

真核生物の初期進化*

統計数理研究所 橋 本 哲 男
昭和大学** 中 村 嘉 宏
総合研究大学院大学*** 足 立 淳
統計数理研究所・総合研究大学院大学 長谷川 政 美

(1994 年 2 月 受付)

1. はじめに

現在の地球上には、われわれヒトから細菌類に至るまで、実に多種多様な生物が生息しており、その数は 3,000 万種にも及ぶと推定されている。これまでに調べられた限りでは、これらの生物の遺伝的な仕組みは基本的にはすべて共通なので、多種多様な生物が、もとをただせば同じ祖先から由来して進化してきたものであると考えられる。したがって、3,000 万種にも及ぶ地球上のあらゆる生物のそれぞれは、一つの巨大な系統樹のどこかに位置づけられるはずである。生物を系統樹の上に位置づけることは、現代生物学のなかでも重要な課題であり、進化の機構を明らかにする上でもまず第一に必要な作業である。さらに、ヒトの生物界における位置づけは、人類の思想・哲学とも深くかかわる問題である。このような分野は、生物系統学と呼ばれるが、これまでは主に化石や現生生物の形態学的な比較を通じて行なわれてきた。

地球の誕生は今から 46 億年前といわれているが、地球上の最初の生命の誕生はおおよそ 35 億年前と考えられている。われわれヒトの祖先をたどっていくと、最終的には 35 億年前の始原生物にまでさかのぼることになる。このような古い進化の歴史をたどるには、従来行なわれてきたような形態の比較では不十分である。この時代の生物は、細菌のような単細胞のものであり、かたちはどれも似たようなものであって、形態学的な研究からは系統に関する情報はほとんど得られないのである。

このような困難に立ち向かうべく登場したのが、DNA や蛋白質などの分子をもとにした分子系統学である。現在の生物がもっている DNA や蛋白質は、それぞれが長い進化の歴史の産物であり、その塩基配列やアミノ酸配列のなかに生物進化の歴史が刻まれている。従って、DNA や蛋白質を解読することによって、進化の歴史を明らかにしようというわけである。形態レベルでは、生きた化石といわれるように、環境が変わらなければ長い間ほとんど変化しなかった生物がいるかと思えば、逆に地質学的には極く短期間に急速な進化を遂げた生物がいるといったふうに、進化速度には一定性がない。ところが、分子レベルの進化では、自然淘汰にとって良くも悪くもない変化がほとんどであり、進化速度も形態の進化に比べるとはるかに一定に近いのである (Kimura (1983))。また形態の比較では、研究者の主観的な判断の入る余地が多かつ

* 本稿は、統計数理研究所 共同研究 (4-共研-57, 5-共研 A-64) の研究成果の一部をまとめたものである。

** 組換え DNA 実験室：〒142 東京都品川区旗の台 1-5-8.

*** 数物科学研究所 統計科学専攻：〒106 東京都港区南麻布 4-6-7.

たが、DNA や蛋白質の比較では客観的な基準に基づいた定量的な議論ができる。このように、進化の歴史を調べるのに、DNA や蛋白質の研究が有望であるという認識が近年になって急速に広まってきたのである。

分子系統学において系統樹推定の手掛かりを与えるのは、DNA における塩基置換や蛋白質におけるアミノ酸置換である。共通の祖先から分かれた後のそれぞれの系統における進化の過程で独立に置換が起こるので、生物種によって配列に違いが見られるわけである。いろいろな生物の配列を比較することによって、系統樹が推定されるのである。

ところで、進化において塩基置換やアミノ酸置換が起こるには、2つの段階が必要である。まず、個体レベルで生殖細胞のDNAに突然変異が起こらなければならない。しかしながら、突然変異が起こっただけでは進化とはいえない。単に変わり者が一個体生じたというだけである。進化というのは集団レベルの現象であり、そのためには、個体で起こった突然変異遺伝子が、繁殖を繰り返して世代を重ねるうちに次第に集団全体に広がり、従来あった遺伝子にとって換わって集団の大部分を占めるようにならなければならない。これを突然変異遺伝子の固定という。これによってはじめてDNAやそれにコードされている蛋白質が進化的な意味で変化したといえるのである。ここでまず、突然変異は予測不可能なランダムな現象である。また、遺伝子の固定も、ランダムな確率過程としてとらえることが妥当である (Kimura (1983))。従って、そのような過程の産物として得られているDNA塩基配列や蛋白質アミノ酸配列などのデータから、系統樹を推定するためには、確率モデルに基づいた統計的な方法が必要である。本論文では、極めて古い時代にまでさかのぼって、遠縁生物種を含む系統樹の推定を行なうが、このような問題を扱うには、DNAレベルよりも蛋白質レベルの解析の方が有用である。もちろん、蛋白質のアミノ酸配列はDNAにコードされているのだから、DNAの塩基配列は蛋白質のアミノ酸配列の情報を含むのであるが、蛋白質レベルで扱う方が、DNAレベルよりもより現実的なモデルを構築しやすいのである。すなわち、アミノ酸置換に関する経験的な遷移確率行列は、Dayhoff et al. (1978) や Jones et al. (1992) によって推定されているため、これらを用いることにより現実的なモデルの構築が可能となるが、DNAレベルでの塩基置換に関しては、これらに匹敵するような経験的遷移確率行列が未だ存在していない。

われわれは、蛋白質レベルでの系統樹の解析を行なうために、アミノ酸配列データから最尤法に基づいて分子系統樹を推定するための方法を開発した (Kishino et al. (1990), Adachi and Hasegawa (1992b))。本論文では、この方法の概要と「真核生物の初期進化」という生物学的にも重要な問題への応用例をレビューする。

2. 研究の背景

細胞内に核をもつ生物を真核生物といい、動物、植物、菌類 (キノコ、カビ、酵母の類) などの高等なものほかに、ミドリムシ、ゾウリムシ、アメーバなどの原生生物がこれに含まれる。一方、細菌のように核をもたない生物を原核生物というが、原核生物はさらに、大腸菌、枯草菌、藍藻などの真正細菌と、イオウ細菌、好塩細菌、メタン細菌などの古細菌とに分かれる。現在では、地球上の生物は全て、これら、真核生物、真正細菌、および古細菌の3つに大別されるものと考えられている。さらに、種々の遺伝子の比較研究から、真核生物は真正細菌よりも古細菌に近いことが明らかになってきた (Iwabe et al. (1989), Miyata et al. (1991))。

図1に典型的な真核生物の細胞 (真核細胞) の模式図を示した。真核細胞では、DNAは蛋白質と複合体を作り、染色体として、膜に囲まれた核内に存在している。ミトコンドリアは酸素呼吸によってエネルギーを得るための器官である。植物細胞には光合成を行なう器官としての葉

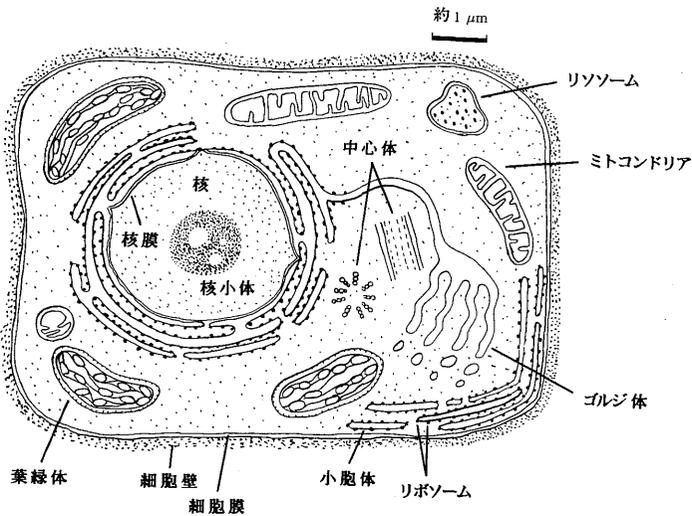


図1. 真核細胞の模式図。代表的な細胞内構造のみを示してある。細胞壁および葉緑体は、原則として植物細胞のみに存在する。

緑体がある。ゴルジ体は分泌器官である。小胞体は膜系で、その一部にリボソームが配列されており、この上でDNAの遺伝情報は蛋白質に翻訳される。リボソームは多くの蛋白質と数種のRNAの複合体であり、大小2つのサブユニットから構成されている。これらの細胞内小器官(オルガネラ)のうちで、ミトコンドリアと葉緑体は核のDNAとは異なる独自のDNAを持っており、自己複製をすることができる。ミトコンドリアと葉緑体のDNAにコードされている蛋白質やRNAの配列はともに真正細菌のものに近いことから、これらのオルガネラは、真正細菌が真核生物の祖先に細胞内共生することによって生じたものであると考えられている。

