

祖先遺伝子推定による古代タンパク質の解析

白井 剛

琥珀に閉じ込められた白亜紀の蚊が吸った血液から、DNAを抽出し恐竜をよみがえらせる。『ジュラシック・パーク』はそういうストーリーのSF小説である。もちろん、これは実現された技術ではないが、遺骸や化石から古代DNAを抽出し塩基配列を決定することは、すでに分子古生物学の分野として確立している¹⁾。ところで、古代DNAの復元を可能にするもう一つの手段がある。それはバイオインフォマティクスの手法を用いて、「計算で再現する」方法である。本稿では、この手法を用いて祖先型の魚類ガレクチン（糖鎖結合タンパク質）を再現し、その立体構造の解明に成功した研究について解説する²⁾。

系統樹による祖先遺伝子の推定

近縁な生物種の相同（進化的に類縁関係にある）遺伝子のDNA配列は類似しており、配列の相違度はその遺伝子の分岐後年代の関数（進化距離）として表すことができる。これを利用して作成されるのが分子系統樹であ

る（図1）。系統樹のブランチ（枝）の長さは遺伝子の進化距離を示し、各ブランチはノード（節）で連結され1つの樹状グラフを形成する。ノードにおけるブランチの分岐は遺伝子の分岐を示す。

分子系統樹の推定法には、最大節約法、近隣結合法、最尤法などさまざまな方法があるが、そのうちのいくつかは、ノードにおけるDNA配列の推定に基づいて進化距離を算出する。系統樹のノードとは、そのノードから分岐した（図ではノードの右側に連なる）遺伝子が同一であった時点、すなわち共通祖先型に相当する。すなわちノードにおけるDNA配列は、かつて存在した祖先遺伝子のものである。すでに分子生物学の発展は、配列情報さえあれば、その遺伝子のコードするタンパク質を実験室で生産し、現存するタンパク質と同様に実験することを可能にしている。これは祖先遺伝子配列を情報推定し、タンパク質の再現が可能である事を意味する。すでにこの技術によって、細菌の翻訳開始タンパク質が年代を遡るほど耐熱性が高く、原始生物は耐熱性であったと推定され

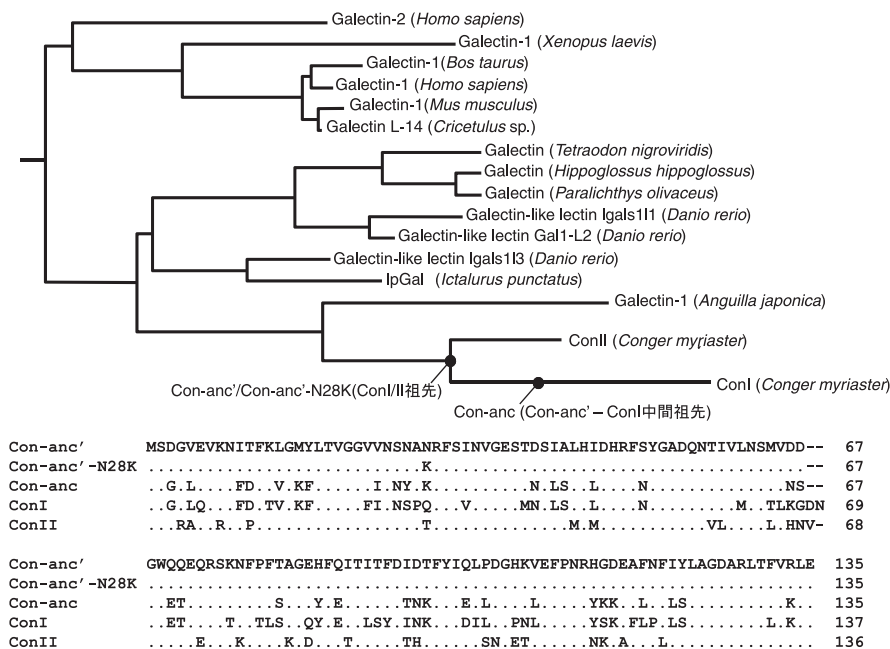


図1. (上) コンジェリン (ConI, ConII) と類縁遺伝子の分子系統樹。Con-anc' と示した黒点がConI とConII の共通祖先 (Con-anc'-N28KはCon-anc'でもっとも不確かな残基Asn28を同程度に確からしいLysに置換した変異体である)、Con-ancで示した黒点はCon-anc'からConIへの進化過程の中間祖先の系統樹上の位置を示す。(下) ConI, ConII, Con-anc'などのアミノ酸配列アライメント。Con-anc'と同じアミノ酸は点で示している。

