総説

霊長類の系統関係と祖先集団の多型

颯田 葉子

総合研究大学院大学先導科学研究科生命共生体進化学専攻

はじめに—霊長類分子系統学の歴史—

最初に、霊長類の分子系統学研究の歴史から紹介する。分子の情報を用いて系統関係を推定する分子系統学の手法で霊長類の系統関係を調べた最初の研究は、1900年代の初頭にまで遡る。1904年ドイツの免疫学者ナッダール (G. H. F. Nuttal) はウサギを用いて抗ヒト血清をつくり、チンパンジー (*Pan troglodytes*), ギリラ (*Gorilla gorilla*), 旧世界猿, 新世界猿, イヌなどの血清と反応させ、その沈殿物の量を測定した (表 1)。血清中に含まれるタンパク質に対する抗体を含む血清を抗血清とよび、血清由来のタンパク質をウサギ

やニワトリに注射して作らせることができる。この抗血清を血清タンパク質と混ぜると抗血清中の抗体とタンパク質 (抗原) が反応して凝集体ができる (免疫沈降反応)。この凝集体の量は、抗体を作る際に用いた血清タンパク質と免疫沈降反応で抗原となった血清のタンパク質の構造が似ている程多くなる。

ナッダールは、その沈殿物の量からこれらの生物とヒト (*Homo sapiens*) との系統の近縁の程度を推定した。その結果、抗ヒト血清は旧世界猿や新世界猿に較べて、チンパンジーやギリラなどの類人猿の血清とより多くの抗原抗体反応を示した。つまり、ヒトはチンパンジーやギリラなどの類人猿に近縁であるという結論を得た。蛋白質の精製技術は十分ではなかったし、定量という点でも粗雑ではあったが、得られた結論は後の分子系統学で得られた結論と一致する。その後の分子系統学の草分け的な研究と位置づけることができる。

1960年代、血清タンパク質の精製法や、凝集体の定量法が進み、免疫沈降により生物間の近縁の程度を推定する方法が改善されていった。1963年グッドマン (M. Goodman) は、血清アルブミンを精製し、アルブミンの構造の違いを免疫沈降法で定量した。抗ヒトアルブミン血清をつくり、チンパンジー、ギリラ、オラウータンのアルブミンと反応させ、その凝集反応の程度の違いからヒトはオラウータ

表 1 ウサギ抗ヒト血清と様々な生物の血液タンパクとの反応の結果

種	免疫血清 ¹⁾		
	1	2	3
ヒト	100	100	100
チンパンジー			130
ギリラ			64
オラウータン	47	80	42
ヒョケザル	30	50	42
ショウハナジログエノン	30	50	
クモザル	22	25	
ネコ	11		3
イヌ	11		3

註 1) ヒトとの沈殿物 (反応物) の量を 100 としたときに、他の生物の沈殿物の量を相対的に表している。

ンよりチンパンジーやゴリラに近縁であると結論した。1967年、カリフォルニア大学バークリー校のサリッチ (V. M. Sarich) とウィルソン (A.C. Wilson) はグッドマンの方法を定量化し、免疫学的距離としてヒト・チンパンジー・ゴリラ・オランウータン・テナガザルの系統関係を定量的に推定した。その結論は、ヒトはチンパンジー・ゴリラに近縁であるというものだった。この論文の焦点は、ヒトと類人猿の系統関係よりも、この系統関係に時間のスケールを加えたことにある。免疫学的距離が時間に比例することを利用して、チンパンジー・ゴリラとヒトの分岐年代を推定した。ヒトと旧世界猿の分岐を今から3000万年前として、チンパンジー・ゴリラとヒトの分岐年代を500万年前と推定した。当時、形態や化石の比較に基づいた研究の多くは、ヒトの分岐は数千万年前と推定していた。

1970年代になると、遺伝子の産物としてタンパク質の種間での比較が可能となっていた。直接タンパク質のアミノ酸配列を決める方法は、手法的に限界があり、大きなタンパク質には使えなかった。それを補う方法として、免疫沈降法の一つ（正確には micro-complement fixation 法）が用いられた。その他に、電気泳動により、二種のタンパク質のアミノ酸が平均的にいくつ異なるかを推定することもできる。この3種類の方法を用いて、1975年にカリフォルニア大学バークリー校のキング (Mary-Clarie King) とウィルソン (A.C. Wilson) はヒトとチンパンジーで44個の構造タンパク質のアミノ酸配列を比較した。これらの方法で、推定したヒトとチンパンジー2種の間での、タンパク質アミノ酸配列の平均的な違いは1000アミノ酸あたり7.2個、0.72%と推定された。つまりこの2種の間では、タンパク質アミノ酸配列は99%以上同一であると推定されたことになる。

