

審査の結果の要旨

氏名 鈴木 慶彦

ゲノム中に反復して現れる配列は、配列間の違いを見分けることが困難なこともあり、解明は遅れている。そのため反復配列とその周辺のゲノム配列は誤りが多く、解決が待たれていた。2011年頃から利用が可能になった長鎖 DNA シーケンシング技術は、反復配列の解読を容易にすると期待されていたが、この技術を適切に活用する情報学が必要であった。2014年より論文提出者は本問題に取り組み、新たなアルゴリズムを研究開発した。反復配列の重要な例である腸内細菌叢とセントロメアに適用し、ゲノム解読の完成度を上げることに成功し、提案手法の有効性を示した。

本論文は4章に分かれている。第1章では、近年の長鎖 DNA シーケンシング技術の進展により、反復配列を解明することに期待がもてるようになった背景を説明している。つづいて、長鎖 DNA 配列を適切に処理する情報学的技術も並行して研究する必要になった動機を解説している。

第2章では、ヒト腸内細菌叢の解読の困難さについて、まず解説している。腸内細菌叢とは細菌の集合体であり、単離が困難な細菌も含んでいるため、細菌の集合体のままゲノムを解読する必要がある。細菌のゲノムの大きさは数百万塩基対であり、ヒトゲノムに比べると3桁小さく、ゲノム解読は容易と誤解されることもあるが、実は難しい。その理由は、細菌の種類が数百に及び、各細菌の存在量が一律でなく多様性に富み、特に存在量の多い細菌由来の DNA 配列が反復して観測・解読されるからである。論文提出者はこの問題を解くため、2014年より世界に先駆けて長鎖 DNA シーケンシングを使った腸内細菌叢ゲノム解読の可能性を考察した。集合体として DNA 配列を解読した後に、配列を異なる細菌へと分けるアプローチを確立した。この方法を使うと各細菌のゲノムだけでなく、細菌に存在するプラスミドやファージのゲノムを解読することも容易になった。ただしプラスミドやファージの宿主細菌を推定することは従来は困難であった。論文提出者は、同一細胞内の DNA メチル化状態が一致することをヒントに宿主細胞を推定することに成功した。この手法をヒト腸内細菌叢 13 検体に応用した結果、新しいプラスミドとファージゲノムを数多く発見し、その多様性と機能的意義を考察できた。

第3章ではセントロメアの配列決定方法について考察している。セントロメアは数百塩基長の基本配列が変異した配列が数千から数万回反復したゲノムの領域である。基本配列は生物種ごとに大きく異なり、未解明な暗黒領域として現在も理解が待たれている。論文提出者は、2019年から普及した塩基精度が99%以上となる長い DNA 断片配列（約2万塩基長）の特長を活用して、セントロメア領域の反復配列を見分けるため1塩基の変異も利用する感度の高い数理的アルゴリズムを研究開発した。この方法を利用すると、反復配列中の異なる位置から由来する DNA 断片配列の違いをも検出することが可能になり、セントロメア解読が大きく前進した。本手法の精度を検証するために、あらかじめ答えが分かっている合成データと、モデル生物・ショウジョウバエから得られた現実のデータを使って、既存のソフトウェアと比較している。

合成データでは既存手法を凌駕する高い予測精度を達成し、ショウジョウバエ由来の現実のデータでは既存手法に比べ著しく長く正確なセントロメア配列を出力することに成功している。

第4章では既存手法に比べ提案手法が優位であることを分析する一方で、今後の改善の方向性についても検討している。

本論文で述べられた研究過程と結果は十分に学位取得に値するものであると判断する。なお本論文第2章は森下真一・服部正平らと、第3章は森下真一・Eugene Myersとの共同研究であるが、論文提出者が主体となって分析及び検証を行ったもので、論文提出者の寄与が十分であると判断する。

よって本論文は博士（科学）の学位請求論文として合格と認められる。

以上1587字