

塩基配列の挿入・欠失による進化と突然変異のメカニズム

斎藤成也

霊長類の塩基配列データを用いて、欠失と挿入の進化を分析した。与えられた系統樹上で、最大節約原理を用いて欠失と挿入の生じた位置を推定するという方法を用いた。核DNAでは、欠失が挿入よりも約2倍高い頻度で生じており、1塩基の欠失・挿入の頻度が最も高かった。欠失・挿入の進化速度はかなり一定であり、核DNAでは1 kbあたり100万年あたり約0.15で、ミトコンドリアDNAはその約10倍速かった。塩基置換速度と比べると、欠失・挿入の進化速度は、核とミトコンドリア双方とも1/10以下であった。欠失と挿入の突然変異を生じるメカニズムについても論じた。

Key words 【欠失・挿入】【突然変異率】【霊長類】【進化速度】

はじめに 突然変異は生物進化における根本である。ここでいう突然変異とは、塩基配列におけるあらゆる種類の変化を指す。したがって、塩基の置換、欠失・挿入、逆転、不等交差、遺伝子変換、遺伝子重複、ゲノム重複などがすべて含まれる。生物進化を真に理解するには、塩基配列レベルという根底で生じている現象を探らなければならない。そのひとつとして、これらさまざまな種類の突然変異の生じる率やその進化速度を知ることは重要である。

進化の過程では、連続的な塩基配列に塩基の欠失や挿入が入ることがある。そのために、複数の塩基配列を整列（アラインメント）すると、ギャップの生じることがある。塩基置換がこれまできわめて詳しく調べられてきたのに比べると、これら塩基の欠失や挿入はそれほど研究されてこなかった。筆者らは、霊長類の核DNAとミトコンドリアDNAの非コード領域の塩基配列データを解析して、欠失と挿入の進化速度を、与えられた系統樹上で最大節約原理を用いて欠失と挿入の生じた位置を特定するという方法を用いて、推定した。

分子進化学においては、近年急速に量が増えてきた塩基配列データを用いて、さまざまな解析が行なわれているが、その中心は塩基置換の分析である。その結果、現在では多数の生物、遺伝子について、塩基置換速度が明らかにされている（たとえば文献1を参照）。一方、蛋白質コード領域では、塩基の欠失・挿入の大多数がフレームシフト突然変異としてその蛋白質のアミノ酸配列を大きく変化させるために、このような突然変異はほとんどの場合生存に不利である。このため、多数の研究が集中する蛋白質コード領域では、塩基の欠失と挿入があまりみられない。TajimaとNei²⁾は塩基の欠失と挿入に基づく進化距離を提案し、それを蛋白質コード領域の塩基配列データに応用したが、長い進化時間にわたって、欠失と挿入に基づく進化距離はあまり増大しなかった。一方、イントロンや5'、3'の非コード領域では、比較された塩基配列は短いものの、コード領域よりも進化距離が著しく大きかった。

Graurら³⁾は、プロセスト型偽遺伝子の塩基配列をヒトとマウスで比較し、マウスの進化系統ではヒトの進化系統よりも塩基の欠失速度が速いと結論した。しか

Naruya Saitou, 国立遺伝学研究所進化遺伝研究部門 (〒411 三島市谷田 1111) [Laboratory of Evolutionary Genetics, National Institute of Genetics, Yata, Mishima 411, Japan]

Evolution of Insertions and Deletions of Nucleotide Sequences and their Mutation Mechanisms

し、彼らの研究は相対的な進化速度の比較にとどまっていた。

Saitou⁴⁾は、ヒト上科4種(ヒト, チンパンジー, ゴリラ, オランウータン)の核DNA非コード領域塩基配列を解析して、塩基の欠失と挿入の絶対進化速度を推定した。最近になってSaitouとUeda⁵⁾は、Saitou⁴⁾の研究を発展させ、霊長類において核DNAとミトコンドリアDNAの非コード領域塩基配列を分析し、欠失と挿入の絶対進化速度にある程度の一定性のあることを示した。本稿では、この研究で得られた成果の紹介を中心に、塩基の欠失・挿入の進化全体についても議論する。

なお、ゲノム内の非コード塩基配列に生じる変化の大部分は淘汰上中立だと考えられるので⁶⁾、これらの領域の塩基配列を分析して得られた進化速度の推定値は、欠失と挿入の自発突然変異率とみなすことができる。

1. 解析に用いた塩基配列データ

筆者らは、以下の6領域の塩基配列データを用いた。そのうち4種類は核DNAであり、2種類はミトコンドリアDNAである。各領域の長さは、比較に用いた塩基配列の平均である。また、これらはすべて塩基配列が整列されており、それらをそのまま用いた。① η グロビン Ψ : β グロビン遺伝子族内の η グロビン偽遺伝子6.4 kb⁷⁾。② β グロビン: β グロビン遺伝子族内の β グロビンスペーサー領域3.1 kb⁸⁾。③Ig- ϵ 3 Ψ :免疫グロブリン ϵ 3プロセスト偽遺伝子2.3 kb⁹⁾。④Ig- α :免疫グロブリン α 遺伝子の非コード領域1.5 kb¹⁰⁾。⑤ミトコンドリアDNA Dループ:ミトコンドリアDNAのいわゆるDループ領域1.1 kb¹¹⁾。⑥ミトコンドリアDNA 1/3ゲノム:ミトコンドリアDNAゲノムの約1/3をおおむね塩基配列中の非コード領域0.1 kb¹²⁾。

