

情報処理

1
2002

VOL.43 NO.1 通巻443号 別刷

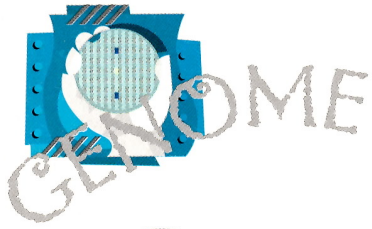


社団法人

情報処理学会

Information Processing Society of Japan

<http://www.ipsj.or.jp>



3 比較ゲノム解析を中心とする 進化ゲノム学の展望

齋藤 成也

国立遺伝学研究所進化遺伝研究部門
nsaitou@genes.nig.ac.jp

地球上には莫大な数の生命が存在し、しかもそれぞれ独自の遺伝子を多数持っている。これら生物や遺伝子の多様性は、長大な期間に蓄積した遺伝子の進化が基礎となって生じてきた。遺伝子はその名のとおり親から子に遺伝してゆく。その間に突然変異が起きて、ゆっくりと生物が変化してきたことが生物進化の核心である。現在に至るまでこの変化は続いており、ヒトの進化も例外ではない。現在「ゲノム」という1つの生物が持っているすべての塩基配列が次々に明らかにされつつある。生物進化の研究でも、ゲノム全体を常に考える「進化ゲノム学」(Evolutionary Genomics) が誕生しつつあるといえよう。本稿は、この進化ゲノム学について、比較ゲノム解析という観点を中心に紹介する。

進化ゲノム学とは

「生き物とは何か」を問題とする生物学は、古代ギリシャのアリストテレス以来、長い間博物学が中心だった。しかし近世になって細胞が生物の単位であることが発見されたころから、どの生物にも共通の性質を追求できるようになった。こうして、細胞より下のレベルへ顕微鏡や生化学の力を借りて下ってゆき、細胞内での染色体の挙動を調べたりタンパク質や核酸などの分子を調べるようになった。こうして細胞から核へ、染色体へ、そしてDNAへと、全生物の共通性を求める旅がはじまった。この大キャラバンの中心にあった遺伝子の研究を振り返ると、メンデルの用いたエンドウマメにはじまって、ショウジョウバエ、アカパンカビ、大腸菌、ファージというように、実験するのにより簡

単な生物を用いた研究へという方向だった。しかしこの旅は1953年、ワトソン・クリックによるDNA二重らせん構造の発見によって目的地に到達したのである。一方、ダーウィン以降の近代進化論が示す生物単一発生論の論理的帰結として、無生物、つまり細胞以前の分子状態の段階を経て生命が誕生したことが明白になった。DNAという生物全体の共通性の存在は、このこととしっかりつながっている。

その後の生物学の発展は、あたかもプールでターンを切るように方向転換をして、生物の多様性に向かって進んでいった。いったん全生物に共通なシステムが発見されれば、興味が各生物群における独自性へと移ってゆくのは当然の流れと言えよう。それは生物学の長い伝統である博物学への回帰である。現在、生物多様性を研究する重要性が叫ばれているが、ある生物が持っている遺伝情報全体を指す「ゲノム」まるごとの研究である「ゲノム計画」も、この多様性探索の波の上に

ある。

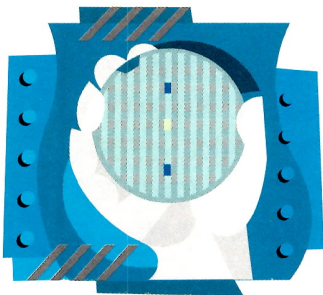
実際に、バクテリア(原核生物)については、インフルエンザ菌というあまり聞いたことのないバクテリアゲノムがまず1995年に決定されたあと、現在では50種以上のバクテリアのゲノム配列が決定されている(DDBJのGIB (<http://gib.genes.nig.ac.jp/>)を参照されたい)。これからますますゲノム配列によってバクテリアの世界の多様性が明らかになってゆくだろう。

一方真核生物においては、これまでのゲノム計画は、パン酵母、エレガンス線虫、ショウジョウバエ、シロイヌナズナ、マウスなどの「モデル生物」を中心に行われてきた。これらは、真核生物、動物、植物、脊椎動物といった広範囲な生物群に共通する遺伝子の特徴を発見することが主要目的である。しかし、この考え方には問題がある。たとえばショウジョウバエのゲノムにはショウジョウバエのグループ独自の遺伝子もあるだろうし、逆にショウジョウバエゲノムへの進化過程で失われた遺伝子群もあるだろう。このため、ショウジョウバエゲノムが動物ゲノムの代表だと結論するのは早計である。当然のことながら、ある大分類群の共通性を抽出するには、そのグループの中で分岐していった複数のゲノムを比較する必要がある。

