

分子系統解析に役立つ高精度な新手法を開発

～生物進化の正確な理解、タンパク質工学、医学、薬学に貢献～

ポイント

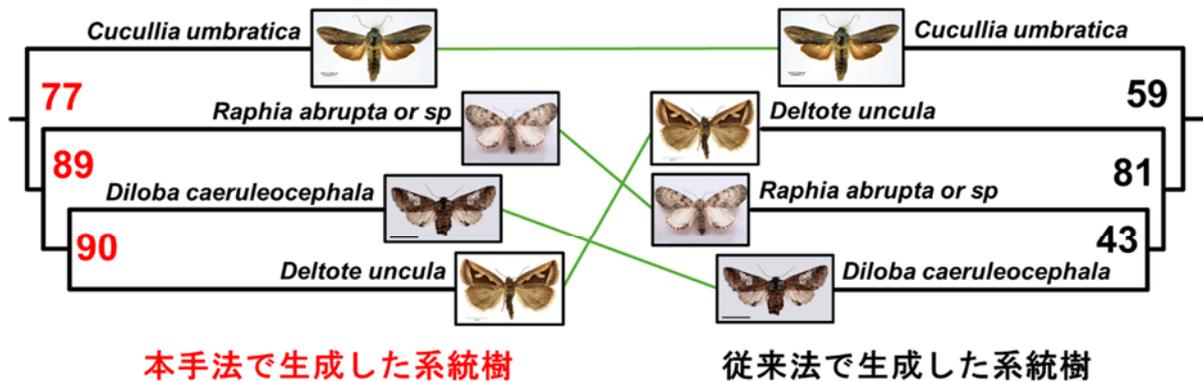
- ・分子系統解析における遺伝的異質性部位を高精度に分割する新手法を開発。
- ・ゲノム規模の系統解析において、進化の歴史をより正確に判定することが可能に。
- ・タンパク質進化の理解や工学への応用に役立つと期待。

概要

北海道大学大学院地球環境科学研究院の小野田晃教授の研究グループは、系統解析^{*1}における遺伝的異質性部位^{*2}を分割する新手法「PsiPartition」の開発に成功しました。この手法はベイズ最適化^{*3}を用いて、遺伝的異質性部位を分割する数と位置の最適解を効率的に決定することができ、分子系統解析において従来法より精度の点で優れています。

分子系統解析では、生物間の遺伝子において異なる部分の進化を説明するために、様々なモデルやパラメーターを用いて系統樹^{*4}を生成します。しかし、遺伝子解析技術の進歩により、膨大なゲノムデータが得られている現在では、系統樹再構築のためのモデルが複雑化しており、従来の手法では経験に依存する部分が多く、最適な系統樹を得ることが容易ではありませんでした。「PsiPartition」では、塩基配列やアミノ酸配列データの差異を記述するために、パラメーター化されたソート指標（Parameterized Soring Index、PSI）を新たに導入し、効率的かつ強力なベイズ最適化と組み合わせ、遺伝的異質性部位の分割の最適化を可能にしました。本手法によって、実際の塩基配列やアミノ酸配列データを用いたシミュレーションにおいて、分割の最適解を迅速に見つけ出し、従来法よりも顕著に高い精度で系統樹を生成することができます。本研究成果は、機械学習^{*5}手法を用いて分子系統解析の精度を大幅に向上させており、本手法によって得られる知見は生物進化について新たな視点を与えうるものです。また、この手法はタンパク質の機能解析や生物工学分野の研究に広く活用されることが期待されます。