II. 解析方法

Saitou⁴⁾は以下のような塩基の挿入・欠失数を推定する新しい方法を考案した。本研究でも同じ方法を用いた。はじめに、霊長類の系統樹をあらかじめ仮定しておく(図1)。こ

の系統樹は、多数の分子データおよび化石データを総合して作成したものである。細かい点においてはまだ議論があるところだが、大きな変化は将来もありえないと考えられる。この系統樹をもとにして、最大節約原理(文献13を参照)を用いて欠失と挿入が系統樹のどの枝で生じたかを推定する。最大節約原理は、系統樹作成でよく用いられる最大節約法で使われるものであり、与えられた系統樹でデータを解釈する場合の最小必要変化量を計算する。通常は、これらの計算を可能なすべての系統樹の樹形について行ない、最小の変化量を与える系統樹を選ぶのであるが、ここではあらかじめ系統樹の樹形が決まっているので、それのみについての検討を行なうのである。この方法で重要な点は、連続した塩基のギャップが常にひとつの欠失または挿入とみなされることである。たとえば、Ig- α 配列の607-621 bp領域にはゴリラとオランウータンの配列で3塩基のギャップが存在するが(図2b)、これらはどちらも1回の突然変異で生じたと仮定するのである。

最大節約原理を用いた本方法を具体例を用いて説明しよう。Ig- α 配列の561-572 bp領域には2個のギャップが存在する(図2a)。この場合、これらのギャップは、それぞれオランウータンにいく枝(1塩基の欠失)とカニクイザルにいく枝にマップされる(図3a)。ここで注意すべきことは、図3(a)の系統樹の根(共通祖先の位置)が実際には無視されていることである。これは、最大節約原理を用いているためである。したがって、図2(a)のカニクイザルの塩基配列にはギャップが存在するが、これが挿入(図3aの枝aにある場合)か欠失(図3aの枝bにある場合)かは判別できない。そのため、図3aには1G(1塩基のギャップ)と示してある。以下では、塩基の欠失か挿入かを区別できな

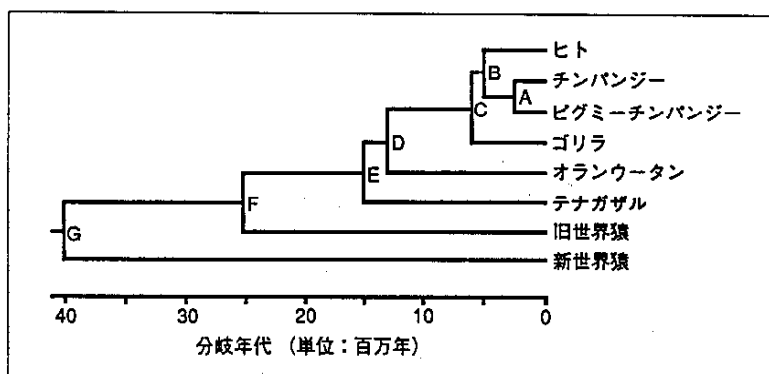


図1 仮定された霊長類の系統樹⁴⁾
分岐点A~Gは順に2.5, 5.0, 6.0, 13.0, 15.0, 25.0, 40.0百万年に対応する

い現象を“ギャップ”とよぶことにする。

あるギャップを系統樹の枝に一意的にマップすることができず、複数の可能性があるときは、すべての可能性を考慮し、それらに等確率でギャップを割り当てる。Ig- α 配列の 607-621 bp 領域には 2 個のギャップが存在するが (図 2b), この場合、欠失と挿入に 2 種類の可能性がある。ひとつは、ヒトとチンパンジーの共

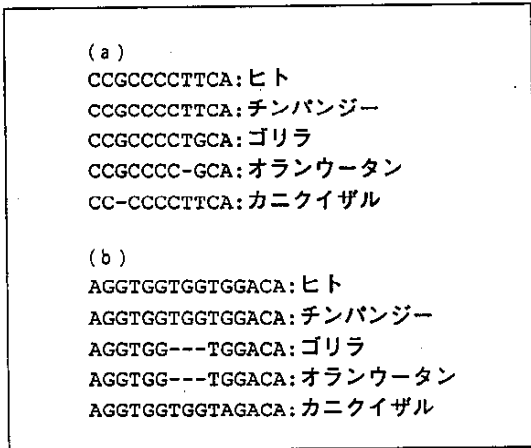


図 2 整列された塩基配列におけるギャップの例⁶⁾
 (a) Ig- α 遺伝子の 561~571 bp 領域。(b) Ig- α 遺伝子の 607~621 bp 領域。

