

生命システムの進化

進化の歴史をたどる

齋藤成也・太田聡史

1 ゲノム遺伝子の多様性

生命の多様性は、遺伝子の多様性である。ここでいう多様性には、2つの側面がある。ひとつは生物の種類の多様性であり、もうひとつは、ひとつの生物のすべての遺伝子セットである「ゲノム」の中の遺伝子の多様性である。生物の多様性については古来から

【用語解説】

エルンスト・ヘッケル：20世紀初頭に活躍したドイツの生物学者、ダーウィン進化論を推し進めた。

真核生物：大腸菌のような細菌（原核生物）と異なり、真の細胞核を持つ生物の総称。動物、植物、真菌類、原生生物が含まれる。

細胞系譜：多細胞生物が1個の卵から発生していく際に、細胞分裂の繰り返しによって定義される各細胞の履歴。

構造遺伝子：ゲノムの中で、RNA やタンパク質を生成する情報を保持している DNA 領域を指す。

突然変異：生物の塩基配列上に生じるすべての変化パターンの総称。塩基置換、塩基の挿入と欠失、組み替え、遺伝子変換、遺伝子重複、逆位、転座、ゲノム重複などの種類がある。

遺伝子重複：ある DNA 領域（遺伝子）が、細胞分裂に伴うゲノム DNA 複製の際にコピーを生み出すこと、元の DNA 領域の近傍にコピーが生じること（直列重複）が多い。

注目されており、博物学として長い伝統があるが、現在では DNA レベルでの研究が細菌（バクテリア）からヒトまで、すべての生物群で進められている。一方、ゲノムの中にある遺伝子の多様性の研究は、最近始められたばかりである。これは、目に見える生物そのものの研究と異なり、「遺伝子」という概念が19世紀にグレゴール・メンデルの遺伝法則発見によってようやく確立し、ゲノム概念にいたっては20世紀になってから提唱されたという遅いスタートを考えれば、仕方のないことであろう。

ゲノムの中にある遺伝子の多様性をとらえようとする、すぐに気づくことがある。それは、多数の遺伝子を含むゲノムには必ず長い進化の歴史がひそんでいる、ということである。ふつうわれわれが進化という概念を頭に思い浮かべるとき、目に見える個体レベルの進化を考えていることが多い。つまり、ステゴザウルスやヒマワリの進化などを思い浮かべるわけである。しかし、生き物というものはたくさんの遺伝子がはたらいだ結果できあがっているわけであるから、遺伝子の進化という視点で生命進化を眺めることができる。このような遺伝子の進化を調べる研究分野を「分子進化学」と呼ぶ。

従来の分子進化学では、データを得ることが大変だったこともあり、ひとつひとつの遺伝子についてだけ調べることが一般的だった。ところが、多数の遺伝子の塩基配列が大量に報告されるゲノム生物学の時代になると、これら遺伝子の進化を同時に解析することが可能になってきた。この新しい局面で、どのようなことがわかるだろうか？ このような新しい試みのひとつとして、器官や組織がどのように進化してきたのかをわれわれは調べている。本稿では、この新分野でわれわれが最近得た研究成果^[1]を中心に紹介する。

2 発生と進化

われわれのからだはたくさんの細胞からできあがっているが、それぞれの細胞は同じというわけではない。筋肉には筋肉特有の細胞が、血管には血管特有の細胞が、脳には脳特有の細胞がある。つまり、細胞はそれぞれ個性を持っており、似たような性質を持った細胞が集まって、ある特定の組織を形づくっている。細胞の個性を決めているのは、その細胞がどのようなタンパク質からできあがっているかである。ある細胞にどんな種類のタンパク質が

現れるかを決めているのは、遺伝子である。つまり、細胞間の違いというのは、それぞれの細胞ではたらいっている遺伝子の違いに他ならない。

エルンスト・ヘッケルは、「個体発生は系統進化を繰り返す」というスローガンを唱えた。細かい点では彼の説は必ずしも正しくはないが、彼の考えが優れた洞察に富んでいるのは確かである。彼は、発生と進化というまったく違った2つの現象の間に類似点を見いだした。では、なぜそんなことが起きるのであろうか。発生と進化の間にはどんな関係があるのであろうか。

