解　説

分子進化とシミュレーション†

舘　野　義　男††

1. はじめに

この解説と以下に続く二つの解説の学問的基礎は分子生物学と分子進化学にある．そこでまず，これらの生物学分野について簡単な説明をしておきたい．その後，分子進化学へのコンピュータシミュレーションの応用について述べることにする．

はじめに次の2点を確認しておきたい．一つは，すべての生物の機能と構造は DNA または RNA のもっている情報に支配されていること．あと一つは，生物の進化は DNA または RNA を唯一の世代間情報伝達手として起こっていることである．もちろんこの両面における情報は同じもので，後でも述べるが，わずか4種の文字で書かれている．ヒトも細菌も皆同じ4種を用いている．このことは，地球上に興亡した全生物が共通の祖先から進化してきたことを強く示唆するだろう．また，生命現象の巧妙さを知れば知るほど，それがただの4種の文字で支配されているのが不思議になる．人間がこの何倍もの文字を用いてもなかなか意志を通じ合えないこと，国が違えば言葉も分からないことなどを思うと，この感慨はいっそう強くなる．

2. 分子生物学略史

分子生物学の歴史は二つの時期に分けることができよう．まず前期は，1953年の Watson と Crick による DNA 構造の解明から1969年ころの Nirenberg らによる遺伝コードの確立までである．この後から現在までが後期となる．前期において DNA の構造とその情報伝達機構の基本が明らかになり，後期では情報伝達機構の詳しい解明，DNA 情報の解読そして遺伝子の構造解明などが成されている．特に後期では，関連技術や装置の発達が著しいこと，分子生物学的考え方や技術が他の生物学の分野はもとより，医学や農学にも浸透してきていることなども無視できない．さらに，前期における実験はほとんどが核のない原核生物（大腸菌などやそれ以上に簡単な生物）を用いているが，後期での実験は核のある真核生物（酵母，ショウジョウバエ，カエル，ニワトリ，マウス，ヒト（細胞）など）にも及んでいる．

少し遡るが，DNA の存在は1869年に Miescher によって初めて確認されている．ダーウィンが“種の起源”を公にしてから 10 年後，メンデルが遺伝の法則を発表してから3年後のことである．そして 1903 年ごろまでには，Kossel らによって，DNA が4つの塩基アデニン(A)，チミン(T)，グアニン(G)，シトシン(C)を含んでいることが分かってきた．4文字の発見である．(RNA ではチミンの代わりにウラシル(U)が使われている．) その半世紀後に構造が明らかになる．Watson と Crick は，DNA が，2本の鎖状の化合物による二重螺旋構造をもつこと，2本鎖が A と T，G と C 2種のみの塩基対で結合していることを発見した（**図-1**）．彼らはさらに，DNA の2本が分離し，おのおのの鎖が鋳型となり，A-T, G-C の塩基対を作ることによって倍化することを予言した．つまり，DNA が自分と同じものを作りその情報を子孫に伝達する基本機構が示されたのである．この予言は1957年 Meselson と Stahl が行った“生物学上で最も美しい”といわれた実験によって実証された．

また，Crick は1956年にセントラルドグマ*なるものを提唱し，DNA の情報が RNA を仲介としてタンパク質に翻訳されることを新たに予言した．生物自身の構造と機能についてはタンパク質が主役を演じているが，それを裏で操っているのが DNA というわけで

† Molecular Evolution and Computer Simulation by Yoshio TATENO (National Institute of Genetics).

†† 国立遺伝学研究所

* 生体内で遺伝子の情報が DNA から RNA を経てタンパク質に流れるという学説．Crick はこの逆の流れはないとしたが，Baltimore と Temin は1970年それぞれ独立に逆転写酵素を発見し，RNA から DNA への情報の流れがあることを証明した．詳しくは三浦謹一郎 (1984)「DNA と遺伝情報」岩波新書 77 頁参照．

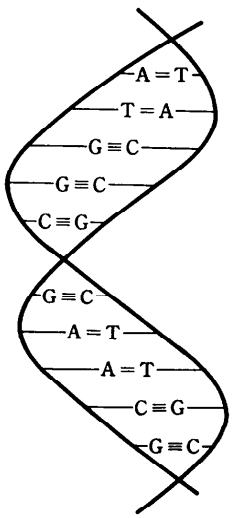

図-1 DNA の構造
2本の鎖状化合物がA-TとG-Cの対を作り，2重螺旋構造状になっている．螺旋の1回転中（軸長34オングストローム）に10対の塩基対が等間隔で存在する．AとTは2本の水素結合で，GとCは3本の水素結合で結ばれている．

ある．ただ，この主役は一代限りで世襲はない．世襲はあくまでもDNAが行う．セントラルドグマは1961年ごろまでに，Zamecnik, Brenner, Jacob, Meselson, Holley, Spiegelman らによって実証された．つまり，DNA の情報がまずメッセンジャー RNA (m-RNA) に写し取られ，それを鋳型として，アミノ酸が1個ずつ繋ぎ合わされてタンパク質ができる機構が明らかになった．アミノ酸を m-RNA に運んでくるもう一つの RNA をトランスファー RNA (t-RNA) というが，t-RNA は m-RNA の（4種の文字で書かれている）情報を三つ区切りに認識して一時的に m-RNA に結合する．この二つの RNA は3文字を単位として会話していることになる．会話の場所は決まっていてリボソームと呼ばれる RNA（リボソーム RNA, r-RNA, という）とタンパク質の混合物上である．m-RNA 側の3区切りをコドン，t-RNA 側のそれをアンチコドンという．ここで大事なことは，特定のアミノ酸は特定の t-RNA によって運ばれてくることである．アミノ酸の種類は20種ある．当然 t-RNA の種類も同じかそれ以上なくてはならない．20種のアミノ酸とコドン（またはアンチコドン）をどう結びつけるかが次の問題となったが，この解決が前期を締めくくる．

