ゲノムから機能へ、配列アラインメントの役割

後藤 修 埼玉県立がんセンター研究所
Osamu GOTOH

昨年未までに線虫 (*Caenorhabditis elegans*) の 97MB にわたる全ゲノム塩基配列の決定がほぼ完了し¹⁾、真核生物としてはパン酵母に続き 2 番目の、多細胞生物としては最初の生命の設計図が明らかになった。すでに線虫では、受精卵から成体に至る細胞分化の運命がすべて知られている。全ゲノム構造の解明は複雑な多細胞生物の発生・分化の過程を遺伝子レベルで理解する確かな礎となり、ひいては生物学、医学、薬学、農学、あるいは環境科学などさまざまな分野の問題解決の糸口になるものと期待される。

一般の真核生物においては、遺伝子に蓄えられた情報は図 1 に示す流れに沿って最終産物（多くはタンパク質）を作りだし、目的の機能を発揮する。各遺伝子のもつ「機能」を知ることがこれからのゲノム研究の中心課題である。基本的には個体にとって必要なすべての情報がゲノム配列に書き込まれているはずである。しかしながら、最も基本的部分に限っても、ゲノムの塩基配列情報が「具体的に」どのように伝わり機能に結びつくかを配列のみから予測することは今なお困難な問題である。遺伝子レベルでの血縁関係（相同性、ホモロジーという）を利用することが、この問題解決のための重要な鍵となる。

酵母で約 6000、線虫で約 19,000、ヒトで 100,000 弱と推定される遺伝子とその産物のすべてについて一つずつその構造と機能を実験的に求めるには膨大な労力を必要とする。しかし、遺伝子の多くが種を超えて共通の機能を保つことや、「ドメイン重複」によっても基本的性質は保存されることから、多数の遺伝子の機能・構造予測が少数例からの類推に基づき可能となる。ここで「ドメイン」とは単独で固有の立体構造を取りうるタンパク質の一部あるいは全領域をいい、ほぼ機能の単位ともみなせる。

「ホモロジー検索」は通常アミノ酸配列の類似性を基に、仲間のドメインをデータベースの中から探し出す計算手段を指す。うまくいけば、配列から機能への最短経路となることが広く知られている。全構造の解明されている 20 近いゲノムの、30~60% の遺伝子について通常のホモロジー検索は有力な手がかりを与えない。しかし、問い合わせを単独の配列ではなくその仲間を集めた「マルチプルアライメント」とすることにより、ホモロジー検索の感度と精度が格段に向上することが最近明らかになってきた²⁾。機能解明のための系統的な実験法の成果と相まって、ほとんどの遺伝子の機能のある程度のレベルで予測できる日も遠くないだろう。

マルチプルアライメントは分子進化、タンパク質構造

図 1 ゲノム配列からタンパク質までの遺伝情報の流れ

予測、遺伝子構造予測などさまざまな応用面での有効性が知られている³⁾。例えば、既知のタンパク質の立体構造を鋳型として構造未知の配列のとり三次構造を予測する「ホモロジーモデリング」にとって重要な役割をもつ。鋳型と目的配列との間の類似性が検出限界に近づく「トワイライトゾーン」において、両者の各アミノ酸残基ごとの対応関係を精度よく推定できる最も信頼性の高い手段だからである。

ゲノム配列から遺伝子機能、さらにはさまざまな生命科学分野への応用に至る道筋には大量の情報の流れが存在し、計算機による迅速かつ知的な処理が欠かせない手段となっている。ホモロジー検索とマルチプルアライメントをその代表として取り上げた。現在急速に発展しているこの分野は「Bioinformatics」と呼ばれ、際だって学際的な活動によって支えられている。3年後ともいわれるヒト全ゲノム構造解明に向け、さまざまな分野の英知を結集したいっそうの研究開発が急務である。

文 献

- 1) The *C. elegans* Sequencing Consortium: *Science*, **82**, 2012 (1998)
- 2) S. F. Altschul, et al.: *Nucleic Acids Res.*, **25**, 3389 (1997)
- 3) 後藤 修: 生物物理, **38**, 52 (1998)

(埼玉県立がんセンター研究所・主任研究員、理博。専門=生物物理学、ゲノム情報学)