遺伝子の系図

国立遺伝学研究所 高 畑 尚 之

The family relationships among a sample of members drawn from a particular generation of populations or species play a central role in describing the genealogical behaviour of generations of individuals. Such a genealogical approach is certainly interesting and important in its own right, but it is also very useful to produce a wide variety of classical results in the mathematical theory of population genetics. The power and elegance of the theory rely on 'equivalence' or 'exchangeability' among individuals in a population and have been best demonstrated in single-locus multiple-allele systems. Here I would like to introduce the basic idea of the theory and some results of its application to molecular taxonomy.

genealogy / coalescence / old polymorphism / phylogeny / population genetics

何年か前に「ルーツ」という映画が評判になった。あるアメリカの黒人が自己の由来に興味をもちその祖先をたどる。そして7世代前の祖先がアフリカの部族の大酋長であった事実を発見して、虐げられた白人社会の中で自己の尊厳を取り返すといった内容であったと思う。この黒人に限らず、多くの人が自己の由来に興味をもつ。しかしその動機は余りはっきりしていないことが多い。それにたとえ何十世代、何百世代も前の祖先のことが分かったとしても、その人の存在には大した意味も与えないであろう。歴史から学ぶことは多いけれども、遠い祖先に対する好奇心を動機づけることはなかなか難しい。にもかかわらず、ヒトはそしてヒトだけが遠い過去に想いを寄せ、繰り返し同じ質問をしてきた。このような知的好奇心の由来を問うことは興味深い。しかしそれはこの小論の目的ではないし著者にその資格もない。ここではなぜか分からないけれど著者はこの好奇心をもっており、また読者の多くの方もそれを共有しているという前提に立って話を進めようと思う。

本論に入る前に、先の「ルーツ」と関係して2, 3の点に触れてみたい。ヒトを初め多くの生物は雌雄異体の二倍体であり、ある個体の祖先というと父系も母系も考える必要がある。従って、祖先をたどっていくとその数は幾何級数的に増加する場合がある。近親婚

がないと7世代前の祖先の数は $2^7=128$ である。その中には大酋長もいたのであろうが、多くは普通の身分であったかも知れない。「ルーツ」の面白さの一つは、過去をさかのぼるときの困難をどう克服するかという点にもあろう。家系図を重んじる風習のある日本でも、余程の名家でない限り戦国時代の混乱期を越えることはそう容易ではあるまい。まして口承伝承の時代までさかのぼることは、普通の人には無理なことである。二倍体個体の祖先図は極めて複雑であるうえ、仮りにそれが記録として留められているとしてもその歴史は高々数十世代に達しない。そこで我々の好奇心を幾分でも満たすためには、質問の仕方を代えざるを得ない。それは一口に言って質問を没個性化させることである。例えば個体という代わりに、その個体をもつ父親由来の遺伝子ということにする。しかしこれでもまだ不十分である。その遺伝子内で組換えが起きるとある部分は父親から残りの部分は母親からということになり、その系図は個体の系図同様大変複雑になる。こうした系図の研究¹⁾は、まだ十分発達しているとはいえない。従って、我々が注目するのは片親由来の、しかも遺伝的組換えが無視できる程強く連鎖した遺伝子またはDNAの領域である。第二の没個性化は、サンプルする個体についてである。サンプルする個体は特定の家系に偏ったものではなく、集団からでたらめに選ばれたものである必要がある²⁾。

上のような前提に立つと、1つの遺伝子の祖先をたどることには意味がない。そのような遺伝子は必ず1個の親遺伝子をもっているし、それがどのような個体

Gene Genealogy

Naoyuki TAKAHATA

National Institute of Genetics, Mishima, Shizuoka-ken 411.

に存在したかということも興味の対象ではないからである。しかし複数のでたために選ばれた相同座位の遺伝子については、それらがいつ最も近い共通祖先遺伝子から由来したかなど、興味ある問いをすることができ。例えば日本人の集団からでたために2個体を選び、それがもつ相同座位の遺伝子の最も近い共通遺伝子はいつ頃であるのかを問うことは、日本人の起源を考える上で有用な情報を与えるであろう。このような質問に答えるため以下では1つの集団内の遺伝子の系図を支配する法則や、あるいは異なる種から選ばれた相同遺伝子の系図と種の系統関係の対応などについて考えてみる。その前に、DNA 配列のデータに基づいた遺伝子の系図そのものを決める方法について簡単に述べておきたい。