前後するが，20種のアミノ酸と4種の文字を使った DNA の情報との対応付けは1956年ごろ Gamov に

表-1 遺伝コード

1文字目 (5′側)	2 文 字 目				3文字目 (3′側)
	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	Term	Term	A
	Leu	Ser	Term	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ileu	Thr	Asn	Ser	U
	Ileu	Thr	Asn	Ser	C
	Ileu	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

塩基3文字が一つのアミノ酸に対応する．Phe＝フェニールアラニン，Leu＝ロイシン，Ileu＝イソロイシン，Met＝メチオニン，Val＝ヴァリン，Ser＝セリン，Pro＝プロリン，Thr＝スレオニン，Ala＝アラニン，Tyr＝タイロシン，His＝ヒスチジン，Gln＝グルタミン，Asn＝アスパラギン，Lys＝リシン，Asp＝アスパラギン酸，Glu＝グルタミン酸，Cys＝システイン，Trp＝トリプトファン，Arg＝アルギニン，Gly＝グリシン，Term＝終止コドン．

よって予測されていた．つまり，DNA の1文字が対応するにしても4種のみの意味しか表せないから少なすぎる，2文字でも 4×4＝16 でこれも不十分，そして3文字ということになった．3文字では 4×4×4＝64 の意味を表すことができて，十分，いや十分すぎる．この十分すぎることが問題となったが，結局1969年ごろまでに，Nirengerg, Ochoa, Lynen, Holley らによってこの3文字説が実証された．アミノ酸と DNA を結びつける遺伝コード（表-1）が完成したのである．遺伝コードでは1種のアミノ酸に，一つの例外を除いて，複数のコドンが対応している．これをコドンの縮重というが，この縮重によって上の十分すぎることが説明される．また，情報の終り用のコドンが三つある．ここで重要なことは，遺伝コードは全ての生物で基本的に同じであることである．このことも，生物が同一の祖先から進化してきたことを強く支持する．

上で述べたように，DNA 中の遺伝子の発現は m-RNA への転写と m-RNA からタンパク質への翻訳の2段階を含む．第一段階の転写について，1961年

Jacob と Monod は，大腸菌が巧妙な仕掛けをもっていることを示した．その理論はオペロン説*とよばれる．この説によると，転写はやはり DNA 中に存在するオペレータといわれる塩基領域で制御される．オペレータにリプレッサという RNA（実は RNA ではなくタンパク質であることが後で分かる）が結合することにより転写を抑え，解離することにより促す．また，一つのオペレータが複数の遺伝子の転写を制御することが多い．実際，彼らの用いた大腸菌のシステムでは三つの遺伝子が制御されていることが分かった．真核生物ではオペロン以外の制御機構も存在すると考えられているが，まだ不明な点が多い．

後期の一つの大きな出来事は，1970 年初頭を飾る制限酵素の発見である．この発見には Smith, Arber Nathans らが加わった．高等生物の DNA は何百万何千万という塩基対がつながった糸状の化合物で，そのままでは取扱いの困難な代物であった．制限酵素は塩基対の特定の並びを見つけてその場所を切断する働きをもっている．たとえば，*Eco* RI と名付けられた制限酵素は GAATTC という配列を探し，そのGとAの間を切る．今までに 160 種以上の制限酵素が発見されているが，これらを使うことにより DNA を特定の場所で切断できるので，数百数千の塩基対の断片として扱うことができるようになった．現在盛んになっている遺伝子工学は制限酵素の発見に端を発するといえる．そして遺伝子工学の中心は，こうして得られた DNA の特定の断片を増産することにある．多くのコピーがなければその解析も利用も不可能だからである．

ヒトなどの高等生物には細菌が寄生しているが，その細菌にも寄生している生物がいる．プラスミドやファージなどである．これらの生物は宿主のなかで自己増殖し，短時間で多くのコピーを作る．そこで，これらの DNA を制限酵素で切断し，その間に特定の DNA 断片を入れてまた繋ぐとキメラプラスミドやキメラファージができる．重要なのはこれらのキメラ（DNA 組替え体と呼ばれる）も宿主のなかで自己増殖することである．このように，生物の基本特徴である自己増殖能を利用して，目的とする DNA 断片を増産することができる．この DNA 組替え技術の基礎は Berg, Cohen, Boyer などによって 1973 年ころまでに確立した．

★詳しくは石川辰夫 (1982)「分子遺伝学入門」岩波新書 158 頁参照．

さて，イギリスに Sanger という人がいる．この化学者は 1950 年代にインシュリンのアミノ酸配列を驚くべき頭脳と根気で決定している．彼の目はその後タンパク質から DNA に注がれた．そして，1977 年 Sanger らによる DNA の塩基配列決定法がその応用結果とともに発表された．彼らはこの配列決定法を用いて φX174 というファージの全 DNA 情報を解読した．このファージは 5,387 の塩基からなる DNA をもっていることが分かったが，驚くことに，このたった一つの情報源を，読み始めの位置を変えることにより，複数の意味をもつようにやりくりしていたのである．何十億年の間転変を繰り返してきた進化情報が，ほんの一部ではあるが，初めて晒され，解読されたのである．同じころアメリカでは，Maxam と Gilbert が別の方法を開発している．ここでは，今はもう Classic となってしまった，Maxam-Gilbert 法の概略を図-2 に従って述べる．

