距離は塩基組成に偏りがあると正確ではないかもしれない。

そこで、K2Pの仮定の上に、さらに塩基組成を考慮したモデルが考案された（Hasegawa et al 1985）。ここではHKY85モデルと呼ぶ（図1C）。マトリックスにはトランジションとトランスバージョンの率の色分けのほかに、4種類の塩基の組成を考慮するという意味でその塩基をまず目に書き込んである。HKY85モデルは現在、mtDNAによる遺伝距離の推定に多用されている。

### TN93モデル

mtDNAの研究が蓄積するにつれ、多くの生物群で、トランジションはトランスバージョンの10倍くらい多いらしいことがわかってきた。すると、A $\leftrightarrow$ GとC $\leftrightarrow$ Tの置換の観察値は統計的に置換率の差を調べるのに十分な数に達しているかもしれない。すると、2種類のトランジションの発生率を別々に推定したほうがより正確な距離の推定ができる。

このような考えに基づいて、塩基置換率を3通り（A $\leftrightarrow$ Gトランジション、C $\leftrightarrow$ Tトランジション、トランスバージョン）推定し、さらに塩基組成も考慮したモデルが考案された（Tamura & Nei 1993）。ここではTN93モデルと呼ぶ（図1D）。マトリックスでは塩基をまず目に書き込んであるほか、A $\leftrightarrow$ GとC $\leftrightarrow$ Tのまず目を塗り分け、それぞれ置換率が違うと仮定していることを示す。

### GTRモデル

これまで列挙した塩基置換モデルはいずれも、より一般的なモデルから、ある部分を省いていったサブセットである。すなわちより一般的なモデルとは、塩基置換のすべてのパターン（A $\leftrightarrow$ Gトランジション、C $\leftrightarrow$ Tトランジション、A $\leftrightarrow$ Tトランスバージョン、A $\leftrightarrow$ Cトランスバージョン、G $\leftrightarrow$ Cトランスバージョン、G $\leftrightarrow$ Tトランスバージョンの6通り）について別々の塩基置換率を仮定し、さらに塩基組成も考慮するというものである（Lanave et al 1984）。これはgeneral time reversibleモデルと呼ばれ、ここではGTRモデルと記す（図1E）。マトリックスでは、対角線をはさんで対称になるようにまず目を6とおりに塗り分け、さらに塩基を書き込んである。

GTRモデルは、たとえばA $\rightarrow$ Gとその逆のG $\rightarrow$ Aの置換率には差がないと仮定している。これは多くのDNA配列で、進化的時間スケールでその塩基組成に変化があまりみられないことから、塩基置換はどちら向きにも同じような率で起きているはず（reversible）だと考えられるからである。

### 系統推定

上記のモデル（あるいは、書ききれなかったが理論的にはGTRモデルからあらゆる組み合わせでパラメータを省略したモデルが可能）のいずれかを選択し、DNA配列間の遺伝距離をすべての1対1組み合わせについて推定すると、その結果を使ってDNA配列間の系統推定を行うことができる。距離マトリックスから系統関係を推定する方法には、広く知られているものに平均距離法（UPGMA）と近隣接合法（NJ）の2種類がある（図2）。

UPGMA法は比較するDNA配列のすべてについて進化速度の一定性を仮定するもので、近いもの同士（A、Bとする）を結びつけた後に、次に近いもの（Cとする）との間の距離をAC、BC間の距離の平均として計算し直して結びつけ、この操作を繰り返すことによりすべての配列を順に結びつけていく方法である（Sokal & Michener 1958）。DNA配列の数が少なければソフトウェアなしでも可能なほど計算が簡単で、なおかつ系統の根が自動的に推定できてしまうので、かつて多用された。

しかし、進化速度の一定性が厳密に成り立つとはとうてい考えられないので、系統間の進化速度の一定性を仮定しない系統関係推定法の1つとして、NJ法が考え出された（Saitou & Nei 1987）。NJ法ではすべての系統が1点から多分岐で分化している状態（星形系統樹）から出発し、その中からどの2つの組み合わせをくくり出すと系統樹の枝の距離の合計がより短くなるか、という計算を繰り返していき、枝長の合計を最小にする組み合わせ（系統樹形）を探す最小進化法（ME）の1つの特殊例である。NJ法はME法より計算が著しく簡単で、UPGMA法よりも正しい系統関係を導き出す可能性が高いと考えられているので、現在ではUPGMA法に代わって多用されている。

