

ゲノムデータに基づく知識発見

Knowledge Discovery Based on Genome Information.

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、配列類似性に基づいたウイルスの分類法について検討した。

近年の配列決定技術の発達により、環境サンプルなどから数多くのウイルスゲノムの配列が決定されるようになったが、病原体の同定や、生態系の理解に役立てるためには得られたウイルスゲノムを正確に分類する必要がある。原核生物では、16S rDNA のような進化的に広く保存された配列を用いた分子系統解析により、生物間の遺伝学的・進化学的距離を比較することが出来る。一方、ウイルスはこうした共通遺伝子配列を持たないため、分類を決定することが難しい。

本研究の目的は、新たなウイルスゲノム配列を高速に分類する手法の開発である。ゲノムを比較する複数の尺度を考案し、ウイルスの各分類群に適した尺度と閾値を設定することで、様々な階級の分類（種・属・科・目など）に柔軟に対応した自動分類器を構築したいと考えている。今年度は平均ヌクレオチド一致度(Average nucleotide identity, ANI)[1]という尺度を用いて、ウイルスの種ごとに種内あるいは種間でどの程度の類似性がみられるのか、ANI の値からウイルスをどの程度分類できるのかを、2種類の ANI 計算法を利用して検討した。

ゲノムネットの Virus-Host DB に収録されているウイルスゲノムの全ての2配列のペアについて ANI を計算し、各種のウイルスについて種内と種間のウイルス配列ペアの ANI の分布を調べた。全ての種内の ANI が種間の ANI より大きい種が最も多く、このような種では ANI の閾値を用いることでゲノム配列を容易に分類出来る。一方で、いくつかの種では、種間の ANI に種内の ANI よりも大きいものがあり、全ての種を ANI のみによって正確に分類することは難しいと考えられる。今後は他のいくつかの尺度も組み合わせて、分類の精度が向上するかを確かめたい。

発表論文 なし

参考文献

[1] Konstantinidis KT, Tiedje JM. (2005) Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci U S A*,102(7), 2567-72.