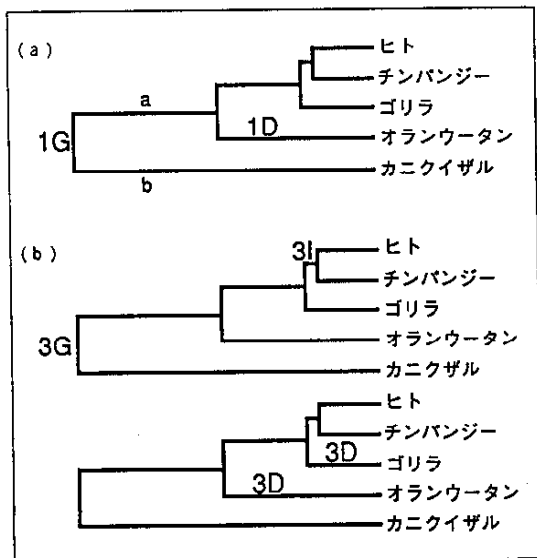


図 3 最大節約原理を用いて図 1 の系統樹における欠失と挿入の位置を推定したもの⁹⁾
 I, D, G は、順に挿入、欠失、ギャップ (挿入または欠失) を示す。これらの記号の前の数字は、それらの長さ (塩基対) である。(a) 図 2(a) のデータの場合。(b) 図 2(b) のデータの場合に考えられる 2 種類の可能性。

通祖先へいく枝で 3 塩基の挿入があり、枝 a または b で 3 塩基のギャップが生じた可能性である (図 3b 上)。もうひとつの可能性は、ゴリラとオランウータンの系統がどちらも 3 塩基の欠失を生じた場合である (図 3b 下)。この場合、これら 2 種類の可能性おのにおに確率 1/2 が与えられる。

最大節約原理では最小変化量しか問題にしないので、実際に生じた変化量を過小推定する危険性が常に存在するが、今回は霊長類という近縁な生物間の塩基配列を比較しているのので、この問題点はかなりの程度まで回避されていると筆者らは考えている (この問題については、後述する)。

III. 欠失と挿入の長さの分布

表 1 は η グロビン Ψ 塩基配列データにおける挿入・欠失の長さの統計を示したものである。数字は、確定的に枝にマップされた各長さの挿入 (I), 欠失 (D), ギャップ (G) の事象であり、かっこ内の数字は、複数の可能性があった場合である。今回用いた 6 種類の

表 1 η グロビン偽遺伝子領域における挿入 (I), 欠失 (D), ギャップ (G) の長さの統計⁹⁾

長さ	I	D	G	合計
1	14(3.83)	29(4.33)	19(0.50)	62(8.67)
2	5(1.67)	11	5	21(1.67)
3	3(0.33)	4	2	9(0.33)
4	3(0.33)	7(0.33)	3	13(0.67)
5	1(0.33)	3(0.33)	2	6(0.67)
6	2	3	2	7
7	—	1	—	1
8	1	1	—	2
11	2	—	1	3
12	1	—	1	2
13	2	1	—	3
14	1	1	—	2
18	—	1	—	1
23	—	1	—	1
24	—	3	—	3
26	—	1	—	1
28	—	1	—	1
38	1	—	—	1
76	—(0.50)	—(0.50)	—	—(1.00)
331	1	—	—	1*
338	—	—	1	1*
339	—	—	1	1*
558	—	—	1	1**
合計	37(7.00)	68(5.50)	38(0.50)	143(13.00)

* A/Iu 配列, ** LI 配列。

データ全体で308個のギャップがあり、そのうち273個が系統樹の中で位置を一意的に特定できた。

η グロビン ψ の塩基配列からは、長い4個のギャップが見いだされた(表1で星印をつけたもの)。これらは *Alu* または *L1* 配列に由来するものであり、他の大多数の短いギャップとは異なるメカニズムで生じた可能性が大きいので、以降の解析からは除外した。

表2は、各データごとの欠失、挿入、ギャップの統計を示している。核DNAでは、塩基の欠失が挿入の約2倍高い頻度で生じており、ミトコンドリアDNAでも弱いながら同様の傾向が見られた。de Jong と Rydén¹⁴⁾ はヘモグロビン偽遺伝子を中心とする塩基配列データを分析して、欠失が挿入よりも高頻度で生じていることを示した。プロセスト型偽遺伝子の塩基配列をヒトとマウスで比較した Graur ら⁹⁾ も同様の傾向を見いだした。各データにおける挿入・欠失の平均長が表2の右

表2 各遺伝子領域における欠失と挿入の数⁹⁾

データ	挿入	欠失	ギャップ	合計	平均長
核DNA					
η グロビン ψ	43.0	73.5	35.5	152	4.6
β グロビン	6.0	24.0	26.0	56	4.2
Ig- α	2.5	11.0	13.5	27	6.5
Ig- $\epsilon\psi$	5.1	6.6	6.3	18	3.8
合計	56.6	115.1	81.3	253	4.6
ミトコンドリアDNA					
Dループ	6.0	12.3	20.7	39	5.2
1/3ゲノム	5.3	3.0	3.7	12	3.1
合計	11.3	15.3	24.4	51	4.7

