

公開講演会要旨

DNA からみた生物の進化

統計数理研究所 長谷川 政 美

(1991 年 11 月 6 日, 統計数理研究所 講堂)

現在, 地球上には大腸菌などの細菌 (バクテリア) からゾウリムシ, ミドリムシなどの原生生物, コウボ, きこのなどの菌類, それにわれわれが日常いつも目にする植物と動物を含めて実に多種多様な生物が生息しており, 分類学上記載されているものだけでも約 300 万種, まだ記載されていないものも含めると 3000 万種を超えるという推定もある. 分子遺伝学の発展に伴って, このような多様性にもかかわらず, これらの生物はもとをたざせばたった一つの共通の祖先から由来して進化してきたものであることが明らかになってきた. なぜならば, バクテリアからヒトに至るまで, 少なくともこれまでに調べられた範囲では, 現在の地球上の生物は基本的な遺伝の仕組みが共通であるからである. C. Darwin が 1859 年有名な「種の起原」を出版して以来, 生物進化の考えは急速に広まり, 1866 年には E. Haeckel はバクテリアからヒトに至る生物を一本の系統樹にまとめてあらわしている. あらゆる生物が共通の祖先に由来して進化してきたものであるという Haeckel の直感は, 結果的には確かに正しかったが, そのことが確かめられたのは最近になって分子遺伝学が進歩してからのことである. このように, 地球上に現在生きている多種多様な生物は, 一つの壮大な系統樹で結ばれているはずであり, そのような系統樹を描き出すことが生物系統学の究極の目標である.

従来の生物系統学は, 主に比較形態学に基礎をおいたものであったが, 動物と植物を比較するにはどのような形態的な特徴を用いたらよいかははっきりしない. さらに, 動物とバクテリアとを比較しようと思ったら, 形態学からはほとんどなんの手掛かりも得られない. ところが, すべての生物のもつ DNA あるいはその情報にしたがってつくられるタンパク質を比較することによって, ヒトとバクテリアを含めた系統樹を推定することができるのである. このように DNA やタンパク質などのデータから系統樹を推定する分野は分子系統学と呼ばれ, 近年になって従来の形態をもとにした系統学を補うかたちで発展してきた. 形態学では判断が主観的になりがちであったが, 分子系統学ではヒトからバクテリアに至るまで共通の客観的な基準で系統関係を論ずることが可能である.

DNA あるいはタンパク質の配列データから分子系統樹を推定するための方法には実にさまざまなものがあるが (例えば, Felsenstein (1988), 根井 (1990)), 従来の方法の多くは ad hoc なものであり, 必ずしも統計学的に基礎づけられているとはいえない.

一般に分子系統学で用いられる DNA あるいはタンパク質の配列データは, 現存生物のものであり (最近では ancient DNA の解析も可能になりつつあるが, ここではふれない), これらは長い進化の結果としてできあがったものであり, そこには生物進化の歴史が刻まれているはずである. 一般には DNA の塩基置換やタンパク質のアミノ酸置換を手掛かりにして分子系統樹を推定するわけであるが, 塩基置換やアミノ酸置換にはランダムな要素がつきものである. そもそも突然変異 (ここでは塩基置換やアミノ酸置換を引き起こす点突然変異だけに限る) はランダムに起こっている. こうして起こったランダムな変異が進化的に意味のあるものになるた

めには、変異遺伝子が生殖を通じて集団全体に広がって固定することが必要であるが、そのような過程においても遺伝的浮動と呼ばれるようにランダムな要素が重要であることが知られている(分子進化の中立説, 木村(1986)). つまり, 分子系統学の手掛かりとなる進化における塩基置換あるいはアミノ酸置換は, ランダムな確率過程としてとらえることができる. したがって, そのような進化過程の結果として得られているDNAあるいはタンパク質の配列データを解析する際には, その点をはっきりと認識しておくことが必要である.

最尤法は, 統計学において広く用いられ, 確率論的に基礎づけられた方法であり, 素過程が確率過程であることを明示的に取り入れている. 置換過程がランダムな変動を伴うものである以上は, そのような過程の結果として得られているデータから推定される分子系統樹が100% 確実なものであることはありえない. したがって得られた系統樹の確からしさ(信頼区間)は, 確率的に表現されることが望ましい. 最尤法ではそのことが容易に可能である. 得られた結果の信頼性を評価せずに, 得られた系統樹を安易に発表することはやめなければならない. そのために一見互いに矛盾した系統樹が異なる論文で発表されるということになり, 多くの混乱を引き起こしている(例えば, Turner et al. (1989), Morden and Golden (1989)). 信頼区間を適切に評価しさえすれば, そのような矛盾はみかけだけのものであり, 単に誤差の範囲内で違っているに過ぎないことがわかるであろう(Kishino et al. (1990)).