現在の真核生物の多くは多細胞生物であるが、その祖先はパン酵母のような単細胞生物だったのであろう。やがて、細胞の数が増えるほど生物の構造は複雑になり、細胞は異なった機能を持つようになる。われわれが「組織」と呼ぶ細胞の集団が、ある時点で出現し、現在見ることのできる複雑な多細胞生物の体を形づくるようになったと考えることは、ごく自然なことである。われわれはこのような現象を、個体発生の分化のアナロジーとして、「進化的分化」と呼んでいる。個体発生の分化において細胞系譜があるように、進化的分化における組織の系統関係が存在するといっているのである。生物の組織には、組織特異的な遺伝子の発現がある。これらの遺伝子を分子進化学的手法で解析すれば、組織の進化を推定できるかもしれない。

これは一見したところうまくいきそうである。しかし、ここに落とし穴がある。われわれが、ある生物の組織の進化を考えたとき、組織特異的な遺伝子は同じゲノム上にあることを忘れてはいけぬ。われわれは「現在」の組織特異的な遺伝子を知っているが、それが進化の過程上ずっと組織特異的であった保証はないのである。もし、ある組織特異的な遺伝子が、過去において違

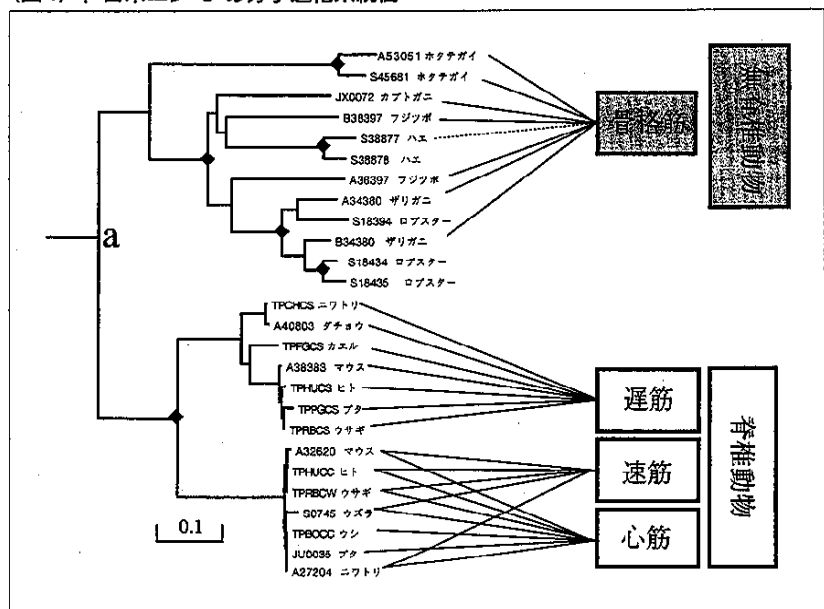
ったパターンで発現をしていたら、組織の系統関係と組織特異的な遺伝子の系統関係の間にはずれが生じてしまう。なぜこんなことになってしまうのであろうか。

それは、進化的分化の様相を知るために、構造遺伝子を使っているからである。確かに組織の個性を決めているのはこの構造遺伝子であるが、その発現パターンを決めているのはもっと別なものである。その別なものは、ゲノム上の調節領域である。この調節領域に、ある種のタンパク質が結合することによって、構造遺伝子の「スイッチ」が入り、発現が始まるわけである。したがって、組織の進化を知りたいければ、むしろゲノム上の調節領域の分子進化的解析をするべきである。しかし、われわれは調節領域についてあまり多くの知識を持ち合わせていない。その配列情報もほとんどわかっていない。では、どうすればよいのだろうか。

