

## 次世代型シーケンシングを生物学へ活かそう

小池 英明

次世代型シーケンシング (NGS) には、もう近未来的な響きはなくなり、生物学の一般的インフラとなったのではないだろうか？ NGSとして普及する小型のシーケンサーでも2500万個のリードが読め、10 Gb程度 (ヒトゲノムの数倍以上) のデータを出力する。これは一般的な糸状菌のゲノム情報 (30 Mbp) の300倍のデータに相当する。得られた配列断片の再構築でもオープンソースプログラムだけを使って数Mbを超える連続配列が得られる<sup>1)</sup>。しかもNGS開発は未だ進行中であり、今後もさらなる革新が望める。シーケンシング価格の下落は、コンピューターの爆発的な発展を表現した「ムーアの法則」を上回り興味を集めた<sup>2)</sup>。しかし、NGSから得られる膨大な配列情報を、生物学の研究に活かしているのだろうか？

この読者の多くは、もう従来型のSangerシーケンシングを経験しない世代になっているのだろうか。筆者がゲノム解析に出会った1990年代は、1 Mb程度の子細菌ゲノムがやっと解明され始めた時代だった。当時1 Mbのゲノム塩基配列をSanger法で決めるのは一大プロジェクトだった。わずか100リードのために大きなポリアクリルアミドゲルを作るだけでも一苦労だが、1日に2ランしてもたかだか100 kbp (500 bp × 100 × 2) しか読めない。しかも既読の部分が増えてくると、新しいリードが未読領域に当たる確率は下がってくるので、結果を見てため息をつく日が増える。ゲノム全体をカバーするまでには長い道のりだった。そのような状況下で、モデル生物として、出芽酵母 (1996年)、糸状菌の赤パンカビ (2003年) のゲノムが黎明期に解明されたことはゲノム科学をさらに推し進めた。その完全さと時代の先取りは一研究室のレベルをはるかに超えていたが、初期のゲノム論文に多数の著者名が並んだことに象徴されるように、莫大な資源があって初めて達成された成果である。関連する研究成果も含めれば、どれほどの資源が必要とされたのか？ゲノム科学の金字塔であり、偉大な共有財産として大事にしていく必要がある。

今ではゲノム配列の決定プロジェクトの主力はNGSである<sup>3)</sup>。初期のNGSの問題点 (短いリード長、高いエラー率) もほぼ解消された。糸状菌ゲノムでは1回の実験で遺伝子の大部分を明らかにできる。数Mbの原核生物であれば、数本以下の連続した配列にまで再構成することが可能である。

これは生物学にとって福音である。NGSによって初めて可能になった研究の中でも“モデル生物以外の微生物”をゲノム塩基配列に基づいて改変できることは、生物学にとって重要である。モデル生物に多数の遺伝子改変を導入するよりも、オリジナルの産業用微生物に

修飾を施すことの方が、はるかに手数が少なく効率的である。発現解析もNGSでできる。NGSの出現によって、新規の微生物を利用する環境が整った。

NGSでゲノム配列決定や発現解析をすることはもはや一般的な作業となった。ただし、そのデータを活かすには、新規の微生物ゲノムから得られた配列データに対して、多様な解析をゼロから実行し、機能に関連するさまざまな情報を蓄積し管理しながら、さらに利用するステップをこなしていく必要がある。しかし、そのデータ量が膨大であるために手を出せない研究者も多い。外注による解析で一部は解決するだろうが、多様な解析が必要なことを考えると定型的な外注だけで完結するとは思えない。しかもカスタマイズした外注は、NGSよりよほど高価である。

医療・医学向けの画一的な解析を実行する商用パッケージは増えてきた。しかし、新規の微生物に対する多様な解析をパソコンで簡単に実行するパッケージはほとんどない。しかも市場が小さいのだから、将来の好転も難しいかもしれない。一方でLinux OSをはじめ、オープンソースの優れたツールは多数ある。NGSの発展に合わせて解析手法も新しいものが次々に出現し、世界中の優秀な情報科学者がツールを次々に提供している。だから、これらを利用するのがもっとも良い戦略だろう。関わる者のIT技術を地道に上げていけば、NGSのデータだけでなく、公共のデータベースにある膨大な情報を活用することができる。そして特に、モデル生物の成果を利用することは、良質で一貫した基準の参照データを使える点で掛け替えがない。モデル生物のデータは莫大な資源が投入された結果であることを思い出して欲しい。モデル生物を参考にしつつ、新しいゲノムを着実に理解する戦略が良い。

筆者は日本の生物学研究においてもっと情報解析の能力が向上して欲しいと (自省も込めて) 願い、身近な研究者には自作のスクリプトやノウハウを伝え、授業などで情報解析の基礎を伝えつつ重要性を力説している。黎明期からゲノム解析に関わってきた研究者として、何とか次世代の研究者へ向けて貢献したいと考える年齢になった。NGSの情報解析が生物学に密に組み込まれている明るい数年後の未来を考え、出向先でベンチャー起業を支援しつつ、ゲノム科学での自身の貢献方法を探っている。

- 1) The assemblathon: <http://assemblathon.org/> (2016/2/25).
- 2) Hayden, E. C.: *Nature*, **507**, 294 (2014).
- 3) 日経バイオ年鑑2016 (日経BP社).