まず，DNA 組替え技術で，目的とする DNA 断片を量産する．量産後にすべてのコピーの一方の端の塩基のリン酸部をその放射性同位元素に置き換えておく．こうして得られたコピーを 4 つのグループに分ける．第 1 のグループは dimethyl sulfate で処理する．この処理によって，塩基 G がメチル化される．他の 3 塩基は影響されない．さらに piperidine で処理すると，メチル化された塩基の直後で切断される．この結果，端がGである断片ができる．もちろん，断片の長さはまちまちである．第 2 のグループではまず熱アルカリ処理する．中和後 piperidine で処理すると，末端がAである多量の断片と末端がCである少量の断片の混合物ができる．第 3 のグループは，まずヒドラジンで処理する．この後 piperidine を加えるとTとCの端をもつ混合物が等量できる．第 4 のグループは濃塩のなかで第 3 グループと同じ処理をする．この場合はC末端をもった断片だけに分解される．これら 4 つの混合物を同じ平板の別々の箇所に一列に置き，電気泳動装置にかける．この装置により，おのおのが長さを規準として分けられる．つまり，短いものは泳動度が大きく早く進み，長いものは遅いので，図-2 のようなパターンができる．しかし，このパターンは目には見えないので，この平板の上に感光紙を置き放射性同位元素により感光させるのである．このために，コピーを放射性同位元素で作る必要があった．感光紙上のパターンを長さの短い順から追うことにより，目的とする DNA 断片の塩基配列を解読できる．

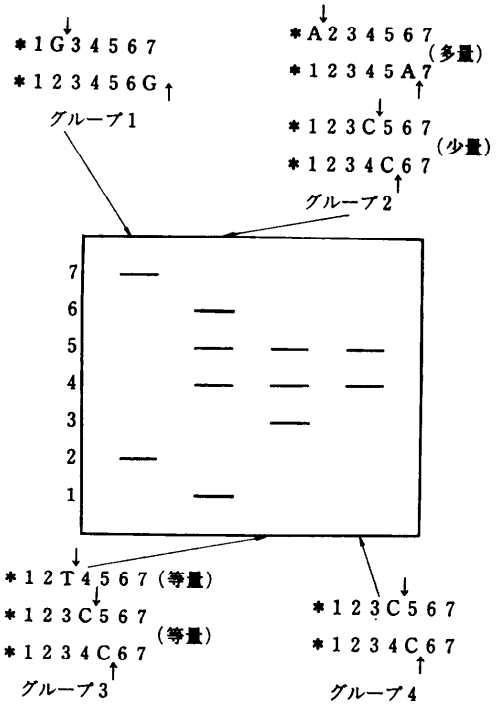

図-2 Maxam と Gilbert 法による塩基配列の決定
図中では7塩基の DNA 断片の配列決定の様子が示してある．＊印は放射性同位元素で置き換えた端を示す．矢印のところで切断され，いろいろな長さの断片になる．電気泳動パターンで1から7までの番号は断片の短さの順を表す．4つのグループ分けの条件を考慮しながら短い順にパターンを読んでいくと塩基配列が分かる．この場合は AGTCCAG となる．

DNA のなかの遺伝子の構造解析が進むにつれてその様相が次第に明らかになってきたが，それらは ϕX 174 の例でも分かるように，それまでに漠然と描かれていた像とは大きく異なるものであった．1977年 Chambon らによって，今度はニワトリのオバルブミン遺伝子の意外な様子が明らかになった．それまでは，一つの意味をもつ遺伝子は DNA 上で連続しているものだと思われていた．ところが，この遺伝子は12の部分に分かれていたのである．そして，それらの部分の間にはそれぞれ意味のない塩基配列が介在していた．ここで，意味をもつ部分をエクソン，もたない部分をイントロンという．つまりこの遺伝子は，12のエクソンと 11 のイントロンから構成されていた．さらに驚くべきは，真核生物でその後明らかになったほとんどの遺伝子が，エクソン-イントロン構造になっていたことである．また，1980年 Efstratiadis らはヒトのヘモグロビン遺伝子が複数重複して存在すること，そしてそのなかのあるものは，もはや遺伝子としての機能を果たしていない偽遺伝子であることを発見した．その後，遺伝子の重複や偽遺伝子は多くの真核生物の遺伝子で見つけられている．

遺伝子の情報がタンパク質に翻訳されるときは，イントロンは障害物になるが，これはどのように除かれる(あるいは無視される)のだろうか．生物はここでも巧妙な手段を用いる．この不可思議な機構は1978年ころ Leder や O'Malley らによって次第に明らかになった．上に述べたように，DNA の遺伝子情報は m-RNA に転写されるが，真核生物ではこれは核内で行われる．この場合には m-RNA といわず，ヘテロジニアス RNA (hn-RNA) などという．hn-RNA はイントロンも含めて転写された情報をもっているが，それが核から細胞質に出る以前にイントロン（に対応する部分というのが正しいが略する）が全て除去されて（スプライシングという）しまう．スプライシングはもちろんイントロンの除去とエクソン同士の接合を含んでいる．スプライシングの正確さは当然生物の死活に係わるが，この切り貼り作業はきわめて正確である．イントロンが二つのエクソンをコドンの3文字の中途で隔離している場合もあるが，スプライシングは1個の塩基対の誤りも犯さず，二つのエクソンをぴたりと合わせる．イントロンが除かれた RNA が m-RNA で，これが鋳型となって細胞質でタンパク質の合成が行われることになる．

以上，分子生物学の歴史を概観するにはあまりにも省略が多過ぎたと思うが，紙面の都合上このくらいにしておく．これくらいのことを知っていれば，以降の分子生物学的な記述はだいたい理解できるのではないかと思う．

