創立90周年記念特別企画
～バイオ技術10年の軌跡～

〔生物工学会誌　第90巻　第12号　764–784．2012〕

特 集

バイオ分野・ライフサイエンス分野のビッグデータ解析

特集によせて

池村　淑道

バイオ分野におけるビッグデータ　バイオ分野は生物・化学・物理・情報科学などを基盤として，分野横断的な総合科学として急速に発展してきた．医学・薬学・工学・農学などの応用科学としての側面も持ち，ますます発展を遂げるだろう産業分野として，国民の期待も大きい．特に，医療や環境分野における重要性から，国の重点分野に位置付けられており，この分野を目指す若手も多い．バイオ分野の現時点での特徴として，相互に関連性を持つ多様なデータ類が超大量化し，ビッグデータ解析が緊急の課題となった．生命の設計図でシナリオであるゲノム塩基配列が大量情報であることは自明であるが，最近になり，そのゲノム塩基配列の解読技術に革命的な進歩がもたらされ，広範な生物種のゲノム配列が解読され，超大量なゲノム配列が入手可能になった．加えて，遺伝子発現やタンパク質の構造解析法，多様な生体分子の高速な分離技術が開発され，相互に関連性を持つハイスループットな測定データ類の集積が著しく，従来の解析手法や高機能計算機でも処理が困難なビッグデータがこの分野の特徴となっている．

この分野の問題点　多様なビッグデータの集積は，バイオ分野のさらなる発展を予感させ，分野の魅力を感じさるが，この分野に携わる人材，特に我が国の人材不足を考えると問題点も大きい．バイオ分野は，その研究対象がヒトを含む生物であり，この分野の大半の研究者や教員は，生物学を中心とした教育を受けてきた．情報解析の十分な知識を要するビッグデータ処理やその教育には困難さが伴う．我が国において，学部や大学院の早い時期から専門分野を限定・特化する教育，近視眼的とも呼べる教育の弊害が，この分野でも大きな問題を引き起こしている．

異分野の研究者の参加　上記の問題の当面での解決には，異分野で教育を受けた人材の積極的な参画が重要になる．バイオ分野で集積しているビッグデータの規模とその増加スピードは，ビッグデータ解析を専門としている研究者にとっても未経験に近い程であり，そこに潜んでいる多様な知識の発見とそのための技術開発は，分野の重要性から「オリンピックゲーム」に近い注目を受けるケースもある．異分野から参入して成功を収めている例が，欧米や中国に多いように見えることは残念である．我が国で，異分野の橋渡しをする人材が不足している結果と考えられる．

異分野の研究者にとって理解しやすいゲノム配列情報　新しい分野に参画しようとする場合，多量な新知識を習得した上でないと研究が開始できない場合には，それは大きな障壁となる．しかしながらゲノム配列の場合には，ATGCの4種類の文字のみで構成されており，異分野の研究者であっても理解が容易である．優れた手法を適切な課題に導入すれば，新しい研究分野を早い時期から開拓できる可能性がある．その意味で，ゲノム配列データは，最初に取り組むのに適した対象の一つといえる．研究を開始する際には，バイオ分野の研究者，特にバイオ情報処理の経験のある研究者との共同研究が重要に思える．筆者らのグループは小規模ではあるが，情報解析とバイオ分野の研究者が，大学の壁を越えて継続的に共同研究を行ってきた．欧州を中心に研究の進んでいたSOM（self-organizing-Map）を，ゲノム情報解析に世界に先駆けて導入し，その有用性を継続的に発表してきた．筆者らの開発した一括学習型のSOM（BLSOM）は，大規模な並列計算に適しており，スパコンを用いたビッグデータ処理に適している．現在では，欧米を中心にSOMがゲノム情報解析に利用され始めている．異分野の研究者の共同研究の例として，本特集の最初の2課題は，BLSOMを用いた研究例を紹介し，加えて他の3課題も紹介する．

本特集で紹介する課題

1）地球環境の保全や浄化における微生物の重要性がますます明らかになってきた．全地球規模での微生物の集団構造を把握する目的で，大規模なゲノム配列解読が進行している．BLSOMはこの解析において世界的に利用され始めた．

2）新技術を社会的に関心の高い課題に適用することは意義深い．BLSOMを用いた解析により，インフルエンザウイルスのゲノム配列の時系列的な変化について，限定された範囲ではあるが，変化の方向性が予測可能なことを明らかにした．変化を予言しておくことで，早い時期に検証が可能になる魅力的な分野である．

3）広範な生物種のゲノムが解読されたことで，過去に存在した祖先生物種の遺伝子配列が推定でき，それを基に，遠い過去のタンパク質の機能や構造解析が可能になっている．

4）多様な分野の情報を体系的に組み合せて，知識発見を行うためのデータベース構築の成功例を紹介する．

5）新型シークエンサーの説明，その大量配列データの収集と公開，ならびにデータの形式と利用方法を紹介する．併せて，この大量情報の処理に適した遺伝研スーパーコンピュータの特徴と利用方法も紹介する．

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_ikemura@nagahama-i-bio.ac.jp