系図の作成

ある遺伝子座の遺伝子を複数個サンプルし、そのアミノ酸配列や DNA 配列を決めると、遺伝子間の配列上の差異を定性的にも定量的にも比較検討することができる。その差異に基づいて遺伝子間の近縁関係を図式化するのであるが、その時用いる基本的な原理はより近縁関係にある遺伝子程配列上の差異が小さいという単純なものである。しかし実際に遺伝子の系図を作成する方法は実に多様である^{3,4,5)}。その内で比較的よく用いられる3つの方法は、(a)距離行列法、(b)最大節約法、(c)最尤法とよばれる。

距離行列法では、配列の決まった相同遺伝子の全てのペアについて進化距離を計算する。進化距離は2つの相同遺伝子間に蓄積した突然変異の総数、あるいは便宜のために塩基座位あたりに生じた突然変異の総数とすることが多い。このようにして得られた距離行列に基づいて系図を作成するいくつかの方法がある。その内で UPGMA とよばれる方法⁶⁾は最も単純である。この方法は、まず進化距離の最も短いペアを系図上で結びつける。結びつけられたペアの分岐点はその進化距離のちょうど中間にとる。一度あるペアの結びつきが行われると対象となっている遺伝子の総数は1つ減る。結びつけられたペアはあたかも1つの遺伝子として改めて距離行列が計算しなおされる。結びつけられた遺伝子のグループ間の距離は、それに含まれる遺伝子間の距離の算術平均とする。この操作を繰り返せば、最後に1つのルートをもつ系図ができる。

距離行列法による系図作成は、上の手続きから推測できるように簡便である。サンプルした遺伝子の数を n とすると可能なルートをもった系図の数は $(2n -$

3) で、 $n=10$ のときそれは 34,459,425 通りある。しかし UPGMA に限らず一般に距離行列法では、上のような天文学的数字の中から比較的容易に1つの系図をほとんど一義的に決めることができる。ただ距離行列法で決められた系図がどの程度信頼のおけるものであるかという点を統計的に議論する必要がある。遺伝子の上に蓄積する突然変異は明らかにある確率法則に従っており、それに基づいて再構成される系図の信頼性もある確率空間の上で議論されねばならないはずである。

図1. ヒトのミトコンドリア DNA に基づく遺伝子の系図。Stoneking et al.⁷⁾を参考に作成。

最大節約法は、与えられた系図でまず祖先遺伝子の配列を推定する。それに基づいて系図全体の中で最小何個の突然変異による変化が要求されるかを数える。可能な系図で同様の操作を行い、その内で最小個数の突然変異を要求する系図を最終的に採択する。この方法は系図の型を最初に決めるが、系図中の枝の長さを決めるには更に別の仮定をする必要がある。最大節約法を用いて描かれた系図の例は多いが、図1にあるヒトのミトコンドリア DNA に関する馬蹄型系図⁷⁾は最も印象的なものの1つであろう。

最大節約法の欠点は、進化が最小の過程を通して起こるという原理そのものにあると思われる。DNA の変化は物理法則と違って必ずしも最小の原理に従ってはいない。更に距離行列法と同じく最大節約法で決められた系図の信頼性を統計的に議論することは大変難しい。

最尤法は、これまでの方法と異なり DNA の変化がどのような確率法則に従っているかをはっきり設定した前ですべての可能な系図の尤度(もっともらしさ)を計算し、その内で最大の尤度を示す系図を選ぶ。それと共に確率モデルに現われるパラメータの最尤値も推定する。従って最尤法を用いるには、DNA の変化に関するモデルを設定しなければならない。これまで多数のモデルが提出されているが、どのモデルが最も現実的かなかなか決め手がないのが現状である。そのため最尤法は非現実的な DNA の変化のモデルに基づいて使用されているという批判がある。しかし DNA 変化についてより現実的な情報を知ることができるになれば、最尤法はそれをいくらかでも取り入れ改良することが可能である。第2の批判は異なる系図に対して得られる尤度を直接比較することはできないのではないかという点である。系図が異なりそこに含まれる推定すべきパラメータの数が異なると、尤度それ自体を比較することは意味がない。パラメータの数を自由に定めることができれば、より多くのパラメータをもつ系図の方がより高い尤度をもつに違いないからである。この自由度と尤度の関係は、赤池⁸⁾の AIC に見事に表現されている。しかし系図の場合にはその型が変わっても自由度は変わらない。従ってどの系図を選択するかは尤度の比較だけから可能である⁹⁾。ただしこのとき、可能な系図の先験的確率は皆等しいことを仮定している。系図作成には未解決の問題が多いが、選んだ系図を他の可能な系図と統計的に比較できるかどうかが最も大切な問題であろう。