3. 分子進化学概説

1965年 Zuckerkandl と Pauling はヘモグロビンのアミノ酸配列を数種の動物間で比較検討し，興味ある発見をした．2種の動物が系統的に離れていれば，アミノ酸配列の違いも大きいのである．つまり，おおまかにではあるが，アミノ酸の数の違いが動物種の分岐時間に比例することが示されたことになる．そして，1967年 Fitch と Margoliash は，チトクローム c*の

＊細菌からヒトまで生物は栄養の補給と呼吸によって生命を維持している．呼吸は代謝的にはエネルギの産生過程であるが，そこに電子伝達系という一連の化学反応がある．この反応で電子を伝達してゆくタンパク質がチトクローム c である．

アミノ酸配列データから，20種の生物の系統関係を示す系統樹を推定した．20種のなかには，酵母，昆虫，魚，爬虫類，鳥類，(ヒトを含む) 哺乳類などが含まれている．つまり，チトクローム c を支配する遺伝子はこれら生物種の祖先種にすでに存在し，このタンパク質の機能を損なうことのないように，それぞれ独立に進化してきたのである．このように共通の祖先から由来し，共通の働きをもつタンパク質を支配する遺伝子を相同遺伝子という．さて，得られた系統樹は，それまで化石，生理学，発生学，形態学などの膨大な研究データから推定されていた系統関係をよく反映していた．ここに分子進化学が急速に発展してきた二つの理由がある．

一つは，チトクローム c のような (110 個くらいのアミノ酸からなる) 小さなタンパク質から得られた結果が，長い時間をかけて種々の生物学分野から論じられていた結果に合うこと．あと一つは，それまでの論点は専門的すぎていて他の研究者には分からないことがしばしばあったが，アミノ酸の違いはだれの目にも明らかであること．前者は結論までの研究経路の短縮を，後者は客観的議論への直結を意味する．

ヘモグロビンやチトクローム c などのアミノ酸配列データの出現が分子進化学の発展の契機にはなったが，その学問的基礎は集団遺伝学に由来する．1968年木村資生は上記 Zuckerkandl と Pauling の研究結果を集団遺伝学理論の立場から次のように捉えた．つまり，タンパク質レベルでの変化 (アミノ酸の違い) が，骨などの表現型の変化と比べて異常に大きく，自然選択によってこの違いが生じたとはとうてい説明できない．生物種がその遺伝的変異を，自然選択によって保有するにはその代償を支払わねばならない．変異のなかには (相対的に) 生存に有利な個体だけでなく不利な個体も含まれ，最も有利な個体と平均個体との間に差が生じる．この差が代償であり，個体数の損失という形で支払うことになる．変異がなければこの差は生じない．木村は，このように大きなアミノ酸変異を自然選択で維持するための代償はあまりにも大きすぎてその生物種は絶滅してしまうと考えた．そこで，彼はアミノ酸の変異は，自然選択に対し有利でも不利でもない，中立な突然変異によって維持されていると結論した．木村の中立説である．中立説によると，分子レベルの進化速度 (k) は年当りの中立突然変異率 (v) に等しい．つまり，ある特定の相同遺伝子に限れば，その遺伝子の進化速度はどの生物種でも等しいことになる．ただ，遺伝子の進化は有限個の遺伝子抽出を世代ごとに繰り返す確率過程なので，この場合の等しさは統計的な等しさを意味することに注意したい．今まで調べられたほとんどの相同遺伝子 (あるいはタンパク質) で，それぞれの遺伝子の進化速度は生物種に関係なくほぼ一定であることが示され，中立説の正しさが実証されている．これによって Zuckerkandl と Pauling の観察に対する学問的裏付けがなされたことになる．

ここで，ダーウィンの自然選択による進化と中立説による進化の違いに触れておきたい．前者では，自然選択に有利な個体 (遺伝子) が他と比べて早く進化し，やがて生物集団全体に広まってゆく．つまり，生存に有利なタンパク質を支配する相同遺伝子はより早く進化することになる．ここでいう生物集団とは，交配または分裂などによって子孫を残せる集団を指す．これに対して後者では，機能的に重要なタンパク質 (またはその部位) を支配する相同遺伝子 (またはその部分) が突然変異によって変わると，自然選択に不利となるので，そのような遺伝子はやがて集団から排除されてしまう．結果的に，機能的により重要な相同遺伝子はより遅く進化することになる．重要なことは，中立突然変異遺伝子は機能がないタンパク質を支配する遺伝子ではなく，変異を起こす前の遺伝子と同等の機能をもったタンパク質を支配する遺伝子を意味することである．二つまたはそれ以上の相同遺伝子が，たとえ DNA レベルで違っていても，機能的に同等のタンパク質を産生するなら，自然選択はそれらを区別できない．つまり，それらの遺伝子はみな中立遺伝子である．ただ，誤解を恐れずにいえば，純粋な中立遺伝子は機能的制約からまったく自由なタンパク質を支配する遺伝子を指す．また，前者では進化速度の絶対的上限が定まらないが，後者では (遺伝子の種類に関係なく) 年当りの純粋な中立突然変異率と定められる．

中立突然変異は現実にはどのようなものだろうか．コドンが縮重していることはすでに述べたが，表-1 の遺伝コードからも分かるように，具体的にはコドンの (主に) 第3番目の塩基が変化しても，それから翻訳されるアミノ酸が変化しないことを意味する．このような塩基の変化が進化上起こるとき同義置換とよぶ．また，アミノ酸を変えるときを非同義置換という．同義置換は産生されるタンパク質の機能に何の影響も及ぼさず，生存の優劣性にほとんど影響を与えないので，中立突然変異に非常に近いといえる．1980 年宮