著者紹介 長浜バイオ大学大学院バイオサイエンス研究科（教授） E-mail: t_shirai@nagahama-i-bio.ac.jp

ることや³⁾、緑色蛍光タンパク質の蛍光波長は緑色が祖先型であり、赤色や黄色はそこから派生してきたことなどが示されている⁴⁾。

近年のゲノム解析や構造ゲノム解析（ゲノムにコードされるタンパク質の網羅的解析）は、生物の遺伝子について以下のことを明らかにしてきた。すなわち、生物進化の過程で新しい遺伝子が獲得されることはしばしばあるが、それはほとんどの場合、既存の遺伝子の重複（遺伝子コピー）と、それに続く遺伝子の分化による⁵⁾。つまり遺伝子進化においてオリジナリティーは稀で、基本は使い回しである。それでは、一つの遺伝子をテンプレートに用いて新たな機能を創出する分子進化過程とはどのような物だろうか？

遺伝子重複進化モデルとしてのコンジェリン

重複遺伝子が機能的に分化する過程は、自然界で長い年月をかけて行われており、その過程を観察し実験することは容易でない。たとえば、マアナゴの体表粘膜にはコンジェリンI（ConI）、コンジェリンII（ConII）と名付けられた相同タンパク質が存在する（図2）。この2つのタンパク質は、アミノ酸配列で27%の一致度を示し、比較的最近に遺伝子重複によって生まれたと推定される⁶⁾。コンジェリンはβ-ガラクトースを含む糖鎖に特異的な糖鎖結合タンパク質ガレクチンの一種であり、生体防御機能を持つと考えられている。それはコンジェリンが2つ

の同一タンパク質（サブユニット）からなる複合体で、サブユニットごとに1つの糖鎖結合部位を持つため、ちょうど抗体のように2価結合能を示し、細菌や寄生生物の胚細胞を凝集させ、成長抑制や細胞死誘導する機能を持つからである。

遺伝子重複で誕生したConIとConIIの間には、現在ではいくつかの違いが見られる。ConIは熱安定であり、50°Cでインキュベートしても細胞凝集活性が失われぬのに対し、ConIIはほぼ完全に失活する。またβ-ガラクトースを含むさまざまな糖鎖に対する結合能も両者で異なっている。ConIとConIIの遺伝子配列から、両者は加速進化している事が示される。加速進化とは、アミノ酸を変化させる非同義（DNA置換）率（ K_A ）と、アミノ酸を変化させない同義置換率（ K_S ）の比 K_A/K_S が高い（ >1 ）事を意味する。ConIとConIIの遺伝子間の K_A/K_S は2.7であり、通常よりかなり高い。これは機能に影響されるダーウィンの進化が、機能に影響されない中立的な進化を大きく上回って起こっていることを示し、遺伝子が強い自然選択に曝されていることを意味する⁷⁾。これは免疫や生体防御に関連する遺伝子にしばしば見られる現象である。

以上の観察から、ConIとConIIは遺伝子重複後に急速に機能分化しつつあると推定され、機能分化による進化を研究するよいモデルであると考えられた。そこで、ConIとConIIの構造をX線結晶解析法で調べたところ、両者は類縁関係にあるので全体構造はよく似ていたが、サブユニットの境界で大きく構造が異なっていることが分かった^{8,9)}。ConIではConIIに比べてアミノ末端のβストランドの位置が異なり、サブユニット間で交換（ストランド=スワップ）を起こしている（図2）。このストランド=スワップによりConIではConIIに比べて、2つのサブユニットを分離することが困難になっており、これがConIの耐熱性が高い理由であると推定された。

祖先型コンジェリンの再現

このようにConIとConIIは、タンパク質のフォールド（ペプチド折りたたみ構造）の変化を伴う、かなりダイナミックな進化を経て機能分化していることが示された。それでは遺伝子重複する前のコンジェリンの祖先（Con-anc'と呼ぶ）は、どのような機能と構造を備えた分子だったのだろうか？ 当然それは、現存するConIともConIIとも異なる、とうの昔に失われた分子である。

