

分子系統樹から進化史における生物の絶滅率を推定する

巖佐 康・久保拓弥

(九州大学理学部)

はじめに

生物進化の傾向に関しては、たとえば、自殖性の植物は他殖性のものより絶滅率が高いのではないかとか、熱帯の貝類では温帯産のものよりも種分化も絶滅も速くて種の入れ換わりが頻繁だ、などという議論がなされる。また、このように、ある生活型をもつものがほかのものに比べて絶滅しやすい、といった事実はどうすれば客観的に確立できるのだろうか？また、そもそも過去に生じた種分化や絶滅といったプロセスについては、現存する生物を調べることによってどこまで読み取ることができるのだろうか？

DNA 塩基配列のデータは急速に蓄積されつつある。高畠（1995）が示すように、ミトコンドリア全ゲノムといった多数の遺伝子を含む長い配列が決定されると、それをもとにして、現存する種の祖先系統樹が、分岐の系統関係（トポロジー）だけでなく枝分かれの時刻も含めて推定することができる。これは進化速度の大まかな一定性、つまり分子時計のおかげであり、形態データには望めない分子系統樹の大きな利点である。

さて、現存生物種のある系統群の祖先系統樹が、分岐時刻を含めて特定できたとしよう。ここでいう系統群と

は、単一の種から種分化によって形成された生物のグループのことであり、「種」と呼ぶ分類単位は属や科でもかまわない。図1はこのような架空の系統樹を示したものである。祖先系統樹の枝数は現在時刻では現存種数と等しいが、過去に遡るにつれて枝数は単調に減少する。このような、化石ではなくて現存生物から得られるデータを比較してつくる祖先系統樹には、現在の時刻までに絶滅した種は現われてこない。

このような系統樹から、過去の種分化率や絶滅率を推定することができるだろうか。また、過去のある時刻にこれらの種が暮らしている環境が悪化して大量絶滅が起こっていたとすると、そのことは生き残った種の系統樹から読み取れるだろうか（Nee *et al.*, 1992; Kubo & Iwasa, 1995; 巖佐ら1994）。

ランダムな進化とは？

ここで考えている問いは、たとえば生物の2つのグループ（熱帯性の種と温帯性の種、または自殖性の種と他殖性の種）の間で絶滅率や種分化率といったパラメータに違いがあるのかどうかである。説明のために、たとえばある期間での2つの生物群の歴史が正確に復元できたと想像してみよう（簡単のために一方から他方に移ったものが無いとする）。その場合でも、種の入れ換わりの少しでも多かったグループの方が回転率が高い、と単純に結論してよいのだろうか？たとえ両者がまったく同じ回転率をもっていたとしても、種数が有限であるのでたまたま一方のグループが他方よりも入れ換わり数が多くなるかもしれない。そのため、違いがあることについて客観的に結論するには、まず種分化や絶滅を含む統計モデルをつくっておいて、その種分化率や種の回転率などのパラメータが2つの生物群で同じだと仮定する帰無仮説を考えて、観測された系統樹の形が帰無仮説では説明できないくらいにずれているというふうに議論する必要がある。そのときに進化過程を表現するモデルは、種がランダムに分化しランダムに滅びるといった仮定を含むことになる（Raup *et al.*, 1973; Gould *et al.*, 1977）。

しかし実際に生物種が減んだプロセスを考えるならば、決してランダムであろうはずはない。病原体が蔓延

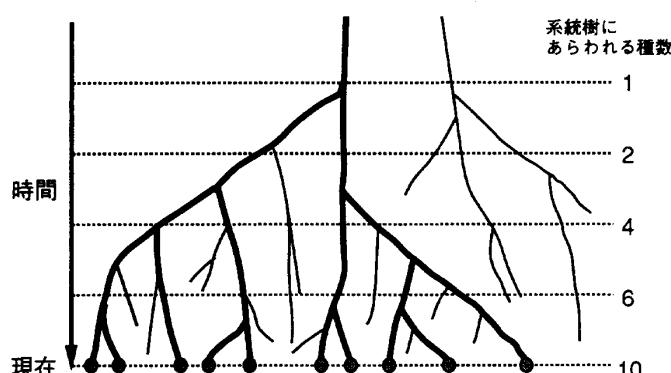


図1. 分子系統樹で観測できる枝（太線）と観測できない枝（細線）(Kubo & Iwasa, 1995)。

現存する生物種に続く枝は観測できるが、子孫が途中で途絶えた系統群の枝は分子系統樹にはあらわれない。その結果、系統樹の枝数（図の右に示す）は単調に増大するよう見える。