一方で、多様性は独自性の重ね合わせである。それぞれの生き物の独自性は、進化によってゲノム内に蓄積されてきた遺伝子の変化が根底にある。ある生物の独自性を浮き立たせるにはどうしたらよいのか？ 答えは、進化的にその生物と近縁な別の複数の生物を調べて、それらを比較すればよいのである。つまり、生物進化の過程で徐々に蓄積していった各生物系統の「独自性」を調べ上げてゆくことが、ゲノム計画の第2段階となるだろう。

特にヒトの場合、その独自性はすなわち「人間とはなにか」という問題である。ヒトゲノムの塩基配列の中には、人間の独自性を規定する遺伝子の変化の証拠が残

されているはずである。その中には、長期の直立二足歩行を可能にした骨盤などの形態変化、犬歯の縮小や歯列弓形態の変化、眉上隆起の退化、体毛の減少、ペニスの巨大化、母指対向性、手足と



胴体の比率の変化などのマクロレベルでの形態であったり、発情期の消失などの生理学的性質、そしてもちろん言語能力をはじめとする脳機能の発達を促した遺伝子の変化もあるだろう。このようなヒト化を特徴付ける遺伝子の変化が解明されれば、人類学・霊長類学をはじめとする生物学分野のみならず、社会科学・人文科学にも大きな影響を与えることは、間違いない。

それぞれの遺伝子単独で比較解析することは、1960年代以来、「分子進化学」で行われてきたことである。たとえば、図-1は私の研究室で行った、Rh式血液型遺伝子の進化系統樹である。ダイヤモンド印は、遺伝子重複を示している。脊椎動物の系統で一度重複が生じてRhおよびRh50という相同遺伝子(この場合、パラログ)が生じた。線虫の系統でも独立に遺伝子重複が生じている。横方向の枝の長さはアミノ酸の変化(置換数)に比例している。また枝の上の数字は、その枝の統計的信頼性を近似的に示す「ブートストラップ確率」(%表示)である。

ゲノム情報が氾濫しつつある現在、このような単独の遺伝子での比較ではなく、ゲノム全体での考察が重要となってくる。それが「進化ゲノム学」であり、分子進化学とゲノム学の融合とも境界領域であるとも言えるだろう。

比較ゲノム解析の方法

進化的に遠い関係および近い関係にある生物種の比較ゲノム解析が「進化ゲノム学」の中心であり、それについてどのような方法が用いられているのかを、ヒトゲノムと類人猿ゲノムの比較解析を例として、この項で解説する。

ヒトゲノムの全塩基配列が明らかになりつつある現在、我々人間自身の特異性を規定している遺伝子の変化を発見することが、現代生物学の射程に入りつつある。それを探るためには、ヒトと最も系統的に近い生物をヒトと比べてみる必要がある。では、どのような生物が我々ヒトに近いのだろうか？ それは「類人猿」である。図-2は、化石などを用いた形態学的研究と分子人類学の成果を総合して、ヒトと類人猿の系統関係を示したものである。最近、言語習得能力や濃密な性行動などの面をとらえて、ボノボが一番ヒトに近いという表現を聞くことがあるが、それは俗説である。チン

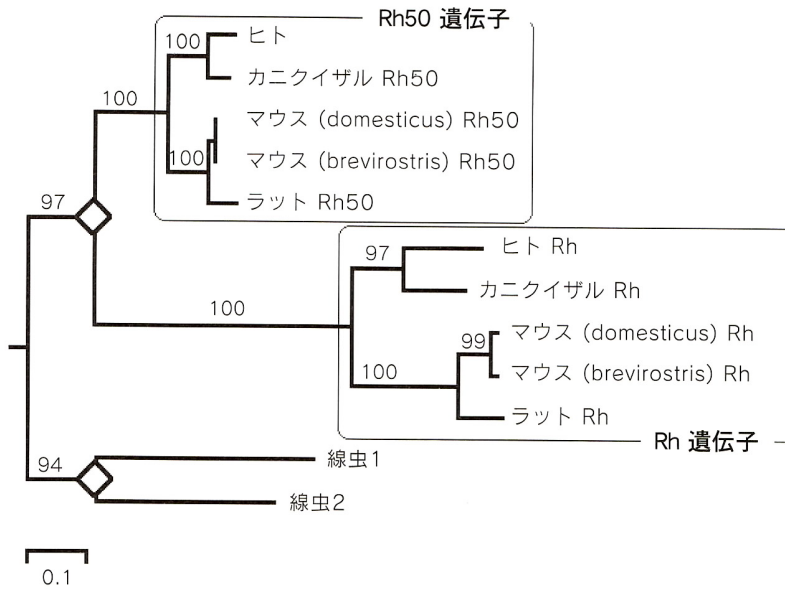


図-1 Rh式血液型遺伝子およびその相同遺伝子の進化系統樹. 文献1) より.