田らはいろいろな遺伝子について同義置換の進化速度と非同義置換のそれを調べ興味ある事実を示した．非同義置換での進化速度は遺伝子によりまちまちであるが，同義置換ではほぼ一定で年当り塩基座当り約2億分の1となる．これは，中立説が予測したことである．中立突然変異による進化速度の上限はだいたいこの値になるであろう．また，タンパク質を産生しない偽遺伝子は純粋中立遺伝子と考えられるが☆，タンパク質を産生する相同遺伝子と比べて進化速度が非常に早くなっている．いい換えれば，偽遺伝子の発見は中立説を支持する強力な証拠をもたらしたことになる．このように，中立説は予測能と自己検証性を具えているが，この両者を併せもつ学説は生物学では少ないのではなかろうか．

さて，進化速度というあまり定義のはっきり示されていない術語がでてきたが，以下でその説明をしよう．まず，二つの相同遺伝子が T 年前に N 個の塩基対をもつ共通の祖先遺伝子から分岐して，図-3 に示すようにそれぞれ独立に進化してきたとしよう．進化は突然変異のみによって起こると仮定する．この二つの子孫遺伝子の間の塩基の違いを D とすると，塩基座当りの違いは D/N となり，これを d とおく．さらに，d を $2T$（二つの相同遺伝子の間の時間は T ではなく $2T$ である）で割ると，この遺伝子の塩基座位当り年当りの進化速度 (k) が求められる．二つの相同遺伝子の間であまり突然変異が多く起こらなければ $k=d/2T$ が進化速度を表すと考えて差し支えないが，突然変異が多くなると，もう少し工夫しなければならない．つまり，突然変異が多くなると，それが同じ塩基座に2回以上起こる（重複置換という）可能性を無視できなくなる．

重複置換を考慮して進化速度を推定する方法は今まで種々開発されてきている．ここでは図-4 に従って代表的な二つの方法について簡単に触れたい．図に示したように，4つの塩基の間の置換様式は 12 ある．この 12 の置換が全て同じ確率で起こると仮定したのが，1969年 Jukes と Cantor が考案した方法である．この方法によれば，塩基座位当りの置換率 K は次の式で求められる．

$$K=-(3/4)\ln\{1-(4d/3)\} \qquad (1)$$

ここで，ln は自然対数である．K を $2T$ で割ることにより k が得られる．もし，d が小さければ，

$$\ln\{1-(4d/3)\}=-4d/3$$

となるので，$K=d$ となり，重複置換を無視したときの値となることが分かる．さて，図-4 のなかでプリン（AとG）間またはピリミジン（TとC）間の置換を transition，プリンとピリミジン間の置換を transversion という．種々の生物のいろいろな遺伝子について塩基置換の様式を調べてみると，一般に transition のほうが transversion より頻繁に起こっていることが明らかになってきた．そこで，木村資生は 1980 年この事実を取り入れて別の方法を開発した．結果だけを示すと，

$$K=-(1/2)\ln\{(1-2P-Q)(1-2Q)^{(1/2)}\} \qquad (2)$$

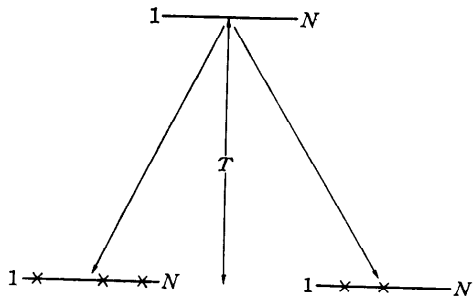

図-3 遺伝子の分岐と進化
祖先遺伝子からの二つの遺伝子の分岐と，分岐時間 T の間の進化を示す．それぞれの遺伝子は N 塩基からなる．子孫遺伝子中の×は突然変異が起きた塩基座である．

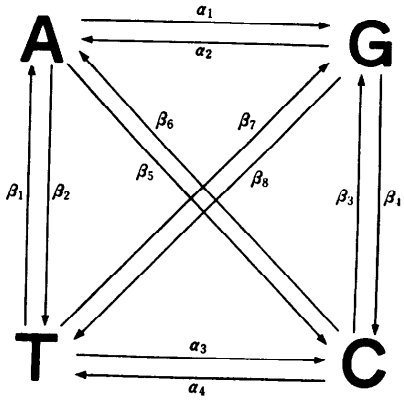

図-4 4つの塩基間の置換
合計12の塩基置換があるが，α_i は transition，β_i は transversion を表す．12全ての置換が等確率で起こる場合が Jukes と Cantor モデルで，transition 同士あるいは transversion 同士が等確率で，それらの間は異なる確率で起こる場合が木村の2パラメータモデルである．

☆ 正常な相同遺伝子が存在しなければ，偽遺伝子は普通有害遺伝子となり中立遺伝子とはならない．遺伝子が正常に機能しているため，それと相同な偽遺伝子は機能的制約から完全に解放されるのである．

となる．ここで，P は比較する二つの相同遺伝子間の transition の割合，Q は transversion の割合である．この場合も P と Q が小さいと，

$$\ln\{(1-2P-Q)(1-2Q)^{(1/2)}\}=-2(P+Q)$$

となり，$K=P+Q$ が導かれる．この方法は transition と transversion の二つの要素を取り入れているので，木村の2パラメータ法とよばれている．

中立説によれば，年当りの進化速度は一定なので，K と k から分岐時間を推定することもできる．この場合，k をあらかじめ求めておかねばならないが，このためには化石などから分岐時間が分かっている生物種に着目する必要がある．たとえば，ヒトとマウスの共通祖先はだいたい8千万年前に遡る．分岐時間が分かっている2種の生物の相同遺伝子について K を求め，これをその分岐時間の2倍で割れば k が推定される．上にも注意したように，中立進化では，k は生物種に関係なく一定である．もちろん遺伝子が異なれば，k は違う値になる．この方法によって，化石が残っていない生物種間の分岐年代が，相同遺伝子の塩基配列を比較することによって推定できる．これも分子進化学の大きな特長である．