した、生息域が気温の変化で縮小した、別の大陸から強力な競争者が侵入したなどと、いちいち理由があるに違いない。種分化のプロセスについても、ランダムではないはずだ。この論文で紹介する議論に含まれる「種がランダムに分化し、ランダムに滅びる」という仮定は、熱帯種であるか温帶種であるという以外には違ひがないような2グループについて絶滅率には統計的に差がないこと、つまり絶滅率と「熱帯種もしくは温帶種である」ととは独立だといっているだけなのである。

2つの系統群の比較ではなくて、1つの群での種分化率や絶滅率を推定するにも、やはり系統発生のモデルが必要である。

系統群を構成する種が互いに等価であると仮定した簡単な系統発生のモデルにおいて、過去の進化についてどの程度の推定ができるのか考えてみる。

モデル1：ランダムな種分化と絶滅

系統発生のモデルとしてまず最初に、種分化・絶滅といった事象が種数などによらずに一定の速度でランダムにおきる場合を考えてみよう。ランダムとはいっても地歴的な状況によって種分化のおきやすい気候や絶滅しやすい地理的状況があるのかもしれないが、一般には種分化率が $b(t)$ 、絶滅率が $c(t)$ のように時間変化してもかまわないとする。しかしそれぞれの時点ではすべての種が同じ種分化率と絶滅率をもつとする。

系統関係はとりあえず無視して、図1の右に数字で示している系統樹の枝数の時間変化に注目しよう。このモ

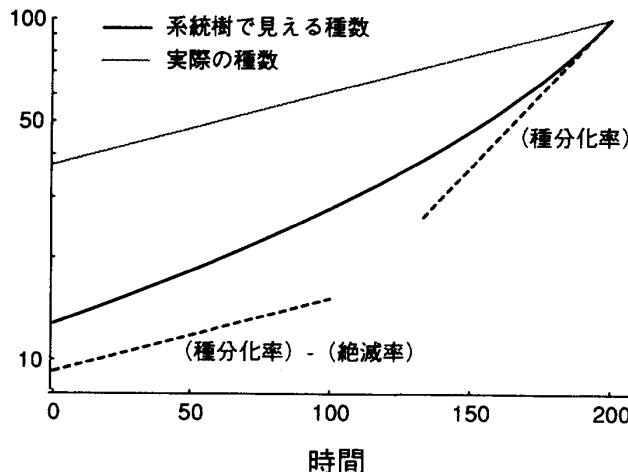


図2. 種分化率・絶滅率が一定値であるとき系統樹の枝数の時間変化 (Kubo & Iwasa, 1995).

横軸は時間を示し、太線は系統樹の枝数の期待値(現存生物の祖先のみ)、その時刻に存在していた種数(後に絶滅する種も含む)を示す。現在時刻に近いところでは系統樹の枝数の対数の増加速度(グラフの傾き)はほぼ種分化率 b に等しく、系統樹の根に近いところでは種分化率と絶滅率の差 $b - c$ に漸近する。

デルに従う種数の時間変化は、分枝過程という確率過程として表現される。その手法を利用して、現在(時刻 T)に存在している種についての祖先系統樹で枝数の期待値の時間変化が計算できる。時刻 t の種数を N_t 、現在時刻を T とすると、

$$\frac{N_t}{N_T} = \frac{1}{1 + \int_t^T b(u) \exp \left[\int_u^T (b(z) - c(z)) dz \right] du} \quad \dots \dots \dots (1)$$

(Harvey et al., 1995; Kubo & Iwasa, 1995)

まず最初に、一番簡単な場合として、種分化や絶滅の速度が時間的に一定であるときに、系統樹の形はどうなるかを図2に示している。横軸は時間で、縦軸は系統樹の枝数の対数であるが、種分化率・絶滅率の大小にかかわらず、単調増加で下に凸のカーブを描いている。とく