なお、本研究成果は2024年12月5日（木）公開のJournal of Molecular Evolution誌にオンライン掲載されました。



*樹形図内の値は分岐群のブートストラップ値を示しており、高いほど優れている。

*写真のクレジットはBert Gustafsson、Jim Moore、JC Schou、及びWikimediaに帰属する。

【背景】

分子系統解析は、遺伝情報に基づいて生物進化の歴史を再構築するために重要であり、隣接結合法 (Neighbor joining 法)、グラフ分割法 (Graph partition)、最尤法 (Maximum Likelihood 法)、ベイズ推定 (Bayesian inference) などが用いられます。これらの中で、最尤法とベイズ推定は最も正確な方法とされてきましたが、ゲノムデータのサイズが増大する中で、分子系統解析における遺伝的異質性部位の取り扱いが課題となっていました。異なる遺伝子は異なる速度で進化するため、ある部位をいくつかの部分集合に分割し、それぞれの部分に異なる進化モデルを適用する分割モデルが用いられます。分割数は重要であり、数が多すぎればモデルに対する過学習のリスクとなります。AIC (赤池情報量基準)、AICc (修正赤池情報量基準)、BIC (ベイズ情報量基準) などの情報基準によって、分割の最適化の程度が判断されますが、既存の分割方法では、経験に基づく事前知識への依存に限界があり、最適な分割であることを保証することはできませんでした。

【研究手法】

本研究では、進化に伴う塩基配列の変異解析において、高精度に進化率を推定するために、Parameterized Sorting Indices (PSI) を導入しました。PSI は、塩基やアミノ酸の種類に応じたパラメーターを導入し、進化的多様性を反映した進化率を推定することができます。この手法では、従来の手法が全ての塩基配列やアミノ酸配列の構成要素を同じ進化速度で扱うのに対し、それぞれに異なる進化速度を割り当てることで、より高い精度で系統樹を再構築することができます。また、PSI では、異質性部位間の比較を通じて進化率を定義し、進化解析においてベイズ最適化を使用しました。この手法では、情報基準 (BIC) を最小化するために、進化解析のパラメーターと分割数を最適化し、過学習を防ぎつつ、より複雑なデータセットに対応できる精度の高い解析を可能にしました。

【研究成果】

新たに提案した分割手法 PsiPartition は、従来の分割手法よりも優れた性能を示しました。実際の塩基配列データセットを用いた解析結果では、本手法の BIC 及び AICc は、他の分割法よりも優位に低減しています (図 1)。特に、新たな手法は最適化パラメーターの重要性を示しており、他の手法とは異なる分割を生成し、異なる最適な分割スキームを効率的に探索することが可能です。また、本手法は、アミノ酸配列データにおいても優れた解析結果を与え、BIC 及び AICc が改善されました (図 2)。PsiPartition は、従来の IQ-TREE⁶ や他の手法と比較しても精度が高く、ベイズ最適化を活用することで、より正確な系統樹の再構築が可能です (図 3)。シミュレーションデータセットにおいても PsiPartition は異質性部位間の進化的多様性を適切に反映し、他の手法よりも優れた系統樹を再構築することができました (図 3)。樹形の統計的支持率はブートストラップ値とよばれ、本手法では、この値が従来法に比べて高くなっています。また、本研究の全ての評価において、ベイズ最適化は約 20 ステップで最高精度に到達します。このようにベイズ最適化が高効率かつ高速で PSI の最適化に適していることが示唆されました (図 4)。

PSI とベイズ最適化を組み合わせることで、ゲノムデータを高精度かつ効率的に分割することができます。この手法は、系統樹再構築の精度を向上させるだけでなく、分割モデル全体の堅牢性を高めており、生物進化の研究に幅広く応用できます。

【今後への期待】

本研究においてベイズ最適化に基づいた PsiPartition 法は、ゲノムデータの最適な分割を効率的か

つ正確に決定できることを実証しました。今後、この手法は他の分割手法にも拡張可能であり、大規模ゲノムデータの系統解析における分割手法として広く活用が期待されます。また、より複雑なデータセットや新たな遺伝的異質性を持つ種の解析に対しても、精度と効率の向上が期待できます。本手法によって得られる知見は生物進化について新たな視点を与えうるものであり、またこの手法はタンパク質の機能解析や生物工学分野に加えて、医学、薬学分野の研究に広く活用されることが期待されます。

【謝辞】

本研究は、北海道大学 DX 博士人材フェローシップ、JST/JICA 地球規模課題対応国際科学技術協カプログラム SATREPS「持続可能な漁業を実現する高付加価値バイオ素材の有効利用 (ReBiS)」(JPMJSA2206)、文部科学省・日本学術振興会科学研究費助成事業 学術変革領域研究 A「化学構造リプログラミングによる統合的物質合成科学創成」(JP24H02213)、基盤研究 B (JP24K01533) の支援のもとで行われました。

