

## 原核生物のレトロエレメント：遺伝子の組み合わせを指標に

逆転写酵素をコードする因子を総称してレトロエレメントと呼ぶ。ただし、レトロエレメントの呼称は、逆転写によって転移する SINE や retrocopy も含んで使われる場合もあるので注意が必要である。また、ここでは転移や増殖をしなくても逆転写酵素をコードしているテロメラーゼなどもレトロエレメントと呼ぶ事にする。真核生物のレトロエレメントには、レトロウイルス、ヘパドナウイルス、カリモウイルス、LTR レトロトランスポゾン、YR レトロトランスポゾン、*Penelope*-like element、non-LTR レトロトランスポゾン、テロメラーゼがある。また、原核生物のレトロエレメントでは、レトロン、グループ II イントロン、diversity-generating retroelement が報告されている。グループ II イントロンは真核生物にも存在するが、そのほとんどは原核生物由来のミトコンドリアと色素体にあるため、原核生物のレトロエレメントとして分類される。他にも AbiA や AbiK など機能が報告されている逆転写酵素があるが、総じて原核生物ではレトロエレメントの報告は少ない。

しかしながら、原核生物のゲノム配列情報が増えるにつれて、大量の逆転写酵素配列が見つかって来ている。そこで Kojima らはこれらを機能的に分類すべく、網羅的な解析を行なった (Kojima and Kanehisa 2008)。原核生物のゲノム配列を見ると、逆転写酵素であるという情報は遺伝子に与えられているのだが、どのような役割を持つのか、既知のどのグループと近いのかといった情報はわからない。従って逆転写酵素だけ見ていては機能の予測は困難である。そこで著者らは逆転写酵素と隣り合う遺伝子、あるいは融合した蛋白質ドメインの機能を推測することで逆転写酵素の役割を解析することにした。

まず、2006 年 7 月の RefSeq データセットを利用して、その中の逆転写酵素遺伝子を網羅的に取得した。PSI-Blast を繰り返し、新しい遺伝子が取れなくなるまで続けた後、HMMER を用いて逆転写酵素のモチーフを持っているものだけに絞り込んだ結果、516 個の配列が得られた。内 7 個は真核生物の配列とそっくりだったのでコンタミとして取り除いた。残る 509 個の両側にある遺伝子を分類し、最終的に逆転写酵素を 12 種類のグループに分ける事ができた。

グループ A は SWIM zinc finger domain を持つ蛋白質を逆転写酵素の隣にコードしている。グループ A は既知のレトロンである Mx65 と Mx162 を含んでいたためレトロンの 1 系統であることがわかった。グループ B は機能未知の蛋白質ドメインである ribosomal S23-like を持つ蛋白質をコードしていた。このグループは既知の diversity-generating retroelement である DGR\_N.punctiforme、DGR\_N.7120-2 を含んでいたため diversity-generating retroelement の 1 系統であるとわかった。グループ C も diversity-generating retroelement の 1 系統で HRDC

(Helicase/RNaseD C-terminal) ドメインを持つ蛋白質を隣にコードしていた。HRDC ドメインは核酸と結合する特徴がある。また、グループ B、C 共に、diversity-generating retroelement の特徴である IR と VR という配列を含んでいた。

A から C 以外のグループは既知の 3 つのグループとは近縁でない事が系統解析で確かめられた。グループ D は DNA polymerase A をコードしていた。この DNA polymerase A は逆転写酵素によって合成された 1 本目の cDNA 鎖を鋳型にして 2 本目の cDNA 鎖を合成する役割を果たしている事が推測できる。

グループ E では真正細菌型の primase と carbon-nitrogen hydrolase に挟まれて逆転写酵素が位置していた。primase は逆転写の際に primer を合成するのに使われると考えられるが、hydrolase がどのような役割をしているのかは不明である。グループ F1 と F2 も hydrolase を持っているが、どちらも逆転写酵素の後ろに融合した形でコードされている。

グループ G は 2 つの逆転写酵素が並んだ形をしている。3 種の原核生物ゲノムから得られたが、お互いの前側同士、後ろ側同士が互いに似ており、同じゲノム中の前後はあまり似ていない。この 2 つの逆転写酵素も何らかの役割分担をしているのだろう。

グループ H1 では逆転写酵素の後ろに cas1 ドメインが融合している。また cas2 遺伝子が周辺に見られる。1 つずつしかメンバーが見つからなかった H2 と H3 でも cas1 が逆転写酵素の後ろに融合していた。また、2 つの逆転写酵素 (グループ H4) が cas1 を隣に別遺伝子として持っていた。これらの持つ cas1、cas2 は CRISPR と呼ばれる特殊な反復配列とセットで見つかり、原核生物の獲得免疫系として働く遺伝子である。(「塩基配列レベルの獲得免疫: CRISPR-Cas system」参照) この CRISPR-Cas system では CRISPR に短い外来 DNA を挟み込み、その配列を含んだ外来 DNA を何らかの形で排除する。逆転写酵素は RNA フェージのゲノム RNA を逆転写し、CRISPR に挟み込む事で、RNA フェージへの免疫を付与するのに働いているのかもしれない。