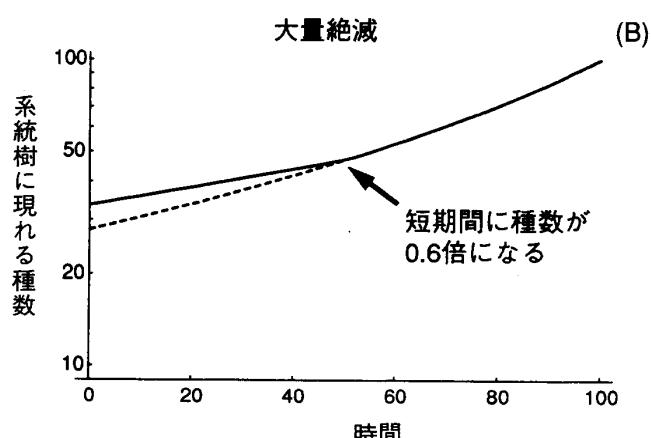
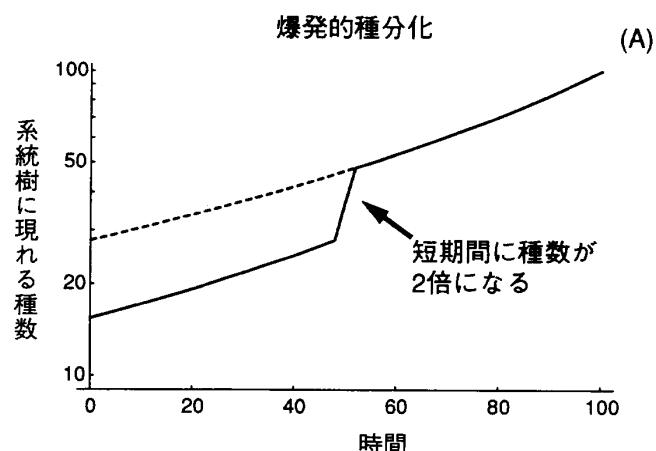


図3. 種数の急激な変動が系統樹に与える影響 (Kubo & Iwasa, 1995).

(A) 時刻50付近で種分化率が一時的に大きくなり種数が2倍に増えた。祖先系統樹の枝数もその時刻で急増する。

(B) 時刻50付近で絶滅率が一時的に大きくなり、種数がもとの60%になったが、その痕跡を系統樹から読みとることは困難である。

に、種分化率が絶滅率より大きいときには、枝数の対数の増加速度（グラフの傾き）は、ごく最近では種分化率 b に等しく、古いところではもっと遅くて種分化率と絶滅率の差 $b - c$ に漸近する。ということは、たとえば実際には種分化や絶滅の率が時間的に一定でも、祖先系統樹の形は、あたかもごく最近に種分化が盛んになったかのように見えるのである。もしこのモデルが正しいならば、種分化の時刻を特定できる系統樹から（途中で絶滅した種は観測できないにもかかわらず）原理的には種分化率 b と絶滅率 c を同時に推定できるはずである。

しかし、このような推定はどの程度まで信頼できるのだろうか？コンピュータシミュレーションによって系統群をつくり、その系統樹から種分化率・絶滅率を（1）式へのあてはめで推定し、その推定値の特徴を調べた。その結果、種数が多いと種分化率 b はかなり正確に推定できた。ところが、絶滅率 c についてはいくら種数を増やしても推定値の分散が小さくならなかった。このことから、種分化率は推定できるが、絶滅率の推定は困難と結論される（Kubo & Iwasa, 1995）。同じ様な考え方にもとづいて、ランダムモデルによってすべての種ではなくて一部をサンプルしてつくった系統樹を解析する場合（Harvey *et al.*, 1994, 1995）や、種分化率と絶滅率の最尤推定法や尤度関数の等高線を用いた解析（Nee *et al.*, 1994），それらのウイルス遺伝子の系統樹へに応用についても議論されている。

種分化率・絶滅率の時間変化については系統樹の枝数から推定できるだろうか？極端な例として、爆発的種分化と大量絶滅がそれぞれ系統樹に与える影響を図3に示す（Kubo & Iwasa, 1995）。絶滅率が一定で種分化率だけが一時的に高くなる爆発的種分化は系統樹からは