その後、分子系統学はタンパク質からDNAの時代に移る。分子生物学のさまざまな技術の発達により、比較的容易なDNAの単離法や

塩基配列の決定法が確立した。また系統樹の作成法や塩基置換数推定法も数多く開発された。1981年、16,000塩基対余のヒトミトコンドリアDNAの全遺伝子の配列が決定され、更にPCR法の確立により核DNAの遺伝子の塩基配列が容易に決定できるようになった。そして、これらの配列情報を用いた分子系統学の研究が全盛となっていった。

1995年、当時国立遺伝学研究所の實来聡博士のグループが、ヒト、チンパンジー、ゴリラ、オラウータン (*Pongo pygmaeus*) のミトコンドリアDNA全塩基配列を決定した。實来らはこの全塩基配列を用いて、これらの種の分岐年代を推定した。その推定によれば、ヒトはチンパンジーに最も近かった。ヒトとオラウータンの分岐を1300万年前とすると、ヒトとチンパンジーの分岐は490万年前と推定された。このミトコンドリアDNAの系統樹は多くの数の塩基置換（ヒトとチンパンジー間ではおよそ1,000塩基置換）に基づくもので、ヒトとチンパンジーが近縁であることは、非常に高い確率で示された。

ところが、核のDNAデータは必ずしも、ヒトとチンパンジーが近いことを支持してはいなかった。1993年ロジャー (Rogers, 1993) は7個の核遺伝子のデータで、ヒト・チンパンジー・ゴリラの分岐関係について調べた。その結果、4個の遺伝子座ではヒトとチンパンジーが近いという系統関係を示したが、残りの2個の遺伝子座では、チンパンジーとゴリラが近縁であることを示唆し、また、1遺伝子座は、3種が同時に分岐したことを示唆していた。これは、核のDNAの進化速度がミトコンドリアDNAに比べ遅く、また、比較する遺伝子の大きさが十分でないことによる誤差のために生じているのだろうか。

2000年以降に、多くの核DNAのデータが蓄積して、2004年までには、252の遺伝子座のデータが集まった。そのうちで、110個の遺伝子座ではヒトとチンパンジーが近縁であることを支持していたが、45遺伝子座と38遺伝

表2 遺伝子座ごとに示す系統関係が異なる

系統関係 ¹⁾	Satta <i>et al.</i> (2000)	Chen and Li ²⁾ (2001)	O'hUigin <i>et al.</i> (2002)	Kitano <i>et al.</i> (2004)	合計 ³⁾
	45	53	51	103	252
(H,C),(G,O)	21	(31)	24	34	11[57]
(H,G),(C,O)	7	(10)	14	14	43[23]
(C,G),(H,O)	7	(12)	9	10	38[20]
(H,C,G),O	10	(0)	4	45	59

註1) Hはヒト, Cはチンパンジー, Gはゴリラ, Oは外群を表す。括弧のなかは, その2種が近縁であることを示している。最下段はヒト, チンパンジー, ゴリラの分岐順序が決められない場合を示している。

註2) 系統関係の解析法が他の3つの論文とは異なるため, 括弧の中に入れて表示をしている。

註3) 角括弧の中の数字は, 最下段を除く三つの系統関係を示す遺伝子座の割合を示している。

子座はそれぞれヒトとゴリラ, ゴリラとチンパンジーが近縁であることを示していた。残りの59遺伝子座では, ヒト・チンパンジー・ゴリラがほぼ同時に分岐した3分岐を示した(表2)。

祖先集団多型が遺伝子分岐に与える影響—種の系統関係と遺伝子の系統関係—

なぜ核DNAでは, ヒトとチンパンジーが近縁であることを示す系統樹だけでなく, ヒト・ゴリラやゴリラ・チンパンジーが近縁であることを示す系統樹が得られるのだろうか?これを説明するために, まず種の系統関係と遺伝子の系統関係について概観する必要がある。