ところが、原生生物のなかには、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ものが存在する。たとえば、*Entamoeba histolytica* (赤痢アメーバ)や *Giardia lamblia* (ランブル鞭毛虫)の細胞にはミトコンドリアはない。もしミトコンドリアをもたない原生生物のうちのあるものが、真核生物の進化のごく初期に他の真核生物に至る系統から分岐したものであるとすれば、その生物は、ミトコンドリアの細胞内共生が起こる以前の真核生物の祖先型に近い生物である、ということになる。一方、ミトコンドリアをもたない原生生物の多くは寄生虫であるが、それらの中のあるものがミトコンドリアをもつ原生生物よりも後から分岐したものであるとすれば、その生物は寄生生活のなかで一度獲得したミトコンドリアを失ったのだとの可能性が出てくる。このように、ミトコンドリアをもたない真核生物の系統的な位置づけを明らかにすることは、真核生物の初期進化の過程を理解するうえで重要である (Cavalier-Smith (1987, 1989, 1991))。

従来、この分野の分子系統学的研究は、小サブユニットリボソームRNA (SrRNA)の塩基配列比較に基づいて行われてきた。Sogin et al. (1989), Sogin (1991)により距離行列法に基づいて推定されたSrRNA系統樹の一部を図2に示す。この図の中でミトコンドリアをもたない生物種は*で示してある。ミトコンドリアをもたない3生物種のうちで、真核生物の進化の最も早い時期にまず *Giardia lamblia*が、つづいて微孢子虫に属する *Vairimorpha necatrix*が分岐しているのに対し、*Entamoeba histolytica*は *Trypanosoma brucei* (トリパノソーマ)や *Euglena gracilis* (ミドリムシ)などのミトコンドリアを有する原生生物の分岐後に分かれた

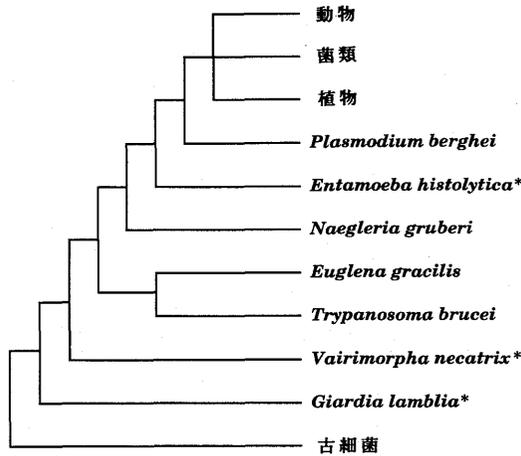


図2. Sogin らによる SrRNA の系統樹. Sogin et al. (1989) より改変したもの. 真核生物の中で*で示した生物種は, ミトコンドリアをもたないものである.

ことになっている. もしこの系統樹が正しいとすると, 前者はミトコンドリアの共生以前の真核生物の祖先型に近い生物であり, 後者は一度獲得したミトコンドリアを失ったものだということになる.

しかしながら SrRNA に基づく系統樹には問題点があると考えられる. それは, SrRNA の塩基組成が, ゲノム DNA の塩基組成の生物種間での偏りの影響を直接受けるため, 近縁の生物種間でも大きく偏ることがある, という点である. 表1 にいくつかの原生生物および古細菌の SrRNA の GC 含有量を示した. これによると, *Giardia lamblia* の GC 含有量は, 74.7% と高いのに対し, *Vairimorpha necatrix*, および *Entamoeba histolytica* のそれはそれぞれ, 37.4%, 38.3% であり, *Giardia* の約半分という極めて大きな偏りを示している. 系統樹推定に際しこのような偏りの影響を考慮に入れることは難しいため, SrRNA の系統樹が誤りを犯している可能性は高い. たとえば, 比較的 GC 含有量の高い古細菌の影響により見かけ上 *Giardia* が真核生物の系統樹の一番外側にきてしまっている, との可能性は否定できない.

表1. SrRNA の GC 含有量.

種名	GC %
原生生物	
<i>Plasmodium berghei</i>	37.1
<i>Entamoeba histolytica</i>	38.3
<i>Naegleria gruberi</i>	47.7
<i>Euglena gracilis</i>	57.0
<i>Trypanosoma brucei</i>	51.4
<i>Vairimorpha necatrix</i>	37.4
<i>Giardia lamblia</i>	74.7
古細菌	
<i>Sulfolobus solfataricus</i>	61.1
<i>Methanococcus vanniellii</i>	57.6
<i>Halobacterium volcanii</i>	56.9

われわれは、こうした点を指摘するとともに、ゲノム GC 含有量の顕著な偏りの影響を比較的受けにくいと考えられる蛋白質アミノ酸配列データを用いて、初期真核生物の分岐の過程を再検討する必要があるのではないかとの提案を行なった (Hasegawa and Hashimoto (1993)). 一方, Elongation factor 1 α (EF1 α), DNA dependent RNA polymerase III, など進化の過程で保存的な蛋白質のアミノ酸配列データを用いて実際の解析を行なうことにより, SrRNA に基づくこれまでの解析結果とは異なるいくつかの新しい知見を明らかにした (Hasegawa et al. (1992a), Hashimoto et al. (1992)). さらに, これらの保存的蛋白質のアミノ酸組成がゲノム DNA の塩基組成の大きな偏りの影響をほとんど受けないことを示し, 保存的蛋白質に基づく解析のほうが SrRNA に基づく解析よりも信頼しうる推定結果を与えるであろうことを示唆した (Hasegawa et al. (1992a), Hashimoto et al. (1992, 1994)).

ところが, 原生生物の保存的蛋白質に関するデータは, ミトコンドリアをもたないものを含め, 現在のところあまり多く報告されていない. そこでわれわれは, 昭和大学・組換え DNA 実験室との共同研究の下で, とくにミトコンドリアをもたない原生生物に注目し, 遺伝子解析により各種保存的蛋白質のアミノ酸配列データを得, それらを分子系統学的解析に供することを目的として, 実験, データ解析両面からの研究を進めてきた. 本稿では, EF1 α のアミノ酸配列データに基づいて解析した, *Giardia lamblia* の系統的位置づけに関する最近の研究成果を紹介する. *Giardia lamblia* は, ヒトを含む高等動物の腸管に寄生し, 下痢などの症状をもたらす寄生虫であり, ミトコンドリアをもたないことから, 近年, 真核生物の祖先型に近い生物ではないかとの注目を集めている. いくつかの鞭毛と, 全く等価と考えられている2つの核をもつ単細胞生物であり, 虫体は左右対称である (図3). 一方, EF1 α は, 蛋白質合成におけるペプチド鎖伸長反応の際に, アミノ酸を結合した tRNA をリボソームに受け渡す役割を担う蛋白質であり, 全ての生物にその相同蛋白質が存在することと, 進化速度が遅く保存的であることから, 真核生物全体にわたる生物種を含めた分子系統学的解析を行なう際には適切な材料であると考えられている (Hasegawa et al. (1993)). 以下, この生物の EF1 α に関する遺伝子解析の経過について簡単にふれた後, そのデータを含めた分子系統学的解析の結果を紹介する. なお, オリジナルデータは既報である (Hashimoto et al. (1994)).

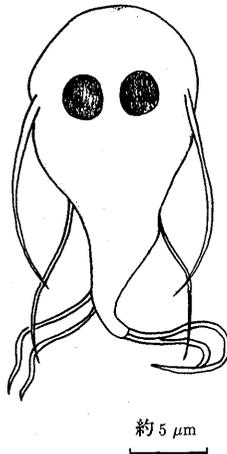


図3. *Giardia lamblia* (ランブル鞭毛虫) の模式図.