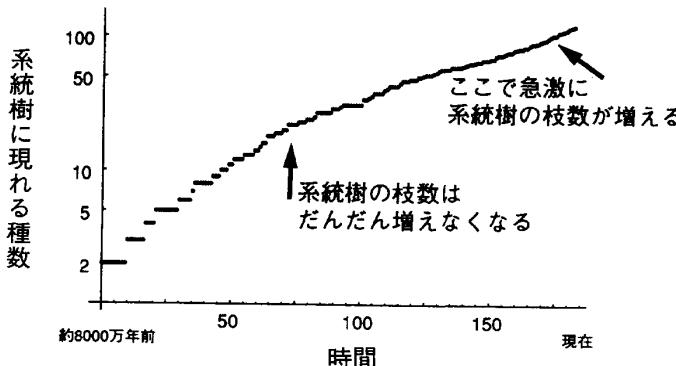


図4. 鳥類のすべての科についてDNAハイブリダイゼーションにより作成した祖先系統樹について、各時刻での枝の数（Kubo & Iwasa, 1995）。縦軸は枝の数の対数、横軸は時間を現わす。右端が現在。祖先系統樹では枝の数は常に増える一方である。Sibley & Ahlquist, (1990)にもとづいて Nee *et al.*, (1992) が描いたもの。

つきり読みとることができる（図3A）。一方、種分化率は一定で絶滅率だけが一時的に高くなる大量絶滅があった場合の結果を、図3Bに示しているが、このような系統樹から大量絶滅の痕跡を見いだすのは困難であろう。

実際の分子系統樹の枝数の時間変化の一例を図4に示す。これは鳥のすべての科の系統関係をDNA-DNAハイブリダイゼーションによって決定し（Sibley & Ahlquist, 1990），その系統樹を整理したものである

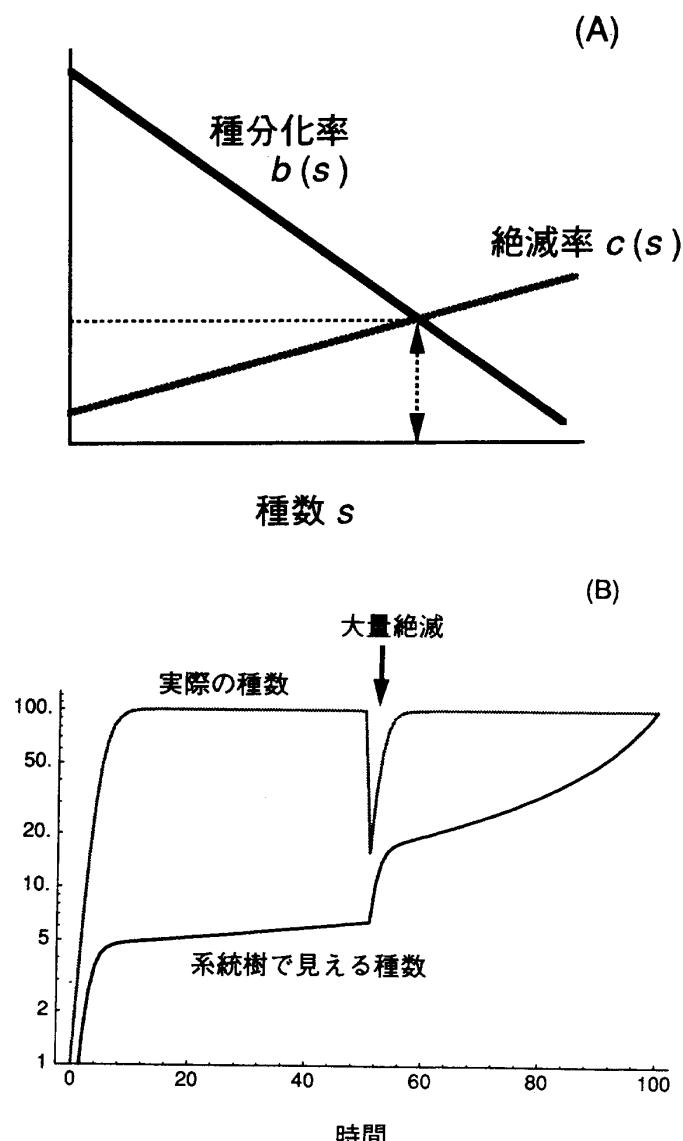


図5. 種数に依存するときの種分化率・絶滅率。

ここでは、例として、種数が増加すると、種分化率が減少し、絶滅率が上昇する場合を考える。種分化率と絶滅率が等しくなるところが平衡種数で、そのときの種分化率・絶滅率を回転率と呼ぶ。回転率の大小は系統樹の形に大きな影響を与える（A）。