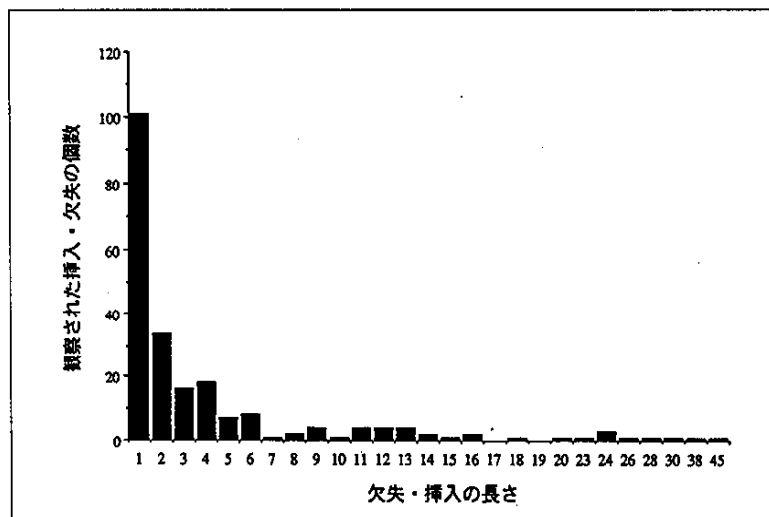


図4 核DNAにおける欠失・挿入の長さの分布⁹⁾

端に示してある。おもしろいことに、平均長は核DNA、ミトコンドリアDNAを問わずかなり一様であり、5塩基弱であった。

核DNAにおける欠失・挿入の長さの全体の分布が図4に示してある。単一塩基の欠失と挿入が最も高頻度(全体の約半分)で見いだされ、ギャップ長が長くなるに従ってその出現頻度が小さくなった。ミトコンドリアDNAでも同様の傾向が見られた。

高頻度の単一塩基ギャップは、Graur ら⁹⁾もヒトとマウスのデータから見いだしている。9種類の単子葉植物の葉緑体DNA非コード領域0.9 kbの塩基配列を調べた Golenberg ら¹⁵⁾のデータでは、56個のギャップのうち23個が単一塩基の挿入または欠失であった。Morita¹⁶⁾は大腸菌 *rpsL* 遺伝子の突然変異パターンを調べたところ、約半分のギャップはやはり単一塩基タイプであった。したがって、単一塩基のギャップが高頻度で出現するのは、原核生物、真核生物を問わず、塩基配列の進化で共通な現象かもしれない。

IV. 欠失と挿入の進化速度

次に、いよいよおのおのの欠失・挿入を図1の系統樹の各枝にマップした。その結果の例が図5に示してある。図5(a)は η グロビン ψ 領域、図5(b)はIg- α 非コード領域、図5(c)はミトコンドリアDNA 1/3ゲノム非コード領域である。欠失と挿入の生じた数字のなかに小数のものがあるのは、どの枝で生じたかの可能性が複数ある場合も含めたからである。

一見して、ギャップの数は系統樹の枝の長さ(進化時間)と高い相関のあることがわかる。このことから、欠失・挿入の進化速度は、霊長類の進化を通じてほぼ一定であることが予想される。そこで筆者らは以下の方法で欠失・挿入の進化速度を推定した。

欠失・挿入の進化速度が一定であれば、進化時間と欠失・挿入数の間に比例関係が成り立つはずなので、横軸に進化時間(系統樹の枝の長さ)を、縦軸に系統樹の各枝で1 kbあたりで生じた欠失・挿入

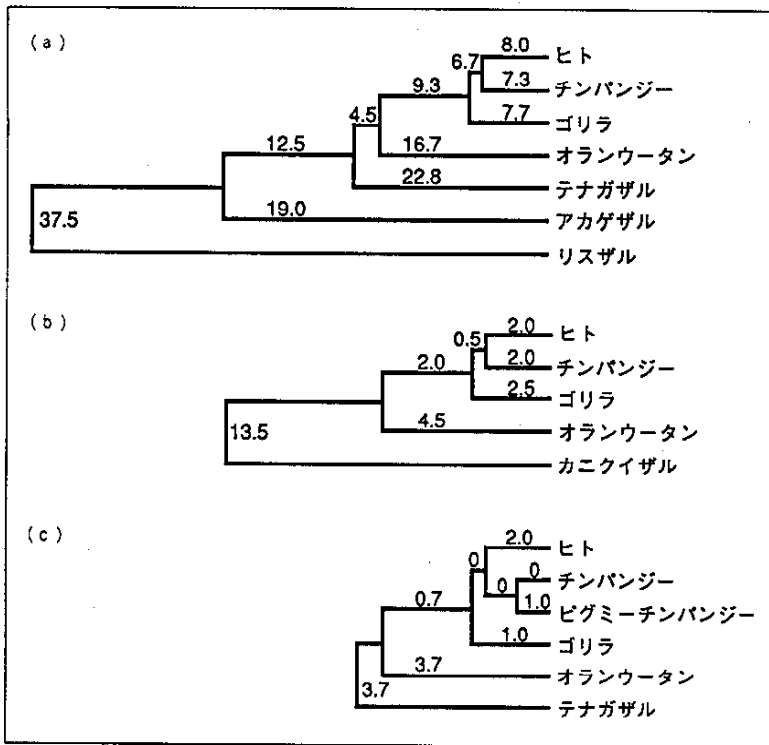


図5 図1の系統樹の各枝にマップされた欠失と挿入の数⁵⁾
 (a) γ グロビン偽遺伝子領域 (文献7のデータより)。 (b) Ig- α 遺伝子非コード領域 (文献10のデータより)。 (c) ミトコンドリア DNA 1/3ゲノム領域 (文献12のデータより)。

数をプロットした。それに原点を通る回帰直線をあてはめて、傾き (進化速度) を求めた結果が表3に示してある。系統樹の各枝はそれぞれ独立に進化しているので、各枝ごとに分析することができるのである。