塩基配列データから最尤法によって分子系統樹を推定するための実用的な方法は, Felsenstein (1981) によってはじめて提案された. 塩基置換の確率モデルと系統樹のトポロジーを与えると, いろいろな生物種のもつDNA の実際の塩基配列データが進化の結果として実現する確率が表現できる(モデルに含まれるいろいろなパラメータの関数として). これを尤度(likelihood)という. このような尤度が最大になるようにパラメータが推定される(最尤推定). そしてなるべく尤度が大きくなるような系統樹のトポロジーが選択される.

この方法は, 進化速度が系統によって異なる場合でも, 真の系統樹をよく再現することが, シミュレーションで確かめられている(Hasegawa et al. (1991a)). 最尤法では仮定がはっきりしているので, 素過程に関する知見が増えて仮定が現実と合わないことがわかってくればそれに応じてモデルを改良していくことができる. その際モデルを複雑にすればするほど, データとの適合度は増すから, 尤度は大きくなる. しかしながら, 限られた量のデータを説明するためのモデルとして, むやみに複雑なものを採用するのはよくない. モデルのパラメータ数を増やすことには, なんらかのペナルティを課することが必要である.

AICはこのことを情報理論的な考察から最尤法の自然な拡張として取り入れたものであり, $AIC = -2 \times \text{対数尤度} + 2 \times \text{パラメータ数}$ と定義される. モデルが複雑でパラメータ数が多ければ, 第2項が大きくなるが, データとの適合度がよくなって尤度が大きくなるので, マイナス符号のついた第1項は小さくなる. AICが最小になるモデルが真のモデルの候補として一番ふさわしい(赤池(1976)). したがってわれわれは, どのようなモデルがデータを説明する上で最もよいかを判断するための客観的な基準をもっているといえる.

このような基準をもとにして, 今後ともわれわれはモデルを改善することに努めなければならないが, それと同時にいかにモデルを改善したとしても必ず現実とのズレが残るということも認識しておかなければならない. したがって, モデルが現実からずれていた場合に, そのことが系統樹のトポロジーの推定に対してどの程度敏感に効いてくるかを調べておくことが必要である. 例えば, 座位ごとの進化速度に差がある場合(現実の遺伝子ではたいていそうになっている), シミュレーションによれば, そのような不均一性を考慮しないモデルでも, 最尤法では真の系統樹をかなり効率よく推定できるのに対して, 従来から分子系統学でよく用いられている方法ではうまくゆかないことがわかってきた(長谷川(未発表)). この意味で最尤法による系

統樹のトポロジー推定は頑健であるといえる。

先に、得られた系統樹の確からしさを確率的に表現することが大切であることを強調した。そのためには、ブートストラップ法が便利である (Felsenstein (1985), Hasegawa and Kishino (1989))。ブートストラップ法は配列データをもとにした系統樹推定法一般に適用できるが、最尤法の場合には簡便法が開発されている (Kishino and Hasegawa (1989), Kishino et al. (1990))。

近縁な生物種間 (科あるいは目の中) の系統を推定するには、進化速度の速いミトコンドリア DNA や偽遺伝子などの塩基配列データが役に立つが、綱、門あるいは界の間の系統関係を論ずるには進化速度の遅いリボソーム RNA の塩基配列や保存的なタンパク質のアミノ酸配列が役に立つ。しかしながら、リボソーム RNA の塩基組成は生物種によってその偏り方が異なり、近縁な種同士でも塩基組成が大きく異なることがある。そのために、リボソーム RNA の塩基配列から推定された系統樹は、時々間違えることがある (Loomis and Smith (1990))。もちろんゲノムの塩基組成の偏りは、タンパク質のアミノ酸組成にも影響を与えるであろうが、その影響は間接的なものであり、タンパク質系統樹の方が信頼できる。

アミノ酸配列データからタンパク質系統樹を推定するための最尤法は、Kishino et al. (1990) によって開発された。これは、Dayhoff et al. (1978) によって調べられた経験的なアミノ酸置換行列にしたがったマルコフ・モデルにもとづいている。この方法により、古細菌の進化的な位置が明らかになった (Hasegawa et al. (1990), Miyata et al. (1991))。地球上のあらゆる生物は一つの共通の祖先に由来して進化してきたものであることは既に述べたが、これらの多種多様な生物は大きく二つのグループに分類できることが知られていた。動物、植物、菌類、原生動物など細胞に核をもつ生物、つまり真核生物と、細菌など核をもたない生物、つまり原核生物の二つである。ところが、1977 年になって、イリノイ大学の C. Woese らは、一見して細菌類と思われる生物のなかで、この二分法のどちらにも属さない一群の生物を発見した。それらの生物は、核をもたないという点では原核生物であるが、生化学的な面や、リボソーム RNA の比較からは、従来から知られていた細菌類とは異なっており、さらに真核生物とも違っていた (Woese (1981))。