そこで、図1に示した系統樹に基づいて、Con-anc'の配列を最尤法により推定した。これは現存遺伝子配列を観察結果と見なして、事後確率が最大になるノード配列

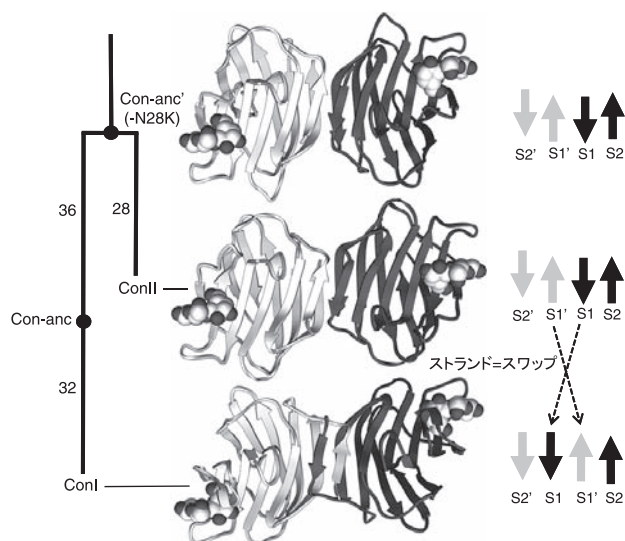


図2. (左) Con-anc', ConI, ConIIの系統樹と(中央)それぞれの結晶構造。それぞれ2つのサブユニットを、白とグレーで示し、結合糖鎖は空間充填模型で示している。(右)それぞれのサブユニット境界（分子モデルの前面中央に相当する）のβストランドの模式図。Con-anc'とConIIは同じトポロジーであるが、ConIではアミノ末端のS1とS1'がサブユニット間で入れ替わっている（ストランド=スワップ）。

を推定する方法である¹⁰⁾。続いて、推定した遺伝子配列に基づいてDNAを合成し、大腸菌発現系を形質転換することで古代タンパク質Con-anc'を実験室で再現した¹¹⁾。

Con-anc'に対しさまざまな生化学的実験を行ったところ、Con-anc'は現存タンパク質(ConIとConII)よりも熱安定性もアポトーシス誘導活性も低いことが示された。また、 β -ガラクトース特異性は維持されていたが、Con-anc'のさまざまな糖鎖に対する結合能はConIともConIIとも異なっていた。

さらに、Con-anc'のX線結晶解析により、1.4Å分解能での立体構造解析にも成功した²⁾。これは祖先配列を持ったタンパク質としては2例目、全長構造としては初の構造解析である。Con-anc'の立体構造は、いくつかの点でConIとConIIの「祖型」を示した。まず、ConIで見られたストランド=スワップによる構造強化は認められない。これは、現時点でこのタイプのフォールドを持つガレクチンがConIだけであり、この進化がConI系統で特異的に起こったとする推定と一致する。また、Con-anc'のサブユニット間の相互作用と糖鎖との間の相互作用は、ConIやConIIに比べて弱いことが示された。これらの観察から、遺伝子重複後ConIとConIIの両系統に、構造安定化と糖鎖結合能力の強化という同じ選択圧がかかっているにも関わらず、その圧に対する遺伝子の応答は異なっていて、構造的にも機能的にも分化しつつあると考えられる。

タンパク質進化による機能分化の解析

コンジェリンの機能構造分化を、定量的に解析した結果を図3に示す。ここでは、データを補うために、Con-anc'と同程度に祖先型らしい1アミノ酸変異体Con-anc'-N28Kと、Con-anc'とConIの中間に相当する祖先タンパク質Con-ancも同様に配列推定・構造解析した結果を加えて解析している。図3は、熱安定性・糖鎖結合能・さまざまな構造パラメータ(サブユニット間相互作用や分子動力学計算から求めた立体構造の安定性など)とCon-anc'を含むコンジェリン間の進化距離の相関係数を求めたものである。この時、進化距離はCon-anc'から測った距離とConIから測った距離の両方を使用している。これは、前者はConIとConIIで共通した進化の傾向を持つ性質と高い相関係数の絶対値を示し、後者はConI系統において強化(または劣化)しつつある性質と高い相関係数の絶対値を示すと考えられるからである。

