

グラフ理論に基づく新規進化解析手法の開発

Development and application of the novel phylogenetic method based on graph theory

東京大学大学院理学系研究科生物科学専攻岩崎研究室 松井求

研究成果概要

分子系統学における今日の難題の一つは「著しく分化した遺伝子群の進化解析」である。特にタンパク質スーパーファミリーレベルの系統樹を再構築可能な方法は未だに存在しない。進化的関係がある遺伝子群であっても、互いの配列距離が著しく遠い場合は informative な多重配列アライメントの作成は難しく、それらを前提とする従来法の適用もまた困難なためである。そこで本研究において、新たな系統解析手法「Graph Splitting 法 (GS 法)」を考案した。GS 法は、まずペアワイズアライメントに基づいて配列類似性グラフを構築し、次にそのグラフを再帰的に分割することで系統樹を得る—つまり部分配列の類似関係を進化情報に変換する方法である。この操作は、多重配列アライメントに起因する従来からの諸問題を回避する。

GS 法と一連の従来手法 (近隣結合法、最大節約法、最尤法、およびベイズ法) の性能を比較するために大規模な進化シミュレーションを行った。この比較解析は、膨大な計算コストがかかる 1) 進化シミュレーション (INDELible など)、2) 配列アライメント (BLAST、Mafft など)、3) 系統樹再構築 (RAxML、MrBayes など) といった各計算素過程をさらに数千回単位で繰り返す行うものであったため、スーパーコンピュータを用いることで初めて実現することができた。

比較解析の結果、進化距離が互いに遠い遺伝子群については GS 法が最良な系統解析手法であることが示された。同時に、これまで系統解析が困難とされてきた条件 (e.g. long branch attraction) において、GS 法が最も頑健な性能を示すことも分かった。また、実データへの適用例として TIM バレルスーパーファミリーの系統解析を行ったところ、高い内部枝支持率を持つ GS 系統樹が得られ、さらにこの GS 系統樹は他の系統樹よりも確からしい系統樹であることが三次構造解析より統計的に有意に示された。

発表論文 (謝辞あり)

なし

発表論文 (謝辞なし)

なし