3 構造遺伝子を使った組織の進化の推定

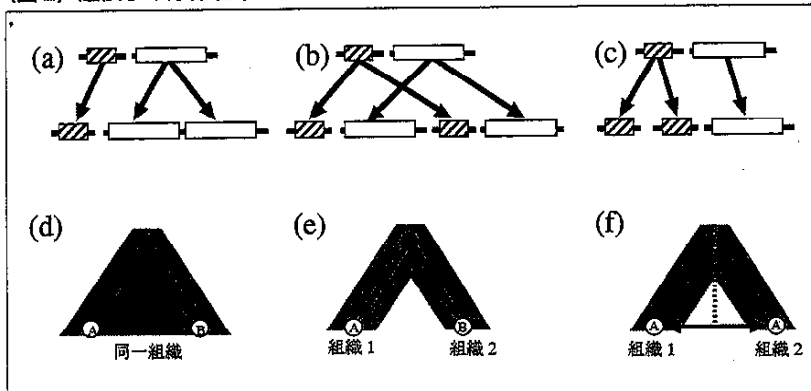
構造遺伝子のデータなら山ほどあるので、ここであきらめずに、なんとか構造遺伝子を使って調節領域の進化の様子を推定してみよう。実は、筋肉組織に発現するトロポニン C というタンパク質のアミノ酸配列情報を用いて、このタンパク質の遺伝子の系統樹を描いてみると、おもしろいことがわかる(図1)。各遺伝子と発現パターンの対応関係を調べてみると、同じ組織に発現する遺伝子は、生物種が違っていても互いに似ているのである。これは決して自明なことではない。なぜなら、前に述べたように、構造遺伝子がどの組織に発現するかを決めているのは構造遺伝子自身ではなく、ゲノムの別の場所にある調節領域だからである。もし、トロポニン C で見られたような現象が、他の遺伝子でも見られるとすると、構造遺伝子と調節領域の間にはある程度、進化的な相関があると考えることができる。

〈図1〉トロポニン C の分子進化系統樹



黒い菱形は遺伝子重複を表す。a は脊椎動物と無脊椎動物の分岐の上限を、スケールバーはアミノ酸の部位あたりの置換数を表す。

(図2) 遺伝子の分化(上)と組織での発現の対応(下)



(a)構造遺伝子は重複したが、その調節領域はそのまま。(b)構造遺伝子とその調節領域が重複した。(c)構造遺伝子はそのままで、その調節領域が重複した。(d)遺伝子重複(黒丸)がAとBという同一の組織に発現する相同な遺伝子をつくり出した。(e)重複した遺伝子AとBが異なった組織1と2に発現した。(f)同じ遺伝子Aが異なった組織1と2に発現した。

実際、同じような現象が他の筋肉組織に発現する遺伝子でも見られるのである。このような遺伝子の発現パターンと遺伝子の進化との関係を説明するには、以下に示すような仮説を導入する必要がある。構造遺伝子と調節領域の考えられる3種類の関係を示しているのが図2a～cである。aでは調節領域がそのままであるが、構造遺伝子だけが重複している。もちろん、もっと複雑な調節が2つの遺伝子にはたらくことも考えられるが、ここでは図示した調節領域がおもなはたらきかけをしていると考えるのである。いったん重複した遺伝子は、突然変異を独立に蓄積していき、異なった遺伝子へと分岐していく。このようなメカニズムがはたらいた結果、われわれは2つの違った遺伝子が同じような調節を受けて、同じ組織に発現するという現象を観察することになる(d)。

bでは、調節領域と構造遺伝子が2つとも何らかの因果関係を持って、あるいはたまたま偶然にか、とにかくほぼ同時期に重複した様子を表わしている。2つの構造遺伝子は異なった調節領域の支配下にあるので、異なった組織に発現することになる。その結果、われわれは2つの違った遺伝子が

別々の調節を受けて、別の組織に発現するという現象を観察することになるのである(e)。この場合、構造遺伝子は確かに組織の進化を反映している。

最後に、cではaとは逆に、調節領域だけが重複して、構造遺伝子はそのままという状況を表わしている。この場合は何が起きるのであろうか。われわれが観察するのは、異なった組織に同じ遺伝子が発現するという現象である(f)。この場合は、2つの組織に同じ遺伝子が発現しているわけであるから、それらの組織は比較的最近になって共通祖先組織から進化したと解釈することができる。