遺伝子の合体

遺伝子の系図は、サンプルした遺伝子間の近縁関係を図式化したもので、それらが共通の祖先遺伝子からいつ、どのような型で分岐してきたかを表わしている。系図を時間軸にそって逆にながめると、サンプルした遺伝子が次々に祖先遺伝子に合体し、その祖先遺伝子が更に古い祖先遺伝子に合体していく様子を知ることができる。このような遺伝子の合体は、それらが単一の有限集団に存在していたときにだけ起こりうる。簡単のために、 $2N$ 個の遺伝子が自己増殖している集団を考えよう。この自己増殖の法則を決めることが、遺伝子の系図を議論する上で基本的なことである。しかし遺伝子の合体が起きる理由に関しては、 $2N$ 個の遺伝子のうちあるものは2個以上のコピーを次世代に残すことがあることを喚起すれば十分であろう。2個以上のコピーを作った親遺伝子はいわば分岐を起こしているわけで、逆にその子孫遺伝子から祖先をたどっていくとこの親遺伝子に合体するようにみえるわけである。合体が起きる時点でのコピー遺伝子は同一集団に属しているはずであることも、以上のことから明らかであろう。図1の147個のミトコンドリア DNA は、究極的には 'Ancestor' に合体しており、その直前の2つの祖先 DNA は同一集団に属していたはずである。このことは他の合体でも同様である。

増殖モデルにはいろいろあるが、その多くのモデルは、時間的に集団の大きさが一定で集団内の $2N$ 個の遺伝子が等価であることを仮定している^{10,11)}。等価とは $2N$ 個の遺伝子の増殖が同一の確率法則に従っているという意味である。従って特定の型の遺伝子がよく増殖するということはない。遺伝子レベルの変異の多くは等価性を満足する¹²⁾と思われるので、以下でもそれを仮定する。

単一の集団から n 個の遺伝子をサンプルしたときの系図法則に関する理論的研究はここ10年間程の間に急速に発展した^{13,14,15)}。系図を決めるには、その合体の型と合体が起こるまでの時間を与えなくてはならない。しかし集団が大きい場合の系図の法則は大変簡単である。まず集団からサンプルした n 遺伝子の系図は増殖モデルの詳細によらなくなる¹⁵⁾。しかも3つ以上の遺伝子が同時に合体を起こす可能性も無視できる。このことと等価性の仮定から、サンプルした n 遺伝子の系図の型はランダムな二分岐を原理として考えればよいことになる。例えば $n=4$ の場合の系図は2通りの型があり、それぞれが起きる確率も容易に計算で

きる¹⁶⁾。問題は n 遺伝子が n' ($1 \leq n' < n$) 遺伝子に合体するまでの時間、あるいは系図中の枝の長さを決める法則を与えることである。 $n = 2$ の場合は単純であるが、本質的なことを理解する上で役立つばかりでなく実的にも重要である。等価性のもとでは子が親を $2N$ 個の中からでたらめにしかも一様を選ぶ権利があることに注意すると、2つの遺伝子が共通の親遺伝子に合体しない確率は $1 - 1/(2N)$ である。従って、世代前まで合体が起こらない確率は $\{1 - 1/(2N)\}^t \approx \exp \{-t/(2N)\}$ 、合体が起こってしまっている確率は $1 - \exp \{-t/(2N)\}$ となる。また合体が起こるまでの待ち時間 t の確率密度は $\{1 - 1/(2N)\}^{t-1} \cdot 1 - 1/(2N) \approx 1/(2N) \exp \{-t/(2N)\}$ となることも理解できよう。

子が親をでたらめにしかも一様を選ぶことができるには、自然選択の他に同類婚や近親婚があってはいけない。自然選択があると特定の遺伝子の分岐がよく起こる。同類婚や近親婚は明らかに $2N$ 個の中から選ぶ親を特定のグループに限定している。この他に、集団が単一でなく部分的な地理的隔離があるとやはり子が親を選ぶ上で制限が生じる。

ある生物種が、複数の集団から構成され移住によって結びつけられているときの遺伝子の系図はどのようになるのであろう。異なる集団からサンプルした2つの遺伝子の合体が起きるには、まず移住によって両方の遺伝子の系統が同一集団に属するのを待たねばならない。しかし同一集団に属したからといって合体が直ちに起きるわけではないから、再び2つの遺伝子の系統は移住によって離ればなれになるかも知れない。単一集団と異なって、集団構造がある場合の遺伝子の系図研究はこのような新しい問題を含んでいる。このときは移住率だけでなく集団の空間的配置や遺伝子のサンプリングの仕方も系図に大きな影響を与える¹⁷⁾。