## モデルの選択

さて、UPGMA法にせよ、NJ法にせよ、系統関係を推定するにはDNA配列間の遺伝距離の計算が必要である。それでは、実際に遺伝距離を計算するにあたって、どのような塩基置換モデルを採用すればよいだろうか。最近ではコンピュータが高速・大容量化し、またいくつかの便利な系統推定ソフトウェアが発表されたので、実際の計算式を知らなくても、複雑なモデルに基づく遺伝距離の計算が簡単にできるようになった。ではどんな場合でも一般的でリアルなモデル（GTRモデル）を用いておけばよいのだろうか。

GTRモデルでは、6通りの塩基置換率と、4種類の塩基についての組成を計算に盛り込む必要がある。比較するDNA配列が短く、また配列数が小さいと、塩基置換の観察数も小さくなる。すると、マトリックスのます目の中に入る例数が統計的に扱えるほどにはならないかもしれない。すなわち、置換率の推定に誤差が出やすくなる。また同様に、塩基組成の推定値にも無視できない誤差が出てしまうかもしれない。そのような誤差を含む推定に基づく計算は信頼できるだろうか？

逆に、JCモデルでは塩基置換率を1通りしか仮定せず、また塩基組成の偏りは考慮しない。マトリックスのます目のすべてを合算するため、観察例数が多くなり、計算上は誤差の少ない推定ができるはずである。しかし、リアリティに欠けるモデルに基づいているため、信頼性に欠けると思われる。

一般に、系群という、種内の問題にかかわるような研究では、比較するDNA配列は互いに近縁なので、塩基置換の観察数は少なく、複雑なモデルでは誤差が生じやすい。これまで多くの研究では、K2Pモデルが用いられてきた。かつてはコンピュータの性能が限られていたし、ソフトウェアも多くのモデルをサポートしていなかったというのが現実的な理由ではあったが、結果的には妥当な選択だったと思う。

しかし、たとえば異なる塩基置換モデルに基づいて計算した遺伝距離から推定した系統推定結果が一致しなかったら、あるいはUPGMA法とNJ法の系統樹形が異なっていたら（分岐の順序にだけ注目し、枝長が一致しないのはこの際無視する）、何を根拠によりよいモデルや系統推定法を選択すればよいだろうか。これまでは複数の方法で推定された系統樹形が異なっていた場合、一致しないところを多分岐で示す（consensus treeという）か、訳の分からない議論でお茶を濁していた。

モデルや系統推定法の選択に理論的根拠を求める向きには、最尤法の枠内でモデルの妥当性を比較することもできる。最尤法というのは、個々のDNA配列がどのような順序とパターンで塩基置換をしてきたか、1つ1つ可能性を計算して積み上げ、確率（尤度）のもっとも高い系統樹形を探すという、これまで述べてきた距離マトリックスに基づく系統推定法とは全く異なる概念に基づく（Felsenstein 1981）。モデルの妥当性を異質な方法で検定してよいか（少なくとも正しい解へ到達する能力は同等か優れていなければならない）、そもそも最尤法でも系統推定はできるのだからはじめから最尤法でやればよいではないか、という問題は残るが、今のところ尤度比検定と、AIC比較の2つの方法のいずれかを利用することができる。

## 尤度比検定

2つのモデルに基づいて得た尤度の比（ $\lambda$ ）を求め、

$$\chi^2 = 2 \cdot \ln \lambda$$

として、モデルの持つ自由度の差を自由度として検定する（Huelsenbeck & Rannala 1997）。たとえば、GTRモデルでは、6通りの塩基置換率と、4種類の塩基についての組成を計算に盛り込み（自由度は $6-1+4-1=8$ ）、K2Pモデルでは2通りの塩基置換率だけ仮定し、塩基組成は考慮しない（自由度は $2-1+0=1$ ）。そこで、GTRとJCモデルの比較の場合、自由度は $8-1=7$ である。UPGMA法とNJ法の比較では、UPGMA法では進化速度の一定性を仮定し自由度が1上昇するので、自由度は1である。

尤度比検定では、自由度の求め方がモデルの持つ固有の値の差（より複雑なモデルでパラメータがいくつ加わるか）となっているところからもわかるように、比較するモデルのどちらかがもう一方のサブセットになっていなければならない。今回挙げた塩基置換モデルでは、前に挙げたものが後ろのサブセットになっており、どの組み合わせにおいても尤度比検定の枠内で検定することができる。

## AIC比較

比較する塩基置換モデルによっては、どちらかが一方のサブセットになっていない場合もある。このようなときにはAkaike情報指標（AIC）の比較を行う。比較する2つのモデル

に基づいて尤度を計算し、

$$AIC = -2 \cdot \text{対数尤度} + 2 \cdot \text{自由度}$$

を用いてAICを求め、その値の小さなほうを選択する (Cao et al 1999) .

