

「特集 分子進化と統計科学」について

長谷川 政美¹・曹 纓² (オーガナイザー)

DNA の塩基配列やそれにコードされている蛋白質のアミノ酸配列のデータから、生物進化の系統樹を推定する分子系統樹法は、生物科学のあらゆる分野で重要なツールになってきた。進化の過程で蓄積する配列の変化は、ランダムな変動を伴うものであるから、そのような進化過程の産物である現存生物の配列データから過去の進化の系統樹を推定する問題は、まさに統計的な推測の問題である。そこには、統計科学の寄与する余地が多い。

21 世紀に入った今日、ポストゲノムの時代とも言われる。ゲノムとはそれぞれの生物のもつ遺伝情報の一揃いのことであるが、ポストゲノムとは、ゲノム解析の時代が終わったという意味ではなく、ゲノム・データの存在が生物学のあらゆる分野で常識となった時代ということである。ゲノム・データの存在が生物学のさまざまな研究の前提になりつつある。ゲノム・データは、遺伝情報の全体ということであるから、個々の遺伝子の情報だけを扱ってきたこれまでの研究とは、質的に全く異なる段階に入ってきたということでもある。10 年前の生物学者の多くは、研究対象となっている生物種のゲノム・データさえ得られれば、生物系統学の問題はたちどころに解決されるはずだと考えていたように思われる。10 年前に未解決だった問題は、当時得られていた配列データが短かったために、サンプリングの誤差が大きくて、確定的な結論が得られないのだと考えられたのである。ゲノム・データが得られれば、もちろん、サンプリングの誤差は問題にする必要がないほど小さくなり、確定的な結論が得られるはずだと考えられるかもしれない。ところが、本特集の西原らの論文で議論されているように、さまざまな生物のゲノム・データが実際に得られてみると、問題はそのように単純なものでないことが明らかになってきた。

分子系統樹解析には、分子進化の素過程(塩基置換)のモデル化が必要であるが、不適切なモデル化は推定に偏りを与える。ゲノム・レベルの解析が可能となってきた今日では、小さな偏りでも、実際の推定に深刻な影響を与えるようになってきたのである。このような問題に対処するためには、さまざまな生物学の実際問題を扱っている生物学者と、統計科学者の密接な連携が必要であろう。本特集では、分子系統学におけるデータ解析手法を開発しておられる統計科学者と、生物系統学のさまざまな実際問題に携わっておられる生物学者に論文を執筆していただいた。この企画がきっかけになって、1 人でも多くの統計科学者が分子系統学の分野で新しいテーマを見つけられ、この分野の発展と統計科学の新しい展開に貢献して下さるようになれば幸いである。

¹ School of Life Sciences, Fudan University, 220 Handan Road, Shanghai 200433, China

² 統計数理研究所：〒106-8569 東京都港区南麻布 4-6-7