ゲノムに刻み込まれた戦略の発見 病原細菌の集団ゲノムデータに基づく新しいアプローチ

次世代シーケンサの技術革新が続き、ヒトゲノム1人分の解読が数十万円で可能な時代が到来した。今日の生命科学を象徴する国際1,000人ゲノムプロジェクトは、現在では世界26カ国の2,500人を対象とするプロジェクトに拡張し、一国の中でも、イギリスで1万人のゲノムを解読するUK10Kプロジェクトが進められ、日本でも1,000人単位のゲノム解読プロジェクトが複数進行中である。ヒト以外の生物にも同様のプロジェクトが存在し、非モデル生物についても、たとえば病原細菌のゲノムを数百個体単位で解読することは一般的となった。こうした潮流の中で、今日の生命科学では、同一種内の集団レベルのゲノムデータの比較に基づいて生物学・医学の問題にアプローチすることが、王道の一つとなった。

それでは、集団レベルのゲノムの塩基配列データを用いてアプローチすべき問題とは何だろうか。さまざまな問題が考えられるが、その一つに、生物が環境に適応し進化するうえでの生存戦略、という問題が挙げられる。生物の生存戦略に関する研究は歴史的に、それが表現型として見えやすい真核生物を中心に、進化生態学と呼ばれる分野で進められてきた⁽¹⁾。これらは、どのような戦略・表現型がどのような環境下で有利となり、それをコードする遺伝子が頻度（コピー数）を集団中で増加させられるのか、に関する研究であり、その多くは、数理モデルを用いた理論的な研究であった。一方、ウイルスのように、自己の生存と拡散のために遺伝子そのものが戦略的に振る舞い進化する（コピー数を増加させる）例が知られるようになり、そうした「利己的」遺伝因子の生存戦略は、進化学・生態学だけでなく分子生物学でも注目を集める研究対象となった⁽²⁾。

こうした生物の生存戦略を、ゲノムの塩基配列とどう結びつけ、塩基配列に基づいてどう理解するのかは、今日の重要な研究課題の一つである。「利己的」遺伝因子に関しては、それ自身と宿主（宿主）ゲノムの塩基配列データの解析による研究が、トップジャーナルに報告されている（たとえば、動く遺伝因子が、植物40種のゲノム間で動き回っているという最近の報告⁽³⁾）。しかし、原核生物・真核生物の場合、生存戦略とゲノムの塩基配列の関係は複雑になる。生物が環境に適応し進化する

うえでの生存戦略は、ゲノムの塩基配列の中にどのように刻み込まれているのだろうか。この問いに答えるために、集団レベルで得られるようになったゲノムの塩基配列データを、どのように生かしたらいいのだろうか。

ここで筆者は、真核生物よりもシンプルな生物として病原細菌を対象に選び、病原細菌が細胞外DNAを自身のゲノムに取り込む「組換え」（図1A）機構に注目した。病原細菌がなぜ、この組換えという機構・戦略を有しているのかについては、真核生物の減数分裂時の組換えの関係と併せて、重要な問題として古くから議論されている⁽⁴⁾。病原細菌の組換えは、特定の種の特定の遺伝子、たとえば髄膜炎菌の膜タンパク質の遺伝子 *tbp2* で、繰り返し高頻度に生じていることが知られている。これは、ヒトの免疫系と直接相互作用するタンパク質であり、その遺伝子に高頻度に外来DNAを取り込むことは、そのタンパク質を多様化させることによって免疫系をくぐり抜けるためのこの菌の戦略だと考えられる。

このように組換えが繰り返し高頻度に生じるゲノム上の領域を、「組換えのホット領域」と呼ぶことにする。組換えのホット領域は一般に、病原細菌が環境に適応し

図1 ■ (A) 病原細菌の組換え機構、(B) 組換えによって生じたゲノムのモザイク構造の推定

進化するための戦略を担っている領域だと考えられる。したがって、それがゲノム内のどこに存在するのかを突き止めることは、病原細菌の感染制御につながる基礎的知見を提供し、かつ、生物の生存戦略を塩基配列に基づいて理解するという一般的意義を有する、重要な課題だと言える。

しかし、突然変異率がゲノム内の特定の領域で高いことはよく知られている一方で、組換えは突然変異よりも検出が難しく、組換えの生じた回数をゲノムに沿って推定すること自体が未解決の難問であった。そのため、ある病原細菌種のゲノムの中に一体どれだけ組換えのホット領域が存在するかは、未解明のままであった。

そこで筆者は、欧州に長期滞在し、集団レベルのゲノムの塩基配列データから、組換えのホット領域を推定する方法の開発に取り組んだ。そのベースになっているのは、インシリコ染色体ペインティング法 (図1B) という、最近開発された手法である。この方法を用いると、ある個体 (ドナー) のDNA断片が組換えによって別個体 (レシピエント) のゲノムに入り込んだ「モザイク」構造を推定することができる。ゲノム全域にわたるSNPのアラインメントとポジションのデータから、レシピエントゲノム上のどの領域が、そのほかのどの個体 (ドナー) に由来しているのかを、塩基配列の類似度に基づいて推定することができる (隠れマルコフモデル)。この方法はもともと、別の目的で考案され、ヒトゲノムに適用されたものであったが、筆者はそれを、初めて病原細菌に応用し、ゲノム全域にわたる組換えの痕跡としてのモザイク構造を明らかにした⁶⁾。

