

論文審査の結果の要旨

氏名 山田智之

本論文は全5章からなり、第1章では本論文を発表するに至った背景について記述している。第2章では大規模ゲノム中の特異領域と反復領域の新規の探索手法を提案している。第3章では提案手法の実データへの応用例について述べている。第4章以降において一連の研究成果に対して考察を加えている。

ゲノム配列解析が多様化・大規模化するにつれて、遺伝子配列のゲノム上での位置を特定するといった単純なアライメントの課題だけではなく、遺伝子の発現量や多様性を解析するという課題が生じたこと、或いは、ゲノム配列や、遺伝子配列においてその配列を特徴づける配列を発見する課題が重要性を増してきたことに基づき、高頻度に出現する配列領域を発見し取り除くという課題や、特異性の高い領域を発見するといった課題の重要性を研究の動機として挙げている。これらは配列全体を概観する際に重要である他、SAGE等のタグ配列の解析や、遺伝子配列に対するプライマーや siRNA等の配列設計を行う際に有用であることが広く認識されているものである。

本論文では大規模な生物配列情報から特異領域と反復領域を効率的に発見するアルゴリズムを提案し、その応用として基礎的な生物実験を設計する方法について述べている。いくつかのエラーが許されたとしても依然として他の配列にアライメントされないような極めて特異性の高い配列を発見するためには、類似の配列を見落としなく列挙する方法が必要である。本論文ではオーバーラップシードを用いた高速な配列アライメントアルゴリズムを提案しているが、このアルゴリズムはその要請を満たしているものである。オーバーラップシードを用いることによって短い配列がクエリーである場合の検索に必要な計算を減らすことで、高速な検索が可能になっている。一方BLAST、SSAHA、BLATやPatternHunterのアルゴリズムは高い確率で類似の配列を発見する方法であるため、一定の割合で類似配列を見落とし、特異領域を発見するアルゴリズムには利用することはできない。

本論文で提唱するオーバーラップシードを用いるアルゴリズムによって、すべての類似配列を高速に列挙することができるようになったが、クエリー配列の長さが長くなると、特異性の正確な指標を計算することは困難となる問題が残っていた。このため、本論文では、特異性の指標の下界を高速に求める方法を検討している。特異性の指標の下界を用いると、全ての特異領域を発見することはできないが、発見された領域が特異領域であることは保証できる。この計算を行うために接尾辞配列等いくつかのデータ構造を導入しているが、このデータ構造は反復領域の発見にも利用することができる。

本論文の後半では、これらのアルゴリズムの生物実験設計への応用例を紹介している。siRNA配列、マイクロアレイ配列、マルチプレックスPCRプライマーを用いる実験において、精度の高い実験を効率よく行うためには対象配列に対しての特異性を保証することは重要であり、本論文のアルゴリズムを活用して設計システムを構築している。これまでも、大規模なゲノム配列が続々と解読されているが、今後は454, solexa, SOLiDなどの新たなシーケンシング技術によって更に解読される配列数は莫大なものになると考えられる。この新たな局面を迎えて、ゲノム科学はゲノム配列の解読からゲノム配列の活用へとその領域を拡大してきている。また、遺伝子配列の決定方法や発現量の測定方法も大きく変わろうとしている。本論文で提案する手法はそのような今後の新しい実験手法にも応用可能なものであると期待される。

なお、本論文第2章の一部は、森下真一、西郷薫、程久美子、内藤雄樹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。したがって、博士(科学)の学位を授与できると認める。