

論文 / 著書情報
Article / Book Information

題目(和文)	ハイスループットDNAシーケンスデータによる全ゲノム配列の構築と解析に関する研究
Title(English)	
著者(和文)	梶谷嶺
Author(English)	Rei Kajitani
出典(和文)	学位:博士(理学), 学位授与機関:東京工業大学, 報告番号:甲第9728号, 授与年月日:2015年3月26日, 学位の種別:課程博士, 審査員:伊藤 武彦,工藤 明,徳永 万喜洋,黒川 顕,山口 雄輝
Citation(English)	Degree:, Conferring organization: Tokyo Institute of Technology, Report number:甲第9728号, Conferred date:2015/3/26, Degree Type:Course doctor, Examiner:,,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	論文要旨
Type(English)	Summary

論文要旨

THESIS SUMMARY

専攻： Department of	生命情報	専攻	申請学位 (専攻分野)： 博士 (理学)
学生氏名： Student's Name	梶谷 嶺		指導教員 (主)： 伊藤 武彦
			指導教員 (副)：
			Academic Advisor(sub)

要旨 (和文 2000 字程度)

Thesis Summary (approx.2000 Japanese Characters)

本論文は「ハイスループット DNA シークエンスデータによる全ゲノム配列の構築と解析に関する研究」と題し、ハイスループット DNA シークエンサが出力する配列データから計算機上で全ゲノム配列を構築 (*de novo* アセンブリ) する手法および、関連するゲノム配列の特徴について述べたものであり、以下の 4 章から構成されている。

第 1 章「序論」では、本研究の背景として、DNA シークエンサのスループットの向上や全ゲノムショットガン法の導入がゲノム配列決定の効率化に寄与してきた経緯について述べている。また、それによって高速で複雑な *de novo* アセンブリアルゴリズムが必要とされていることを説明した上で、ゲノム解読計画の対象がモデル生物から非モデル生物へと広がる過程で、ゲノムのヘテロ接合性が *de novo* アセンブリの障害として顕在化したことを指摘している。これらの背景を踏まえ、高ヘテロ接合性サンプルのデータに対応し、更に汎用性も備えたハイスループットシークエンスデータ用の *de novo* アセンブリアルゴリズムを開発し、全ゲノム配列解析を効率化することを本研究の目的として述べている。

第 2 章「真核生物ドラフトゲノム配列の構築」では、新規開発された *de novo* アセンブラ *Platanus* のアルゴリズムを説明し、構造的な差異を含む相同染色体の配列を統合することで、高ヘテロ接合性サンプルから長いゲノム配列を構築する方法を新たに示している。*Platanus* の有効性を検証するため、線虫 *Caenorhabditis elegans* のシークエンスデータを基にヘテロ接合度のシミュレーションテストを行い、既存ソフトウェアの性能が低下する条件においても *Platanus* はアセンブリ結果の長さや精度の両方の観点で優れた性能を発揮することが明らかにされている。高ヘテロ接合性サンプルの実データによるベンチマークとしては、線虫 *Strongyloides venezuelensis* と牡蠣 *Crassostrea gigas* のシークエンスデータの *de novo* アセンブリを行い、シミュレーションデータだけでなく実データにおいても本手法が効果的であることを示している。このベンチマークでは、fosmid クローニングにより得られたゲノムの部分配列を用いて *Platanus* のアセンブリ結果が高精度であることを確認し、さらに相同染色体間の構造的な差異を統合することで既存手法ではアセンブリが困難なゲノム領域が構築されていることを確認している。本章の後半では、ゲノムサイズが比較的大きい (1 Gbp 前後) 3 生物種のデータを用いたベンチマークにおいても *Platanus* が高い性能を発揮することを確認し、実際のゲノム解読計画への応用例としてシーラカンス *Latimeria chalumnae* のゲノム配列決定・解析の詳細を述べることで、*Platanus* が高い汎用性を持つことを示している。