ただし、染色体ペインティング法では最近1回の組換えしか検出できず、過去の履歴を考慮できていない。つまり、ある領域に組換えの生じた回数を考慮できず、したがって組換えのホット領域を推定することもできない。筆者は、これらの点を解決し、組換えの強度 (過去に生じた組換え回数に比例する指標) を1塩基単位で推定できる方法を開発した⁶⁾。

この方法によって、食中毒の主要な原因である病原細菌 (カンピロバクター) について、種内のさまざまな系統からサンプリングされ、次世代シーケンサで解読された200本のゲノムを解析した。その結果、ゲノム内に3つの組換えのホット領域が存在することが明らかになった (図2)。1番目の領域は、著しく多様化していることが知られ、膜タンパク質関連の遺伝子をコードしている

図2 ■ 食中毒病原菌 (カンピロバクター) ゲノムの組換え強度の分布と、組換えのホット領域

ホット領域の遺伝子には、それぞれの個体において、そのいろいろな場所に、さまざまなDNA断片が取り込まれている。

領域であった。3番目の領域も同様で、そこにコードされた外膜タンパク質は、この細菌の腸粘膜への接着に必要であり、免疫系からの強い選択を受けるものであった。2番目の領域は、細菌細胞に亜鉛を取り込むトランスポーター (多くの細菌種が宿主への定着と病原性に必須としているもの) をコードしている領域であった。

この方法において、入力として必要なのは、ゲノムワイドなSNPとそのポジションの情報のみである。ほかの関連する手法^{7, 8)}と異なり、組換えの影響を除外した系統樹をあらかじめ推定し、入力として与える必要がない。さらに、種としての平均組換え率が低い種から高い種まで、適用可能である。また、大半の計算を計算機クラスター上で並列化しているため、従来は不可能であった100本を超えるゲノムの解析が可能である。プログラムは一般公開しており計算機クラスターを使えるユーザーであれば、誰でも利用可能である。

現在、海外の共同研究者の協力を得て、ほかの病原細菌種のゲノムデータも、この新しい手法によって同様に解析可能な状態である。その解析により、病原細菌がゲノムの特定の領域で組換え強度を上昇させるという戦略の全容と、その生物学的意義に関する仮説の検証 (たとえば、組換え強度の上昇は病原性に関連するかどうか) が進むことが期待される。

- 1) 酒井聡樹, 高田壮則, 東樹宏和: “生き物の進化ゲーム—進化生態学最前線: 生物の不思議を解く 大改訂版”, 共立出版, 2012.
- 2) A. Burt & R. Trivers: “せめぎ合う遺伝子—利己的な遺伝因子の生物学”, 共立出版, 2010.
- 3) M. El Baidouri, M. C. Carpentier, R. Cooke, D. Gao, E. Lasserre, C. Llauro, M. Mirouze, N. Picault, S. A. Jackson & O. Panaud: *Genome Res.*, **24**, 831 (2014).
- 4) R. E. Michod & B. R. Levin: “The Evolution of Sex: An Examination of Current Ideas,” Sinauer, 1988.

- 5) K. Yahara, Y. Furuta, K. Oshima, M. Yoshida, T. Azuma, M. Hattori, I. Uchiyama & I. Kobayashi: *Mol. Biol. Evol.*, **30**, 1454 (2013).
- 6) K. Yahara, X. Didelot, M. A. Ansari, S. K. Sheppard & D. Falush: *Mol. Biol. Evol.*, **31**, 1593 (2014).
- 7) X. Didelot, D. Lawson, A. Darling & D. Falush: *Genetics*, **186**, 1435 (2010).
- 8) S. R. Harris, I. N. Clarke, H. M. Seth-Smith, A. W. Solomon, L. T. Cutcliffe, P. Marsh, R. J. Skilton, M. J. Holland, D. Mabey, R. W. Peeling *et al.*: *Nat. Genet.*, **44**, 413, S1 (2012).

(矢原耕史, 久留米大学バイオ統計センター, College of Medicine, Swansea University)

プロフィール

矢原 耕史 (Koji YAHARA)

＜略歴＞2006年に東京大学大学院新領域創成科学研究科メディカルゲノム専攻修士課程修了／2011年IT企業に勤めながら久留米大学大学院医学研究科バイオ統計学群で博士号取得／2012年日本学術振興会特別研究員 (PD, 東京大学)／2014年久留米大学バイオ統計センター講師, 現在に至る
＜研究テーマと抱負＞病原細菌ゲノムの組換えのホット領域と適応的アミノ酸置換の関係, ホット領域の種間比較, 病原細菌のゲノムワイド関連解析に関する国際研究を継続発展させる一方, 医師との共同による臨床データ解析に取り組む. さらに, 学内での次世代シーケンサ導入に向けて, 次世代シーケンサのデータ解析全般に興味をもつ. 生物学, 医学, バイオ統計学, バイオインフォマティクスの経験を組み合わせた新たな仕事の方向性も模索している＜趣味＞音楽, 野球, 旅行, 映画, 街歩き＜所属研究室ホームページ＞<http://www.biostat-kurume-u.jp/about/staff.html>

Copyright © 2015 公益社団法人日本農芸化学会