

論文 / 著書情報
Article / Book Information

題目(和文)	環境中の微生物叢解析のための超高速な配列解析ツールの開発
Title(English)	
著者(和文)	矢野雅大
Author(English)	Masahiro Yano
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第10498号, 授与年月日:2017年3月26日, 学位の種別:課程博士, 審査員:山田 拓司,黒川 顕,伊藤 武彦,山口 雄輝,中島 信孝
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第10498号, Conferred date:2017/3/26, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	審査の要旨
Type(English)	Exam Summary

(博士課程)

論文審査の要旨及び審査員

報告番号	甲第	号	学位申請者氏名		矢野 雅大	
論文審査 審査員		氏 名	職 名		氏 名	職 名
	主査	山田 拓司	准教授	審査員	中島 信孝	准教授
	審査員	黒川 顕	特任教授			
		伊藤 武彦	教授			
		山口 雄輝	教授			

論文審査の要旨（2000 字程度）

本博士論文は、環境全般の動態を理解および制御することを最終的な目的として、環境中の微生物叢についての知見をより容易に得られるようにするために、メタゲノム解析とメタ 16S 解析における大規模な配列相同性検索と配列クラスタリングを、感度や精度を維持したまま高速化する手法についての研究をまとめたものである。本博士論文は、序論と総括を含む次の 4 つの章によって構成されている。

第一章では、序論として研究の背景について述べている。第一に生物学における問題の計算可能性が計算機環境とアルゴリズムにより左右されることに言及し、生物学研究を推進する上での計算機環境の適切な活用とアルゴリズムの適切な設計の意義を明記している。第二に環境中の微生物叢は代謝活動を通じて環境と相互に作用を及ぼしていることに言及し、環境中の微生物叢について知見を得る意義を明記している。第三に環境中の微生物叢についての知見を得る手段としてのメタゲノム解析とメタ 16S 解析について説明したうえで、メタゲノム解析とメタ 16S 解析においては配列相同性検索の計算量の増大が解析を困難にする要因となっていることを提起している。第四に配列相同性検索の計算量を削減する手段として、科学技術計算用途での GPU の適用と配列クラスタリングの現状について概説している。これらより論文提出者は GPU の活用と新しい配列クラスタリングアルゴリズムの開発により、メタゲノム解析とメタ 16S 解析での配列相同性検索と配列クラスタリングの高速化が達成されれば、環境中の微生物叢についての知見がより容易に得られるようになる」と説明している。

第二章では、論文提出者が開発した GPU を用いた超高速な配列相同性検索ツール CLAST について説明および検討している。CLAST の設計の特徴として、第一に一度に複数のデータベース配列と複数のクエリ配列の間で同時に配列相同性検索を行うことで GPU の性能を引き出せるという点、第二にデータベースの事前のインデックス化を行わないため、大規模なデータベースの頻繁な更新に即応できるという点、第三にアライメントを行う際にクエリ配列の端から端までデータベース配列と比較するモード (Global モード) と最もビットスコアが高くなる範囲を計算するモード (Local モード) を持つことで、モチーフ解析用途にも系統推定用途にも使用できるという点について説明している。CLAST の性能評価試験の結果として、第一に CLAST の速度は Bowtie 2 と同等であったという点、第二に CLAST の感度は BLAST および BLAT の感度と同等であったという点を確認している。以上の結果から、GPU を使用することで配列相同性検索は感度を維持したまま高速化できることを示している。

第三章では、論文提出者が開発した高速な配列クラスタリングツール Saturn について説明および検討している。Saturn は高速なクラスタリングを行うためにグリーディ・アルゴリズムを用いてクラスタリングを行う。Saturn の設計上の特徴として、第一に同一長かつ完全に一致している配列 (重複配列) の本数を数え、その本数が多い順に配列を真のクラスタ中心と推定し、その配列に対してユーザーが指定するよりも高い相同性が認められた配列をクラスタリングしていく点、第二にクエリ配列の集合から仮想的なクラスタ重心配列を推定し、そのクラスタ重心配列からの距離を基準にしてクラスタリングを行う点について説明している。Saturn の性能評価試験における結果として、第一に Saturn の速度は UCLUST の速度の最大で 1.39 倍高速であったという点と、第二に Saturn の精度は UCLUST、DNACLUSt、CD-HIT それぞれの精度よりも高かったという点を確認している。以上の結果から、重複配列の本数を用いることで精度を維持したまま高速な配列クラスタリングが可能になることを示している。

第四章では、以上の結果を統合し、生物学研究のための計算機環境とアルゴリズムについての研究という観点から、CLAST と Saturn が達成した革新の意義について説明している。

以上を要するに、本博士論文は環境中の微生物叢を大規模に解析するための計算手法について新たなアルゴリズムを提案し実装したものであって、工学上貢献するところ大である。よって、本論文は博士 (工学) の学位論文として十分な価値があるものと認められる。

注意：「論文審査の要旨及び審査員」は、東工大リサーチリポジトリ (T2R2) にてインターネット公表されますので、公表可能な範囲の内容で作成してください。