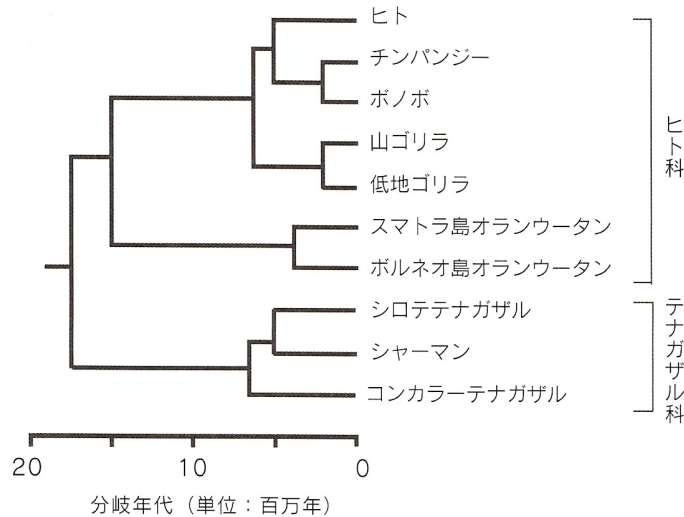


図-2 ヒトと類人猿の系統関係

パンジーとボノボはどちらも系統的には同じようにヒトに近いことが、系統樹から読みとれる。

では、ヒトに至る進化系統における遺伝子の独自の進化を知るには、具体的にどうしたらよいのだろうか。ヒトに最も近い種はチンパンジー（およびボノボ）であるが、それとの共通祖先Xから分岐したあとの、ヒトの

独自の進化は、チンパンジーのゲノム配列を調べることによって初めて知ることができる。しかしそれでもまだ不十分である。我々は現生種であるヒトとチンパンジーのゲノム配列しか知ることができないので、図-3のようにヒトがA、チンパンジーがTであった場合、違いの生じた方向は分からない。

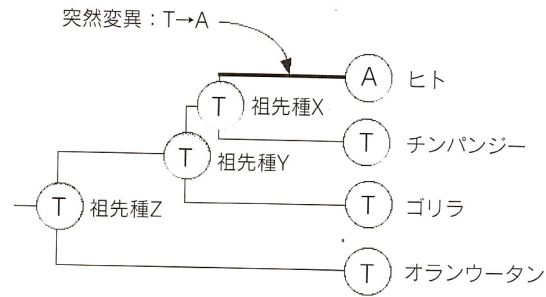


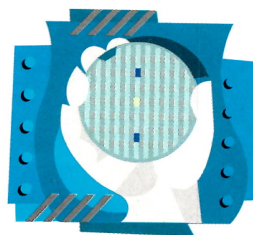
図-3 ヒトの系統独自の塩基変化を抽出する方法

ところが、チンパンジーよりも少し進化的に離れている近縁種のゴリラやオランウータンについても、同じ遺伝子の塩基配列を決定すれば、最大節約原理（必要とされる変化を最小とする進化経路を選ぶ）によって、祖先種Xがこの塩基サイト（進化的に相同な塩基の位置）で持っていた塩基を推定することができる。図-3では、ヒトだけが塩基Aであり、他の近縁種はみなTなので、祖先種X、Y、Zはともに塩基Tであり、ヒトへの系統が祖先種Xのところでチンパンジーへの系統と別れたあとに、TからAへの塩基置換が生じたと推定される。このような推定の精度を高くするには、比較する近縁種はなるべく近いものを選ぶ必要がある。

ヒトの独自性を決定する遺伝子

具体的にどのくらいの数の遺伝的変化が人間性を規定しているのだろうか？ チンパンジーやゴリラは系統的にヒトに近いが、彼らの行動パターンや外部形態にはヒトと異なる点が多数存在することはよく知られている。また細胞組織レベルでもさまざまな種特異性があるだろう。これらそれぞれの形質に変化を及ぼした遺伝子変化が存在するはずである。形態に関与する遺伝子については、多面発現の可能性も考慮しなければならない。脳機能の発達を促した遺伝子の変化の中には、言語能力に直接かかわるものもあるだろう。

