

「特集 DNA 配列の統計解析」について

長谷川 政美[†] (オーガナイザー)

2001年2月にはヒト全ゲノムのドラフト配列が発表され、現在そのほかの多くの生物種についても加速度的な勢いでゲノムのデータが蓄積しつつある。また、ゲノム全部でなくても多くの遺伝子の塩基配列データも爆発的な勢いで蓄積している。しかしながら、研究の終局的な目標は、このようなデータを蓄積していくことではなく、そこから生物学あるいは医学、農学などに役立つ情報を取り出すことにあることは当然である。ところが残念ながら、このようなデータの洪水のなかにあつて、そこから意味のある情報を取り出すためのデータ解析の技術が立ち遅れているのが現状である。DNA配列データの解析にあたっては、統計的な方法が重要である。なぜならば、現在の生物のもつ生体情報分子は、長い進化の産物であり、分子進化の過程においては、ランダムな確率的な要因が重要な働きをしてきたからである。ランダムなノイズのなかから、意味のある情報を取り出すためには、統計的な方法が必須であり、統計学者が活躍する余地の多い分野である。「統計数理」の今回の企画は、「DNA配列の統計解析」というタイトルで、この分野の研究の一端を紹介し、より多くの統計学者がこの分野に積極的に進出することを促したいという思いから生まれたものである。

今回執筆をお願いしたのは、長谷川が代表をつとめている科学研究費・基盤研究(B)「分子系統学のためのソフトウェアの開発」(平成12年度-14年度)に参加されている方々であり、そのために分野が分子進化に偏ったものになっている。DNA配列の統計解析という分野は、実際にはもっと広範囲のものであるが、そのなかでも分子進化的な解析は中心的な役割を果たすものである。ゲノムの解析においては、たとえばヒト・ゲノムの完全な配列データが得られても、それだけでは、そこから得られる情報はきわめて限られていると言わざるを得ない。チンパンジー、マウス、魚、植物、原生動物、バクテリアなどほかの生物のゲノムと比較することによってはじめて、得られる情報量は格段に多くなるのである。従って、比較ゲノムの方法は、DNA配列の統計解析の中心であり、そのことは、進化的な枠組みのなかでデータを扱わなければならないことを意味する。実際、生物学におけるほとんどの問題は、進化的な枠組みのなかではじめて意味をもつものである。なぜならば、現在生きている生物のもつあらゆる構造と機能は、それぞれの生物の祖先がたどってきた長い進化の歴史の産物だからである。

ゲノムデータは、研究の目標ではなく、ポストゲノムと呼ばれるこれからの時代においては、生命科学のあらゆる分野の研究の出発点であるともいえる。今回の特集がきっかけになって、一人でも多くの統計学者がゲノムデータの解析に興味をもって、この分野の研究に参加されるようになれば、企画者としての大きな喜びである。

[†] 統計数理研究所：〒106-8569 東京都港区南麻布 4-6-7