以上、3つのパターンを念頭に、注意深く構造遺伝子の分子進化系統樹を見ていけば、組織の進化を推定できる。しかし、ある遺伝子を調べて、たまたまbのパターンに出会えば理想的であるが、必ずしも都合のよいデータばかりではない。cのパターンは、2つの組織が系統樹の中でお互いに近い関係であるということしか言ってくれない。aのパターンに至っては、何の情報ももたらしてはくれないのである。つまり、ある遺伝子系統樹を発現パターンに表したからといって、すべてがただちに明らかになるわ

けではないのである。

4 遺伝子系統樹の重ね合わせ

以上のようなことを踏まえ、われわれはひとつひとつの分子進化系統樹がもたらす情報が断片的であることを予想して、分子進化系統樹から推定された、不完全な系統樹を重ね合わせるという方法をとることにした。もし組織の系統樹が、あるひとつの樹形を強く支持するようであれば、われわれは調節領域と構造遺伝子の進化には何らかの相関があると主張することができる。また、ある系統樹が持っていない情報を、別の系統樹が補完してくれるかもしれない。

しかし、系統樹を重ね合わせることは、いったいどうことなのであろうか？ 要するに、これは木構造を持つグラフを重ね合わせていくということであるが、いちばん最初に思いつくのは、木の節の対応関係を調べることである。こう考えれば、系統樹の重ね合わせは、一次元の節の整列の問題に帰着できる。

しかし、このやり方には問題がある。まず、どの節とどの節が対応しているかを、どのように知ることができるのであろうか。つまり、何を評価基準にして、節の整列を最適化したらよいのであろうか。しかも、われわれがやろうとしている系統樹の重ね合わせは、ペアワイズ(2組)の重ね合わせではない。3つ以上の重ね合わせをやろうとしているのである。したがって、上の整列の問題は、多重整列の問題となる。さらに、われわれの望む重ね合わせにおいては、系統樹の葉の数や節の数は互いに違っている可能性がある。このような場合を許さなければ、ある系統樹が持っていない情報を別の系統樹が補完するということが期待できないので、それほど簡単ではな

い。

考え方を考えてみよう。われわれが知りたいのは、複数の組織の系統樹がどのような樹形を支持するかということである。つまり、その複数の系統樹といちばんよく似ているある系統樹を発見したいのである。似ている・似ていないを距離で表現すると、複数の系統樹のそれぞれからいちばん近い距離にある系統樹を見つければよいことになる。すでに、グラフ理論を用いて樹形距離というものが提案されている。樹形距離とは、最低何回の操作（枝の交換）で、ある系統樹が別の系統樹と同じ樹形になるかという尺度である。もし同じ系統樹どうしなら当然、樹形距離は0となる。逆に全然違っている系統樹は（枝の数－葉の数）×2の樹形距離を持つことになる。

では、3つ以上の系統樹それぞれから、いちばん樹形距離の短い系統樹を見つけるにはどうしたらよいのであろうか。われわれは候補の樹形をあらかじめつくっておいて、それと複数の系統樹の間の樹形距離の合計をとるという方法を用いることにした。いちばん小さな樹形距離の合計を持つ系統樹が「重ね合わされた」系統樹というわけである。また、候補の系統樹は、可能なすべての樹形を生成することになっている。つまり、完全探索である。

それでは、実際の推定方法を見ていこう。われわれは筋肉組織の進化について調べてみた。生物の組織は、筋肉に着目すると、脊椎動物では心筋、骨格筋（遅筋と速筋にさらに分ける）、平滑筋、および非筋に、無脊椎動物ではもっと簡単に、横紋筋、平滑筋、非筋に分けることができる。これらの組織は、発生学的に別々の細胞系譜を持っていると考えられている。これら8個のクラスを葉に持つ系統樹は、一見簡単そうであるが、実際にはこの樹形のパターンは10万通り以上もあり、