ヒトとチンパンジーのゲノム全体での差は1.4%程度であるが、遺伝子の情報を担っている部分では、「淘汰上の制約」



のために違いが小さいはずである。実際に、最近我々がヒトと類人猿で決定した30種類の遺伝子配列では、ヒトとチンパンジーで意味のありそうな違いは0.2%にすぎなかった。この違いがゲノム全体における「遺伝子」の平均だと仮定しよう。すると、今度はこれら「遺伝子」領域の割合が問題となる。というのは、霊長類ゲノムの95%以上はいわゆる「がらくたDNA」と考えられているからだ。これらは「遺伝子」としての情報を載せていないとみなされている。この見方が正しければ、遺伝子はがらくたDNAという大海に散らばる群島のようなものである。そこで、仮にゲノム全体における「遺伝子」の割合を4%としよう。ヒトゲノム30億塩基中の遺伝子数は3万個～4万個と推定されているので、これが正しいとすると、1遺伝子あたりの塩基配列長は、平均で3,000から4,000になる。

ヒトとチンパンジーの違いは、ヒトとチンパンジーの進化系統それぞれに生じた変化の和だから、ヒトの系統のみで生じた変化は、進化速度にヒトとチンパンジーで大きな違いがないとすればどちらの種への系統でも同じくらいの変化が蓄積したことになる。つまり、ヒトの系統だけを考えると、2種の間の違いの半分となる。したがって、ヒトの系統で生じた「遺伝子」の変化のうちで表現型に変化を与える可能性のあるのは、12

万塩基（ $= 0.2\% \times 1/2 \times 4\% \times 30$ 億塩基）と推定される。これらの中には、実際には遺伝子の機能に明確な変化を及ぼさないものもあるだろうから、ヒト独自性を与えた変化はそのうちの1割程度、つまりおよそ1万個ぐらいではなからうか。さらに、我々が通常興味を持つヒトと類人猿



図4 類人猿ゲノム計画Silverのトップページ

の間の目に見える表現型の違いに絞れば、それらを与える基盤となる塩基の変化は1,000個前後だろうと私は推定している。

類人猿ゲノム計画

私の研究室では、1999年から小規模ながら類人猿ゲノム計画を開始した。類人猿ゲノムは英語でApe genome, 略してAgである。Agは銀の元素記号でもある。そこで「銀」の英語名(silver)をこの研究のコードネームとした。この計画のWebサイトは<http://sayer.lab.nig.ac.jp/~silver/>である。図-4はトップページである。ここにある「Silver Projectの主要データベース」の中の「Silver計画で塩基配列を決定したものを含む遺伝子」には、上述した300余の遺伝子の一覧が出てくる。各遺伝子について、我々が配列決定してDDBJ/EMBL/GenBank国際塩基配列データベースに登録したエントリー

のリストのあとに、多重配列結果と比較した塩基配列をもとにした系統樹が示されている。

我々は現在、動物の前後軸を決定するのに重要なHoxAクラスター(図-5)の塩基配列をチンパンジー、ゴリラ、オランウータンで決定している。これらの大部分が遺伝子間領域であるにもかかわらず、ヒトと類人猿の違いはゲノム全体での違いより少し小さくなっている。これは、HoxAクラスターの遺伝子間領域が、遺伝子の発現調節に重要であるために淘汰上の制約が働いていることを示している。

一方、理化学研究所ゲノム科学総合研究所の榊佳之プロジェクトディレクターは、チンパンジーの全ゲノム解読計画を進めている。すでにヒトの21番染色体に対応するチンパンジー染色体については、我々国立遺伝学研究所のグループを含めた日本、韓国、中国、ドイツの研究所との間で共同研究が始まっている。これは、2001年3月に理化学研究所と国立遺伝学研究所の共催により東京で開催された類人猿ゲノムに関する国際ワークショップGEMINI(Webサイト<http://sayer.lab.nig.ac.jp/GEMINI/>)から生まれたものである。また、東海大学医学部の猪子英俊教授を中心とするグループは、さまざまな脊索動物のMHC(主要組織適合性抗原遺伝子複合体)配列決定の一環として、チンパンジーのMHCについてもその塩基配列決定を進めている。このように、類人猿ゲノムでは日本が世界をリードしているといえよう。