尤度比検定, AIC比較ともに, それらのもとになる尤度は系統樹形ごとに独立に計算されるので, モデルを系統樹形の両方を同時に検定することはできない. 異なるモデルにより推定された系統樹形が異なるときには, それぞれの系統樹形ごとに独立に検定し, よりすぐれた, または有意に劣っていないモデルが選択されたら, そのモデルのもとで遺伝距離と系統樹形の推定をやり直すという手順を繰り返して最終的な結果が得られる (Huelsenbeck & Rannala 1997) .

## 引用文献

- Cao Y, Kim K, Ha J, Hasegawa M. 1999. Model dependence of the phylogenetic inference: relationship among carnivores, perissodactyls and cetartiodactyls as inferred from mitochondrial genome sequences. *Genes Genet Syst* 74:211-217.
- Felsenstein J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. *J Mol Evol* 17:368-376.
- Hasegawa M, Kishino H, Yano T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* 22:160-174.
- Huelsenbeck JP, Rannala B. 1997. Phylogenetic methods come of age: testing hypotheses in an evolutionary context. *Science* 276:227-232.
- Jukes TH, Cantor CR. 1969. Evolution of protein molecules. In Munro HN (ed): *Mammalian Protein Metabolism*, p21-132, Academic Press, NY.
- Kimura M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16:111-120.
- Lanave C, Preparata G, Saccone C, Serio G. 1984. A new method for calculating evolutionary substitution rates. *J Mol Evol* 20:86-93.
- Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406-425.
- Sokal RR, Michener CD. 1958. A statistical method for evaluating systematic relationships. *Univ Kansas Sci Bull* 38:1409-1438.
- Tamura K, Nei M. 1993. Estimation of the Number of Nucleotide Substitutions in the Control Region of Mitochondrial DNA in Humans and Chimpanzees. *Mol Biol Evol* 10:512-526.