回転率が高い系統群は系統樹をみて大量絶滅の時刻が特定できる。種分化率・絶滅率が（B）のようになっている系統群では、大量絶滅の直後に爆発的種分化がおこり、しかもそのときに生じた種はなかなか入れ換わらない。そのため現存生物の系統樹をつくって、枝数グラフが折れまがりから大量絶滅が起ったことがわかる。

(Nee *et al.*, 1992; Kubo & Iwasa, 1995). この図4の横軸は時刻に対応しており、左端は約8千万年前、右端は現在にあたる。縦軸は系統樹の枝数を示しているが、その増え方は、はじめ急速に増え、次第にその速度が鈍り、現在に近いところではまた急に増えるように見える。どのような種分化率や絶滅率を想定すればこのようなパターンが生じるだろうか。その答えはじつは無数にあるのだ。種分化率が一定値で絶滅率が次第に上昇するというのでも、逆に絶滅率が一定値で種分化率が減少するのでも、またそれらの組み合わせでも、図4に示されるものとまったく同じ祖先系統樹が再構築できる(Kubo & Iwasa, 1995)。つまり分子系統樹の形をみただけでは、どの組み合せが本物であるかは判別できない。

以上のようにランダムな種分化と絶滅とがまったく独立に生じるモデルでは、系統樹から絶滅率を推定することはできない。

モデル2：種数とともに種分化率が減少する

種間の競争を考えると、種数が多くてニッチのほとん

どが占められていれば、新たな種分化は成功しにくくなり、すでにいる種は絶滅しやすくなるかもしれない。そこで、種分化率が系統群の種数とともに減少し、絶滅率が種数とともに増大する系統発生のモデルを考える(図5 A)。このような一般出生死過程において、系統樹の枝数の時間変化がどうなるかを調べた。

ランダムモデルでは大量絶滅の時刻が特定できなかつたが、このモデルでは大量絶滅によって種数が減少すると同時に種分化率が高くなり、引き続いて爆発的種分化がおこり、その影響は祖先系統樹の形にも現われる(図5 B)。つまり大量絶滅をその後の爆発的種分化から知ることができるるのである。

ただし、種分化率と絶滅率がつりあったときのそれらの値、すなわち種の回転率(図5 A参照)が重要である。回転率が大きいときには、つりあいの状態でも種がどんどん入れ替わるのに対して、回転率が低いと、いったん生じた種はなかなか入れ替わらない。回転率が遅い場合には大量絶滅の影響が系統樹にはっきり現れるが、回転が速い場合には大量絶滅だけでなく爆発的種分化さえも系統樹に影響を残しにくく、読みとることが困難になる。