このようにして推定された進化速度を見ると、推定値の標準誤差がきわめて小さいことがわかる。これは、塩基の欠失・挿入の進化速度に一定性があることを示

表3 各DNA領域における欠失と挿入の進化速度の推定値⁵⁾

データ	ギャップ 総数	配列長 (kb)	進化速度±標準誤差 (/kb/MY)	自由度
核DNA				
γ グロビン ψ	152	6.4	0.12±0.04	10
β グロビン	56	3.2	0.14±0.01	8
Ig- α	27	1.5	0.24±0.01	6
Ig- $\epsilon\psi$	18	2.3	0.16±0.03	4
合計	253	13.4	0.15±0.01	31
ミトコンドリアDNA				
Dループ	39	1.1	2.01±0.38	4
1/3ゲノム	12	0.1	2.31±0.26	8
合計	51	1.2	2.27±0.21	13

している。Ig- α 遺伝子非コード領域の進化速度 (0.24±0.01/kb/MY) が、他の3領域の速度とは統計的に有意に高くなっている。これは、領域によって欠失・挿入の進化速度が異なっている可能性を示唆するものである。しかし、今回用いた各データセットはそれほど大きいものではなく、また核DNAとミトコンドリアDNAでは、それらグループ内での速度のばらつきに比べるとグループ間の違いがきわめて大きいので、各グループのデータを合体して核DNAとミトコンドリアDNAそれぞれで回帰直線を求めた (図6)。その結果、核DNAの進化速度は0.15±0.01/kb/MY、ミトコンドリアDNAの進化速度は2.27±0.21/kb/MYと推定された。明らかにミトコンドリアDNAにおける欠失・

挿入の進化速度は核DNAよりもはるかに高い。

Saitou⁴⁾はヒト上科の核DNAにおいて、4種類の非コード領域塩基配列を分析して、欠失・挿入の進化速度を0.18/kb/MYと推定した。この値は、今回の方法とは多少異なる方法で得られたものであり、また、使

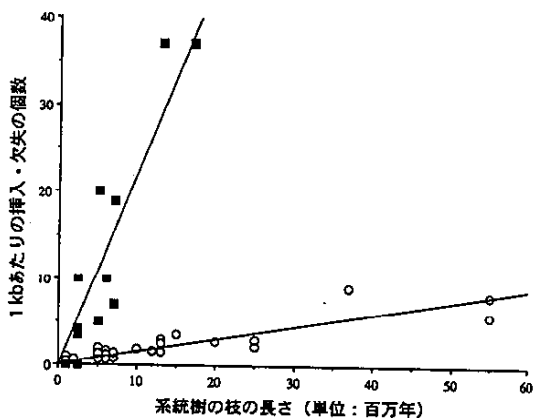


図6 核DNA (○) とミトコンドリアDNA (■) における欠失と挿入の進化速度の比較⁵⁾

表 4 図 1 の系統樹の各枝での核 DNA における欠失・挿入の進化速度の比較^{a)}

枝 ^{a)}	欠失・挿入の進化速度 (/kb/MY)				
	η グロビン ψ	β グロビン	lg- α	lg-e3 ψ	平均
C-B(1)	1.05	0.47	0.33	0.61	0.75
B-ヒト(5)	0.25	0.28	0.27	0.30	0.27
B-A-チンパンジー(5)	0.23	0.13	0.27	0.40	0.24
C-ゴリラ(6)	0.20	0.10	0.28	0.17	0.18
D-C(7)	0.21	0.11	0.19	—	0.18
D-オランウータン(13)	0.20	0.12	0.23	—	0.18
D-E-F-旧世界猿(37)	0.13	0.11	0.24	—	0.14
F-G-新世界猿(55)	0.11	0.15	—	—	0.12

a) 図 1 の系統樹の枝。かつこ内の数字は各枝の長さ (MY)。

われたデータセットも一部分重なっているだけだが、今回の推定値 (0.15 \pm 0.01) とよく似ている。

本研究では最大節約原理が各欠失・挿入を系統樹の枝にマップするための方法として用いられている。このため、前述したように、ある程度以上離れた配列を比較した場合に、過小推定を得るおそれがある。塩基置換の場合と同様に、同じ位置で欠失と挿入がくり返し生じている可能性が出てくるからである。そこで、系統樹の枝の長さが欠失・挿入の進化速度の推定に及ぼす影響を調べた。表 4 は、図 1 の系統樹のさまざまな枝ごとに推定した欠失・挿入の進化速度の推定値を示している。枝の長さ (進化時間) が長くなるほど進化速度の推定値が小さくなる傾向が明らかに認められる。したがって、核 DNA の欠失・挿入の真の進化速度は、すべての枝のデータを総合して得られた今回の推定値 (0.15/kb/MY) より少し大きい可能性がある。Golenberg ら¹⁵⁾も、単子葉植物の族間および族内の種の塩基配列を比較して、同様の傾向を見いだしている。この問題点は、今後もっとデータを増やして検討することが必要なことであろう。