図1には, 種の系統関係と遺伝子の系統関係の模式図が示してある。大事なことは, 異なる種に由来する相同遺伝子(オーソログ遺伝子と呼ぶ)は, 祖先集団で共通祖先から分

図1 オーソログの遺伝子系図と種の系統関係1。

種1と種2に由来するA, B, 2遺伝子座のオーソログ(A1/A2とB1/B2)の系統関係を示している。最も外側の太線では種(集団)の系統関係, 点線は種分岐の時間をあらわしている。オーソログの分岐時間(t)は, 種の分岐時間(t_0)と祖先集団中での分岐時間(τ)の和($t=t_0+\tau$)となる。

岐したこと、そして、その遺伝子の分岐は必ず種分岐がおこる前に起きているという二つの点である。このことは、遺伝子の分岐時間を推定すれば、それは種の分岐時間よりも長くなることを意味している。ここでは、種分岐と比べて遺伝子の分岐が長くなるその時間を τ (タウ) という変数であらわす。

“変数” という言葉を使ったのは、図 1 に示した様に τ は比較する遺伝子により異なるからである。例えば、 τ の値をゲノム中の全ての遺伝子座について決めることができたとして。そうすると、この τ の値が遺伝子座によりどのように変化するかを知ることができる。 τ の値が 0.005 から 0.01 の間にある遺伝子が全体の 30% で、0.01 から 0.015 の間にある遺伝子は全体の 50% にあたるという様にある。異なる τ の値にそれぞれどれくらい割合の遺伝子座 (あるいは領域) が相当するか、別の言葉で表現すると、ある特定の τ の値の遺伝子座 (あるいは領域) がどの程度の確率であられるかを表現したものを「 τ の分布」という。この τ の分布は現存するある種 (集団) から複数の遺伝子座の遺伝子をランダムに 2 個とった時その 2 個の配列が祖先配列から分岐した時間の分布と同じである。この分布は集団遺伝学の遺伝子系図学 (gene genealogy) という分野で研究が進んでいる (この分布に興味のある方は、Tajima 1983, Kingman 1982 を参照いただきたい。)

オーソログの分岐時間 (t) は、種の分岐時間 (t_s) と祖先集団中での分岐時間 (τ) の和 ($t=t_s+\tau$) となる (図 1)。繰り返しになるが、オーソログの分岐時間 (t) をそのまま種の分岐時間としてしまうことは、あまり好ましくない。特に種の分岐時間 (t_s) に対して、祖先集団中の分岐時間 (τ) が無視できない程度に大きい時には、注意が必要である。

上述した遺伝子系図学に基づけば、 τ の分布は集団の有効な大きさ (effective population size : N_e) に依存して決まる。そして、 τ の平均値は $2N_e$ 世代である。ここで気をつけるべ

きことは時間の単位である。集団の中で起こる出来事は基本的に世代単位で起きる。一方、種の分岐時間は習慣的に物理時間で記述される、つまり年単位である。そこで、オーソログの分岐時間を表現するには、単位の統一が必要で、そのために世代時間 (g) を用いる。この g は一世代の長さが何年に相当するかを表している。ここでの世代とは、個体の一生の長さ、つまり誕生してから死ぬまでの時間 (いわゆる寿命) ではなく、個体が次世代を残すよう (つまり生殖可能) になるまでの年齢をさす。現生人類の場合はおよそ 20 年と考えてよいだろう。この g と祖先集団の有効な大きさ (N_a) を考えるとオーソログの分岐時間 (t) は $t=t_s+2N_ag$ となる。そして、 t_s に対して相対的に $2N_ag$ が大きいとき、例えば、 $t_s=2N_ag$ となるような場合は、オーソログの分岐時間 (t) を種の分岐時間 (t_s) と見積もってしまうと、分岐時間の過剰推定となることが明らかである。

実際のヒトとチンパンジーの種分岐について考えてみよう。ヒトとチンパンジーの種分岐時間を 600 万年前 ($t_s=6 \times 10^6$ 年) とする。上記の $t_s=2N_ag$ が成り立つような祖先集団の有効な集団サイズはどのくらいだろうか。ここで、 g を 10 年として、計算してみる。 $t_s=6 \times 10^6$ 年 $= 2N_ag = 20N_a$ となり、祖先集団の有効な集団サイズはおよそ 30 万となる。現存チンパンジーの塩基レベルでの多型を示すデータが蓄積しつつあるが、それによれば、現在の多型を作り出した有効な集団サイズは 10 万のオーダーであることを示しているので、祖先集団が 30 万であったことは、想像に難くない。