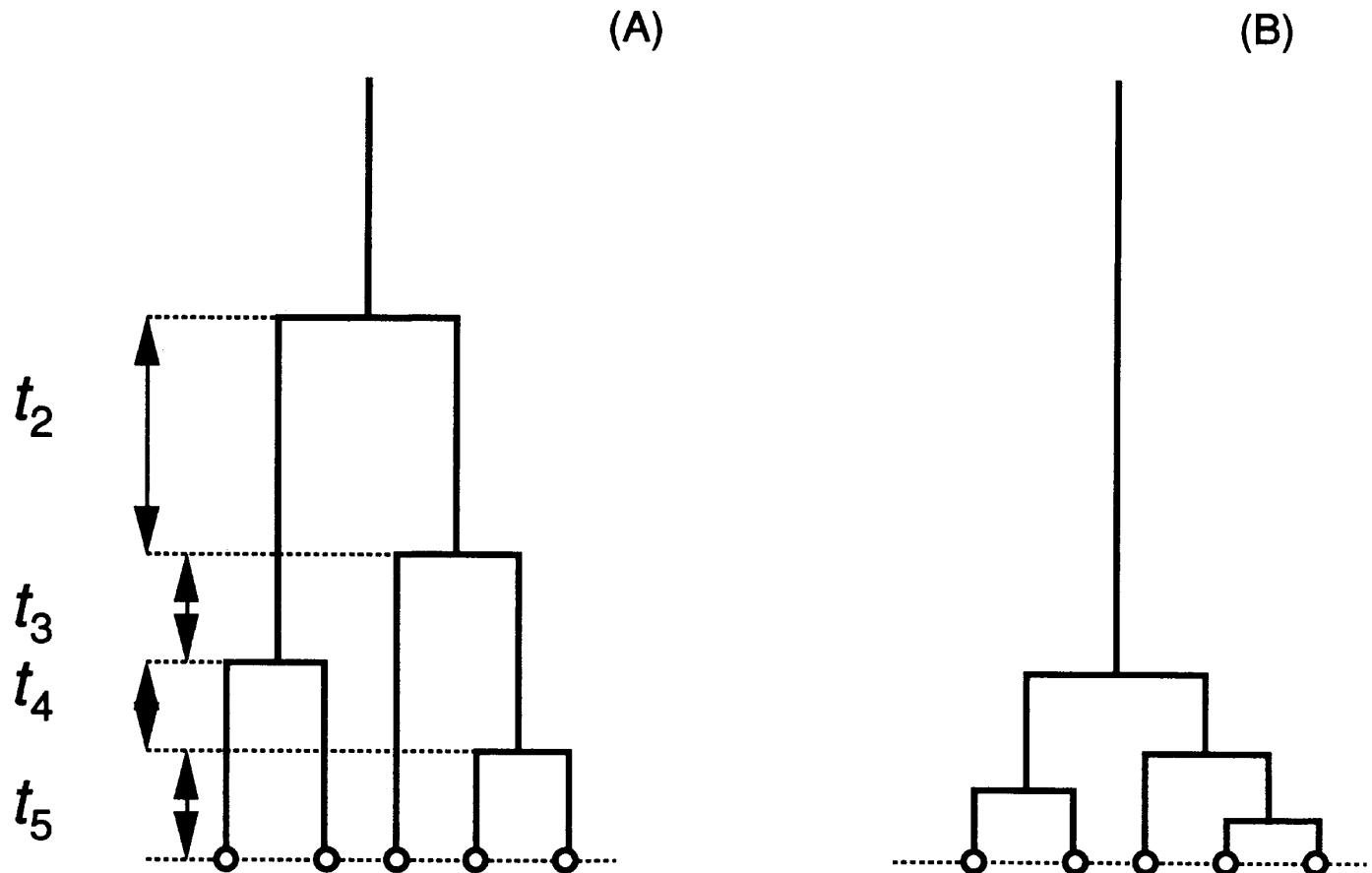


図6. 2つの系統樹の違いから回転率の差を結論する方法。

現存する分類群の祖先系統樹をたどり、それらが1つのまとまるまでの間の系統樹の枝の長さを合計したものを系統樹の枝の全長と呼ぶ。これは、Heyモデルでは回転率の十分統計量であり、推定値や検定もこの枝の全長を用いるものである。2つの種のグループ(A)と(B)について、後者の方が絶滅・種分化が頻繁に生じていると結論しそうだが、この図の差異は統計的に有意ではない(片側2.5%で検定)。

モデル 3：絶滅と種分化が同時に起こる

上に述べたように、現存種の分子系統樹から種分化率と絶滅率を独立したパラメータとして推定すると、とくに絶滅率に関してはその推定値の分散がきわめて大きい。進化過程に関する推定を行なおうとすればもっと単純でパラメータ数の少ないモデルを用いる必要がある。たとえば系統群の種分化率と絶滅率が強くカップリングしていると考えてみよう。

ここでは極端な場合として、系統群に含まれる種数は常に一定値に維持されていると仮定する。系統群に属する1つの種がある時刻にランダムに滅びると、生き残っている任意の1種から直ちに新しい種が種分化して、絶滅した種の生態学的地位を占めるとする(Hey, 1992)。ここではこのような種の入れ換わりの速度、「回転率」を系統樹の形から推定することを考える。(この確率過程は、種を遺伝子、系統群を集団と読み変えると、集団遺伝学の Moran モデルと同じである)。系統群からランダムに n 種を選び、それらのすべての共通祖先までたどるときの系統樹の枝の長さの和 B を分子系統樹の「枝の全長」と呼ぶ(図 6)。このモデルを仮定すると回転率の最尤推定値は $(n-1)/B$ で与えられる(Hey, 1992)。モデルに従って生成される分子系統樹の枝の全長はガンマ分布に従い、それは回転率の十分統計量になっている。一様最強力不偏検定、一様最小不偏推定区間なども枝の全長 B から計算する(Iwasa, Y., T. Kubo & H. Matsuda, 未発表)。