V. 欠失と挿入の進化速度からわかること

Li と Sadler¹⁷⁾は、ヒトの核 DNA の同一遺伝子座で複数の配列が決定されている塩基配列をデータベースから拾い出し、各データについて綿密な検討を行なった。その結果、エクソンの 3' および 5' 非コード領域の計 23.4 kb のなかで、4.4 個のギャップを見いだした (小数点があるのは、3 本以上の配列を比較した場合があるため)。今回筆者らが得た塩基の挿入・欠失の進化速度の推定値 (0.15/kb/MY) をこのデータにあてはめると、ヒトの 2 個の相同遺伝子間の違いは、進化時間にして (4.4/23.4)/(2 \times 0.15)=0.63 MY、すなわち 63 万

年となる。この分岐年代の推定値は、Takahata¹⁸⁾が推定した、ヒト集団において中立に進化している核 DNA 遺伝子の分岐年代 (約 80 万年) と似通った値となっている。これは、欠失・挿入の進化速度が、ある程度一定であることを支持する結果である。

Saitou¹⁹⁾は 1 万塩基以上の塩基配列データを分析して、

ヒトとチンパンジーの間の非コード領域の核 DNA における塩基置換数を、塩基サイトあたり 0.0139 と推定した。これら 2 種の分岐年代を、図 1 に従って 500 万年と仮定すると、塩基置換の進化速度は塩基サイトあたり・年あたり 1.39×10^{-9} (1.39/kb/MY) となる。したがって、霊長類の核 DNA 非コード領域においては、塩基置換は欠失・挿入の約 8 倍高い頻度で生じていることになる。

ミトコンドリア DNA の非コード領域における欠失と挿入の進化速度は今回 2.3/kb/MY と推定された。つまり、欠失・挿入はミトコンドリア DNA では核 DNA よりも 10 倍以上高い頻度で生じていることになる。Kondo ら²⁰⁾は霊長類におけるミトコンドリア DNA の同義置換速度を 23.7/kb/MY と推定したが、これは核 DNA よりも約 17 倍高い値である。したがって、核 DNA と同様にミトコンドリア DNA でも、塩基置換は欠失・挿入よりもずっと高い頻度で生じていることになる。

このように、ミトコンドリア DNA でも核 DNA でも、塩基の置換が高頻度で欠失・挿入は低頻度であり、また塩基置換と欠失・挿入どちらの場合でも、ミトコンドリア DNA では高速度で核 DNA では低速度という対照的な結果は、塩基置換と欠失・挿入の進化速度ないし突然変異率の間に共通のメカニズムが存在することを示唆するのではなからうか。可能性として、哺乳類のミトコンドリア DNA で DNA 複製効率が核 DNA よりも低いのは、塩基の置換だけでなく、欠失・挿入にも影響していることが考えられる。

もしこの仮説が哺乳類において正しいとすれば、ショウジョウバエでは欠失・挿入の中立進化速度 (自発突然変異率) が核 DNA とミトコンドリア DNA でほぼ等しいと予想することができる。ショウジョウバエでは同義塩基置換速度が核 DNA とミトコンドリア DNA

でほぼ似ているからである(たとえば文献21を参照)。近い将来、ショウジョウバエの塩基配列データが多数発表されれば、筆者らの予言を確かめることができるかもしれない。

VI. 塩基の欠失と挿入に基づく進化距離の分析

筆者は一部のデータについて、TajimaとNei²⁾の塩基の欠失と挿入に基づく進化距離を計算した。整理された2本の塩基配列XとYにおいて、 S_X と S_Y をそれぞれXとYの塩基数、 S_{XY} をXとYの間でギャップを除いた共通部分の塩基数とすれば、この進化距離(D_{XY})は以下のように定義される。

$$D_{XY} = -2 \log(S_{XY}) / \sqrt{S_X S_Y}$$

これら欠失と挿入に基づく進化距離行列から近隣結合法²³⁾を用いて系統樹を作成した(図7)。図7(a)(b)とも、新世界猿を外群に加えて、系統樹の根の位置を与えてある。どちらの系統樹も系統関係(樹形)は図1のものと同じとなっているが、枝の長さがかなり異なっている(これら2つの系統樹は同一スケールで描いてある)。Koopら²³⁾の η グロビン Ψ 領域2.0kbの部分配列データを用いた図7(a)の系統樹で、アカゲザルの系統とヒト上科が分岐する以前の系統において、進化距離(枝の長さ)が大きくなっている。しかし、ヒト上科の系統内では、 η グロビン Ψ 領域でも β グロビン領域でもあまり違いがないように見える。

そこで、ヒト上科4種間の進化距離を縦軸に、図1の系統樹の進化時間を横軸にとって、図6と同様の分析を試みた。すると、進化速度はおよそ 0.5×10^{-9}

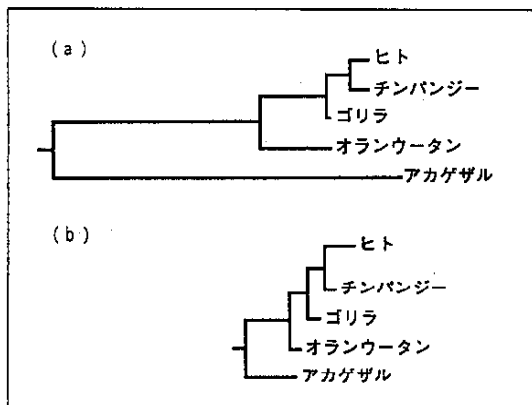


図7 欠失・挿入に基づく進化距離行列から近隣結合法²³⁾を用いて作成した系統樹⁴⁾

(a) 文献23のデータより。(b) 文献8のデータより。

(年あたり・塩基あたり)と推定された。向井と田嶋²⁴⁾は、TajimaとNei²⁾が発表した欠失と挿入に基づく進化距離をもとにして、遺伝子領域ごとの進化速度を推定したが、非コード領域では、およそ 1.0×10^{-9} であった。これは、筆者が推定した上記の速度とほぼ同じ桁の値である。もっとも、彼らの用いた進化距離は必ずしも進化時間に比例して増加しているわけではない¹⁾。