分子進化学のなかであと一つの重要な問題は，タンパク質やDNAなどの配列データから関与する生物種または遺伝子の系統樹（分子系統樹という）を推定することである．特に生物種の系統樹を種系統樹，遺伝子の系統樹を遺伝子系統樹ということがある．分子系統樹を推定する方法はいろいろ開発，改良されてきているが，ここでは紙面の都合で一番頻繁に用いられている Unweighted Pair-Gruop Method using Arithmetic Averages (UPGMA) という方法を簡単に紹介する．この方法の原型は1958年 Sokal と Michener によって，もともと形態学的データに利用する目的で考案された．1973年 Sneath と Sokal はこの方法を詳しく論じている．また，根井正利は1975年この方法を分子系統樹の推定に改訂し，今日にいたっている．ただ，一番頻繁に利用されているといっても，後で述べるように，一番信頼性の高い方法ではないので断っておきたい．

さて，UPGMA で遺伝子系統樹を推定するとしよう．まず，関与する相同遺伝子の全ての組合せについて上でのべた K（遺伝距離ともいう）を求めておく．**表-2** は5つの相同遺伝子間の遺伝距離を示す遺伝距離行列である．ここでは系統樹の単位は一つの相同遺伝子であるが，一つの生物種あるいは集団の場合もある．そこで，1973年 Sneath と Sokal はこれらを全て包含する系統単位として Operational Taxomic Unit (OTU) なる概念を提唱している．ここでも以降 OTU を用いることにする．UPGMA では，お互いに距離の近い OTU 同士から順々に結びつけて系統樹を作り上げていく．いま，表-2 のなかで D_{12} が最小とする（もちろん他の距離でも差し支えない）．すると OTU 1 と OTU 2 が結合し新たな OTU となる．これを OTU(1, 2) とおく．次に，OTU(1, 2) と他の OTU との間の距離を計算する．これらの距離は次の式で求められる．

表-2 遺伝距離行列

OTU	1	2	3	4
2	D_{12}			
3	D_{13}	D_{23}		
4	D_{14}	D_{24}	D_{34}	
5	D_{15}	D_{25}	D_{35}	D_{45}

OTU が5の場合の距離行列で，D_{ij} は OTU i と OTU j の間の遺伝距離である．

$$D_{(1,2)X}=(D_{1X}+D_{2X})/2 \qquad (3)$$

ここで，X は OTU 3, 4, 5 である．ここで，新たに計算された D_{1X} と D_{2X} はどちらも $D_{(1,2)X}$ に等しくなることに注意したい．つまり，UPGMA では結合された OTU のおのおのから他の OTU への距離は全て等しいと仮定している．別のいい方をすれば，二つの OTU の共通祖先からの距離はそれぞれ等しいことになる．これを等距離仮定とよぶことにする．中立遺伝子がこの仮定を満たすことに気づかれたと思う．

次に，新たに計算された距離も含めて $D_{(1,2,4)}$ が最も小さいと仮定すると，新たに OTU(1, 2, 4) が作られ，また D_{34} が最小になると仮定すると，OTU(3, 4) が作られる．この操作を全ての OTU が一つのグループの中に入るまで繰り返すと**図-5** のような系統樹ができる．この図では D_{34} が最小になると仮定してある．この場合 OTU(1, 2) と OTU(3, 4) の間の距離は式(3)と同じように，

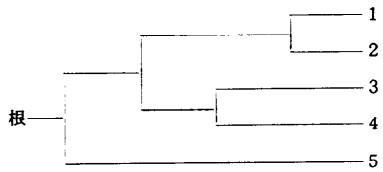

図-5 UPGMA 法で求められた系統樹
番号は5つの OTU を表す．この方法は等距離仮定をもっているため，根の位置を決めることができる．

$$D_{(1,2)(3,4)}=(D_{13}+D_{14}+D_{23}+D_{24})/4 \quad (4)$$

と求められる．UPGMA は以上のように簡単な方法であるが，この方法の最大の特徴は等距離仮定にある．この仮定によって系統樹全体の共通祖先（根ともいう）の位置が決められることに注意したい．この他の方法では 1967 年の Fitch と Margoliash による方法，1972 年の Farris 法，1982 年の筆者らによる Modified Farris 法，1987 年の斎藤成也と根井正利による近隣結合法など多くの方法がある．また，1981年 Felsenstein が開発した方法や 1990 年筆者が考案した方法は距離行列からでなく，塩基配列から直接分子系統樹を推定する．これらの方法はいずれも等距離仮定をもっていないため，普通の手段では根を決めることができない．

4. シミュレーションによる分子系統樹推定法の評価

地球上の生命の起源は 35 億年以上に遡るといわれている．生物の進化はこのように，人の一生からみれば，はるかに膨大な時間をかけて，現在地球に生息している全ての生物を産み出してきた．進化学と他の生物学との違いを際立たせている点がここにある．（進化の）再現実験ができないことである．ところが，DNA やタンパク質レベルでの進化が解明されてくるにつれて，進化をもっと簡単に捉えることが可能となってきた．分子進化に関する限り，形態や生理などの複雑な様相に捉われることなく，塩基やアミノ酸の置換として進化を論じることができるからである．分子進化学の特長といえよう．つまり，分子進化では，進化モデルを比較的簡単に設定することができ，それを利用して千万年あるいは億年規模の進化のシミュレーションが可能となる．もちろん，シミュレーションによる進化は実際の進化に比べ単純にならざるをえないので問題は残るが，モデルや仮定をはっきり認識していれば学問的に十分価値のある議論に結びつく．