3. *Giardia lamblia* の EF1 α 遺伝子の構造と推定アミノ酸配列

Giardia lamblia Polish strain を仔牛血清, ウシ胆汁, L-cysteine を含む Diamond's 培地, TYI-S-33, にて細胞数が 1×10^6 /ml に達するまで培養し, 超遠心により total RNA を分離した後, mRNA を精製した. さらに, reverse transcriptase を作用させることにより, mRNA : cDNA duplex を合成し, これを PCR (polymerase chain reaction) 法による遺伝子増幅の鋳型とした. PCR 法を用いると, 鋳型とする微量の核酸 (DNA, RNA) をもとにして, 2つのオリゴヌクレオチドプライマーではさまれた特定の遺伝子領域を大量に増幅することができる. しかし, *Giardia lamblia* の EF1 α 遺伝子の塩基配列は全く未知であるため, 特異的なプライマーを設計することはできない. そこで, 他の真核生物および古細菌の EF1 α に関する既知アミノ酸配列を比較し, 分子の後半に相当する領域の中から高度に保存されている部分, PVGRVETG および DMRQTVAV (図6参照) を選び, これらのアミノ酸配列に対応する可能な全ての組み合わせの塩基配列を含む混合オリゴヌクレオチドプライマー, Primer A1 および Primer A2, を合成し, これらを用いて PCR 反応を行なった. 幸いある反応条件の下で, これら2つのプライマーではさまれた部分に対する約 0.5 kb (kb; kilo basepairs) の遺伝子断片を増幅することができた (図4(A)). この断片をゲルから切り出して精製し, 大腸菌のプラスミド DNA に組み込んだ後, そのプラスミド DNA の大量調製を行ない, ジデオキシ法により塩基配列を決定した. その結果, この増幅断片が確かに EF1 α をコードしていることが明らかとなった. 次に分子の先頭に近い領域中の高度保存部分, VIGHVDSGK (図6参照) に対して合成した混合オリゴヌクレオチド, Primer B1, と既に塩基配列の明らかとなった部分に対して合成した, Primer B2 (図5参照), とではさまれる部分を同様の条件で増幅し, 約 0.75 kb の断片を得 (図4(B)), 塩基配列を決定した. こうして EF1 α をコードしている遺伝子 (mRNA) 全体の9割以上に及ぶと考えられる全長 1,188 bp の塩基配列データを得, 遺伝暗号表に基づき 396 残基のアミノ酸配列を推定した (図5). 推定されたアミノ酸配列は, 他の真核生物の EF1 α と高い類似度を示し, 増幅遺伝子断片は確かに *Giardia* の EF1 α に対応するものであると考えられた. 図5に示した塩基配列のなかで, アミノ酸に対応する各コドンの3番目については, GC含有量が 99.5% ときわめて高く全 396 座位中で A および T は 1箇所ずつ (それぞれ 1,089 および 357 番目の塩基) でしか観測されなかった. この値は, 配列既知の *Giardia* の他の蛋白質と比較しても非常に極端なものである. たとえば, DNA dependent RNA polymerase III でのコドン3番目の GC含有量は 52.5%, Glyceraldehyde 3-phosphate dehydrogenase でのそれ

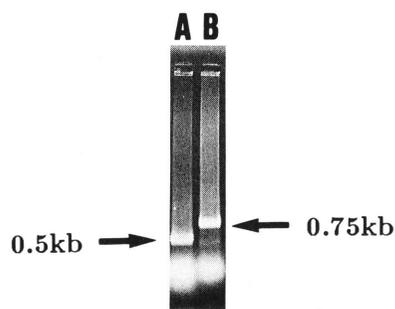


図4. PCR 法によって増幅された遺伝子断片. (A) Primer A1 および Primer A2 による増幅断片, (B) Primer B1 および Primer B2 による増幅断片. PCR 産物をアガロースゲル上で電気泳動し, エチジウムブロマイドで染色後写真撮影した.

(Val Ile Gly His Val Asp Ser Gly Lys)

Primer B1

TCG ACG CTG ACG GGC CAC CTC ATC TAC AAG TGC GGC GGG ATC GAC CAG CGC ACG ATC GAC	60
Ser Thr Leu Thr Gly His Leu Ile Tyr Lys Cys Gly Gly Ile Asp Gln Arg Thr Ile Asp	
GAG TAC GAG AAG CGC GCG ACG GAG ATG GGC AAG GGC TCC TTC AAG TAC GCG TGG GTC CTC	120
Glu Tyr Glu Lys Arg Ala Thr Glu Met Gly Lys Gly Ser Phe Lys Tyr Ala Trp Val Leu	
GAC CAG CTC AAG GAC GAG CGC GAG CGC GGG ATC ACG ATC AAC ATC GCG CTC TGG AAG TTC	180
Asp Gln Leu Lys Asp Glu Arg Glu Arg Gly Ile Thr Ile Asn Ile Ala Leu Trp Lys Phe	
GAG ACG AAG AAG TAC ATC GTC ACG ATC ATC GAC GCC CCG GGC CAC CGC GAC TTC ATC AAG	240
Glu Thr Lys Lys Tyr Ile Val Thr Ile Ile Asp Ala Pro Gly His Arg Asp Phe Ile Lys	
AAC ATG ATC ACG GGG ACG TCC CAG GCC GAC GTC GCG ATC CTC GTC GTC GCG GCG GGC CAG	300
Asn Met Ile Thr Gly Thr Ser Gln Ala Asp Val Ala Ile Leu Val Val Ala Ala Gly Gln	
GGC GAG TTC GAG GCC GGG ATC TCG AAG GAC GGC CAG ACG CGC GAG CAC GCG ACC CTT GCG	360
Gly Glu Phe Glu Ala Gly Ile Ser Lys Asp Gly Gln Thr Arg Glu His Ala Thr Leu Ala	
AAC ACG CTC GGG ATC AAG ACG ATG ATC ATC TGC GTC AAC AAG ATG GAC GAC GGC CAG GTC	420
Asn Thr Leu Gly Ile Lys Thr Met Ile Ile Cys Val Asn Lys Met Asp Asp Gly Gln Val	
AAG TAC TCG AAG GAG CGC TAC GAC GAG ATC AAG GGC GAG ATG ATG AAG CAG CTC AAG AAC	480
Lys Tyr Ser Lys Glu Arg Tyr Asp Glu Ile Lys Gly Glu Met Met Lys Gln Leu Lys Asn	
ATC GGC TGG AAG AAG GCC GAG GAG TTC GAC TAC ATC CCG ACG TCC GGC TGG ACC GGG GAC	540
Ile Gly Trp Lys Lys Ala Glu Glu Phe Asp Tyr Ile Pro Thr Ser Gly Trp Thr Gly Asp	
AAC ATC ATG GAG AAG TCC GAC AAG ATG CCC TGG TAC GAG GGC CCG TGC CTG ATC GAC GCG	600
Asn Ile Met Glu Lys Ser Asp Lys Met Pro Trp Tyr Glu Gly Pro Cys Leu Ile Asp Ala	
ATC GAC GGG CTC AAG GCC CCG AAG CGC CCG ACC GAC AAG CCC CTC CGC CTC CCG ATC CAG	660
Ile Asp Gly Leu Lys Ala Pro Lys Arg Pro Thr Asp Lys Pro Leu Arg Leu Pro Ile Gln	
GAC GTC TAC AAG ATC TCG GGC GTC GGG ACC GTC CCC GCG GGC CGC GTC GAG ACG GGC GAG	720
Asp Val Tyr Lys Ile Ser Gly Val Gly Thr Val Pro Ala Gly Arg Val Glu Thr Gly Glu	

(Pro Val Gly Arg Val Glu Thr Gly)