さて、ここでもう一度、なぜ核 DNA では、ヒトとチンパンジーが近縁であることを示す系統樹だけでなく、ヒト・ゴリラやゴリラ・チンパンジーが近縁であることを示す系統樹が得られるのかという疑問に戻る。祖先集団の多型は、種の分岐時間の推定だけでなく、種の系統関係と遺伝子の系統関係が一致しないことにも大きな影響を与える。チンパンジー

図2 オーソログの遺伝子系図と種の系統関係2。

図の底辺にある四角で示したのは、現存する3種。種の系統関係は、種1と種2がより近縁であるとする。1と2の祖先種の集団を共通祖先集団1、1, 2, 3の祖先種の集団を共通祖先集団2とする。赤、青、緑の丸印で示したのはそれぞれの種の遺伝子。各種からは2本ずつの相同遺伝子を単離したとする。 τ は共通祖先集団1で、種1と種2に由来する遺伝子が共通祖先へ遡る時間、 T_s は二回の種分岐の間隔時間をあらわす。

(種1)・ヒト (種2)・ゴリラ (種3) の3種の系統関係はチンパンジーとヒトがより近縁である。いっぽう、これらの種からサンプルした遺伝子の系統関係には、種分岐の間隔 (T_s) と祖先集団のサイズが関係する。

ここでは、この遺伝子の系統関係を普通とは逆に現在から過去へ時間をさかのぼって見ていく。たとえば、チンパンジー・ヒト・ゴリラのそれぞれの種から2個ずつ相同遺伝子の配列を得たとしよう。時間をさかのぼっていくと、チンパンジー・ヒト・ゴリラの遺伝子はまずそれぞれの種内で共通祖先遺伝子にたどり着く。そして、チンパンジー・ヒト・ゴリラに由来する3つの系統になる。多くの場合、図2Aに示した様にチンパンジーとヒトの共通祖先集団 (共通祖先集団1) でチンパンジーとヒトに由来する遺伝子は、共通祖先にたどり着く。そして、更に時間をさかのぼるとゴリラとの共通祖先集団 (共通祖先集団2) で、ゴリラの遺伝子との共通祖先にたどり着く。このような遺伝子の系統関係は、種の系統関係を正しく反映している。しかし、図2Bではヒトとチンパンジーの遺伝子が祖先集団1

では共通祖先にたどり着かない。祖先集団1ではなく、祖先集団2までさかのぼる。この場合は、この祖先集団2のなかで、ゴリラの遺伝子とも共存することになる。祖先集団2の中では、チンパンジーとヒトの遺伝子が先に共通祖先にたどり着かなければならない理由は何もない。チンパンジー・ヒト・ゴリラのうち、どの二つが先に共通祖先にたどり着くかは、全くの等確率で起こりえる。即ち、図2Cに示す様に、3種類の場合がありうることになる。もし、ヒト・チンパンジーの遺伝子が先に共通祖先にたどり着けば、遺伝子の系統関係と種の系統関係は一致する。しかし、そうでない場合、つまりヒトとゴリラ、あるいはチンパンジーとゴリラが共通祖先に先にたどり着いた場合には、種の系統関係とは異なる遺伝子の系統関係となる。上述の252の遺伝子座の例では、110遺伝子座でチンパンジーとヒトが近縁であることを示していた。そして、およそ40遺伝子座がそれぞれヒトとゴリラ、あるいはチンパンジーとゴリラが近縁であることを示している。チンパンジーとヒトが近縁であることを示していた110遺伝

子座の中でも、40 遺伝子座程度は、図 2C のような 3 種の祖先集団で共通祖先にたどり着いていたと推測される。ヒトとゴリラ、あるいはチンパンジーとゴリラが近縁であることを示した遺伝子座も加えると、系統関係が確定した遺伝子座の半数以上が祖先集団 1 ではなく、祖先集団 2 で共通祖先にたどり着いていることが推測される。