Woese らが見つけた第三の生物は、高温、高塩、あるいは強酸などといった異常環境に生息する細菌であり、二酸化炭素をメタンに還元するこの細菌の代謝は、原始地球上に存在したと思われる型の大気、つまり二酸化炭素に富み、酸素が少ないという大気に適したものである。このため、これらの生物は古細菌と名付けられ、従来から知られていた細菌はこれに対して真正細菌と呼ばれるようになった。リボソーム RNA の塩基配列からは、古細菌、真正細菌、そして真核生物は、互いに進化的に等距離にあるようにみえたので、Woese らは、地球上の生物は、二分法ではなく、これら三つの生物界に分類されるべきであると主張した。

ところが、実際の進化においては、これら三つの生物界が同時に分かれたというよりは、ある特定の順番で分かれたと考える方が自然であろう。つまり、古細菌と真正細菌とが系統的に近い関係にあるか、真正細菌が真核生物と近い関係にあるか、あるいは古細菌が真核生物と近い関係にあるかのいずれかであろう。このことは、地球上のすべての生物の共通の祖先がどのようなものであったかを探る上で最も大切な点であるし、われわれヒトを含む真核生物の起原を知る上でも基本的に重要な問題である。

一般に系統樹を推定する際には、進化速度の一定性を仮定しないと系統樹の根元を決めることができないが、これら三つの大きくかけ離れた生物界の間で進化速度が一定であるという保証はない。そのために分子系統学では、系統関係を知りたい生物群に対して、あらかじめそれらよりも遠い関係にあることがわかっている生物をアウトグループとし、前者の生物群とアウ

トグループとの間に系統樹の根元があると考えられるわけである。ところが、古細菌、真正細菌、真核生物といった地球生物の最も大きな分類単位の間系統関係を知らうとした場合、アウトグループに相当する生物が存在しないのである。そのために、古細菌の発見以来10年以上にわたって、さまざまな論争があったものの、これら三者の間系統関係に関しては決着がつかなかった。

1989年になってようやく、九州大学(現在京都大学)の宮田隆と筆者らの研究グループによって明確な答が与えられた。われわれの基本的なアイデアは、次のようなものである。三つの生物界が分かれる以前に遺伝子重複を起こし、すべての生物に重複して存在する遺伝子を使って複合系統樹(重複遺伝子の両方を含む系統樹)をつくれれば、あらゆる生物を含む系統樹の根元を決めることができるであろう。例えば、ATPアーゼには、明らかに遺伝子重複によって分化し、互いに相同な配列をもつ二つのサブユニット、つまり触媒サブユニットと非触媒サブユニット、がある。すべての生物はこの二つのサブユニットをもつから、遺伝子重複が起きたのは三つの生物界が分かれる以前であると仮定することはきわめて合理的である。ATPアーゼの非触媒サブユニットを古細菌、真正細菌、真核生物の触媒サブユニットの間系統関係を定めるためのアウトグループとして用い、逆に触媒サブユニットを非触媒サブユニットのアウトグループとして用いることができるわけである。われわれは、このような重複遺伝子のいくつかを用いて解析した結果、古細菌と真核生物が明らかに近縁であることを示した(Iwabe et al. (1989), Hasegawa et al. (1990), Miyata et al. (1991))。

このようにして、いわゆる古細菌が真核生物に近いことが明らかになった。このことは、真核生物が古細菌に近いものから進化してきたものであることを示している。ただし、ここでいう真核生物とは、厳密には真核生物の核のことである。実は真核生物の細胞は、由来を異にするいくつかの遺伝的な要素のモザイクなのである。普通の真核生物の細胞内にはミトコンドリアと呼ばれる膜につつまれた小器官があり、植物にはそのほかに葉緑体がある。分子系統学的な研究から、ミトコンドリアや葉緑体は、真核生物の祖先細胞内にある種の真正細菌が取り込まれ、共生関係が結ばれて進化してきたものであることが明らかになってきた。ところが最近、真核生物であるにもかかわらず、ミトコンドリアなどの細胞内小器官をもたない生物の存在が明らかになり、これらを総称してアーケゾアと呼んでいる。アーケゾアがミトコンドリアをもたないのは、この生物がミトコンドリアが共生する以前の真核生物の古い系統を代表しているからなのか、あるいはもともとはミトコンドリアをもっていたものが進化の途中でそれを失ったことによるのかという問題がある。このことは、真核生物の起原を探る上で重要な問題である。