Primer A1

Primer B2

CTC GCG CCC GGG ATG AAG GTC GTC TTC GCC CCG ACG TCC CAG GTC TCG GAG GTC AAG TCC	780
Leu Ala Pro Gly Met Lys Val Val Phe Ala Pro Thr Ser Gln Val Ser Glu Val Lys Ser	
GTC GAG ATG CAC CAC GAG GAG CTC AAG AAG GCC GGG CCC GGG GAC AAC GTC GGC TTC AAC	840
Val Glu Met His His Glu Glu Leu Lys Lys Ala Gly Pro Gly Asp Asn Val Gly Phe Asn	
GTC CGC GGG CTC GCC GTC AAG GAC CTC AAG AAG GGC TAC GTC GTC GGG GAC GTG ACG AAC	900
Val Arg Gly Leu Ala Val Lys Asp Leu Lys Lys Gly Tyr Val Val Gly Asp Val Thr Asn	
GAC CCG CCC GTC GGC TGC AAG AGC TTC ACC GCC CAG GTC ATC GTC ATG AAC CAC CCG AAG	960
Asp Pro Pro Val Gly Cys Lys Ser Phe Thr Ala Gln Val Ile Val Met Asn His Pro Lys	
AAG ATC CAG CCC GGC TAC ACG CCC GTC ATC GAC TGC CAC ACC GCG CAC ATC GCG TGC CAG	1020
Lys Ile Gln Pro Gly Tyr Thr Pro Val Ile Asp Cys His Thr Ala His Ile Ala Cys Gln	
TTC CAG CTC TTC CTC CAG AAG CTC GAC AAG CGC ACG CTC AAG CCC GAG ATG GAG AAC CCG	1080
Phe Gln Leu Phe Leu Gln Lys Leu Asp Lys Arg Thr Leu Lys Pro Glu Met Glu Asn Pro	
CCC GAC GCA GGC CGC GGC GAC TGC ATC ATC GTC AAG ATG GTC CCC CAG AAG CCC CTG TGC	1140
Pro Asp Ala Gly Arg Gly Asp Cys Ile Ile Val Lys Met Val Pro Gln Lys Pro Leu Cys	
TGC GAG ACG TTC AAC GAC TAC GCG CCC CTC GGC CCG TTC GCC GTC CGC	1188
Cys Glu Thr Phe Asn Asp Tyr Ala Pro Leu Gly Pro Phe Ala Val Arg	

(Asp Met Arg Asn Thr Val Ala Val)

Primer A2

- 図 5. *Giardia lamblia* の EF1 α 遺伝子 (mRNA) の構造と推定アミノ酸配列。アミノ酸の表現は 3 文字表記法による。PCR 反応に用いたプライマーを下線で示してある。() 内のアミノ酸配列は、混合オリゴヌクレオチドプライマーの設計のために用いた配列である。今回配列決定を行なった領域の最も上流の塩基からの番号を右端に示してある。

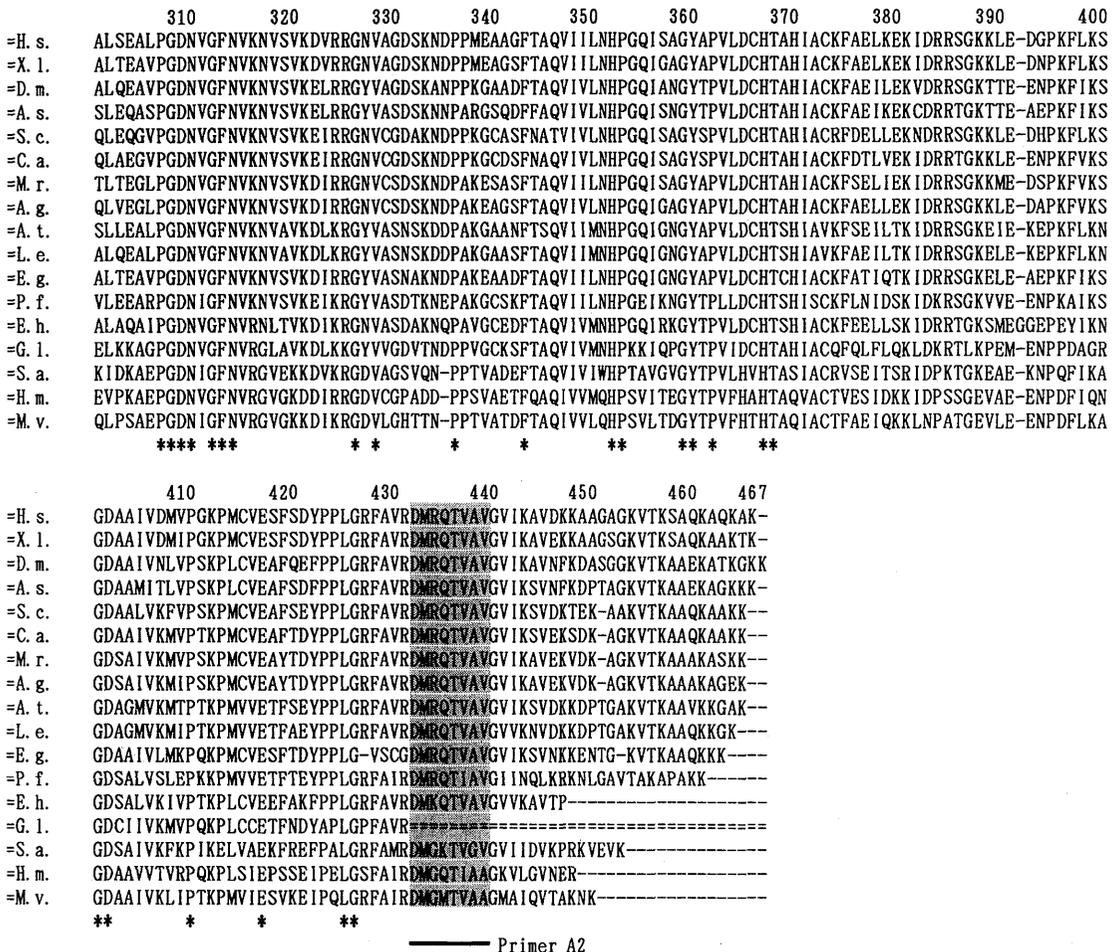


図6. EF1 α アミノ酸配列のアライメント。アミノ酸の表現は1文字表記法によった。アライメント上部にアミノ酸座位の番号を示してある。アライメント内部の-はギャップを、=は今回の *Giardia* データの未決定部分を示す。アライメント下部の*印は、その座位が全て同一のアミノ酸で占められていることを示す。PCR 反応に用いた混合オリゴヌクレオチドプライマーを設計した領域をアライメント下部に下線で示し、その部分に対応するアミノ酸配列には影をつけてある。G1~G4 は、GTP 結合蛋白質ファミリーに属する蛋白質で共通に保存されている領域である。アライメント左部分に示した=の右の記号に対応する生物種名は以下のとおりである：[動物] H.s., *Homo sapiens* (ヒト)；X.l., *Xenopus laevis* (アフリカツメガエル)；D.m., *Drosophila melanogaster* (ショウジョウバエ)；A.s., *Artemia salina*, [菌類] S.c., *Saccharomyces cerevisiae* (酵母)；C.a., *Candida albicans*；M.r., *Mucor racemosus*；A.g., *Absidia glauca*, [植物] A.t., *Arabidopsis thaliana*；L.e., *Lycopersicon esculentum* (トマト), [原生生物] E.g., *Euglena gracilis* (ミドリムシ)；P.f., *Plasmodium falciparum* (マラリア病原虫)；E.h., *Entamoeba histolytica* (赤痢アメーバ)；G.l., *Giardia lamblia* (ランブル鞭毛虫), [古細菌] S.a., *Sulfolobus acidocaldarius* (イオウ細菌)；H.m., *Halobacterium marismortui* (好塩細菌)；M.v., *Methanococcus vannielii* (メタン細菌)。それぞれのオリジナルデータの文献は、Hashimoto et al. (1994) に示してある。

は77.5%である。

次に *Giardia* の EF1 α アミノ酸配列を他の真核生物13種および古細菌3種の既報 EF1 α 配列と比較し、アライメント、すなわちアミノ酸座位の対応づけ、を行なった。図6に *Giardia* を含む全17生物種からなる EF1 α のアライメントを示した。分類群および種名はキャプションに示してある。EF1 α は、非常に保存的な蛋白質であるため、進化的に遠い生物種、たとえば、ヒト (*Homo sapiens*) と古細菌との間でさえも同一アミノ酸座位が50%以上にのぼる。そのため、挿入もしくは欠失(ギャップ)となる座位も比較的少なく、分子の先頭(N末端側)および終わり(C末端側)の部分を除けばアライメントにはほとんど曖昧さを伴わなかった。とくに、N末端側に存在するG1からG4までのモチーフは、全ての真核生物、古細菌を通じて高度に保存されている。この部分は、EF1 α がアミノ酸を結合したtRNAをリボソームに受け渡す際に必要なエネルギーを供給する物質、GTP、が結合する領域であるため、機能的な重要性から高度に保存的になっているものと考えられる(Bourne et al. (1991))。このアライメントから、ギャップの生じている座位を取り除き、合計382座位を以下の分子系統学的解析に用いた。

