

審査の結果の要旨

氏 名 安田 知弘

本論文は、近年のゲノム解析において、次世代シーケンサーの発展とともに大きな脚光を浴びているゲノム構造多型の解析をはじめとして、ゲノム解析のアルゴリズムならびにその情報理論的な側面について理論、計算機実験の両面から多面的に論じたものである。近年のゲノム解析においては、データ量の爆発的な増加にともない、それらを高速かつ高精度に解析するための総合的な情報科学技術が必要とされる。本論文は、そのような観点から、ゲノムの構造多型を高精度に求めるためのアルゴリズムを提案するとともに関連した情報科学問題の計算量について論じ、さらに計算機実験によってその精度を検証している。さらに、構造多型のいくつかの種類解析の際の精度向上を行う手法について提案するとともに理論的な解析と計算機実験による検証を行っている。また、ゲノム配列検索技術の並列化、新しい配列解析の解析アルゴリズムの提案など、精度だけではなく、ゲノム解析の高速化を図る手法の提案を行っている。

本論文は六章からなり、第一章では、本論文の背景と動機として、次世代シーケンサー技術や、それによって算出される構造多型情報をはじめとする様々な大量のゲノム情報について解説している。さらに、本論文の貢献とその構成について述べている。

第二章では、ゲノムの構造多型のうち、転座という減少に着目し、それを解析する問題について論じている。本章では、この問題のいくつかのバリエーションを示し、そのうちのひとつの問題がNP困難に属する一方で、多項式時間で解くことが可能な定式化も存在することを示している。さらに、計算機実験を通して、その多項式時間アルゴリズムが、実際に効果があることを示している。

第三章では、ゲノムの構造多型のうち、欠失部のより高精度な決定法を提案している。さらに、提案した決定法が、従来の半分のデータから同等の精度を達成できることを理論的に示すとともに、計算機実験を通してその効果を検証している。

第四章では、ゲノムの構造多型を検出するにあたって非常に重要な技術であるマッピングアルゴリズムのメニーコアプロセッサ上での実装に関して議論を行っている。そして計算機実験を通してその効率を検証している。

第五章では、構造変異やミスアセンブリによって引き起こされる配列同士の比較する新しい効率的なアルゴリズムを提案している。さらに、計算機実験を通してその有効性を検証している。

第六章では、本論文を総括し、本論文における貢献をまとめるとともに、将来の展望について議論を行っている。

このように、本論文は、次世代シーケンサー技術の進展にともなって爆発的に増加したゲノムデータ、特に今脚光を浴びている構造変異に関連したデータの高精度かつ高速な解析技術を、情報理論、統計理論、組合せパターンマッチング、並列化技術等のきわめて多面的かつ網羅的なアプローチによって、より高度な解析を実現し、ゲノム解析、あるいはバイオインフォマティクスといった分野に対して大きな貢献を行っている。

よって本論文は博士（情報理工学）の学位請求論文として合格と認められる。