

配列類似性に基づいたウイルスゲノム分類法の開発

Development of viral genome classification method based on sequence similarity

京都大学化学研究所 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、配列類似性に基づいたウイルスゲノム分類法の開発を行った。

配列解析技術の進展によって、生体あるいは環境からウイルスの完全長ゲノム配列が高速かつ容易に検出・決定できるようになった。その結果、配列情報のみが得られているウイルスが増加しており、ウイルスゲノム配列からウイルスを正確に分類する必要性が高まっている。同じ種のメンバーである 2 つのウイルスゲノム間の配列類似度は種によって変動するが、種によって異なるゲノム類似度閾値(適応的閾値)を用いたゲノム自動分類法はこれまで開発されていなかった。そこで、Virus-Host DB<sup>1</sup>に登録されているウイルスゲノム配列を用いて、新規ウイルスゲノム配列を適応的閾値に基づいて正確に分類する手法を開発した。

Virus-Host DB(2018 年 12 月版)に含まれるウイルスのうち、ゲノム配列が 2 つ以上登録されている分類群のゲノム配列を配列解析ツールである Mash<sup>2</sup>を用いて比較し、各分類群の配列類似度閾値を学習した。学習した適応的閾値を基に既知ウイルスゲノム配列の分類予測を行い、実際のウイルス分類と比較した。

適応的閾値を採用した分類方法は単純な方法よりも高い適合率でゲノム配列を分類することができた。データベースに含まれていない種のゲノム配列を分類した場合、適応的閾値によって高い精度でそれが新種であることを判定出来た。今後さらに多くのウイルスゲノム配列が公開データベースに蓄積されていくと考えられる。これにより閾値計算の元となる配列データと閾値の決定できる分類群が増加し、新規ウイルスゲノム配列の分類精度をより向上できることが期待できる。

発表論文(謝辞あり)

発表論文(謝辞なし)

なし

参考文献

1. Mihara T, Nishimura Y, Shimizu Y, et al. Linking virus genomes with host taxonomy. *Viruses*. 2016;8(3):66.
2. Ondov BD, Treangen TJ, Melsted P, et al. Mash: Fast genome and metagenome distance estimation using MinHash. *Genome Biol.* 2016;17(1):132-016-0997-x.