筆者らは 1986 年コンピュータを用いた分子進化のシミュレーションによって分子系統樹推定法の評価を行ったが，ここでは主にその研究について解説する．まず，この研究の意義を述べてみたい．前にもちょっと触れたが，進化の原動力は DNA レベルの突然変異であり，それは確率事象として捉えられる．また，生物集団は常に有限であるため，進化そのものが確率過程であると考えられる．さらに，与えられた OTU の数に対する系統樹の形の種類は，OTU が増えるにつれて急激に増加することが知られている．たとえば，OTU が 10 の場合，その形の数は 34,459,425 となる．これらの事実は，分子系統樹の推定もやはり確率的作業であることを意味する．その上，系統樹推定法のすべてに一長一短あり完全な方法はないことを考え合わせると，系統樹の推定は多くの研究者が思っているほど確実なものではないことが分かるだろう．そこで，種々の方法がどれほど正しい結果を出すか統計的に調べることが必要になってくる．この場合何を規準にして正しいか否かを判定するか．現在得られる塩基配列やアミノ酸配列がその進化過程を正しく反映しているという保証はなく，それらのデータ量がいまだ不十分である事実を考えると，別の規準を設定しなければならない．シミュレーションはそのための一つの方法なのである．もちろん，シミュレーションの結果はそのモデルに依存するし，モデルは現実の近似にほかならない．近似の善し悪しは少数良質のパラメータをいかに選ぶかに係わっている．

さて，進化のモデルにはモデル遺伝子とその進化経路を示すモデル系統樹が必要となる．モデル遺伝子は 300 の塩基からなり，その配列は A, T, G, C がランダムに並んでいるものとした．また，モデル系統樹は **図-6** に描かれているように，8 つの OTU を含んだものを使用した．図中で M は系統樹の枝の長さの単位となるもので，モデル遺伝子ごとの塩基置換数で表される．今回は M は 2 としてある．つまり，このモデル系統樹の二つの OTU 間の遺伝距離は塩基座位当り 0.093 (14 M/300) から 0.013 (2 M/300) までの値をとる．これに上に述べた宮田らの同義置換速度，塩基座位当り年当り 2 億分の 1 を適用すると，モデル系

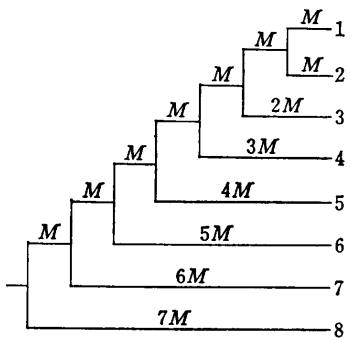

図-6 シミュレーションに用いたモデル系統樹
8 つの OTU からなる．M またはその倍数は枝の長さで，遺伝子ごとの塩基置換数で表してある．ここでは M は 2 とした．モデル遺伝子は一番左端の根から右方向へ進化する．

統樹は1860万年から260万年の進化規模を表すことになる．もちろん，モデル遺伝子も系統樹もコンピュータのなかで作られる．

次に，モデル遺伝子をモデル系統樹に沿って進化させるために，おのおのの枝において突然変異を起こす．枝ごとの突然変異の数はポアソン分布に従うものとした．たとえば，枝の長さが$3M(6)$の場合は，平均が6のポアソン分布からのランダム抽出によって実際の突然変異数を決めた．ここで重要なのは，このランダム抽出によって確率的な遺伝子の進化を真似ることができることである．抽出した突然変異はモデル遺伝子にランダムに配布した．また，図-4からも分かるように，塩基が変異する場合三つの可能性があるが，おのおの等確率で起こると仮定した．たぶん気付かれたと思うが，塩基置換は先に述べたJukesとCantorの様式に従っている．このようにして，モデル遺伝子に順次突然変異を起こさせながら進化させることによって，系統樹の末端で8つの子孫遺伝子を得る．そして，これらの子孫遺伝子の塩基配列の全ての組について，JukesとCantorの方法（式(1)）で遺伝距離を計算し距離行列を求めた．シミュレーションはここで完了するが，これを500回試行した．

ここで，シミュレーションで得られた距離行列を用いてUPGMA，Farris法，Modified Farris (MF)法の信頼性を検定する．つまり，分子進化に必然的に付随する確率誤差の影響に抗して，おのおのの方法がどれほど正しくモデル系統樹を再構築するか調べるわけである．念のためにいえば，500の距離行列は塩基置換の確率誤差によって（同じモデル系統樹から得られたのであるが）それぞれ違ったものになり，さらに，おのおのは必ずしもモデル系統樹の進化経路を反映しているとは限らない．三つの方法がそれぞれ推定した500の系統樹のおのおのとモデル系統樹を比較する場合，二つの要素が考えられる．一つは系統樹の形であり，あと一つは枝の長さであるが，ここでは形の比較のみについて述べる．形を定量的に比較するのはそれほど簡単な問題ではないが，一つの方法として1981年のRobinsonとFouldsの方法を適用する．彼らの方法の概略は図-7を使って次のように説明できる．図中(a)はモデル系統樹で(b)と(c)の一番上に示すものが推定された系統樹である．系統樹(b)をモデル系統樹に戻すためには，矢印の枝を長さがなくなるまで圧縮し，OTU2とOTU3を交換し，圧縮された枝を元に伸張すればよい．この場合圧縮と伸張が一つずつ必要なので，系統樹(b)に2の歪度（d_T）を与える．系統樹(c)の場合は矢印のところの伸張だけでモデル系統樹に戻るので，d_Tは1となる．モデル系統樹の形を正しく推定できれば，d_Tはもちろん0で

