

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

タンパク質スーパーファミリーの進化過程は、現存するあらゆるタンパク質の進化の歴史を含んでおり、もしそれがまるごと解明できれば、タンパク質の諸機能と進化を結ぶ統合的知見が得られると考えられる。しかしながら、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)には根本的な弱点があり、タンパク質スーパーファミリースケールの信頼性の高い系統樹これまで得られていなかった。そこで、我々は既存手法の代わりとなる Graph Splitting (GS) 法, Edge Perturbation (EP) 法, Mixed Inference (MI) 法という三つの新規手法を開発した。GS 法は配列類似性グラフに基づいて系統樹再構築を行う方法, EP 法は同じく配列類似性グラフに基づいて内部枝支持率を計算するための方法, MI 法はペアワイズアライメントとマルチプルアライメントの両方の情報を用いて距離行列を計算するための方法であり、それぞれ既存手法の弱点を克服するために、グラフを利用した新たなアルゴリズムに基づいて開発された。

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、タンパク質スーパーファミリーのデータベースである SCOP から得られた全てのデータに対し、GS 法・EP 法・MI 法を網羅的に適用した結果、190 のタンパク質スーパーファミリーについて内部枝が支持された信頼性の高い系統樹を初めて得ることができた。特に T-fold, DNA/RNA polymerase, Toxins' membrane translocation domains の各スーパーファミリーは系統樹全体にわたって内部枝が強く支持されており、代謝、転写・複製、膜輸送の初期進化を照らす新たな知見が示唆された。

発表論文(謝辞あり)

なし

発表論文(謝辞なし)

なし