4. 蛋白質のアミノ酸配列データに基づく分子系統樹の最尤推定

進化における遺伝子配列の変化やそれに伴う蛋白質アミノ酸配列の変化には確率過程としての側面が強い。そのため、遺伝子・蛋白質の進化、ひいては生物の系統進化の道すじをたどるためには、確率モデルに基づいたデータ解析が必要となる。蛋白質のアミノ酸配列データから最尤法によって進化系統樹を推定するための方法は、Kishino et al. (1990) によって定式化された。以来われわれのグループは、さまざまな現実的問題に対するデータ解析を通じて、モデルの比較、改良、評価を行ってきた(Hasegawa et al. (1992b), Adachi and Hasegawa (1992a), Adachi et al. (1993), Hashimoto et al. (1993), Hasegawa et al. (1993))。

例として、図7(a)のように1~5の5系統からなる根なし系統樹を考える。現存する5つの生物種(系統)のある蛋白質に関するアミノ酸配列データから、5つの系統の進化的位置関係を推定する問題を考える。データとして与えられた現存生物のアミノ酸配列がアミノ酸置換のある確率モデルに基づいて実現する確率をパラメータの関数としてみたものを尤度という。図7(b)に示すように、5つの系統に関して考える全ての系統樹のトポロジー(モデル)は15通りある。これらのそれぞれについて、尤度が最大となるように各枝におけるアミノ酸置換数(枝の長さ)などのパラメータを推定(最尤推定)する。こうして得られる各系統樹の尤度を比較して、その値が一番高いものを真の系統樹のもっとも良い候補(最尤系統樹)として選択する。

いま、図7(a)における各枝の長さを未知パラメータとし、

$$\theta = (t_1, t_2, \dots, t_7)^T$$

とおく。各分岐点から次の分岐点(または枝の先端)への進化は独立に起こるものと仮定する。時間 t の間にアミノ酸 i からアミノ酸 j に置換が起こる確率(遷移確率)を $P_{ij}(t)$ とする可逆な定常マルコフ過程を考えると、ある座位 h でアミノ酸の観察値が、

$$X_h = (x_{1h}, x_{2h}, x_{3h}, x_{4h}, x_{5h})^T$$

となる確率、すなわちある座位 h における尤度は、

$$f(x_{1h}, x_{2h}, x_{3h}, x_{4h}, x_{5h} | \theta) = \sum_i \sum_j \sum_k \pi_i P_{ix_1}(t_1) P_{ix_2}(t_2) P_{ij}(t_6) P_{jx_3}(t_3) \\ \times P_{jk}(t_7) P_{kx_4}(t_4) P_{kx_5}(t_5) \quad (i, j, k = 1, \dots, 20)$$

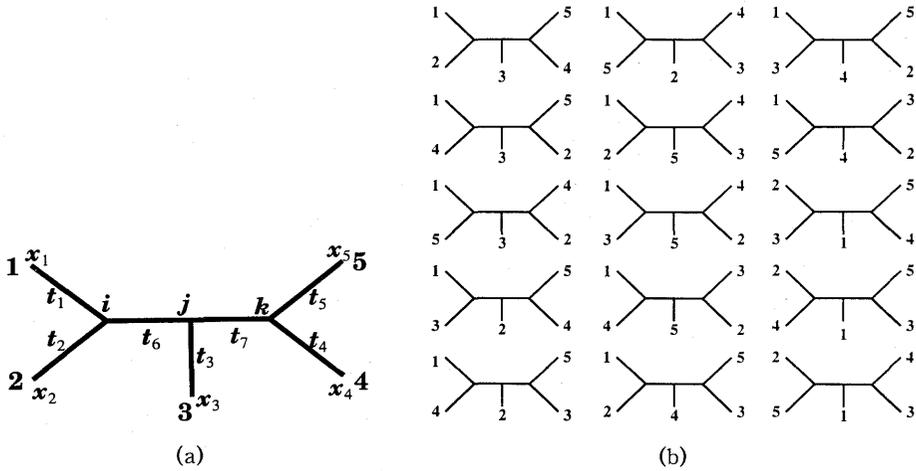


図7. (a) 5つの系統に対する根なし系統樹。1~5: 現存生物種, $x_1 \sim x_5$: 各座位における現存生物種のアミノ酸の観測値, i, j, k : 祖生物種のアミノ酸の状態, $t_1 \sim t_7$: 枝の長さ。
 (b) 5つの系統に対する15通りの系統樹のトポロジー。

と表わせる。ここで、 $\pi_i (i=1, \dots, 20)$ は各アミノ酸の組成値である。さらに、各座位が互いに独立に同一の確率法則（独立同分布）にしたがって進化すると仮定すると、 n 個の座位に対する全尤度 L は各尤度の積、すなわち、

$$L = \prod_{h=1}^n f(X_h | \theta)$$

となる。よって、この対数尤度、

$$\log L = l(\theta | X) = \sum_{h=1}^n \log f(X_h | \theta)$$

を最大にするように θ を推定すればよいことになる。

現在われわれは、遷移確率に対するモデルとして以下の6つを仮定し、パラメータ数の異なるモデル間の比較はAICに基づいて評価している。

- 1) Dayhoff et al. (1978) によって推定された経験的な遷移確率行列を用いる。これは主としてヘモグロビンとチトクロムcの配列データから推定されたものである (Kishino et al. (1990), Adachi and Hasegawa (1992b)) (Dayhoff モデル)。
- 2) 解析の対象とするデータセットのアミノ酸組成値により Dayhoff らの遷移確率行列を計算し直したものをを用いる (Dayhoff (F) モデル)。
- 3) Jones et al. (1992) によってより多くのアミノ酸配列データに基づいて推定された遷移確率行列を用いる (JTT モデル)。
- 4) 解析の対象とするデータセットのアミノ酸組成値により Jones らの遷移確率行列を計算し直したものをを用いる (JTT (F) モデル)。
- 5) 微小時間 dt の間にアミノ酸 i がアミノ酸 j に置き換わる確率を、 $P_{ij}(dt) = u\pi_j dt$ (ただし、 u は定数、 π_j は j の組成値) とおくモデルで、アミノ酸の遷移確率は「行き先」の組成値 π_j のみに依存するものと仮定する (Hasegawa et al. (1992b)) (Proportional モデル)。

ル)。

- 6) さらに単純に $P_{ij}(dt) = u(1/20)dt$ とおくモデルで、アミノ酸置換の起こりやすさは20種のアミノ酸相互間で全て同等であると仮定する (Hasegawa et al. (1992b)) (Poissonモデル)。

これらの解析を行なうためのプログラムは、PROTML という名前でパッケージ化されており (Adachi and Hasegawa (1992b)), 広く国内外に配布されている。

5. EF1 α のアミノ酸配列データからみたミトコンドリアをもたない真核生物 *Giardia lamblia* の系統的位置

EF1 α のデータに基づき、*Giardia lamblia* の真核生物内部での系統的位置づけをするために、最尤法による解析を行なった。解析に用いた生物種は、図6のアライメントに示した17種である。全17種に対する可能な系統樹のトポロジーは膨大な数にのぼるため、関係が明らかである種間については、あらかじめその関係を仮定するものとし、最終的に以下の6つの系統に絞って解析を行なった。すなわち、1) *Giardia lamblia* (ランブル鞭毛虫)、2) *Entamoeba histolytica* (赤痢アメーバ)、3) *Plasmodium falciparum* (マラリア病原虫)、4) *Euglena gracilis* (ミドリムシ)、5) 高等真核生物、および6) 古細菌の6系統である。1)~4) は原生生物に属する生物種である。5) の高等真核生物については、Hasegawa et al. (1993) の解析結果に基づき、動物と菌類が近いという関係、すなわち(植物、(菌類、動物))を、6) の古細菌内部の関係については、Miyata et al. (1991) により、(イオウ細菌、(好塩細菌、メタン細菌))という関係をそれぞれあらかじめ仮定した。さらに、動物および菌類内部の関係についても Hasegawa et al. (1993) に従った。なお古細菌は、真核生物の系統樹の根もとを決めるために用いたアウトグループである。以上6系統に対する可能な系統樹のトポロジーは全部で105通りあり、そのそれぞれについて、前述の6つの確率モデルによる枝長の最尤推定を行ない、得られた対数尤度に基づいて系統樹間の比較を行なった。