アメーバ性赤痢の病原体であるエントアメーバ *Entamoeba* が実は真核生物であるにもかかわらずミトコンドリアをもたないアーケゾアの一員である。われわれは、伸長因子のアミノ酸配列データを解析した結果、エントアメーバが真核生物の進化の初期にほかの系統から分かれたものであることを明らかにし、この生物がミトコンドリアが共生する以前の真核生物の系統を代表するものであることを示した(Hasegawa et al. (1991b))。さらに、動物界、菌界、植物界など間系統関係に関しては、以前はアウトグループとして細菌類という非常に遠い関係にある生物しか使えなかったためにはっきりしなかったものが、エントアメーバがアウトグループとして使えるようになり、動物界と菌界とが近縁な関係にあることが明らかになった(Hasegawa et al. (1991b))。

このようにして、地球上の生命がたどってきた進化の歴史に関して、DNAやタンパク質の配列データを統計的に解析することによって、さまざまな新しい知見が得られているのである(長谷川(1989, 1991))。

参 考 文 献

- 赤池弘次 (1976). 情報量基準 AIC とは何か, 数理科学, **153**, 5-11.
- Dayhoff, M.O., Schwartz, R.M. and Orcutt, B.C. (1978). A model of evolutionary change in proteins, *Atlas of Protein Sequence and Structure*, Vol. 5, Supplement 3 (ed. M.O. Dayhoff), 345-352, National Biomedical Research Foundation, Washington, D.C.
- Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach, *Journal of Molecular Evolution*, **17**, 368-376.
- Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap, *Evolution*, **39**, 783-791.
- Felsenstein, J. (1988). Phylogenies from molecular sequences: inference and reliability, *Annual Review of Genetics*, **22**, 521-565.
- 長谷川政美 (1989). 『DNA からみた人類の起原と進化——分子人類学序説(増補版)』, 海鳴社, 東京.
- 長谷川政美 (1991). 『DNA に刻まれたヒトの歴史』, 岩波書店, 東京.
- Hasegawa, M. and Kishino, H. (1989). Confidence limits on the maximum-likelihood estimate of the hominoid tree from mitochondrial-DNA sequences, *Evolution*, **43**, 672-677.
- Hasegawa, M., Iwabe, N., Mukohata, Y. and Miyata, T. (1990). Close evolutionary relatedness of archaeobacteria, Methanococcus and Halobacterium, to eukaryotes demonstrated by composite phylogenetic trees of elongation factors *EF-Tu* and *EF-G*: eocyte tree is unlikely, *Japanese Journal of Genetics*, **65**, 109-114.
- Hasegawa, M., Kishino, H. and Saitou, N. (1991a). On the maximum likelihood method in molecular phylogenetics, *Journal of Molecular Evolution*, **32**, 443-445.
- Hasegawa, M., Hashimoto, T., Adachi, J., Iwabe, N. and Miyata, T. (1991b). Early divergences in the evolution of eukaryotes: ancient divergence of *Entamoeba* that lacks mitochondria revealed by protein sequence data, *Journal of Molecular Evolution* (submitted).
- Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S. and Miyata, T. (1989). Evolutionary relationship of archaeobacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes, *Proc. Nat. Acad. Sci. U.S.A.*, **86**, 9355-9359.
- 木村資生 (1986). 『分子進化の中立説』, 紀伊國屋書店, 東京.
- Kishino, H. and Hasegawa, M. (1989). Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea, *Journal of Molecular Evolution*, **29**, 170-179.
- Kishino, H., Miyata, T. and Hasegawa, M. (1990). Maximum likelihood inference of protein phylogeny and the origin of chloroplasts, *Journal of Molecular Evolution*, **30**, 151-160.
- Loomis, W.F. and Smith, D.W. (1990). Molecular phylogeny of *Dictyostelium discoideum* by protein sequence comparison, *Proc. Nat. Acad. Sci. U.S.A.*, **87**, 9093-9097.
- Miyata, T., Iwabe, N., Kuma, K., Kawanishi, Y., Hasegawa, M., Kishino, H., Mukohata, Y., Ihara, K. and Osawa, S. (1991). Evolution of archaeobacteria: phylogenetic relationships among archaeobacteria, eubacteria, and eukaryotes, *Evolution of Life: Fossils, Molecules, and Culture* (eds. S. Osawa and T. Honjo), 337-351, Springer, Tokyo.
- Morden, C.W. and Golden, S.S. (1989). *psbA* genes indicate common ancestry of prochlorophytes and chloroplasts, *Nature*, **337**, 382-385.
- 根井正利 (1990). 『分子進化遺伝学』(五條堀孝・斎藤成也 訳), 培風館, 東京.
- Turner, S., Burger-Wiersma, T., Giovannoni, S.J., Mur, L.R. and Pace, N.R. (1989). The relationship of a prochlorophyte *Prochlorothrix hollandica* to green chloroplasts, *Nature*, **337**, 380-382.
- Woese, C.R. (1981). 第三の生物“古細菌”, サイエンス, **11**(8), 84-98.