

ゲノムデータに基づく知識発見

京都大学化学研究所 バイオインフォマティクスセンター 化学生命科学研究領域 黒西 愛

研究成果概要

本研究では、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、ウイルスゲノム配列の分類法の検討を行った。ウイルスの分類は、ウイルスの形態、宿主域、ゲノムの構成 (DNA/RNA、1 本鎖/2 本鎖)、複製機構、核酸・アミノ酸配列の類似性など、さまざまな基準の組み合わせから決定されている。病原体の同定や、生態学的な解析を行う際に、ウイルスの分類は不可欠な第一歩である。最近、ウイルスの分類・命名の国際組織である国際ウイルス分類委員会 (ICTV) では、メタゲノム解析から見つかった配列情報のみが見つかったウイルスも公式の分類体系に組み入れられるようになり、配列情報に基づいた正確なウイルス分類の必要性は高まってきている。

新たなウイルス配列を得たとき、既存のウイルス配列との配列比較から、そのウイルスと最も良く似たウイルス配列を見つけることが出来る。しかし、そのウイルスを分類しようとした時、既知ウイルスと同じ種のウイルスである、または同じ属の違う種である、ということは最もよく似ていることだけから判断することは出来ない。そこで我々は、既存のウイルス配列の種、属、科、目それぞれの類似性の閾値を設定して、正確な分類を実現することを考えた。また、ウイルスはかなり多様なため、このときの類似性の閾値はウイルスグループごとに異なると考えられる。

本研究では、Virus-Host DB (<http://www.genome.jp/virushostdb/>)に収録されている既知ウイルスの 8198 ゲノムの全配列 vs 全配列の配列類似性を計算し、ウイルスグループごとの配列類似性の閾値 (下限値) を算出した。次に、8198 ゲノムからランダムに選んだ 20% の配列をクエリ、残りの配列をレファレンスとして、クエリと最も高い配列類似性を持つレファレンス配列 (ベストヒット) の分類情報をクエリに割り当てる方法と、ベストヒットの分類情報を配列類似性がレファレンスウイルスの種・属・科・目のそれぞれの閾値を満たす場合のみ割り当てる方法を比較した。その結果、閾値を用いた方法ではベストヒットのみを用いた場合より正確に分類することが出来た。

発表論文 (謝辞あり) なし

発表論文 (謝辞なし) なし