そのうちのたったひとつの樹形を推定しようというのである。しかも、これは各組織が単系統である場合の話で、実際にはもっと多くの可能性を考えなくてはならない。

われわれは図1で例に示したトロポニンCのほかに、筋肉で必須であるタンパク質として、ミオシン軽鎖（必須タイプと調節タイプ）、ミオシン重鎖、およびアクチンの計5個の遺伝子族を用いて、それぞれの遺伝子系統樹を推定した。これらの間での共通点は、発現する組織が重なっているということである。遺伝子系統樹は近隣結合法や最尤法を用いて推定したが、ここでは詳しく述べない。興味のある方は、参考文献を見ていただきたい。次に、復元した遺伝子系統樹から、組織の系統樹を推定する。ほとんどの場合、発現パターンを描くことにより、組織の系統樹はただちに推定できる。これら遺伝子族の発現パターンは、Web上 (<http://thinker.lab.nig.ac.jp/express/express.html>) で参照できる。

5 系統樹重ね合わせプログラム

われわれは系統樹の重ね合わせをするため、super というプログラムを開

発した。このプログラムは、ICOT（第五世代コンピュータ開発機構）で開発されたKL1言語で書いた。KL1は、KLICというセルフプリコンパイラによってC言語のソースコードに翻訳されるために、すぐれた可搬性を実現しており、また並列処理がきわめて容易に行える。もともと人工知能の研究用に開発された言語だけに、記号処理に強いのも特長である。KL1ではPrologと同じようにネストしたリストデータを簡単に扱うことができる。

たとえば、図1の脊椎動物のトロポニンCの筋肉組織の樹形は、[根、速筋、[遅筋、心筋]]と表現できる。superでは、樹形のリストを生成して、各遺伝子系統樹の樹形データと比較を行い、結果を樹形距離でソートして返してくれる。その結果が表1である。樹形距離の合計であるスコアの小さい上位12個の樹形を示した。この結果から、スコア4の樹形（図3aを参照）が支持されることがわかる。

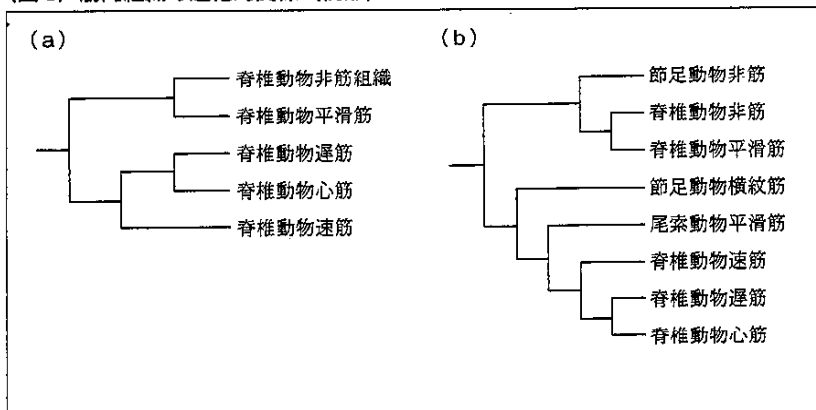
この系統樹で注目すべきは、根（共通祖先）の位置である。脊椎動物の平滑筋が、脊椎動物の非筋組織とクラスターをつくっている。このことは、脊椎動物の平滑筋が、他の筋肉組織とは独立に進化してきたことを意味してい

〈表1〉プログラム super を用いた結果

スコア	樹形	TROC	MEL	MRL	MHC	ACT
4	((0,((3,4),5)),1,2)	0	0	0	0	4
8	(0,1,(2,((3,4),5)))	0	2	2	2	2
8	((0,(3,(4,5))),1,2)	2	2	2	0	2
10	(0,(1,((3,4),5)),2)	0	2	2	2	4
10	((0,((3,5),4)),1,2)	2	2	2	0	4
10	((0,5),(3,4)),1,2)	0	2	2	0	6
10	((0,(3,4)),5),1,2)	0	2	2	0	6
12	(0,1,(2,(3,(4,5))))	2	4	4	2	0
14	(0,1,(2,((3,5),4)))	2	4	4	2	2
14	(0,1,((2,5),(3,4)))	0	4	4	2	4
14	(0,1,((2,(3,4)),5))	0	4	4	2	4
14	(0,(1,(3,(4,5))),2)	2	4	4	2	2