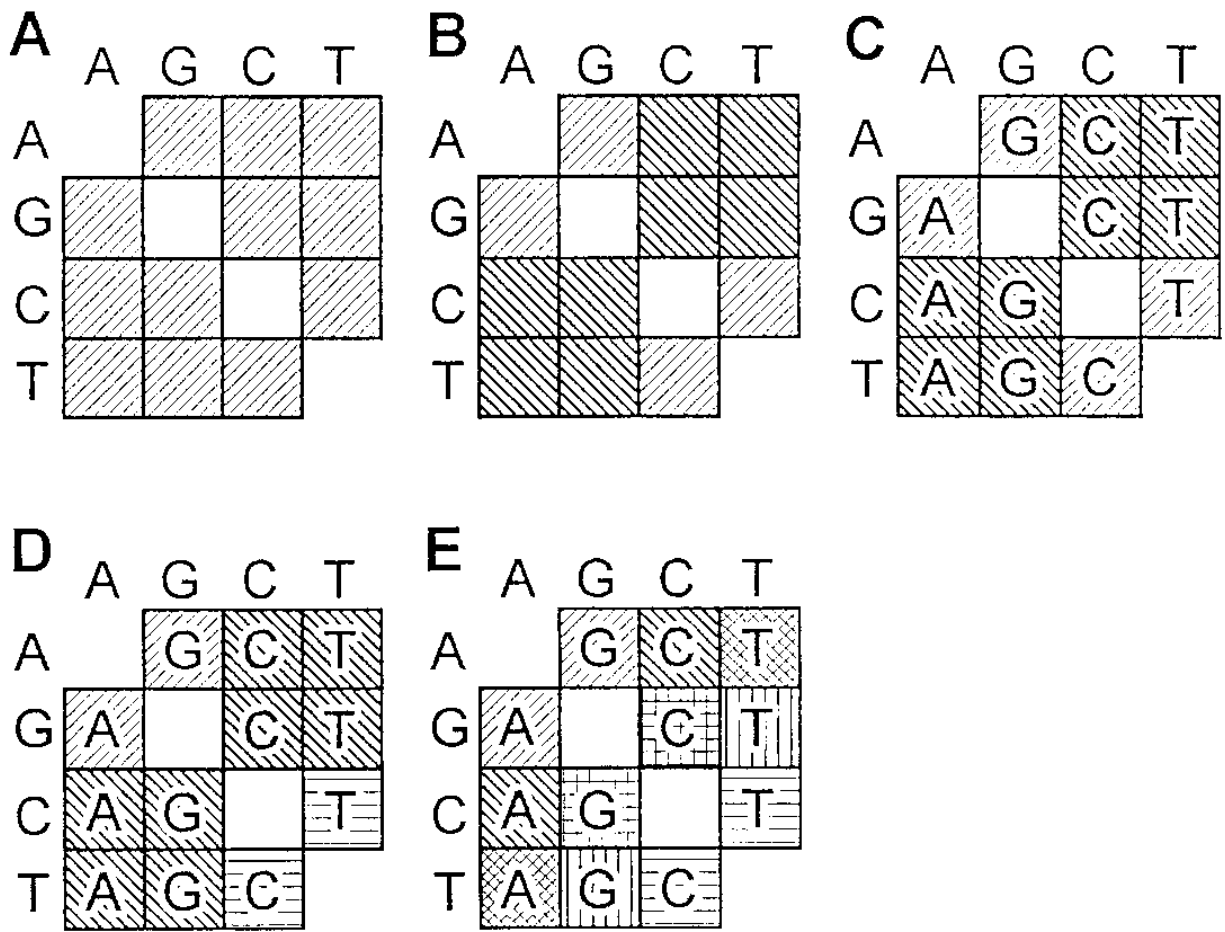


図1. 塩基置換モデルの例. A, JCモデル (自由度=1-1=0). B, K2Pモデル (自由度=2-1=1). C, HKY85モデル (自由度=2-1+4-1=4). D, TN93モデル (自由度=3-1+4-1=5). E, GTRモデル (自由度=6-1+4-1=8).



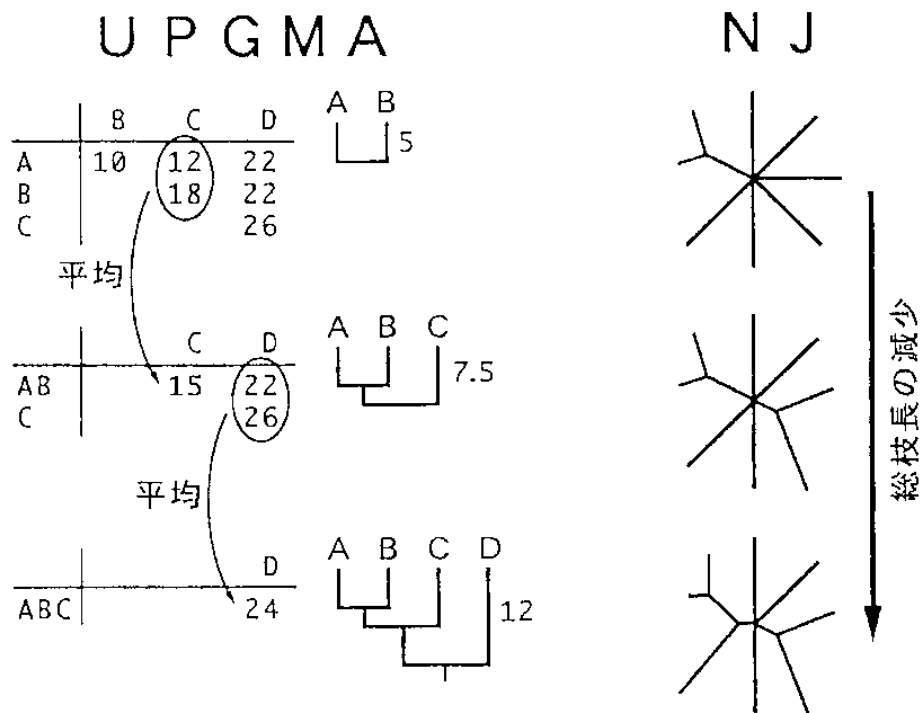


図2. 距離マトリックスに基づく代表的な系統樹構築法.