グループ D、E、F1、F2、G は既知のどのレトロエレメントとも近縁ではなく、しかも複製あるいは機能に関与すると思われる蛋白質を隣に持っている。著者らはこれらのレトロエレメントが転移するかどうかを明らかにするため、その周囲 40kb の配列を近縁ゲノムから探索し、orthologous な位置に挿入があるかどうかを調べた。

グループに関わらず、レトロエレメントの挿入は 2 タイプに分けられた。1 つはプロフェージの上に乗っている場合であった。プロフェージの両側にできる直列反復配列 TSD も見つかる事から、プロフェージに乗ってレトロエレメントが挿入された事がわかった。しかし、プロフェージの中に、レトロエレメント

がどのように挿入されたかを示す情報は得られなかった。この状態はレトロンや diversity-generating retroelement でも知られている。

もう1つのタイプでは、超多型座位にレトロエレメントが見られた。超多型座位では、種や株ごとに全く異なった配列が見られる。配列に類似性は見られず、これらは外来 DNA が何らかの機構で挿入されてできたものと考えられる。すなわち、超多型座位は挿入のホットスポットであり、宿主側の機構、あるいは性質、によって外来 DNA が取り込まれていると考えられる。そのためか、レトロエレメントの両側に TSD は見られない。超多型座位への挿入は diversity-generating retroelement やレトロンでも観察された。

どちらのタイプでも、グループ II イントロン以外のレトロエレメントが自立的に転移していることを示す結果は得られなかった。むしろ宿主やファージの機構によって偶然取り込まれていると考えた方が自然である。

グループそれぞれのレトロエレメントがどの生物のゲノムに存在しているかを見ると、近縁種に挿入されているものは少数で全く系統の離れた数種で見られるものが多い。また逆転写酵素の配列は互いに似ているとはいえ、それほど相性は高くない。上記の観察結果も加えて著者らは、これら原核生物のレトロエレメントは染色体外 DNA あるいは RNA として存在し、たまたまゲノムに挿入されたものが見つかったのではないかと考えている。

Simon らも原核生物のレトロエレメントの網羅的な解析を Kojima らよりも遅れて発表している (Simon and Zimmerly 2008)。彼らの逆転写酵素の絞り込みが Kojima らよりも弱いこと、使用した配列データセットが新しくかつ情報の重複を許したもの (2007 年 7 月更新の GenBank) であることから、彼らは 1,049 個の逆転写酵素配列を得ている。GenBank には重複して登録されている配列があるので実際の数はずっと少ないはずである。この配列をアラインメントし、前後に接続された蛋白質配列の情報も利用して、分類を行なった。Kojima らの報告していないグループは 9 つあるが、これらは他に特徴的なドメインを持っていないため、蛋白質配列上の特徴から分類されたものである。Kojima らも発見しているが論文には含めていない。

残念ながらほとんどの新規レトロエレメントの機能や増殖機構は未知のままである。既知の原核生物レトロエレメントの大グループであるグループ II イントロン、レトロン、diversity-generating retroelement はどれも RNA の構成要素を持っている。グループ II イントロンではリボザイム、レトロンでは機能未知の DNA と RNA の複合体である msDNA、diversity-generating retroelement でも IR と TR は RNA レベルで働くと考えられている。従って、上記の蛋白質配列を基にした解析では見つかって来ない、たとえばテロメラーゼのように逆転写酵素と RNA だけで働くレトロエレメントが原核生物には多数存在しているのかもしれ

ない。また、見つかったものでも RNA の構成要素が重要な役割を果たしている可能性は否定できない。原核生物のレトロエレメントはグループ II インترون以外では詳細な解析がなされているとは言いがたい。Kojima らの考えるように見つかって来たものがたまたまゲノムに取り込まれたレトロエレメントであるならば、今後も更に新しいレトロエレメントが見つかるはずである。そのような新規レトロエレメントの探索と並行して、見つかったものの機能解析がなされていく事を期待したい。

Kojima KK, Kanehisa M.

Systematic survey for novel types of prokaryotic retroelements based on gene neighborhood and protein architecture.

Mol Biol Evol. 2008 Jul;25(7):1395-404. Epub 2008 Apr 7.

Simon DM, Zimmerly S.

A diversity of uncharacterized reverse transcriptases in bacteria.

Nucleic Acids Res. 2008 Dec;36(22):7219-29. Epub 2008 Nov 12.

2009/05/13

小島 健司 著  
禁 無断複写転載