

アライメント手法を見直し、新たな分子系統解析手法を開発する

Revisiting alignment methods, and developing the novel phylogenetic method

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

最大節約法, 最尤法, ベイズ法はいずれも Multiple sequence alignment (MSA) に基づく手法であり, 良い MSA を構築できない問題には適用が困難である。一方で, 近隣結合 (NJ) 法は MSA の代わりに pairwise sequence alignment (PSA) に基づく系統樹推定が可能だが, ブートストラップ法による内部枝評価ができないという問題があった。この問題を解決するために, 我々は PSA に基づく系統推定法「Graph Splitting (GS) 法」と, ブートストラップ法に代わる内部枝評価法「Edge Perturbation (EP) 法」を開発してきた。EP 法は NJ 法と組み合わせることのできるため, NJ+EP 法と GS+EP 法を組み合わせることで広範囲の問題を解決できると期待できる。

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, 一連の標準的な手法と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。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。一方で, 進化距離が小さな場合でも, 配列数が多い場合や矛盾を多く含む MSA については, 最尤法やベイズ法, 最大節約法を適用することは難しい。したがって, 本研究で提案する NJ 法と GS 法の組み合わせは, 今後様々な未解決問題を解決できるのではないかと考えている。

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*, 2022, **2569**:327–342

Nanako Kanno, Shingo Kato, Moriya Ohkuma, Motomu Matsui, Wataru Iwasaki, and Shinsuke Shigeto, **Nondestructive microbial discrimination using single-cell Raman spectra and Random Forest machine learning algorithm**. *STAR Protocols*, 2022, **3**(4):101812

Motomu Matsui, バイオインフォマティクスでポストコッホ生態系を解き明かす, *日本微生物生態学会誌*, 2022, **37**(1), 37–39