相関係数の絶対値で対角化クラスタリングを行うと、熱安定性・尿素変性に対する耐性などの構造安定性に関するパラメータは、主にConIからの進化距離と相関している。すなわち、ConIとConIIは主に構造安定性において分化しつつある。一方、機能(アポトーシス誘導活性)についてはConI, ConII両系統で上昇しつつある。また、構造安定性に関与するパラメータの中でも、サブユニット間相互作用(水素結合と接触面積)については

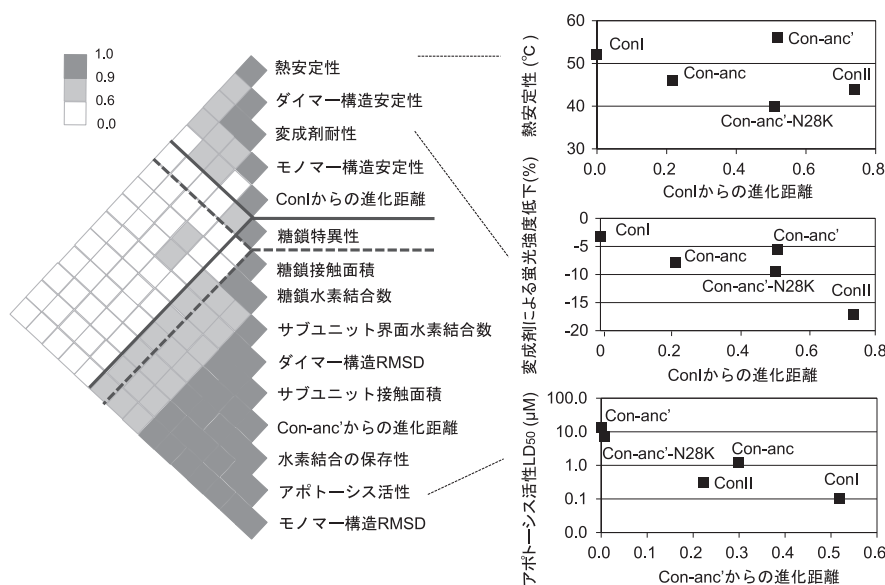


図3. (左) ConI, ConII, Con-anc', Con-anc'-N28K, Con-ancのさまざまな機能・構造パラメータ間の対角クラスタリングされた相関係数マトリックス。パラメータ間の相関係数は絶対値として、左上のグレースケールに従って表示されている。実線・破線はそれぞれ、これらのパラメータを2分する第一・第二境界を示す。(右) 機能・構造パラメータのうち、(右上) 熱安定性と(右中) 変性剤による蛍光強度低下とConIからの進化距離の相関、(右下) アポトーシス活性LD₅₀値のCon-anc'からの進化距離の相関をグラフで示す。

どちらの系統でも増大している。全体構造の微妙な差を表す指標（構造重ね合わせを行った際のC α 原子間距離の根平均自乗距離RMSDと水素結合パターンの保存性）は、この進化距離と相関し、ストランド=スワップなど不連続で大きな変化を別とすれば、構造は進化に従って連続的に変化することを示している。

興味深いのは、2つのクラスターの境界付近に糖鎖特異性が位置する点である。コンジェリンは生体防御機能を持ち、結合する糖鎖は外敵を認識するためのマーカーの役割を持つと考えられる。免疫系のタンパク質が加速進化しやすいのは、外敵の分子が急速に変化するのに対抗して、自身も変化し続ける「軍備競争」が果てなく続く結果と考えられている。この場合、コンジェリンは外敵が変化させる細胞表面糖鎖や、新たな外敵の糖鎖に絶えず対応する必要がある。つまり、糖鎖認識は定向進化し難いと推定されるが、解析結果はこの推定を支持していると考えられる。