これまで本稿で論じてきた進化速度(λ_e)は欠失・挿入の生じた数をもとにしていて、上記の進化距離から推定した進化速度(λ_d)は、塩基サイトあたりの欠失・挿入による速度である。したがって、両者を同一に扱うことはできないが、おおまかにいって、 λ_e に欠失・挿入の平均塩基長をかけると、 λ_d に対応するはずである。筆者らが推定した核DNAの λ_e は $0.15/\text{kb}/\text{MY} = 0.15 \times 10^{-9}$ であり、欠失・挿入の平均塩基長は4.6塩基なので、この値をもとにすると、 $\lambda_d = 4.6 \times 0.15 \times 10^{-9} = 0.69 \times 10^{-9}$ となり、進化距離から推定した λ_d (およそ 0.5×10^{-9})とあまり差がない。もっとも、 λ_d と λ_e を厳密に比較するには同一の塩基配列データを用いる必要があるので、これら2種類の進化速度の性質については今後いつその検討を行なうべきであろう。

VII. さまざまな種類の欠失・挿入とその突然変異のメカニズム

塩基の微少な欠失・挿入が生じるメカニズムとして、Streisingerら²⁵⁾のモデルがある。これは、DNAに損傷が起きると、その近傍に“AAA”などのような同一塩基が並んでいる配列がずれて(slipped mispairing)欠失・挿入が生じるというものである。しかしこのモデルでは欠失と挿入が同一頻度で生じることになる。欠失が挿入よりも起こりやすいことを説明するために、de JongとRydén¹⁴⁾は、配列のずれる際に生じる小さなループが1本鎖特異的ヌクレアーゼで切断されるという過程を上記のモデルにつけ加えることを提唱している。

本稿では、もっぱら塩基の微少な欠失と挿入に的をしぼって話を進めてきた。しかし、塩基の欠失と挿入を広義に解釈して、「塩基数が変化する突然変異」と考えると、それにはいくつかのタイプが存在する(表5)。各タイプによって、突然変異メカニズムや発生率、進化速度が大きく異なっているだろうが、現象の詳しい分析はまだ十分に行なわれてはおらず、今後の研究が待たれる。

表 5 塩基数の変化する突然変異のタイプの種類

タイプ	特 徴
微少な欠失・挿入 (本稿で扱ったもの)	コード領域ではフレームシフト突然変異などをひき起こすことが多い
微少な直列重複のコピー数変化	STR, VNTRなどが含まれる; コード領域では遺伝病発症と関連することがある
外来 DNA によるもの	ウイルス, トランスポゾンなどによる
遺伝子重複	遺伝子族の誕生をうながす; 不完全重複は遺伝子の伸長をひき起こすことがある
ゲノム重複	ゲノム塩基量の倍増
遺伝子の再構成	体細胞における Ig 遺伝子などの場合に限られる

これらのなかでも STR (short tandem repeat) および VNTR (variable number of tandem repeat) などの名称でよばれる、微少な直列重複がコピー数を変化させる現象は最近とくに注目されている。これらの突然変異は、どうやら塩基の微少な欠失と挿入とは異なるメカニズムで生じているらしい。STR は 1 塩基から 5 塩基程度の短いリピートユニットがくり返す構造をもつが、リピート数がある閾値をこえるとリピート数が急増し、それが遺伝病の発病につながる例が、ハンチントン病²⁶⁾などで知られている。

Richards と Sutherland²⁷⁾ は、この閾値が岡崎フラグメントの長さ (真核生物では約 100~200 塩基) ではないかと提唱している。STR 配列のリピート数が突然変異で増加して岡崎フラグメントの長さをこえると、1 個の岡崎フラグメントがすべて STR 配列で構成されることが起こりうる。このとき、著しい配列のずれが生じて、STR 配列のリピート数急増が起こるというわけである。

一方、VNTR はいわゆるミニサテライト DNA の多型であり、STR よりもリピートユニットの塩基数が多い。Jeffreys ら²⁸⁾ は精子における VNTR を詳しく調べて、減数分裂時に生じる高率の遺伝子変換が深く関与している可能性を指摘している。このような新しい知見は、塩基のくり返し構造の変化を生じるメカニズムが複数ある可能性を示している。

VIII. 塩基配列の整列の問題

本稿では、観察値としての塩基配列のギャップがすでに与えられたものと仮定して話を進めたが、分岐時間の大きい配列間の比較になると、話はそう簡単ではない。配列の整列 (アラインメント) に恣意性が生じるからである。Tyson²⁹⁾ は、2 本の配列を整列する際のギャップ・ペナルティー変数を変化させると、同一の