では、ヒトとチンパンジーの遺伝子が共通祖先集団 1 で共通祖先遺伝子にたどり着くことができないのはどのような場合か？ヒトとチンパンジーが共通祖先遺伝子にたどり着く時間の平均値 (τ) は、祖先集団 1 のサイズを N_{a1} とすると $\tau = 2N_{a1}$ 世代 ($2N_{a1}g$ 年) であることは先に述べた。もしこの τ が種分岐の間隔 (T_s) より長くなると、ヒトとチンパンジーの遺伝子の共通祖先は祖先集団 2 にまで遡ることになる。おおまかにいって、この $2N_{a1}$ 世代 ($2N_{a1}g$ 年) が種分岐の間隔 (T_s) と同じくらいのときは、種の系統関係と遺伝子の系統関係が一致しないことが起きる。(一致する場合と一致しない場合のそれぞれの確率は、 T_s と N_{a1} により決まる。この確率を正確に知りたい方は、Nei & Kumar 2002, “Molecular Evolution and Phylogenetics (Oxford Univ. Press)” の 77 ページを参照いただきたい。) チンパンジー・ヒト・ゴリラの場合、もし、共通祖先集団のサイズが 10 万のオーダーであれば、 $\tau = 2N_{a1}g = 200$ 万年となる。この 3 種の種分岐の間隔がおおよそ 100 万年から 200 万年とすると、チンパンジー・ヒト・ゴリラの場合には、種の系統関係と遺伝子の系統関係が一致しないケースは十分あり得るということになる。

祖先集団サイズと種分岐の時間の同時推定法

ここまでは、ヒトとチンパンジーの祖先集団の大きさをおおよそ 10 万と仮定して説明を進めてきた。ここからは、実際のデータとこの仮定とに整合性があるのかどうかを調べ、現

図 3 ヒトとチンパンジーの BES58,158 座位の塩基置換の程度ヒストグラム。(Satta *et al.* 2004, JME 改変)

横軸は塩基置換の程度を、縦軸は、各塩基置換のレンジの領域の頻度を示す。図中の棒グラフは、実際のヒトとチンパンジーのゲノム配列の比較に基づくデータ。折れ線グラフは、コンピュータシミュレーションに基づく値。

在の配列データから祖先集団サイズの大きさを求める方法について解説する。

ヒトとチンパンジーの祖先集団の大きさの仮定が正しいかどうかを確かめるには、オーソログ間の分岐時間の分布を調べ、実際の分布と仮定した条件での分布を比べてみればよい。オーソログの分岐時間は直接測ることができないので、分岐して以降蓄積した塩基置換の数から推定する。図 3 には、ヒトとチンパンジーのゲノム DNA を比較して、塩基あたりの違いの程度 (p) とそのような p の値を示す領域がどれくらいあるかを示してある。比較した i 番目の領域の長さ (bp) を L_i 、そこで観察される塩基の違いの数を K_i とすると、この領域の塩基あたりの違いの程度 (p_i) は $p_i = K_i/L_i$ となる。図 3 の横軸にはこの p の値の幅を、縦軸には、その p の値が観察された頻度を表している。このデータは、2001 年藤山ら (Fujiyama *et al.*, 2001) が発表した、チンパ

ンジーの BES (BAC end sequences) とヒトのゲノムとの比較である。58,158 組の BES とヒトのゲノム配列の比較で得た p_i の分布を調べた。図 3 の 3 種類の折れ線は、3 種類のモデルで、BES のデータと同じ長さを持つゲノムの断片に蓄積した塩基置換をコンピュータシミュレーションで得て、その分布を表示したものである。1 つめのモデルは、分布に祖先集団の多型が影響しない場合 ($\tau = 0$) で、塩基置換は一定の時間 (t_s) に一定の突然変異率 μ の割合で起きたときのポアソン分布である。また 2 つめのモデルは、1 つめのモデルと同様に、分布に祖先集団の多型が影響しない場合だが、突然変異率は領域によって異なることを仮定している負の二項分布である。最後のモデルは、 $\tau = 2N_a \mu$ $g = 0.25\%$, $t_s \mu = 0.36\%$ が平均となる幾何分布とポアソン分布の和の分布になっている。このとき、 t_s を 600 万年とすると、 N_a はおよそ 20 万となる。