遺伝子型と表現型をつなぐことの難しさ

類人猿がケモノにとどまっておらず、ヒトだけがなぜこのようなヘンな生き物になってしまったのか? 論理的に考えて、その鍵はヒトゲノムと類人猿ゲノムの違いのどこかにひそんでいるはずである。おそらく脳の働きの何かであろうが、脳だから、ソフトウェアは難しすぎて簡単には解明できないと逃げる必要はないだろう。コンピュータが人間という「神」の創り出した合目的な機械であり、ハードウェアとソフトウェアが明確に区分できるのに対して、生物は長い進化を経てできあがってきた、優れて歴史的な産物である。ここでは、ハードウェアとソフトウェアの区別は明瞭ではない。両者は渾然一体としているのではなからうか? とすれば、ハードウェアたる神経細胞群を構築するの

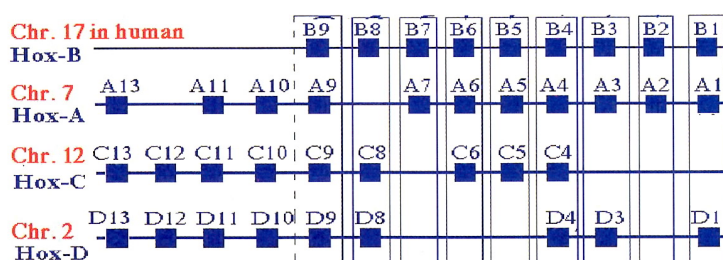


図-5 哺乳類の4個のHox遺伝子クラスター

に必要な遺伝子セットを見つけたら、脳の不思議はほぼ解明できたに等しいかもしれない。

しかしここに、ゲノム配列という「遺伝子型」と目に見える形質である「表現型」をどのようにしてつなげるかという、遺伝学における大問題が存在する。DNAを考えてみよう。その構造は二重らせんであり、4種類のヌクレオチドの非周期的結晶だ。そのエッセンスはA,C,G,Tという4文字のつらなりで表すことができる。この意味で、ゲノム配列は明快である。ところが遺伝子の機能が具現した表現型はどうだろう。これはそう簡単ではない。

我々が森羅万象を把握しようとするとき、それはモノとコトという2つの視点に分けて考えることができる。モノとは物質そのもの。コトは情報、あるいは、物質間の相互関係だ。コトは論理的に記述できる世界であるのに対して、モノはそれが簡単ではない。また、構造はモノで、機能はコトであると考えたくなるが、論理的な記述のしやすさを考えると、構造の方が簡単であり、その意味では相互関係の記述であるコトに近い。機能が漠然としている点は、モノの記述がきわめて困難であることにつながっているように見える。これはどう考えたらよいのだろうか？ ゲノムの塩基配列はデジタルであり、明確かつ簡単に記述できる情報、つまりコトである。これに対して、ゲノム配列の中に埋まっている遺伝子の機能を情報のレベルあるいは論理構造で記述しようとする、著しい困難に出会ってしまう。だいたい、「遺伝子」の「機能」とは何なのだろうか。実はどちらもあいまいな概念なのである。ゲノムの塩基配列はたしかに膨大なデータだが、そこから遺伝子の機能を探ることは、そう簡単ではない。

細胞の中では、転写システムがDNAの上を滑って次々と遺伝子の発現を行っているのだろうが、我々はまだその全体像をつかんではいないのである。もしすべての転写制御システムが解明されたら、その知識をもとにして転写系のシミュレータを作ることができるだろう。そこにゲノム配列を放り込めば、遺伝情報が正しく認識されて、たちどころに発生が始まり、多細胞生物が浮かび上がってくるだろう。しかしこれはまだ夢物語である。

現代生物学は膨大で複雑な物質交代の知識を有しているが、それらは人間がつけた物質の名前の間の相互関係として記述されている。数種類の分子間の反応だけが存在する試験管内の単純な系ならば、それらなりに真理を含んでいると思う。しかし、ゲノム中の多数の遺伝情報を使って繰り広げられている細胞全体の物質交代の振る舞いを記述しようすると、無理が出てくるのではなからうか？ 遺伝子の知識はあらゆる生物現象の解明に使われるようになった。きわめて複雑で錯綜した現象を記述する新しい「ゲノム文法」とでも呼ぶような生命知識の表現法を編み出さなければならない時期に我々は到達している。これは生物学研究だけでなく、情報処理の世界においても、自然言語処理に類似した大きな課題であろう。

参考文献

- 1) 斎藤成也, 北野 誓: ABO式およびRh式血液型遺伝子の進化, 細胞工学, Vol.18, No.7, pp.1039-1047 (1999).
- 2) 榎 佳之, 小原雄治編: ゲノムから個体へ, 中山書店 (2001).
- 3) 斎藤成也: 遺伝子は35億年の夢を見るーバクテリアからヒトの進化までー, 大和書房 (1997).

(平成13年12月5日受付)