論文情報

論文名	PsiPartition: Improved Site Partitioning for Genomic Data by Parameterized Sorting Indices and Bayesian Optimization (PsiPartition: ゲノムデータ解析のためにパラメータ化されたソート指標とベイズ最適化により改良された部位分割化手法)
著者名	許 仕傑 ¹ 、小野田晃 ^{1,2} (¹ 北海道大学大学院環境科学院、 ² 北海道大学大学院地球環境科学研究院)
雑誌名	Journal of Molecular Evolution (分子系統学の専門誌)
DOI	10.1007/s00239-024-10215-7
公表日	2024年12月5日(木)(オンライン公開)

お問い合わせ先

北海道大学大学院地球環境科学研究院 教授 小野田晃 (おのだあきら)

T E L 011-706-2257 F A X 011-706-2257 メール akira.onoda@ees.hokudai.ac.jp

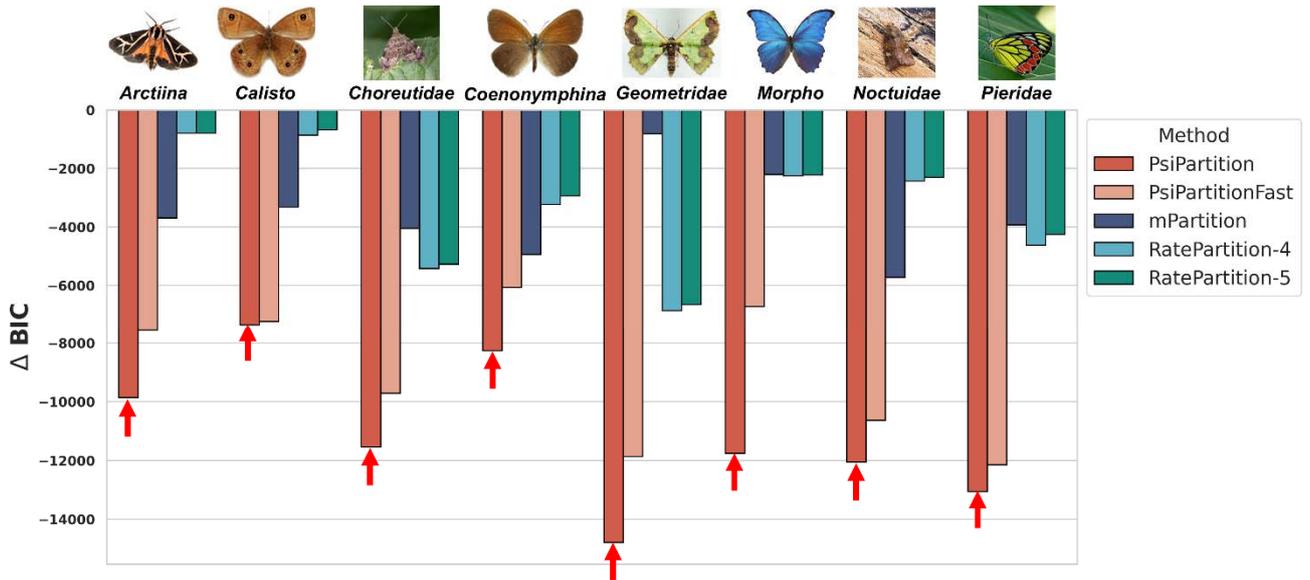
U R L <https://onoda-lab.jp>

配信元

北海道大学社会共創部広報課 (〒060-0808 札幌市北区北8条西5丁目)

T E L 011-706-2610 F A X 011-706-2092 メール jp-press@general.hokudai.ac.jp

【参考図】



*図のクレジットはCATERPILLAR LAB INC.、Wikipedia、British Butterfly Aberrations、及びbutterflyhouse.com.auに帰属します

図 1. 実際の塩基配列データセットにおける PsiPartition と他の手法の比較。 ΔBIC が小さいほど優れている。赤い矢印は本手法であり、優れていることを示している。