図 4 には、図 1 で示した種の分岐と遺伝子の分岐の時間の関係をあらわした系図に塩基

置換の蓄積の過程を加えた。種 1 と 2 からそれぞれ、 a と a' あるいは b と b' というオーソロジーの関係にある二つの遺伝子座の塩基配列を得たとしよう。それぞれの遺伝子の中立な変異を起こせる塩基の数を L_A と L_B とする。この遺伝子対の分岐時間は、種分岐の時間 (t_s) と祖先集団内の遺伝子の分岐時間 (τ) とに分けられる。ここで、それぞれの遺伝子対に蓄積する塩基置換数について考える。種分岐の時間 t_s は遺伝子対によらず常に一定、つまり定数である。 a と a' あるいは b と b' のような 1 対の遺伝子間に t_s の時間におこる塩基置換の数の平均値は、突然変異率 (塩基置換率) を年当たり塩基当たり μ とすると、それぞれ $y_A = 2t_s L_A \mu$ と $y_B = 2t_s L_B \mu$ となる。実際に起る塩基置換の観察値はこのそれぞれを平均とするポアソン分布に従う。一方、祖先集団での、遺伝子の分岐時間 τ は、指数分布である。時間が指数分布となると、その時間内に起きる塩基置換の数は幾何分布となる。 τ の平均値が $2N_a$ (祖先集団の有効な大きさが

図 4 オーソロジーの遺伝子系図と種の系統関係 3。
オーソロジーの遺伝子系図に塩基置換のプロセスをくわえたもの。

N_a) なので、塩基置換数の平均は $x = 4N_a L_A \mu g$ である。ここで、 g は上述したように一世代の長さ、世代時間となる。

もし、多くの遺伝子座のデータが手に入れば、塩基置換数がどのような分布を示すかを知ることができる。このときのデータの分布と理論的にもっともよく合う t_s と N_a の値を決めることができる。このような方法を最尤法といい、このようにして求めた t_s と N_a の値を最尤推定値と呼ぶ。

ここで、 t_s と N_a の推定値をもとめる際に、系統による世代時間の違いがどのような影響をおよぼすか考察する。確かに、世代時間は系統により異なり、また、同じ系統でも時間を遡った時、世代時間が同じである保証はない。しかし、この推定値は常に $x = 4N_a \mu g$ という突然変異率 μ と世代時間 g が込みになった『積』で表されている。 μ や g の値をどうとるかにより、確かに N_a の大きさは違ってく

x と y の値の尤度分布

図5 ヒトとチンパンジーの BES58,158 座位に基づく祖先集団多型 (x) と種分岐時間 (y) の最尤推定値。図の横軸は x の値を、縦軸は y の値を表している。図中の等高線は、尤度分布を示したものである。図中の赤の矢印が、最大尤度を示す x と y の位置を示している。

図6 霊長類の系統関係と推定値

ヒト (H)、チンパンジー (C)、ゴリラ (G)、オラウータン (O)、旧世界ザル (M)、新世界ザル (N) の系統関係を示している。X は祖先集団の多型の程度を、Y は種分岐の時間に関する推定値。それぞれの記号に付してある、X の下つきの文字は関連する種を表している。たとえば、XHC はヒト (H) とチンパンジー (C) の祖先集団の多型を示している。また YHC、YHG、YHO、YHM、YHN はそれぞれ、チンパンジー、ゴリラ、オラウータン、旧世界ザル、新世界ザルの系統がヒトの系統と分岐した時間を表している。

るが、 x の推定値は μ や g の値には影響されない。また、種分岐の時間についても、実は $y = 2t_s\mu$ を推定している。 μ の値により、 t_s の推定値は違ってくるがやはり y の推定値は影響を受けない。

ヒトに至る系統での祖先集団サイズと種分岐の時間の推定

多くの遺伝子座のデータが手に入った場合、様々な t_s と N_a の値について、観察されている塩基置換数の分布が得られる確率が求められる。この確率がもっとも大きくなる t_s と N_a の値が最尤推定値である。図 5 には、先ほど示した BES の 5 万余の断片間の塩基置換数の分布に基づく、様々な t_s と N_a の値の組み合わせでの、確率が等高線で示してある。図中で矢印で示してある点が、最も尤度の高い点、つまり t_s と N_a の最尤推定値である。

