

グラフに基づく分子系統解析手法の開発と適用

Development and application of the graph-based phylogenetic method

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

進化距離の大きな遺伝子群の系統樹再構築は、例えばタンパク質の初期進化といった”古い”問題だけでなく、「病原菌・ウイルス・癌の短い時間スケールにおける高速な進化」といった“モダンな”問題を解決するためにも必要な技術である。我々はこれまでに Graph Splitting (GS) 法という新たな系統解析手法を開発し、GS 法こそが進化距離の大きな問題を解決できる手法であることを示している。そして、その手法開発の中で、(1) GS 法と既存手法を使い分けるための基準の制定、(2) 塩基配列の進化解析への適用、(3) 無根系統樹から root を決定する手法の開発、といった新たな研究課題が見出された。今年度は、京都大学化学研究所スーパーコンピュータシステムを利用し、大規模進化シミュレーションを繰り返して都度精度検証を行いながら、とくに「(1)GS 法と既存手法を使い分けるための基準の制定」に取り組んだ。

様々な条件のもと、一連の既存手法、すなわち近隣結合 (NJ) 法、最節約 (MP) 法、最尤 (ML) 法、ベイズ (BI) 法と、我々が開発した GS 法の性能を比較した。その結果、進化距離が十分小さい時は GS 法のようにペアワイズ配列アライメント (PSA) に基づく NJ 法を使用し、大きい時は GS 法を使う、というように使い分けることで常に最良の系統樹を推定できることがわかった。このことは、現在広く使用されている ML 法や BI 法ではなく、距離法に分類される NJ 法と GS 法の組み合わせが最良の系統樹推定法であることを示唆しており、分子系統学にとって大きなインパクトのある発見であると考えている。また我々は、配列類似性グラフの密度がある一定の値を上回る時は NJ 法、下回る時は GS 法のように使い分けることが、最善の手法選択であることも見出した。さらに、PSA に基づく NJ 法には「内部枝支持率を計算する良い手法が存在しない」という問題があったため、その解決を試みた。PSA から擬似的な多重配列アライメントを構築する方法や、ランダムに PSA をシャッフルする方法、といった古典的な手法とともに、GS 法を開発する際に合わせて開発した内部枝支持率計算手法である Edge Perturbation (EP) 法を NJ 法に適用した結果、EP 法を使用することで、適切な内部枝支持率を算出できることが明らかになった。

以上をまとめると、進化距離の小さな場合は PSA+NJ+EP という組み合わせで、進化距離の大きな場合は PSA+GS+EP という組み合わせで系統解析を行うのが最善であることが示された。

発表論文

投稿準備中