105通りの系統樹のうち *Giardia lamblia* が真核生物の進化の最も早い時期に分岐したとする15通りに関する解析結果を表2に示した。最尤系統樹の対数尤度を()内に示してある。各系統樹の対数尤度は、最尤系統樹のそれとの差として表現しており、 \pm はその標準誤差(SE)である (Kishino and Hasegawa (1989))。P_iの欄には各系統樹に対するブートストラップ確率を示してある。これは、解析に用いた382個の座位から382個の標本をランダムに復元抽出して仮想的なデータセットを多数回(10,000回)作って最尤推定を繰り返したとき、各系統樹が、最尤系統樹として選ばれる頻度である。実際に全てを実行すれば、膨大な計算時間が必要となるので、ここではRELL法という簡便法 (Kishino et al. (1990)) によっている。この方法は、各座位をリサンプリングして最尤推定を繰り返す代わりに、オリジナルデータセットに基づいて計算された各座位の尤度をリサンプリングしてその都度全体の尤度を計算するという単純なものであるが、最尤推定を繰り返す方法の非常に良い近似となっていることが確かめられている (Hasegawa and Kishino (1994))。

表2の解析結果より、JTTおよびJTT(F)モデルでは系統樹1が、Dayhoff, Dayhoff(F), Proportional, Poissonの4つのモデルでは系統樹2が最尤系統樹となった。系統樹1は、真核生物の系統のなかで最も早い時期にまず *Giardia* が分岐し、続いて *Plasmodium*, *Entamoeba* の順に分岐が起こったとするものであり、系統樹2は、系統樹1において *Entamoeba* と *Plasmodium* の順番を入れ替えたものである。JTT, JTT(F)モデルにおける系統樹2、および

表2. 古細菌をアウトグループとしたときの真核生物5系統の関係。

系統樹のトポロジー	Dayhoff モデル		Dayhoff(F) モデル		JTT モデル		JTT(F) モデル		Proportional モデル		Poisson モデル	
	Δl_i	P_i	Δl_i	P_i	Δl_i	P_i	Δl_i	P_i	Δl_i	P_i	Δl_i	P_i
1.(G,(P,(En,(Eu,H))))	-0.6 ± 8.2	.3310	-2.4 ± 7.5	.2282	(-6232.6)	.4392	(-6223.5)	.3604	-8.8 ± 10.0	.1285	-8.7 ± 10.6	.1431
2.(G,(En,(P,(Eu,H))))	(-6239.4)	.4088	(-6222.4)	.4285	-2.1 ± 8.4	.3002	-0.7 ± 8.1	.3237	(-6676.3)	.6924	(-6842.3)	.6898
3.(G,((En,P),(Eu,H)))	-4.6 ± 7.0	.0918	-3.5 ± 7.4	.1451	-5.2 ± 7.6	.0838	-3.7 ± 7.3	.0974	-9.9 ± 10.0	.0946	-11.0 ± 10.1	.0687
4.(G,(Eu,((En,P),H)))	-14.8 ± 10.7	.0111	-13.2 ± 11.2	.0232	-12.6 ± 10.3	.0269	-11.3 ± 10.4	.0333	-25.0 ± 14.9	.0141	-23.7 ± 14.7	.0188
5.(G,(P,(Eu,(En,H))))	-15.6 ± 13.6	.0353	-18.6 ± 13.7	.0261	-14.2 ± 9.4	.0375	-15.5 ± 10.1	.0318	-32.3 ± 16.5	.0062	-27.5 ± 16.4	.0140
6.(G,((Eu,(En,P)),H))	-15.9 ± 10.4	.0014	-14.5 ± 10.9	.0020	-14.3 ± 10.1	.0016	-13.3 ± 10.1	.0018	-27.9 ± 14.2	.0000	-26.9 ± 13.9	.0001
7.(G,(En,((P,Eu),H)))	-12.8 ± 7.5	.0163	-13.8 ± 8.2	.0167	-15.7 ± 12.0	.0104	-15.3 ± 12.0	.0093	-20.4 ± 9.5	.0045	-18.4 ± 9.1	.0064
8.(G,(En,(Eu,(P,H))))	-13.7 ± 7.1	.0012	-14.7 ± 7.7	.0015	-16.5 ± 11.8	.0008	-16.1 ± 11.8	.0007	-20.6 ± 9.4	.0021	-18.4 ± 9.0	.0054
9.(G,(P,((En,Eu),H)))	-18.9 ± 12.8	.0017	-22.3 ± 12.8	.0014	-17.2 ± 8.5	.0016	-19.2 ± 9.0	.0007	-34.5 ± 16.0	.0009	-29.2 ± 15.9	.0040
10.(G,(Eu,(P,(En,H))))	-23.5 ± 12.8	.0002	-24.6 ± 13.2	.0006	-23.2 ± 11.6	.0002	-23.3 ± 12.0	.0003	-39.3 ± 16.6	.0005	-35.3 ± 16.5	.0009
11.(G,((P,Eu),(En,H)))	-23.6 ± 12.8	.0005	-24.7 ± 13.2	.0002	-23.4 ± 11.6	.0001	-23.6 ± 11.9	.0002	-40.0 ± 16.0	.0001	-37.0 ± 15.9	.0001
12.(G,(Eu,(En,(P,H))))	-29.1 ± 11.3	.0000	-29.5 ± 11.8	.0000	-27.8 ± 11.2	.0000	-27.9 ± 11.4	.0000	-43.9 ± 15.6	.0000	-40.6 ± 15.4	.0000
13.(G,((P,(En,Eu)),H))	-28.8 ± 11.4	.0000	-30.0 ± 11.7	.0000	-28.1 ± 10.8	.0001	-28.9 ± 10.9	.0000	-43.3 ± 15.5	.0000	-40.3 ± 15.3	.0000
14.(G,((En,Eu),(P,H)))	-28.8 ± 11.4	.0000	-30.0 ± 11.7	.0000	-28.1 ± 10.8	.0000	-28.9 ± 10.9	.0000	-43.3 ± 15.5	.0000	-40.3 ± 15.3	.0000
15.(G,((En,(P,Eu)),H))	-29.4 ± 11.1	.0000	-30.1 ± 11.5	.0000	-29.1 ± 11.0	.0000	-29.5 ± 11.2	.0000	-45.4 ± 14.4	.0000	-43.1 ± 14.2	.0000
合計		.8993		.8735		.9024		.8596		.9439		.9513

トポロジーの表現に用いた記号は, G, *Giardia*; En, *Entamoeba*; P, *Plasmodium*; Eu, *Euglena*; および H, 高等真核生物である。

105通りの可能な系統樹の組み合わせのうち, *Giardia*の分岐が最も早いとするもの15通りのみを示してある。

Δl_i は系統樹*i*の対数尤度と最尤系統樹のそれとの差, ±はその差に対する1SEを表わす。最尤系統樹の対数尤度は()内に示してある。

ブートストラップ確率(P_i)は, Kishino et al. (1990)のRELL法に基づき, 10,000回のリサンプリングによって推定した。

Dayhoff, Dayhoff(F) モデルにおける系統樹1の対数尤度は、それぞれの最尤系統樹の対数尤度とほとんど差がなく、その差も全て1SEの範囲内に入っている。これは、系統樹1および2に対するブートストラップ確率の値が、それほど大きく異なっていないことに対応している。Proportional, Poisson モデルでは、系統樹1と最尤系統樹(系統樹2)の差はやや大きくなっているが、いずれも1SEの範囲内に収まっている。系統樹3は、*Giardia*の次に*Entamoeba*と*Plasmodium*の共通祖先が分岐したというものであるが、この対数尤度もPoissonモデル以外のモデルにおいては最尤系統樹のそれとの間で1SEを越える差を示していない。これに対し、表2に示していないものも含めて上記3つ以外の全ての系統樹の対数尤度は、いずれのモデルにおいても最尤系統樹のそれと1SEを越える差を示している。これらのことより、上記3つの系統樹の中に真の系統樹が存在する可能性は高いと考えられる。