表 3 には、いくつかの文献で推定されたヒト、チンパンジー、ゴリラ、オラウータン、旧世界ザル、新世界ザルの 6 種 (図 6) につい

での推定値をまとめてある。

Kim *et al.* (2010) と Satta *et al.* (2004) はここで紹介した最尤法を用いて推定値をもとめている。一方、Burgess & Yang (2008) は、Markov Chain Monte Carlo (MCMC) 法と呼ばれるコンピュータシミュレーションにより、データによく合うパラメーターを探すという方法を用いている。表に掲載した三つの論文では用いているデータはそれぞれ異なる。Satta *et al.* (2004) では、20 遺伝子座の intron のみを用いているが、ヒトとチンパンジーの比較についてだけは、上記の BES の配列データ (58,158 座) を用いている。また、Burgess & Yang (2008) は Patterson *et al.* (2006) のデータを用いており、データ量としては最大である。一方、Kim *et al.* (2010) では、過去の様々な論文に掲載されたデータを集め、83 領域 (イントロンや non-coding 領域も含める) を用いている。三者それぞれ異なるデータを用いているが、Satta *et al.* (2004) の X_{HO} 、 X_{HOWN} 、 X_{HNWM} を除いて推定値は大変よく似たものとなった。これらの x の値が小さいのは、用い

表 3 霊長類の種分岐時間と祖先集団の有効大きさの推定値の比較

解析に用いた領域数		Kim <i>et al.</i> (2010)	Burgess & Yang (2008)	Satta <i>et al.</i> (2004)
		83	14,663	20 ¹⁾
H-C	X_{HC}	0.35	0.65	0.33
	Y_{HC}	0.41	0.41	0.45
H-G	X_{HG}	0.39	0.34	0.21
	Y_{HG}	0.55	0.65	0.55
H-O	X_{HO}	0.52	0.61	0.03
	Y_{HO}	1.3	1.5	1.4
H-OWN	X_{HOWN}	1.0	1.2	0.65
	Y_{HOWN}	2.4	2.7	2.6
H-NWM	X_{HNWN}	2.7	N.A.	0.51
	Y_{HNWN}	4.0	N.A.	5.0

註 1) ヒトとチンパンジーの比較のみ、58,158 の BES の配列を用いた。

図7 祖先集団多型 (x) と種分岐時間 (y) の推定値の比較。異なる3つの論文による推定値の比較。祖先集団多型 (x) の推定値では現存のヒト集団の体系の程度を1に、種分岐時間 (y) の推定値では、ヒトとチンパンジーの推定値を1とした時の相対値を表している。

た領域の数が小さいことに依る可能性が考えられる。

図7には、表3のデータを図示している。ここでは、祖先集団の多型 (x) については、現存のヒト集団の多型の程度を1として、また種分岐時間 (y) についてはヒトとチンパンジーの分岐を1とした時の相対的な割合で示している。Burgess & Yang (2008), Kim *et al.* (2010) の推定値に共通にみられる傾向として、種分岐時間 (y の推定値) が大きくなるにつれて祖先集団の多型の程度 (x の推定値) も大きくなる傾向が見られる。これが、推定方法に依る systematic な error に由来することではないことを明らかにするために、祖先集団の多型の程度を一定にした種分化のモデルを構築し、コンピュータシミュレーションにより生成された DNA 配列を元に、最尤法、MCMC 法を用いて、結果を比較したところ、シミュレーションデータでは、観測されたような y と x の推定値の間の相関は見られなかった。しかし、観測データに基づく推定値は、祖先集団の多型は分岐年代が大きくなるほど大きくなっていることを示している。旧世界ザル、あるいは新世界ザルとの共通祖先集団は、現在のヒト集団 (0.08%) の10倍から30倍もの多型を維持していたことになり、y の推

定値についてもヒトとチンパンジーの分岐を600万年とすると、ヒトと旧世界ザル、あるいは新世界ザルの分岐は、3600万年、あるいは6000年以上となる。これらの推定値は、化石のデータや生物地理学的な推定値とは必ずしも一致しない。この相違が最尤法の推定方法によるエラーで生じたのではないとしたら、生物学的な要因が存在するはずである。

集団の多型の程度が大きくなる理由としては、突然変異率の上昇や、祖先集団での集団構造などが考えられる。突然変異率の上昇は、y の推定値の上昇も説明できる。このことは、従来から指摘されている類人猿での突然変異率の低下 (hominoid slowdown) とも合致する。

今後は、突然変異率が時間により変化することを取り入れた最尤法の開発とその結果の尤度比較による類人猿での突然変異率の低下がデータをよりよく説明できるかどうかの検定もおこなっていききたい。また、本稿には、取り入れることができなかったが、テナガザルの含めた解析、原猿の祖先集団の大きさの推定なども行うことで、霊長類の系統解析に新たな情報を付け加えられると考える。