移住があると、系図の解析から集団がいつ分岐したかという問いにははっきり答えられなくなる。図2は図1の1部を拡大したものである。ニューギニアには今から3~5万年前にアジアから移住が起こったとされている。ニューギニアの3つの型のミトコンドリアDNAはこの系図で1つのクラスターを作っているが、ニューギニアからアジアへ逆移住した可能性があるとしてニューギニアにヒトが入ったのは図2のA以前と考えても矛盾しない。もしこのような逆移住がなければ、ニューギニアにヒトが入ったのはA以後である。しかしその時点がB以前でもあると主張する⁷⁾ことはできない。複数の異なるミトコンドリアDNAの系統が

図2. 遺伝子の系図と人種の系統関係。多型の遺伝子が古くから存在する可能性や移住の可能性によって両者の系図は必ずしも一致しない。Stoneking et al.⁷⁾を参考に作成。

もともと存在していたと考えることができるからである。この“古い”多型の存在は移住が無視できる程離れた種間からサンプルした相同遺伝子の系図の場合にも問題になる。

種分化前の祖先集団

完全に生殖的隔離が完成した2つの種から相同遺伝子をサンプルした場合の系図は、種の分岐をある程度反映したものになる。しかし遺伝子の合体が種の分岐時と一致する保証はない。このことは集団構造がある場合と同様で、合体が1つの集団内で起きるまで時間がかかるからである。例として図3にあるように、2種から1つずつ相同遺伝子をサンプルしたとしよう。この遺伝子の祖先は種が分化した時点まで独立にたどることができ、その間では合体は決して起こらない。しかしこの種分化の時点を超えると2つの遺伝子は単一の祖先集団のメンバーであり、ここで初めて共通の

図3. 2種からサンプルした相同遺伝子の合体。

図4. *C. simulans* とその近縁種 *D. mauritiana* のミトコンドリア DNA の系図. Satta et al.¹⁰⁾ のデータを参考に作成.

祖先遺伝子から由来した可能性が生ずる。祖先集団での2遺伝子の合合法則についてはすでに述べたとおりであり、この合体までの平均待ち時間は $2N$ 世代となる。もし祖先集団が大きいと祖先集団における合体までの待ち時間が長くなり、種分化後経た世代時間に比べて、無視できなくなる。

図4はその例である。この図は、*D. simulans* と *D. mauritiana* というキイロショウジョウバエに近縁な2種のミトコンドリア DNA を制限酵素で解析した結果¹⁰⁾に基づいて描いたものである。普通、種はそれ固有の遺伝子をもっていると考えがちであるが、近縁種間を比較するとむしろ逆に類似の遺伝子を共有していることが多い。今の場合、*D. simulans* の集団をしらべると2種類の型のミトコンドリア DNA (S型とM型としよう)が多型的に存在するが、その1つは*D. mauritiana*の型(M型)と極めて高い相同性を示すのである。図4は、2種が分化する以前にすでにS型とM型が祖先集団に存在しており、種分化の時点で*D. mauritiana*はS型を偶然にもたない個体ばかりから形成される場合を示している。もちろん他の説明も可能であるが、塩基配列やM型の地理的分布の情報は“古い”多型の存在を示唆している。この点を更にはっきりするため最近この2種が棲息する島の調査¹¹⁾を行った。その解析は現在進行中だが、おそらく古い多型の存在を支持するものと思われる。

遺伝子はその分岐以後独立に突然変異を蓄積する。2節で述べたようにこの突然変異で生じた配列上の差

異が遺伝子の系図作成を可能にする情報であった。ある時間間隔に、1つの遺伝子あたり何個の塩基に突然変異が生じるかに関する直接の観察結果はほとんどないが、普通、その個数はポアソン分布に従うと仮定する。この法則では、分岐後 T 時間たった2つの遺伝子間の塩基が k 個違う確率は、1つの遺伝子あたりの突然変異率を ν とすると、 $P_k = (2\nu T)^k \exp(-2\nu T) / (k!)$ となる。問題は T が種分化後へた時間に加えて祖先集団で遺伝子が合体するまでの時間を含んでいることである。祖先集団が大きいと合体までの時間も長く、その間に多くの突然変異を蓄積することが可能である。従って塩基配列の比較から再構成される遺伝子の系図と種の分岐関係は一致しないのである。こうした理論的考察と数多くの遺伝子の塩基配列のデータから、種分化が起こる以前の集団の大きさを推定することも可能だと思われる³⁰⁾。それにはまだ解決しなければならない問題が残されている²¹⁾が、こうした大胆な試みも進化学という歴史性の強い学問分野には必要なことではないだろうか。