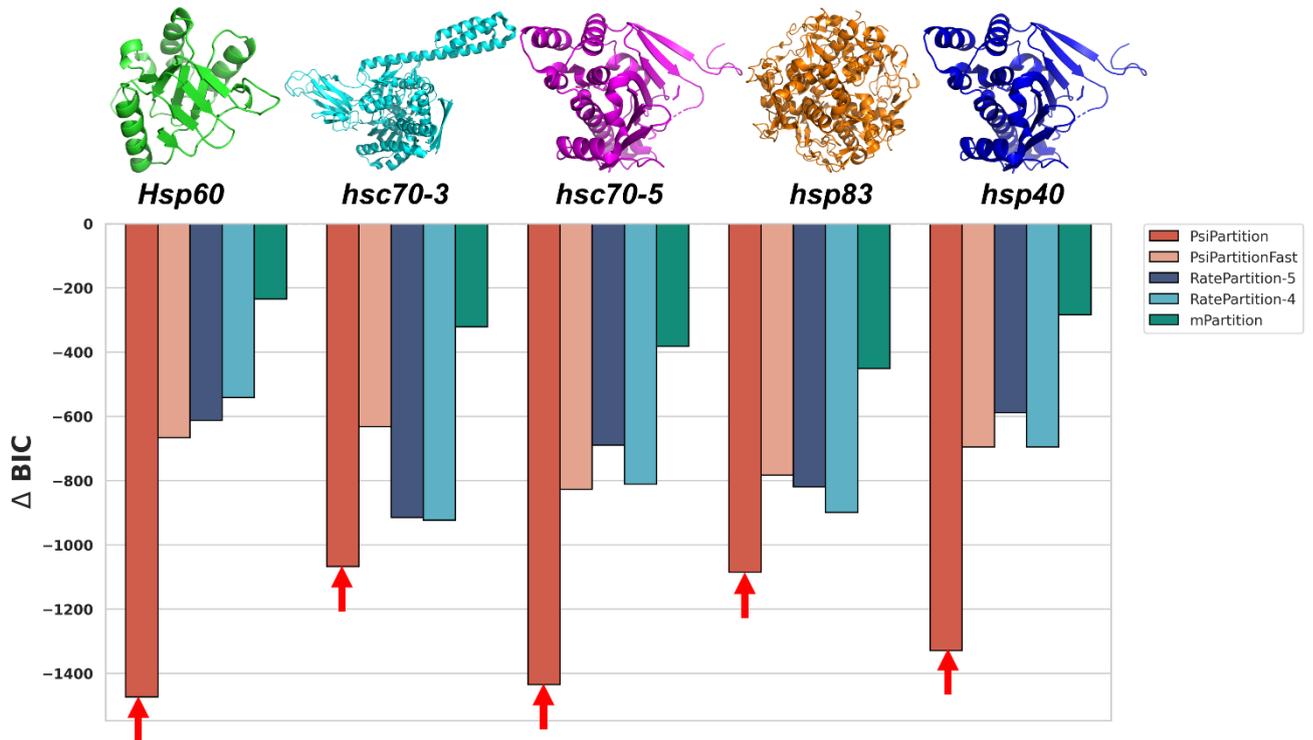
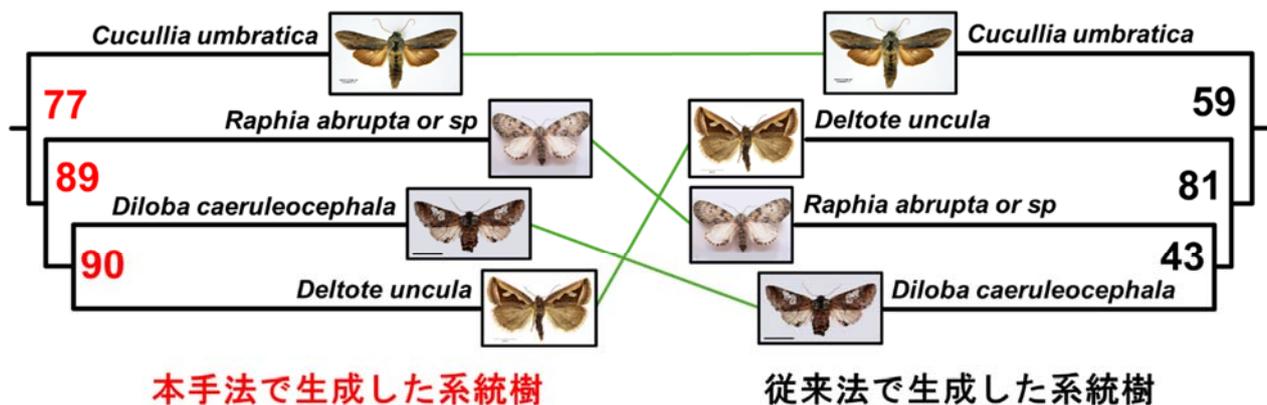