樹形の0,1,2,3,4,5は、それぞれ根、非筋、平滑筋、遅筋、心筋、速筋を意味する。TROCはトロポニンC、MELとMRLは2種類のミオシン軽鎖、MHCはミオシン重鎖、ACTはアクチンを指す。

〈図3〉 筋肉組織の進化的関係 (樹形)



枝の長さは進化時間に比例していない。(a)もっとも可能性の高い脊椎動物の筋肉組織の進化的関係、(b)もっとも可能性の高い動物の筋肉組織の進化的関係。

る。もうひとつの重要な点は、脊椎動物の遅筋（骨格筋のひとつ）が心筋とクラスターをつくっていることである。われわれは骨格筋をひとくりに扱っているが、分子レベルの進化では、骨格筋は単系統ではないということになる。

プログラム super で重ね合わせを行った結果、いちばん支持される組織の系統樹は b のようになる。この系統樹のもっとも特徴的なことは、節足動物の組織の位置である。節足動物と脊椎動物は7億年ほど前に分かれたと考えられるが、にもかかわらず脊椎動物の平滑筋・非筋組織は、節足動物の非筋組織と祖先を共有している。同様に、脊椎動物の骨格筋・心筋のグループも、節足動物と祖先を共有している。つまり、われわれの平滑筋・非筋組織と、骨格筋・心筋組織は、7億年以上独立に進化してきたことになる。もうひとつのおもしろい点は、尾索動

物の平滑筋の進化上の位置である。尾索動物（ホヤの仲間）は、われわれ脊椎動物とともに脊索動物門のメンバーだが、b の系統樹は脊椎動物の骨格筋や心筋と尾索動物の平滑筋が祖先を共有していることを示している。つまり、尾索動物の平滑筋と脊椎動物の平滑筋は由来が別物ということになる。

6 おわりに

ゲノム生物学の時代は、生物進化の研究にも新しい光を投げかけている。機能がどの生物でも似かよっていて、相同であることが明らかな遺伝子のグループだけを調べていた分子進化学の幼年期は過ぎ去りつつある。というよりも、それは、もはや遺伝子を扱う生物学では常套手段となりつつある。ゲノムの中に遺伝子として貯えられている情報が、生物のどのようなシステムを背景に、進化という長大な時間の中

で織られているのか。システムが何もない状況から、なぜ「われわれ」というシステムの固まりのような人間が生じてきたのか？ これらの疑問を解き明かすためには、生物進化の歴史を背負った遺伝子の進化を明らかにすることが必須である。つまり、ゲノム生物学は、すなわち「ゲノム進化学」なのである。

参考文献

- [1] Oota and Saitou : *Molecular Biology and Evolution*, 印刷中, 1999.
- [2] 斎藤成也 : 遺伝子は35億年の夢を見る—バクテリアから人間の進化まで, 大和書房, 1997.
- [3] 宮田隆編 : 分子進化—解析の技法とその応用, 共立出版, 1998.
- [4] Lewin 著, 斎藤成也監訳 : DNAから見た生物進化, 日経サイエンス社, 1998.

● 斎藤成也
(さいとう・なるや)
国立遺伝学研究所
本稿では、太田聡史君の博士論文の一部を紹介しました。私の研究室では、遺伝子の進化をさまざまな側面から研究しています (URL: <http://smiler.lab.nig.ac.jp/>)。

● 太田聡史
(おあた・さとし)
国立遺伝学研究所
進化学は生命の歴史を対象としています。しかし、この一過性の時間の流れの中にこそ、私たちが真理と呼ぶものの正体が隠されているのではないのでしょうか。