第 3 章「原核生物の全ゲノム、環境ドラフトゲノム配列の構築」では、*Platanus* が原核生物のシークエンスデータに対しても適用可能であり、真核生物と比較してゲノムサイズの小さいバクテリアサンプルに関してはギャップ部分の無い完成ゲノム配列が構築可能であることを示している。大腸菌 *Escherichia coli* の 2 株のシークエンスデータを用いたベンチマークでは、データ量、シークエンスライブラリ構成、DNA サンプル調整方法等を変えた様々な条件でアセンブリを行い、完成ゲノム配列決定のための効率的な方法について考察している。本章の後半では、*Platanus* を基に新規開発された環境メタゲノム用 *de novo* アセンブラ *MetaPlatanus* について述べている。*MetaPlatanus* は部分配列の組成や配列毎のシークエンス量の情報を用いて配列をクラスタリングする機能を持ち、メタゲノムデータから生物種毎の配列クラスタを出力するよう設計されている。複数種のバクテリアを含むメタゲノムデータを用いたベンチマークでは、*MetaPlatanus* の結果配列の精度と長さについて既存ソフトウェアより優れた性能を持つことが示され、環境 DNA サンプルから難培養性の生物種を含むドラフトゲノムが構築可能であることも示唆された。

第 4 章「総括」では各章の結果をまとめ、研究の意義と開発された手法の応用可能性について論じている。

以上を要するに、本論文はハイスループット DNA シークエンスデータによる全ゲノム配列の構築・解析方法を確立し、非モデル生物を含む生物種のゲノム解析を効率化することに成功しており、遺伝学・ゲノミクス上貢献するところが大きい。よって、本論文は博士 (理学) の学位論文として十分な価値があるものと認められる。

備考：論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 1 部提出してください。

Note: Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1 copy of 800 Words (English).

注意：論文要旨は、東工大リサーチリポジトリ (T2R2) に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。

Attention: Thesis Summary will be published on Tokyo Tech Research Repository Website (T2R2).

(博士課程)
Doctoral Program

論文要旨

THESIS SUMMARY

専攻 : Department of	生命情報	専攻	申請学位 (専攻分野) : Academic Degree Requested	博士 (理学) Doctor of (Science)
学生氏名 : Student's Name	梶谷 嶺		指導教員 (主) : Academic Advisor(main)	伊藤 武彦
			指導教員 (副) : Academic Advisor(sub)	

要旨 (英文 300 語程度)

Thesis Summary (approx.300 English Words)

Whole-genome shotgun sequencing is the efficient method to determine an entire genome sequence. However, construction of genome sequences *in silico* (*de novo* assembly) is the challenging issue due to high throughput of DNA sequencers and repetitive sequences in genomes, and a more efficient and accurate algorithm is required. In addition, it has been reported that heterozygosity of non-model organisms increased complexity of *de novo* assembly, resulting in low contiguity of assembled sequences.

In this study, I developed novel *de novo* assembler, called Platanus, to construct genomes of diploid organisms including highly heterozygous ones. Platanus firstly transforms sequenced DNA fragments (reads) into de Bruijn graph and generates contiguous sequences (contigs). Next, it determines the order of contigs using paired-ends information. Here, heterozygous genomic regions including structural difference between homologous chromosomes are merged in the novel function. In the benchmarks, Platanus exhibited higher accuracy and contiguity for both simulated heterozygous and real-world data from diploid organisms than existing *de novo* assemblers. Structural differences between homologous chromosomes merged by Platanus were confirmed using fosmid sequences, and it is shown that the assembly result of Platanus is able to substitute that of the hierarchical shotgun method, which is more cost-consuming compared to whole-genome shotgun.

Based on Platanus, I also developed the *de novo* assembler, named MetaPlatanus, for metagenomic data. Assembly of the sequences from environmental DNA samples is effective to analyze the genomes of uncultured organisms. MetaPlatanus is implemented to handle with data whose coverage depths are uneven, and it generates clusters of assembled sequences using the information of sequence composition. The accuracy of the result of MetaPlatanus was evaluated using the data consisting of multiple bacteria, and its performances exceeded those of existing tools. Consequently, the methods I developed in this study are expected to accelerate genome projects regardless of highly heterozygous or uncultured organisms.

備考 : 論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 1 部提出してください。

Note : Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1copy of 800 Words (English).

注意 : 論文要旨は、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。

Attention: Thesis Summary will be published on Tokyo Tech Research Repository Website (T2R2).