祖先遺伝子推定の問題点

祖先遺伝子推定法に問題がない訳ではない。最大の問題は、祖先配列の推定は確率に基づいたものに成らざるを得ないということである。100アミノ酸から成るタンパク質があるとして、そのタンパク質の祖先配列が、各アミノ酸サイト0.99の尤度で特定されたとする。この場合、タンパク質のアミノ酸配列全体の尤度は $0.99^{100}=0.366$ に過ぎない。全アミノ酸サイトで0.99を上回る尤度での推定は、通常では期待できない高水準である。また、祖先配列の推定に用いる遺伝子は、現存あるいは過去に存在した配列の、ほんの一部に過ぎないのが普通である。そのため、新たな観察（新規に決定された配列）により、推定祖先配列は大きく影響を受ける可能性がある。

よって、祖先配列をどこまで信頼してよいかという問題は常につきまとう。この問題を検討するために、Con-anc'について以下の様な検定法を考案した。将来に観測される配列がどのようなものかは予測不能である。そこで、もし現在の系統樹（図1）に用いられている配列が未知であったらと仮定し、無作為に1個ずつ配列を排除して祖先配列を推定し、祖先配列がどの程度変動するかを調べた。系統樹から遺伝子を排除していくと、当然の結果として、推定される祖先配列はCon-anc'配列から排除遺伝子数におおむね比例して離れてゆく（図4左）。また、1遺伝子（総計n+1）排除による、前回推定配列（n遺伝子排除）からの変動も、排除遺伝子数に

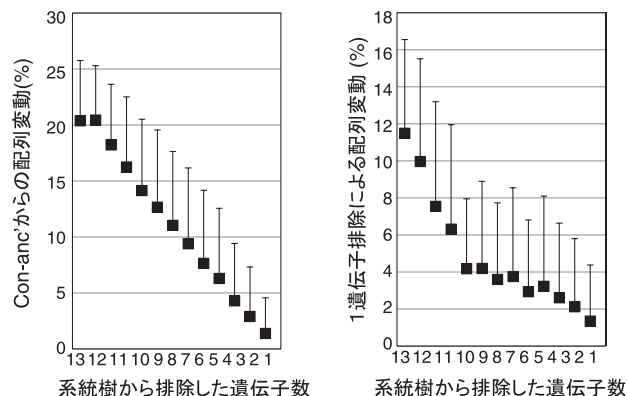


図4. 図1に示す系統樹から順次遺伝子配列を排除しConIとConIIの祖先配列を推定した場合の、(左) 排除遺伝子数（横軸）に対する祖先配列のCon-anc'に対する配列一致度の変動と、(右) 1遺伝子を排除した時の祖先配列一致度の変動を示す。

依存して増大する（図4右）。ここで、今後新たな遺伝子配列が決定され系統樹に取り込まれる事態は、図4の結果の外挿に相当し、Con-anc'からの変動はせいぜい2%以下であると想定される。つまり、Con-anc'配列は十分に収束しており、よほど意外な配列が発見されない限り、現在の配列から大きく変化しないと考えられる。

まとめ

以上のように、配列推定による祖先遺伝子の再現は、実物を実験できるという点で大きな魅力がある。また、分子生物学と情報生物学の融合により新たに手に入った、大過去を覗き込む手段というロマンもある。実用的な面では、ヒト適応する以前のウイルスタンパク質を再現し、実験するといった応用も期待できる。これは、ウイルスの宿主免疫回避の戦略を解明するために強力な手段となるだろう。

文 献

- 1) Rizzi, E. et al.: *Genet. Sel. Evol.*, **44**, 21 (2012).
- 2) Konno, A. et al.: *Structure*, **19**, 711 (2011).
- 3) Gaucher, E. A. et al.: *Nature*, **425**, 285 (2003).
- 4) Ugalde, J. A. et al.: *Science*, **305**, 1433 (2004).
- 5) Wolfe, K. H. and Li, W. H.: *Nat. Genet.*, **33**, 255 (2003).
- 6) Ogawa, T. et al.: *Biosci. Biotechnol. Biochem.*, **63**, 1203 (1999).
- 7) Shirai, T. et al.: *Tanpakushitsu Kakusan Koso*, **48**, 1913 (2003).
- 8) Shirai, T. et al.: *Structure*, **7**, 1223 (1999).
- 9) Shirai, T. et al.: *J. Mol. Biol.*, **321**, 879 (2002).
- 10) Yang, Z.: *Mol. Biol. Evol.*, **24**, 1586 (2007).
- 11) Konno, A. et al.: *Mol. Biol. Evol.*, **24**, 2504 (2007).