

## 生物情報科学科 高木研究室

情報エンジニアリングが拓く、新しい生物学。

自分は将来、どんな病気になる可能性があるのか——ゲノムの解析技術が飛躍的に向上した現代、人は自分の遺伝情報「パーソナルゲノム」を事前を知り、人生設計を立てることが可能になった。最近では、女優アンジェリーナ・ジョリーががん抑制遺伝子のひとつ「BRCA1」に変異が見つかり、乳房の予防切除を行ったことが話題となった。「不運の死」をもたらす病気はもはや、予報可能な「悪天候」になりつつあるのかもしれない。その根幹を支えるのが生物情報学、「バイオインフォマティクス」だ。

「コンピュータを使い、生物学のための情報の技術をつくるのが従来のバイオインフォマティクスでした。しかし最近では、いかにコンピュータの理論を使って生物を研究するか、というアプローチが既存のバイオインフォマティクスと両輪構造をしています」と高木利久教授は話す。

2003年にヒトゲノムが解読されるまで

### 学生 ▶ 教授 逆評定

「学生や先生という立場を越えて議論できるフラットさが好きです」(D2・尾崎遼さん)

は、ヒトゲノムの全貌が掴めず、病気の原因遺伝子を「犯人探し」するにも、手探りだった。しかし解読後のバイオインフォマティクスは、完成した地図を手「犯人探し」ができるという「データドリブン」の世界になった。加えて、ゲノム解析装置が低価格化・高性能化し、「テクノロジードリブン」の効果もたらされた。かつては数千億円が必要だったゲノム解析の費用は、今や数万円だ。それによってパーソナルゲノムの解析が可能となった。

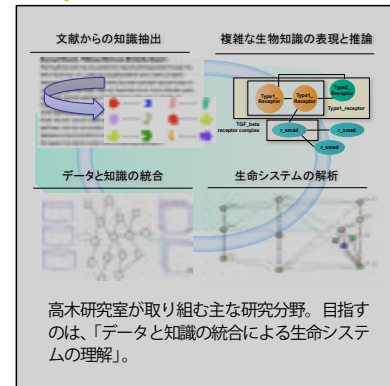
こうした背景から、生物学へのアプローチはより工学的なものとなった。「研究の究極的な目標は、コンピュータの中で生命を再構築すること」と高木教授は言う。

高木教授は、実験で得られたゲノムデータを効果的に意味付けるための、データと知識を関連付ける研究を専門とする。

たとえば2つの長大な論文があるとすると、ひとつの論文にはAとBの遺伝子の関係が書かれており、別の論文にはBの遺伝子とCの病気の関係が書かれている。この2つの論文の知識を合わせれば、遺伝子AからCの病気に関連性があること、それにBの遺伝子が関わっていることがわかる。

だが現実には、関連する論文数は何万件にもほり、同じ事象でも異なる言語、異なる用語で記述されていることがある。そのため、ひとりの研究者が関係性を見抜くのは困難を極める。高木教授はこうしたデータの標準化を推進し、世界中と協力して生物に関する全データをつなぎ、因果関係を

### Study



明らかにするデータベースの構築に取り組んでいる。

「生物学にテクノロジーとコンピュータがもたらされ、生物を「システムとして捉える」ようになり、大量のデータを扱わなければならないになりました。そこからデータベースの構築や品質・工程管理など、エンジニアリング的なアプローチが必要となってきたのです」と高木教授。

そもそもゲノム自体がA（アデニン）、T（チミン）、G（グアニン）、C（シトシン）の塩基記号で記述されたデジタルテキストである。求められるのは情報処理のエンジニアリングだ。これからの生物学では、研究者（リサーチャー）の席のいくつかに「エンジニア」が座ることになるかもしれない。

高木研究室では、自由度の高い研究が重んじられる。従来の生物学者とは異なる情報の視点で生物を捉え、新たな発見をするためだ。尾崎遼さん（D2）は、次世代シーケンサから得られるデータをもとに、魚の比較ゲノム進化解析を行う。「魚が海水や淡水という環境にどう適応していったのか。シーケンサの性能がどんどん向上して、新しい仮説を発見できるのが面白い」と言う。

未来の生物学のエンジニアたちは、今日もコンピュータを通し、生命の神秘に出会い続けている。

高木利久 教授  
新領域創成科学研究科 情報生命科学専攻

1976年東京大学工学部計数工学科卒業。九州大学情報処理教育センター助教授、東京大学医科学研究所教授等を歴任。2014年4月に東京大学大学院理学系研究科生物科学専攻教授に着任。工学博士。科学技術振興機構/バイオサイエンスデータベースセンターセンター長、情報・システム研究機構国立遺伝学研究所DBJセンター長等兼務。