表3には、最尤系統樹に対するAIC値をモデル別に示した。AICの規準でみると、JTTモデルが最も良く、次はDayhoffモデルである。さらに、わずかの差でDayhoff(F), JTT(F)の順となっている。JTT(F), Dayhoff(F), およびProportionalモデルでは、解析に用いたデータからアミノ酸組成値を推定しているためパラメータ数が19個多くなっている。表2の対数尤度でみた場合、JTT(F), Dayhoff(F)モデルの方が、JTT, Dayhoffモデルよりも適合が良くなっているが、表3のAIC値からみると、パラメータを増やしたことに對するペナルティーを相殺できるほどには良くなっていないということになる。おそらくこれは、今回解析に用いたEF1 α のアミノ酸組成値が、JTT, Dayhoffモデルにおける経験的な組成値とそれほど大きく異なっていないことを反映しているのであろう。Proportional, Poisson両モデルでは、上記4つのモデルよりもはるかに適合が悪い。これは、経験的なアミノ酸置換行列をモデルとする場合に比べて、モデルが単純すぎるためであると思われる。それにもかかわらず、経験的なモデルと比べて、表2の各系統樹間での対数尤度の値の動き方に大差はない。系統樹のトポロジーの選択という観点からみて、各種のモデル間でその判断に大差がないという点は重要である。

一方、表2に示した15通りの系統樹に対するブートストラップ確率の合計は、いずれのモデルにおいても85%を越えている。これは、今回解析に用いた原生生物のなかで、*Giardia*が最も早く分岐したとする可能性の高いことを示唆している。しかし、*Giardia*の次の分岐が、*Entamoeba*であるか*Plasmodium*であるかについては、系統樹1, 2, 3相互の対数尤度の差およびそれらに対するブートストラップ確率の値からみて全く明らかではない。前述のSoginらのSrRNAの系統樹(図2)と比較してみると、*Giardia*の分岐が最も早いことは、SrRNAの解析結果と今回のEF1 α の結果とで同じであるが、*Euglena gracilis*の位置づけが異なっている。SrRNAの系統樹では、*Euglena*は*Entamoeba*や*Plasmodium*よりも先に分岐したことになる。今回のわれわれの解析でその可能性に對する系統樹は、表2の系統樹4, 10, および12であるが、これらの系統樹に対するブートストラップ確率の合計はいずれのモデルによる解析においても、5%未満であり、*Giardia*の次に*Euglena*が分岐したとの可能性は低い。

SoginらのSrRNA系統樹にはその信頼性を示す指標が添えられていないため、この食い違いについてはっきりしたことはいえないが、SrRNAのGC含有量の原生生物種間での極端な偏り(表1)のことを考慮すると、われわれのEF1 α の解析結果の方が信頼しうものと思われる。図8(a)は、今回の解析で用いた原生生物4種のEF1 α 382座位の塩基組成値を各コドン別に示したものである。コドン3番目についてみると、*Giardia*のGC含有量は非常に高い(99%)のに対し、*Entamoeba*, *Plasmodium*のそれは、それぞれ19%, 22%と低く、極めて大きな偏りを示している。遺伝暗号表からも明らかなように、コドン3番目の変異がアミノ酸の変異をもたらす場合はあまり多くないため、3番目の塩基組成値はゲノムDNAの塩基組成値の影響を直接受けて大きく偏るものと考えられる。ところが、コドン1番目, 2番目の塩基組成

表 3. 最尤系統樹に対するモデル別 AIC 値.

モデル	最尤系統樹	AIC	Δ AIC
JTT	系統樹 1	12,527.2	0
JTT (F)	系統樹 1	12,547.1	+19.9 \pm 21.9
Dayhoff	系統樹 2	12,540.9	+13.7 \pm 39.9
Dayhoff (F)	系統樹 2	12,544.9	+17.7 \pm 42.4
Proportional	系統樹 2	13,452.5	+925.3 \pm 90.8
Poisson	系統樹 2	13,746.6	+1,219.4 \pm 100.8

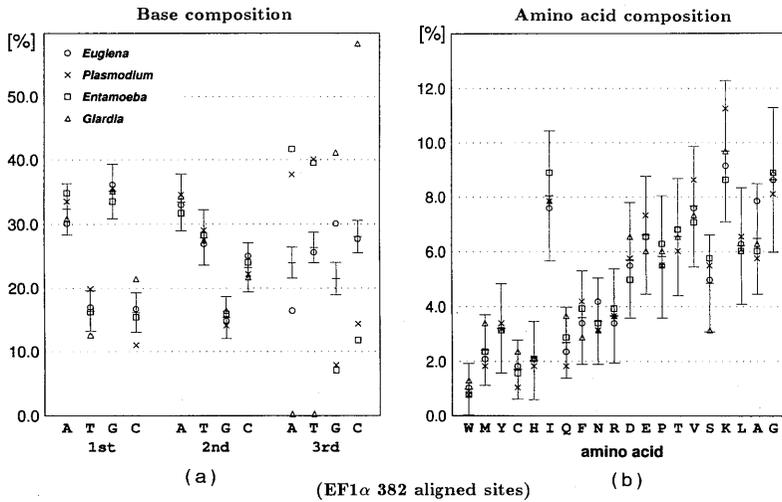


図 8. 原生物 4 種の EF1 α の (a) 塩基組成および (b) アミノ酸組成の分布. 多項分布の仮定のもとでの平均と 2 倍の標準誤差を \pm により幅をつけて示した (Hasegawa and Kishino (1989) 参照).

値については、1 番目の C および T を除けば、誤差の範囲内に収まっている。とくにコドン 2 番目の偏りは小さい。2 番目の変異は物理化学的性質の異なるアミノ酸への変異をもたらすため、強い機能的制約がかかり、保存的になっているものと考えられる。このことは、図 8(b) に示した 4 生物種の EF1 α 382 座位のアミノ酸組成値の分布からも明らかである。全てのアミノ酸に関して、4 生物種の組成値は誤差の範囲内に収まっており、ゲノム DNA の塩基組成値の偏りは、EF1 α のアミノ酸組成値にほとんど影響していない。さらに、DNA dependent RNA polymerase III, Glyceraldehyde 3-phosphate dehydrogenase, Elongation factor 2 などわれわれがこれまでに解析してきた他の保存的蛋白質についても同様の事実が明らかとなっている。これに対し、SrRNA はアミノ酸配列をコードしていないため、おそらくその塩基組成はゲノム DNA の塩基組成の偏りの影響を受けやすいであろう。したがって、保存的蛋白質に基づく解析の方が、これまでの SrRNA に基づく解析よりも頑健な推定結果を与えるものと考えられる。

以上の解析結果に基づいて JTT モデルにより推定した真核生物全体の系統樹を図 9 に示した。前述のように *Entamoeba* と *Plasmodium* の分岐の順番については、はっきりしたことがいえないため、その部分については多分岐の関係を用いている。もし、ミトコンドリアの細胞内

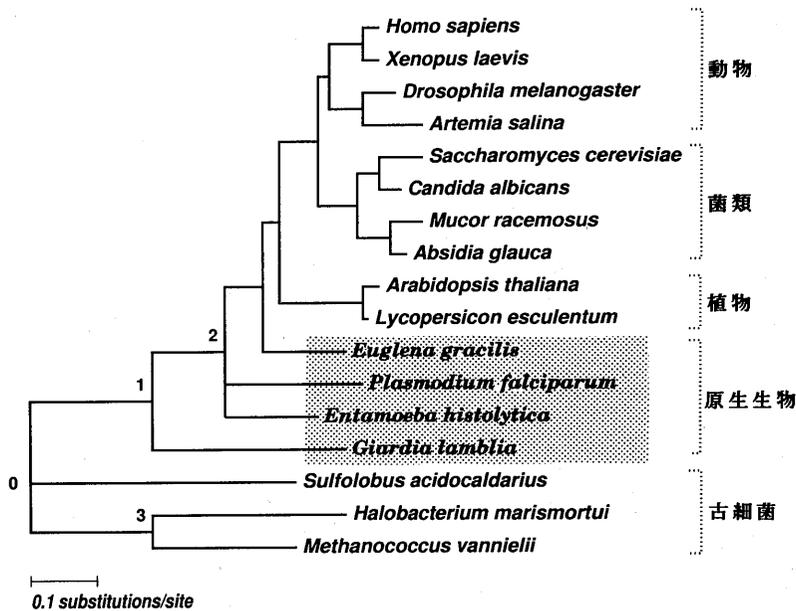


図9. EF1 α による真核生物全体の系統樹(JTTモデル). 系統樹の根が, 0と1の間にあるか(Miyata et al. (1991)), 0と3の間にあるか(Rivera and Lake (1992))はまだはっきりしていない.

