

論文審査の結果の要旨

専攻名 システム創成工学専攻

氏名 遠藤友基

ゲノム解析の分野においては、次世代シーケンサと呼ばれる低コストで大容量の塩基配列を読み取る装置の登場により、解析するデータの規模が巨大化している。ゲノム解読において必要となるDNAの塩基配列の決定においても、次世代シーケンサにより読み取られた、大量のリードと呼ばれる塩基配列の断片を繋ぎ合わせることで元のDNA配列を決定する。このリードから元の塩基配列を決定する技術は、アセンブリアルゴリズムと呼ばれ、そのうち、未知である塩基配列を決定するものは、de novo アセンブリアルゴリズムと呼ばれる。これまでにVelvetやSOAPdenovoなど多数のde novoアセンブリアルゴリズムを実装したソフトウェアが公開されており、様々な種のDNA塩基配列の決定に用いられている。de novoアセンブリは、大量のリードを読み込み、リード間の重複関係を調べて、グラフ表現で表す必要があるため、非常に多くのメモリを消費することで知られている。したがって、ヒトよりも更に大規模なゲノムにおけるDNA全塩基配列を決定する場合、大容量メモリを搭載したスーパーコンピュータを用いても、メモリ不足となってしまうため、メモリ消費量の少ないde novoアセンブリアルゴリズムの開発が望まれている。本論文では、こうした背景を踏まえ、大規模なゲノムに対してもde novoアセンブリが可能となるように、消費メモリ量の少ないde novoアセンブリアルゴリズムを提案している。提案手法では、リード間の重複関係を表すグラフの辺の情報は保持せずに、ノード間に辺が存在するかは、必要な時に効率的に調べることにより、大幅なメモリ削減を実現している。

以下に本論文の構成と各章の内容について述べる。

第1章では、本研究の背景と、研究の目的および本論文の構成について述べている。

第2章では、DNAの構造、塩基配列とゲノムとの関係について概説している。さらに、ゲノムの持つ役割とゲノム解読の有用性について述べている。

第3章では、ゲノム解読における最初のステップとなる、塩基配列の決定方法の手順について述べている。まず、リードを読み取るためのシーケンシングの原理について述べ、次世代シーケンサの特徴について説明している。次に、リードの重複関係を調べて繋ぎ合わせる処理となるde novoアセンブリについて述べている。

第4章では、提案手法のベースとなる、de Bruijnグラフを用いたde novoアセンブリアルゴリズムの原理と課題について説明している。多くのde novoアセンブリアルゴリズムでは、de Bruijnグラフと呼ばれるグラフを用いてアセンブリが実現されており、提案手法においても、de Bruijnグラフをベースとして、アセンブリを実装している。

第5章では、提案手法である、消費メモリ量を大幅に削減したde novoアセンブリアルゴリズムについて述べている。提案手法では、de Bruijnグラフ構築時にグラフの辺の情報を保持しな

いなど、最低限の情報のみをメモリ上に保持することで大幅な消費メモリの削減を実現している。

第6章では、実際に次世代シーケンサから得られたリードを用いて、提案手法の有効性を確かめるための実験を行っている。現在広く用いられているde novoアセンブルソフトであるVelvetとSOAPdenovo2との比較を行い、メモリ削減による提案手法の有効性を示している。

第7章は本論文のまとめであり、本研究で得られた成果をまとめ、今後の課題・展望について述べている。

本論文については、2015年2月16日に911講義室において、審査委員全員及びこの分野の研究者の出席のもとに公聴会が開催され、研究内容の発表及び質疑応答が行われた。公聴会終了後ただちに学位審査委員会を開催し、本論文の内容について詳細に検討した。その結果、本研究成果はゲノム解析分野への学術的寄与が高く、実用分野での貢献も期待できると判断された。また、研究内容の学術的水準と独創性においても極めて優れていると評価された。

よって、本論文は博士（工学）の学位論文に値するものと認める。