3種以上の種からそれぞれ1つずつ相同遺伝子をサンプルしその系図を作成すると、その型が種分化の順序と一致しないことが起こる。その理由もこれまでに述べたことから明らかだと思うが、図5には3種の場合における3通りの遺伝子の系図とそれが起こる確率が示してある³¹⁾。第一番目と第二番目の種分化の時間が近接していると、遺伝子の系図は全くの偶然以上に種分化の順序を反映することができなくなる。このとき遺伝子の系図は種の系統関係を明らかにする情報を全くもっていない。

遺伝子の系図に関連した未解決の問題はまだまだ多い。系図作成法にしても遺伝子の合体に起因する確率

図5. 3種からサンプルした相同遺伝子の可能な系図とそれぞれの起こりうる確率。Nは種XとYの祖先集団の大きさ、tはこの祖先集団の存続時間。Nei³¹⁾を参考に作成。

的要素を取り入れる必要がある。また遺伝子の合体法則を、一般的な集団構造や近親婚の存在で定式化することも近縁種間の進化を考える上で意義がある。とりわけ自然選択の効果を系図上にどう表現するかは最も興味深い。そうした理論とDNA配列の情報から、分子進化のメカニズムを再吟味してみることも遺伝子の系図学における今後の重要な問題である。

文 献

- 1) Hudson, R. R. (1983) *Theor. Pop. Biol.*, 23, 203-217.
- 2) Sauners, I. W., Tavaré, S., Watterson, G. A. (1984) *J. Appl. Prob.*, 16, 471-491.
- 3) Felsenstein, J. (1982) *Quart. Rev. Biol.*, 57, 379-404.
- 4) Nei, M. (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- 5) Saito, N. (1986) *Theoretical studies on the methods of reconstructing phylogenetic trees from DNA sequence data*. Ph. D. Thesis, University of Texas.
- 6) Tatenio, Y., Nei, M., Tajima, F. (1982) *J. Mol. Evol.*, 18, 387-404.
- 7) Stoneking, M., Bhatia, K., Wilson, A. C. (1986) *Cold Spring Harbor Symp. Quant. Biol.*, 51, 433-439.
- 8) Akaike, H. (1973) *2nd Inter. Symp. on Information*

- Theory*. (Petrov, B. N., Csaki, F., eds.) 716-723, Adademiai Kiado, Budapest.
- 9) Hasegawa, M., Yano, T. (1984) *Bull. Biomet. Soc. Japan*, No. 5, 1-7.
- 10) Wright, S. (1981) *Genetics*, 16, 97-159.
- 11) Moran, P. A. P. (1958) *Proc. Cambridge Philos. Soc.*, 54, 60-71.
- 12) Kimura, M. (1968) *Nature*, 217, 624-626.
- 13) Felsenstein, J. (1971) *Theor. Pop. Biol.*, 2, 391-403.
- 14) Griffiths, R. C. (1980) *Theor. Pop. Biol.*, 17, 37-50.
- 15) Kingman, J. F. C. (1982) *Stochastic. Process. Appl.*, 13, 235-248.
- 16) Tajima, F. (1983) *Genetics*, 105, 437-460.
- 17) Takahata, N. (1987) *Genet. Res., Camb.* submitted.
- 18) Satta, Y., Toyohara, N., Ohtaka, C., Tatsuno, Y., Watanabe, T.-K., Matsuura, E.-T., Chigusa, S.-I., Takahata, N. (1987) *Genet. Res., Camb.* In press.
- 19) 石和貞男, 深民玲之, 布山喜章, 高畑尚之 (1987年7月8日~8月1日) 文部省海外学術調査「異種間溶透と多型細胞質性を示す南インド地域ショウジョウバエ群の学術調査」
- 20) Takahata, N. (1986) *Genet. Res., Camb.*, 48, 187-190.
- 21) 高畑尚之: 第58回日本遺伝学会, 第9回日本分子生物学会合同年会シンポジウム (名古屋, 1987年12月)

たかはた なおゆき

国立遺伝学研究所集団遺伝部門, 助教授 (理博). 70年の大学紛争から3ヶ年京大理学部生物物理教室の寺本研究室に他の多くの仲間と居候したのち, 九大理学部生物教室の松田研究室で博士課程を過す。これが契機となつてはからずも集団遺伝学の道を歩むことになった。集団遺伝学は現在1つの節目にあると思っている。将来の方向を模索中。趣味は30才の時に始めたテニスと西伊豆での海づり。

☎0559-75-0771 内線545