異なる系統群での回転率の違いを検定する

この論文の最初に挙げた問題にもどってみよう。たとえばすみ場所の違いが種分化と絶滅に影響を与えるという経験則がある。島にいる種は大陸の種よりも滅びやすい、あるいは逆に種分化しやすいといわれている。また、化石による研究では熱帯産の貝類は温帯産のものに比べて種分化(正確には属分化)の速度も絶滅率も高いことがわかった(Jablonski, 1993)。このように2つの系統群での回転率が異なっていることを確立するためにはどのようにすればよいだろうか(Iwasa, Y., T. Kubo & H. Matsuda, 未発表)？

比較している2つの系統群(系統群1と2)があるとき、上記の「種数一定のモデル」を仮定して、それぞれの回転率を最尤推定する。系統群 i ($i=1, 2$) の回転率の最尤推定値 \hat{c}_i は、

$$\hat{c}_i = (n_i - 1)/B_i \quad \dots \dots \dots (2)$$

で与えられる(これは不偏推定値ではない)。 n_i , B_i は

それぞれ系統群 i のサンプル種数と分子系統樹の枝の全長である。ここで以下の2つの仮説

仮説1：比較している系統群はどちらも同じ回転率を持つ($\hat{c}_1 = \hat{c}_2$)

仮説2：比較している系統群はそれぞれ異なる回転率を持つ($\hat{c}_1 \neq \hat{c}_2$)

を評価し、このうち一方をよりよい仮説として選択しよう。その方法として、まず尤度比検定を考えてみる。それぞれの仮説のもとで、自由に選べるパラメータを尤度が最大になるように選んでおいて、それぞれの尤度の比を計算し、それがある値よりも大きいという条件で棄却域の形を決めるのだ。2つの種のグループにおける回転率が等しいという仮説が棄却されるのは、

$$\frac{B_1}{B_2} < x_- \text{ or } \frac{B_1}{B_2} > x_+ \quad \dots \dots \dots (3)$$

という形をしている。いま、有意水準を片側2.5%と設定したすると、枝の全長の比 B_1/B_2 が従う確率分布の計算から、実際には等しいにも関わらず上記の不等式の後者をみたす確率が0.025となるような値 $x_+^{0.025}(n_1, n_2)$ が求められ、表1に示してある x_- については、次の式から求められる。

$$x_-^{0.025}(n_1, n_2) = \frac{1}{x_+^{0.025}(n_2, n_1)} \quad \dots \dots \dots (4)$$

これとは異なる考え方として、赤池の情報基準量(以下、AIC)を用いて、仮説1と2のどちらがよいモデルであるかを比較する方法もある。これを用いるときには有意水準といった考えをしない。データが与えられる

表1. 系統群1が系統群2よりも回転率が小さいと結論できるための枝の全長の比の最小値。

系統群2の種数

	2	5	10	20	50
系統群1の種数	2	37.49	1.515	0.5066	0.2143
	5	151.3	4.433	1.336	0.5364
	10	341.1	9.076	2.596	1.007
	20	720.7	18.28	5.054	1.907
	50	1860	45.82	12.36	4.551

左側の数字は系統群1の種数、上の数字は系統群2の種数。この数字よりも枝の全長の比が大きいときに系統群1が系統群2よりも回転率が、片側2.5%の検定で有意に低いと結論できる。たとえば系統群1も2もともに20種を含む場合には、枝の全長の比率が約2倍以上だと、系統群1が系統群2よりも有意に回転率が遅いといえる。

とき、それぞれのモデルについて、

$$AIC = -2 \times \log(\text{尤度の最大値}) + 2 \times (\text{自由なパラメータ数})$$

として定義される量を計算しそれが小さい方が選択すべきモデルであるとするのだ。今の場合には、パラメータが2つある仮説2の方が最大尤度が高くなるが、その相対的な差がパラメータ数の差である1より大きいかどうかを調べるというのが、AICを使う方法の意味である。仮説1と2のそれについてAICを計算し、その差を整理すると、

$$\left(1 + \frac{B_2}{B_1}\right)^{n_1-1} \left(1 + \frac{B_1}{B_2}\right)^{n_2-1} > \frac{e^{(n_1+n_2-2)^{n_1+n_2-2}}}{(n_1-1)^{n_1-1}(n_2-1)^{n_2-1}} \dots (5)$$

