

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

進化距離が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。

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ながら、**GS** 法と**NJ** 法を融合した方法の開発を試みた。様々な条件のもと、近隣結合(**NJ**)法、及び一連の標準的な手法、すなわち最節約(**MP**)法、最尤(**ML**)法、ベイズ(**BI**)法、と**GS** 法の性能を比較した結果、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いる。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が、**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。現在、これらの新たな手法は実装済みであり、検証結果をまとめながら投稿論文を準備しているところである。

発表論文(謝辞なし)

Takao K Suzuki, Motomu Matsui, Sira Sriswasdi, and Wataru Iwasaki, **Lifestyle Evolution Analysis by Binary-State Speciation and Extinction (BiSSE) Model**, *Methods in Molecular Biology*, in press

Motomu Matsui, **Frontier of Inferring Phylogenies**, *JSBi Bioinformatics Review*, 2021, **2**(1), 30-57

Nanako Kanno, Shingo Kato, Moriya Ohkuma, Motomu Matsui, Wataru Iwasaki, and Shinsuke Shigeto, **Machine learning-assisted single-cell Raman fingerprinting for in situ and nondestructive classification of prokaryotes**. *iScience*, 2021, **24**(9):102975