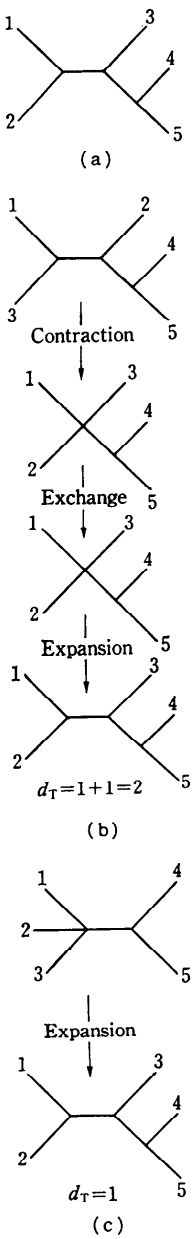

図-7 二つの系統樹間の形の比較
(a)モデル系統樹，(b)，(c)推定された系統樹．
d_Tは二つの系統樹間の形の違いを定量的に表す歪度で，同じなら0，離れれば離れるほど大きくなる．
この場合，(b)のd_Tは2，(c)のd_Tは1となる．

ある．結果は図-8 に示されている．

図-8(a)は共通祖先(根)の位置を決めた場合，図-8(b)は決めない場合である．上で述べたが，UPGMA (UPG) 以外は根を決めることができないので，他の2法の根は系統樹の一番長い経路の中点とした．根を決めた場合，3方法の間で d_T に統計的に有意な差はない．しかし，UPGMA がモデル系統樹を正しく推定する確率 ($d_T=0$ の割合) は他の2法の約半分になっている．また，稀にではあるが，Farris 法と MF 法はモデル系統樹とかなり違った系統樹を推定することがある．UPGMA と他の2法のこのような違いは，等距離仮説の有無と枝の長さの推定法の違いに由来すると考えられる．根を決めない場合は UPGMA と他の2法との間で d_T に有意な差がみられた．前者のほうが劣るのである．さらに，d_T が0になる確率は後者の半分未満になっている．これは，Farris 法や MF 法の根の決め方が良くないことを表している．以上の結果，UPGMA は一番頻繁に利用されている方法ではあるが，一番信頼すべき方法ではないことが分かる．また，斎藤成也と根井正利（前述）もシミュレーションにより同様な結果を得ているが，彼らの近隣結合法が特に信頼性の高いことを示している．

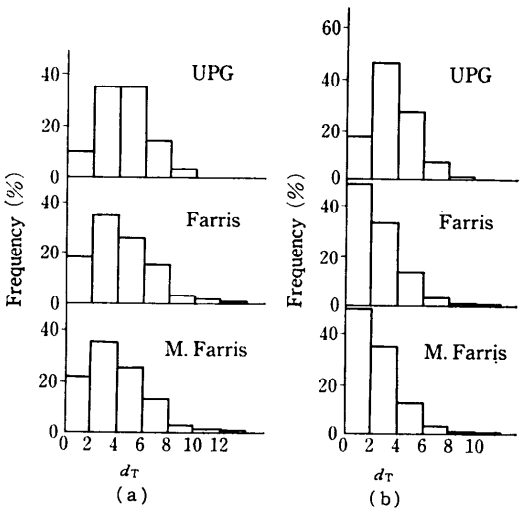

図-8 シミュレーションによる系統樹作成法の評価結果
UPG は UPGMA, M. Farris は Modified Farris 法の意味である．500 回の試行の結果をヒストグラムで示してある．図(a)は根の位置を決めた場合，図(b)は決めない場合である．

5. 終 り に

生物が DNA のなかにもっている情報の量は大腸菌で 420 万字，酵母で 2000 万字，ショウジョウバエで1億4000万字，そしてヒトでは33億字といわれている．これらの情報は全て 35 億年以上にわたる進化の産物である．つまり，その解釈には進化学的な考えが必須となる．また，このように膨大な量のデータを処理するにはコンピュータと情報解析ツールがやはり必須となる．分子生物学，分子進化学ならびに集団遺伝学の発展はこれからも目覚しいものがあると期待されるが，同時にコンピュータ，周辺機器，オペレーティングシステム，情報解析ツールなどもいっそう発達してゆくことであろう．そして，この二つの流れに精通している人あるいはグループが生物情報解析の発展に大きく貢献することになると思われる．

謝辞 国立遺伝学研究所の永田恭介博士から分子生物学略史の部分について有益な助言をいただきました．感謝いたします．

参 考 文 献

英文原著のものは日本語訳だけ記す．

分子生物学関係

J. D. ワトソン・N. C. ホプキンス・J. W. ロバーツ・J. A. ステイツ・A. M. ワイナー著，松原謙一・中村桂子・三浦謹一郎監訳：遺伝子の分子生物学（第4版），トッパン (1988)．

F. H. ポーチュガル・J. S. コーエン著，杉野義信・杉野奈保野訳：DNA の一世紀，岩波書店 (1980)．

H. F. ジャドソン著，野田春彦訳：分子生物学の夜明け，東京化学同人 (1982)．

分子進化学関係

木村資生著，向井輝美・日下部真一訳：分子進化の中立説，紀伊國屋書店 (1986)．

木村資生著：生物進化を考える（岩波新書），岩波書店 (1988)．

根井正利著，五條堀孝・斎藤成也訳：分子進化遺伝学，培風館 (1990)．

シミュレーション関係

木村資生編：分子進化学入門(7章)，培風館(1984)．分子進化遺伝学（上記）．

（平成2年4月5日受付）