最後に、この最尤法の限界について触れておきたい。最尤法は、進化に中立な座位での塩基置換数を用いる。そのために、実際の塩

基の違いから、塩基置換の数の推定（補正）が正しく行われることが大事な条件となる。現在いろいろな補正の方法があるが、これらの方法はすべて塩基置換のモデルに依存する（Nei & Kumar, 2000）。どの塩基置換のモデルを使うかによって、塩基置換数が大きく変わる様になれば、塩基置換数の推定値の確からしさが損なわれるので、このような状態は、中立な座位での塩基置換数に基づく最尤法利用の限界といわざるをえない。Nei & Kumar (2000) の Figure 3.1 にあるように、塩基の違いの程度が、30% を超えるようになると、補正のモデルの影響が表れる。おそらく、霊長類目の種間比較では、塩基の違いはたかだか20%であることを考えると、霊長類目内の比較には問題ないものと思われる。

この総説の執筆の機会を与えてくださった古市剛史先生、最尤法の開発と解析に協力をいただいた高畑尚之先生、データ解析に協力して下さった現在ペンシルバニア州立大学にいる金慧琳さんに感謝する。

引用文献

- Burgess R, Yang Z 2008: Estimation of hominoid ancestral population sizes under Bayesian coalescent models incorporating mutation rate variation and sequencing errors. *Mol Biol Evol* 25: 1979-1994.
- Fujiyama A, Watanabe H, Toyoda A *et al.* 2002: Construction and analysis of a human-chimpanzee comparative clone map. *Science* 295: 131-134.
- Goodman M 1963: Serological analysis of the systematics of recent hominoids. *Human Biol* 35: 377-436.
- King MC, Wilson AC 1973: Evolution at two levels in humans and chimpanzees. *Science* 188: 107-116.
- Horai S, Hayasaka K, Kondo R, Tsugane K, Takahata N 1995: Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proc Natl Acad Sci USA* 92: 532-536.
- Kingman JFC 1982: The coalescent. *Stoch Proc Appl* 13: 235-248.
- Kim HL, Igawa T, Kawashima A, Satta Y, Takahata N 2010: Divergence, demography and gene loss along the human lineage. *Philos Trans R Soc Lond B Biol Sci* 365: 2451-2457.
- Nei M, Kumar S 2000: *Molecular Evolution and Phylogenetics*. Oxford University Press.
- Nuttall GHF 1904: *Blood Immunity and Blood Relationship*, Cambridge.
- Rogers J 1993: The phylogenetic relationships among Homo, Pan and Gorilla: a population genetics perspective. *J Hum Evol* 25: 201-215.
- Sarich VM, Wilson AC 1967: Immunological time scale for hominid evolution. *Science* 158: 1200-1203.
- Satta Y, Hickerson M, Watanabe H, OhUigin C, Klein J 2004: Ancestral population sizes and species divergence times in the primate lineage on the basis of intron and BAC end sequences. *J Mol Evol* 59: 478-487.
- Tajima F 1983: Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105: 437-460.

(Summary)

Primate Phylogeny and the Ancestral Polymorphism

Yoko SATTA

Department of Evolutionary Studies of Biosystems, The Graduate University for Advanced Studies

As genomic studies have progressed, it has become easier to get the nucleotide sequences at several nuclear loci from many species. Molecular phylogenetic studies have provided the branching (diverging) order and divergence time of many primates. Recent phylogenetic studies have revealed that the most closely related species to humans is the chimpanzee, not the gorilla or orangutan. However, there remain arguments about the time of divergence of many species, including the hominoids. In fact, estimates of the divergence time show a large variation among different loci and this may be due to a large extent on ancestral polymorphism. In this article we present the method to estimate the effective size of an ancestral population and the divergence time simultaneously. We showed both estimates in the lineage leading to extant humans. We will discuss the biological meaning and problems of estimating effective sizes of an ancestral population and divergence times in primate evolution.

Key words: polymorphism, species divergence, gene tree

颯田葉子 総合研究大学院大学先導科学研究科生命共生体進化学専攻
〒240-0193 神奈川県三浦郡葉山町湘南国際村

Yoko Satta Department of Evolutionary Studies of Biosystems, The Graduate University for Advanced Studies
Shonan Village, Hayama, Kanagawa
240-0193 Japan
e-mail: satta@soken.ac.jp