配列データから異なる整列結果の得られることを示している。この、整列における恣意性はやっかいな問題であり、分子進化学の今後の課題であろう。

さらに、与えられた塩基配列データから恣意性をできるだけ回避して最良の整列結果が得られたとしても、そうし

て得られたギャップが、進化の過程で実際に生じたものかどうかという問題がある。たとえば、Thorne と Kishino³⁰⁾ はコンピュータシミュレーションで塩基配列に塩基の置換、欠失・挿入を生じさせ、そうして得られた 2 本の配列をあるアルゴリズムを用いて整列させた結果と、シミュレーションで実際にギャップが挿入されていった結果 (真の整列) とを比較して、それらがかなり異なっていることを示した。これは形質が“相同”とはどういうことかという、進化学で古くからの大問題とも直接関係する。塩基配列の変化が生物進化の根底にあるのは疑いないことだが、底に着地したからといっても、まだまだ解決しなければならない問題は多いようである。

文 献

- 1) 根井正利 (五條堀孝・斎藤成也共訳): 分子進化遺伝学, 培風館, 東京 (1990)
- 2) Tajima, F., Nei, M.: *Mol. Biol. Evol.*, 1, 269-285 (1984)
- 3) Graur, D., Shuali, Y., Li, W.-H.: *J. Mol. Evol.*, 28, 279-285 (1989)
- 4) Saitou, N.: *Population Paleo-Genetics* (ed. Takahata, N.), pp.335-343, Japan Science Society Press, Tokyo (1992)
- 5) Saitou, N., Ueda, S.: *Mol. Biol. Evol.*, 11, 印刷中 (1994)
- 6) 木村資生 (向井輝美・日下部真一共訳): 分子進化の中立説, 紀伊國屋書店, 東京 (1986)
- 7) Bailey, W., Hayasaka, K., Skinner, C. G., Kehoe, S., Sieu, L. C., Slighton, J. L., Goodman, M.: *Mol. Phylogen. Evol.*, 1, 97-135 (1992)
- 8) Maeda, N., Wu, C. I., Bliska, J., Reneke, J.: *Mol. Biol. Evol.*, 5, 1-20 (1988)
- 9) Ueda, S., Watanabe, Y., Saitou, N., Omoto, K., Hayasida, H., Miyata, T., Hisajima, H., Honjo, T.: *J. Mol. Biol.*, 205, 85-90 (1989)
- 10) Kawamura, S., Tanabe, H., Watanabe, Y., Ku-

- rosaki, K., Saitou, N., Ueada, S. : *Mol. Biol. Evol.*, 8, 743-752 (1991)
- 11) Foran, D. R., Hixson, J. E., Brown, W. M. : *Nucl. Acids Res.*, 16, 5841-5861 (1988)
 - 12) Horai, S., Satta, Y., Hayasaka, K., Kondo, R., Inoue, T., Ishida, T., Hayashi, S., Takahata, N. : *J. Mol. Evol.*, 35, 32-43 (1992)
 - 13) 三中信宏・斎藤成也 : 新生化学実験講座 第16巻 分子進化実験法 (五條堀孝ら共編), pp.380-395, 東京化学同人, 東京 (1993)
 - 14) de Jong, W. W., Rydén, L. : *Nature*, 290, 157-159 (1981)
 - 15) Golenberg, E. M., Clegg, M. T., Durbin, M. L., Doebley, J., Ma, D. P. : *Mol. Phylogen. Evol.*, 2, 52-64 (1993)
 - 16) Mo, J.-Y., Maki, H., Sekiguchi, M. : *J. Mol. Biol.*, 222, 925-936 (1991)
 - 17) Li, W. H., Sadler, L. A. : *Genetics*, 129, 513-523 (1991)
 - 18) Takahata, N. : *Mol. Biol. Evol.*, 10, 2-22 (1993)
 - 19) Saitou, N. : *Am. J. Phys. Anthropol.*, 84, 75-85 (1991)
 - 20) Kondo, R., Horai, S., Satta, Y., Takahata, N. : *J. Mol. Evol.*, 36, 517-531 (1993)
 - 21) Sharp, P. M., Li, W. H. : *J. Mol. Evol.*, 28, 398-402 (1989)
 - 22) 斎藤成也 : 新生化学実験講座 第16巻 分子進化実験法 (五條堀孝ら共編), pp.400-410, 東京化学同人, 東京 (1993)
 - 23) Koop, B. F., Goodman, M., Xu P., Chan, K., Slightom, J. L. : *Nature*, 319, 234-238 (1986)
 - 24) 向井輝美・田嶋文生 : 化学, 42, 543-548 (1987)
 - 25) Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E., Inouye, M. : *Cold Spring Harb. Symp. Quant. Biol.*, 31, 77-84 (1966)
 - 26) Huntington Disease Collaborative Research Group : *Cell*, 72, 971-983 (1993)
 - 27) Richards, R. I., Sutherland, G. R. : *Nature Genet.*, 6, 114-116 (1994)
 - 28) Jeffreys, A. J., Tamaki, K., MacLeod, A., Monckton, D. G., Neil, D. L., Armour, J. A. L. : *Nature Genet.*, 6, 136-145 (1994)
 - 29) Tyson, H. : *Genome*, 35, 360-371 (1992)
 - 30) Thorne, J. L., Kishino, H. : *Mol. Biol. Evol.*, 9, 1148-1162 (1992)