図 2. 実際のアミノ酸配列データセットにおける PsiPartition と他の手法の比較。 ΔBIC が小さいほど優れている。赤い矢印は本手法であり、優れていることを示している。



*樹形図内の値は分岐群のブートストラップ値を示しており、高いほど優れている。
 *写真のクレジットはBert Gustafsson、Jim Moore、JC Schou、及びWikimediaに帰属する。

図 3. 本手法を用いて生成した系統樹。赤い矢印は本手法であり、従来法より優れていることを示している。

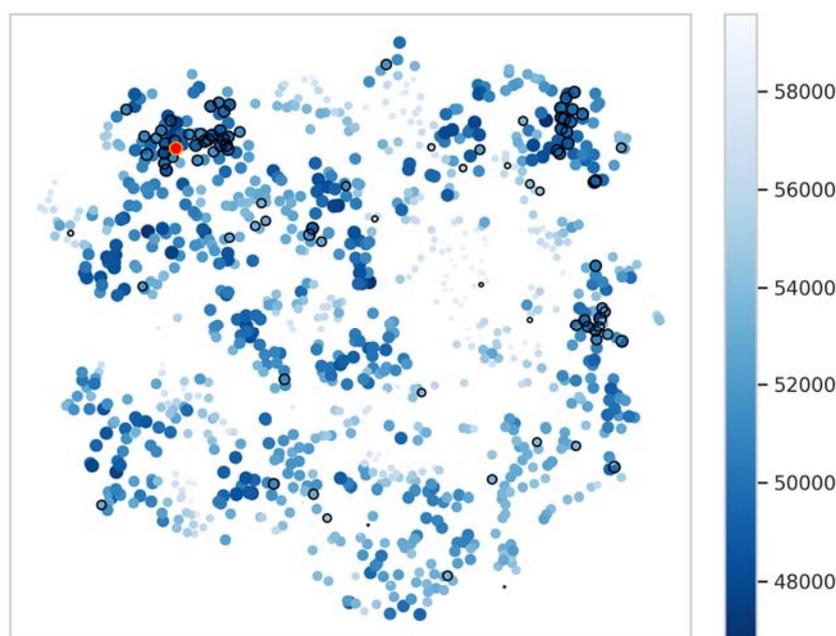


図 4. 蝶の塩基配列データセットにおけるベイズ最適化の軌跡。各点は再構築された系統樹を表し、色の濃さは Δ BIC を示す。最適な系統樹は赤で示しており、最適解付近に解が集まっていることから、効率的な最適化であることが分かる。

【用語解説】

- *1 系統解析 … 生物化の歴史を調べ、それぞれの種がどのように関係しているかを明らかにすること。
- *2 遺伝的異質性部位 … 生物の遺伝情報における遺伝子が異なっている特定の位置や部分のこと。例えば、DNA やタンパク質の中で、特定の塩基やアミノ酸が異なっている場所。
- *3 ベイズ最適化 … 少ない試行で最適な解を効率よく見つける方法。過去の結果を元に次に試すべき最適なパラメータを予測し、高精度な結果を短期間で得ることができる。
- *4 系統樹 … 生物の進化的な関係を示す図。異なる種やグループがどのように進化してきたかを示し、共通の祖先を持つグループ同士を枝分かれさせて表現される。

- *5 機械学習 … コンピュータが大量のデータを使って、パターンや法則を学び、自分で予測や判断をする技術。
- *6 IQ-TREE … 塩基配列やアミノ酸配列を用いた最尤法による系統樹探索プログラムの一つで、高機能的かつ高速で解析できるため、広く活用されている。