

スペクトラルグラフ理論に基づく新規系統解析手法の開発

Developing the novel phylogenetic method based on the spectral graph theory

東京大学大学院 新領域創成科学研究科 先端生命科学専攻 松井求

研究成果概要

分子系統学の分野では近年、AlphaFold2 の登場を契機に、タンパク質の立体構造情報を系統樹推定法に取り込むという取り組みが活発に行われている。しかしながら、タンパク質立体構造の進化過程に関する理解がまだ不十分なことから、進化モデルをどのように構築すべきか？立体構造の情報を既存の系統樹推定のスキームにどのように接続すべきか？など基礎的な知見の蓄積すら十分ではない。

本研究では、これまでに開発してきた Graph Splitting (GS) 法、Edge Perturbation (EP) 法、および PA+NJ+EP (PANJEP) 法をタンパク質立体構造も利用できるようにアルゴリズムを改善し、実データに適用することで有効な手法であるか検証した。アルゴリズムの開発と検証は、京都大学化学研究所スーパーコンピュータシステムにおいて、AlphaFold DB に登録されている膨大な立体構造データセットに上記の新手法を適用することで行った。ベンチマーク解析の結果、これまでアミノ酸配列情報からでは系統解析が困難だったタンパク質スーパーファミリーについて、十分な精度の立体構造情報が蓄積していれば、立体構造に基づく GS+EP 法や PANJEP 法は内部枝がよく支持された良い系統樹を推定できることが分かった。タンパク質立体構造予測の手法開発は日進月歩で進んでおり、その進捗を積極的に分子系統学に取り込んでいくことは、機能未知遺伝子の機能予測やその進化解析など幅広い波及効果も期待できることも相まって、重要課題となっている。本成果は、その嚆矢になりうるものと考えている。

発表論文(謝辞なし)

Hiroataka Matsumoto, and Motomu Matsui, **NeTaGFT: A similarity network-based method for trait analysis**, *Methods in Ecology and Evolution*, 2024, **15**(1):153–163