となる領域で、仮説2の方がよいモデルとして選択される。計算してみると、これは尤度比検定の1種であることがわかる。有意水準に対応する確率を求めてみると片側2.5%よりもかなり高い。つまり同じデータに対して、AICを用いた方法の方が、2.5%の有意水準の検定よりも2つの系統群の回転率の最尤推定値に差があると結論しやすい傾向がある。

具体的に計算してみると、この回転率の最尤推定値が2つの系統群の間で回転率が違うと結論するためには、各系統樹に含まれる種数がかなり多くなければならないことがわかる。たとえば、同数の種を含む2つの系統樹で一方の枝の全長が他方の2倍である時には、回転率の最尤推定値も2倍違うことになる。このように推定値に大きな違いがあったとしてもそれが有意かどうかは別である。このとき2つの回転率が有意に違うと結論するためには、各系統樹には18種以上含まれねばならない。もし最尤推定値が一方が他方の3倍であれば、8種以上が含まれねば有意ではないし、50%の違いならば49種以上がそれに含まれねばならない（AICをもちいると、それぞれ10種、5種、26種となり、回転率の違いを結論しやすい）。

おわりに

生活型、生息場所、配偶システムなどによって種分化・絶滅などの率が異なるという仮説を考えて、それを現存種の分子系統樹から結論するためには、統計モデルが必要である。かなり簡単な統計モデルを仮定して相当な数の種を含む系統樹を調べないと、統計的に有意だとする結論を導くことは難しい。しかし、この方法は、近い将来ますます有望になるだろう。枝分かれ時刻の推定を含めた分子系統樹が多数の分類群について得られることは間違いないからである。

引用文献

- Gould, S.J., D.M. Raup, J.J. Sepkoski Jr., T.J.M. Schopf & D.S. Simberloff 1997. The shape of evolution: a comparison of real and random clades. *Paleobiol.* 3 : 23-40.
- Harvey, P.H., E.C. Holmes, A. ϕ . Mooers & S. Nee 1994. Inferring evolutionary processes from molecular phylogenies. In : R.W. Scotland, D.J. Siebert & D.M. Williams (eds.), *Models in Phylogeny Reconstruction*, p. 313-333. Systematics Association Special Volumes Series 52, Oxford University Press, Oxford.
- Harvey, P.H., R.M. May, & S. Nee 1995. Phylogenies without fossils : Inferring lineage birth and death rates. *Evolution* (in press)
- Hey, J. 1992. Using phylogenetic trees to study speciation and extinction. *Evolution* 46 : 627-640.
- 巖佐庸・久保拓弥・松田裕之 1994. 生物集団の絶滅はどのように起こるのか：分子系統樹から絶滅率を推定する *科学* 64 : 675-683.
- Jablonski, D. 1993. The tropics as a source of evolutionary novelty through geological time. *Nature* 364 : 142-144.
- Kubo, T. & Y. Iwasa 1995. Inferring the rates of branching and extinction from molecular phylogenies. *Evolution* 49 : 694-704.
- Nee, S., A. ϕ . Mooers & P.H. Harvey 1992. Tempo and mode of evolution revealed from molecular phylogenies. *Proc. Nat. Acad. Sci. USA* 89 : 8322-8326.
- Nee, S., R.M. May, & P.H. Harvey 1994. The reconstructed evolutionary process. *Phil. Trans. Roy. Soc. (Lond.) B* 339 : 139-146.
- Raup, D.M., S.J. Gould, T.J.M. Schopf & D.S. Simberloff 1973. Stochastic models of phylogeny and the evolution of diversity. *J. Geol.* 81 : 525-542.
- Sibley, C.G. & J.E. Ahlquist 1990. *Phylogeny and Classification of Birds*. Yale Univ. Press, New Haven, CT.
- 高畠尚之 1995. 遺伝子の系図と靈長類の系統樹 種生物学研究 19 : 31-32.