共生が, *Giardia*の分岐の後, すなわち図中の2に至る枝, で起こったものであるとすれば, その後に分岐した *Entamoeba* は一度獲得したミトコンドリアを失ったものということになる. 一方, 表2の系統樹2が正しくて, *Entamoeba*の分岐後にミトコンドリアが共生したものであるとすれば, *Giardia*, *Entamoeba*はともに, ミトコンドリアの共生以前の真核生物の祖先型に近い生物である, ということになる. さらに, ミトコンドリアを有するが *Giardia*よりも先に分岐するような生物の存在が明らかとなれば, *Giardia*, *Entamoeba*はともに, 寄生生活の中で一度獲得したミトコンドリアを失ったことになる. 残念ながら, 現在のところミトコンドリアをもたないものを含めて原生生物に関する保存的蛋白質のデータはあまり多くないため, 現時点でこれらのうちどの可能性が高いのかを明らかにすることはできない. 今後, Elongation factor, RNA polymerase, ATPaseなど各種保存的蛋白質のデータの蓄積を待って, この問題をより総合的に判断する必要がある.

謝 辞

実験技術面でのご指導, ご援助をいただいた, 東京工業大学生命理工学部・岡田典弘, 広島大学原爆放射能医学研究所・大鷹英子, 昭和大学医学部・岡本謙一, 同歯学部・後藤延一の各教授, および各研究室の先生方, 院生, 技術員の方々に深謝いたします.

参 考 文 献

- Adachi, J. and Hasegawa, M. (1992a). Amino acid substitution of proteins coded for in mitochondrial DNA during mammalian evolution, *Japanese Journal of Genetics*, **67**, 187-197.

- Adachi, J. and Hasegawa, M. (1992b). MOLPHY: programs for molecular phylogeny I—PROTML: maximum likelihood inference of protein phylogeny, *Computer Science Monographs*, No. 27, Institute of Statistical Mathematics, Tokyo.
- Adachi, J., Cao, Y. and Hasegawa, M. (1993). Tempo and mode of mitochondrial DNA evolution in vertebrates at the amino acid sequence level: rapid evolution in warm-blooded vertebrates, *Journal of Molecular Evolution*, **36**, 270-281.
- Bourne, H.R., Sanders, D.A. and McCormick, F. (1991). The GTPase superfamily: conserved structure and molecular mechanism, *Nature*, **349**, 117-127.
- Cavalier-Smith, T. (1987). Eukaryotes with no mitochondria, *Nature*, **326**, 332-333.
- Cavalier-Smith, T. (1989). Archaeobacteria and archezoa, *Nature*, **339**, 100-101.
- Cavalier-Smith, T. (1991). The evolution of cells, *Evolution of Life: Fossils, Molecules, and Culture* (eds. S. Osawa and T. Honjo), 271-304, Springer, Tokyo.
- 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, Suppl. 3 (ed. M.O. Dayhoff), 345-352, National Biomedical Research Foundation, Washington D.C.
- Hasegawa, M. and Hashimoto, T. (1993). Ribosomal RNA trees misleading?, *Nature*, **361**, p. 23.
- Hasegawa, M. and Kishino, H. (1989). Heterogeneity of tempo and mode of mitochondrial DNA evolution among mammalian orders, *Japanese Journal of Genetics*, **64**, 243-258.
- Hasegawa, M. and Kishino, H. (1994). Simple methods for estimating bootstrap probability of a maximum likelihood tree, *Molecular Biology and Evolution*, **11**, 142-145.
- Hasegawa, M., Hashimoto, T. and Adachi, J. (1992a). Origin and evolution of eukaryotes as inferred from protein sequence data, *The Origin and Evolution of the Cell* (eds. H. Hartman and K. Matsuno), 107-130, World Scientific, Singapore.
- Hasegawa, M., Cao, Y., Adachi, J. and Yano, T. (1992b). Rodent polyphyly?, *Nature*, **355**, p. 595.
- Hasegawa, M., Hashimoto, T., Adachi, J., Iwabe, N. and Miyata, T. (1993). Early branchings in the evolution of eukaryotes: ancient divergence of *Entamoeba* that lacks mitochondria revealed by protein sequence data, *Journal of Molecular Evolution*, **36**, 380-388.
- Hashimoto, T., Adachi, J. and Hasegawa, M. (1992). Phylogenetic place of *Giardia lamblia*, a protozoan that lacks mitochondria, *Endocytobiosis and Cell Research*, **9**, 59-69.
- Hashimoto, T., Otaka, E., Adachi, J., Mizuta, K. and Hasegawa, M. (1993). The giant panda is closer to a bear, judged by α - and β -hemoglobin sequences, *Journal of Molecular Evolution*, **36**, 282-289.
- Hashimoto, T., Nakamura, Y., Nakamura, F., Shirakura, T., Adachi, J., Goto, N., Okamoto, K. and Hasegawa, M. (1994). Protein phylogeny gives a robust estimation for early divergences of eukaryotes: phylogenetic place of a mitochondria-lacking protozoan, *Giardia lamblia*, *Molecular Biology and Evolution*, **11**, 65-71.
- 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.
- Jones, D.T., Taylor, W.R. and Thornton, J.M. (1992). The rapid generation of mutation data matrices from protein sequences, *CABIOS*, **8**, 275-282.
- Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*, Cambridge University Press, Cambridge.
- 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.
- 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), 271-304, Springer, Tokyo.
- Rivera, M.C. and Lake, J.A. (1992). Evidence that eukaryotes and eocyte prokaryotes are immediate relatives, *Science*, **257**, 74-76.

- Sogin, M.L. (1991). Early evolution and the origin of eukaryotes, *Current Opinion in Genetics and Development*, **1**, 457-463.
- Sogin, M.L., Edman, U. and Elwood, H. (1989). A single kingdom of eukaryotes, *The Hierarchy of Life* (eds. B. Fernholm, K. Bremer and H. Jörnvall), 133-143, Elsevier, Amsterdam.

Early Divergences of Eukaryotes

Tetsuo Hashimoto

(The Institute of Statistical Mathematics)

Yoshihiro Nakamura

(Laboratory of Gene Manipulation, Showa University)

Jun Adachi

(Department of Statistical Science, The Graduate University for Advanced Studies)

Masami Hasegawa

(The Institute of Statistical Mathematics and
The Graduate University for Advanced Studies)

A partial region of the mRNA encoding a major part of elongation factor 1α ($EF1\alpha$) from a mitochondria-lacking protozoan, *Giardia lamblia*, was amplified and sequenced, and the phylogenetic relationship among lower eukaryotes was inferred by the maximum likelihood (ML) method of protein phylogeny. The ML methods on the basis of six different stochastic models of amino acid substitutions consistently demonstrated that *G. lamblia* is the earliest offshoot of the eukaryotic tree among the four protozoan species being analyzed. Although the *Giardia* $EF1\alpha$ gene showed an extremely high GC content as compared with those of other protozoa, it was concentrated only to the third codon positions, resulting in no remarkable differences of amino acid compositions from other species. This clearly suggests (a) that the amino acid compositions of conservative proteins are free from the drastic bias of genome GC content which is a serious problem in the widely used tree of ribosomal RNA, and (b) that protein phylogeny gives a robust estimation for the early divergences in the evolution of eukaryotes.