

令和 元年 6 月 10 日現在

機関番号：32658

研究種目：若手研究(B)

研究期間：2017～2018

課題番号：17K18045

研究課題名(和文) ナノポアシーケンサーの解析を野外で可能にするクラウド型解析環境の構築と公開

研究課題名(英文) Development a cloud-based analysis environment that enables bioinformatics analysis for nanopore sequencer in the field

研究代表者

志波 優 (SHIWA, Yuh)

東京農業大学・生命科学部・助教

研究者番号：00647753

交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：Oxford Nanopore Technologies社のナノポアシーケンサーであるMinIONは、フィールドでのシーケンスを可能とするポータブルシーケンサーとして注目され、幅広い応用が期待されている。本研究計画では、バイオインフォマティクス解析の容易化・効率化を目指して、IlluminaとMinION用の各種ツールをGUIで使用できるように、WebアプリケーションのGalaxyを利用した解析環境を構築した。本ツールの利用方法はホームページ (<https://github.com/youyuh48/pitanano-galaxy>) から閲覧可能である。

研究成果の学術的意義や社会的意義

実験研究者にとってシーケンスデータ取得後のバイオインフォマティクス解析は依然として敷居が高い。本研究では、WebアプリケーションのGalaxyをベースとしたMinION向けのクラウド型解析環境の構築を行った。本解析環境はコンテナ型環境のDocker上で動作し、ユーザはノートパソコン上でGalaxyサーバをDockerコンテナとして起動することで、Webブラウザを利用してグラフィカルな操作系でデータ解析が可能となり、バイオインフォマティクス解析の高度化に資すると期待される。

研究成果の概要(英文)：MinION, a nanopore sequencer from Oxford Nanopore Technologies, has increasingly attracted attention as a portable sequencer that is capable of sequencing in the field, and novel applications are expected. In this study, in order to facilitate bioinformatics analysis for non-bioinformatician, we developed an analysis environment on the web-based platform Galaxy so that various tools for Illumina and MinION can be used in GUI. The usage of this tool can be viewed from <https://github.com/youyuh48/pitanano-galaxy>.

研究分野：バイオインフォマティクス

キーワード：微生物ゲノム バイオインフォマティクス ナノポアシーケンサー MinION Galaxy

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

1. 研究開始当初の背景

Oxford Nanopore Technologies 社のナノポアシーケンサーである MinION は、フィールドでのシーケンスを可能とするポータブルシーケンサーとして注目され、幅広い応用が期待されている。しかしながら、フィールド研究者にとってシーケンスデータ取得後のバイオインフォマティクス解析は依然として敷居が高い。現状では用途に応じ、複数のソフトウェアを組み合わせる必要があり、そもそも最適なソフトウェアの組み合わせを検証することから始めなければならない。また、これらのソフトウェアはコマンドライン・インタフェースで操作するため、インストール作業からある程度のコンピュータスキルが要求されることも多い。加えて、既存ソフトウェアで不足している機能はプログラミングの必要があり、さらに解析のハードルを上げている。バイオインフォマティクス解析環境の改善は、MinION の応用範囲の拡大に不可欠の課題である。

2. 研究の目的

本研究計画では、バイオインフォマティクス解析の容易化・効率化を目指して、MinION 向けのクラウド型解析環境の構築と公開を行う。ユーザは Web ブラウザを利用してグラフィカルな操作系で容易にデータ解析が可能となる。また、解析用途に応じた最適なソフトウェアの組み合わせをパイプラインとして提供することで、解析の高度化に資することが期待される。

3. 研究の方法

(1) 解析ソフトウェアの比較検討と選定

本研究で対象とする解析は、参照配列と比較し変異を検出するリシーケンス解析、新規に配列を決定する *De novo* アセンブリ解析、細菌叢解析のメタゲノム解析の3種類とした。それぞれの解析において一般的なツールを評価し、解析系に組み込むツールを選定した。各解析ごとに、以下のソフトウェアを検討した。

得られたリード配列の品質評価(QC; Quality control)や品質不良配列のフィルタリング用に、poretools, Porechop, NanoPlot 等のソフトウェアを検討した。リシーケンス解析では、得られた塩基配列を参照配列へマッピング用に BWA-MEM を検討した。

De novo アセンブリ解析では、得られた塩基配列の重なりを利用して新規配列を決定する。Canu/nanopolish, miniasm/racon, Unicycler 等のソフトウェアを検討した。メタゲノム解析では、得られた塩基配列から菌種組成の解析用に Centrifuge を検討した。

(2) クラウド型解析環境の構築

Web アプリケーションの Galaxy (<https://galaxyproject.org>) を利用することで、コマンドラインで操作する様々な解析ソフトウェアを Web ブラウザからグラフィカルな操作系で利用することが可能となる。まず、計算機上に Galaxy の環境を構築する。構築した環境を容易に他の計算機に移植できるよう、バーチャル・マシン (VM) と呼ばれる仮想環境上に構築する。VM 環境としては VirtualBox や Docker を検討する。続いて、選定した解析ソフトウェアを Galaxy 環境上で実行できるように実装する。

4. 研究成果

(1) 初年度

乳酸菌の全ゲノムシーケンスデータを MinION で取得した後、*De novo* アセンブリ解析を通じて、既存の解析プロトコルの調査、得られたリード配列の品質評価、アセンブル用のソフトウェアの比較、Illumina MiSeq から得られたアセンブリデータとの精度比較を行った。アセンブル用のソフトウェアである miniasm/racon、Canu、nanopolish について比較検討を行い、得られた最適なソフトウェアの組み合わせや解析パラメーターを、解析プロトコルとしてシェルスクリプトに実装した。

続いて、クラウド型 GUI 解析環境の構築に用いる Web アプリケーションである Galaxy の開発環境を構築した。開発環境を容易に他の計算機に移植できるよう、仮想環境である VirtualBox 上に Galaxy を構築した。また、Galaxy 環境で利用可能なバイオインフォマティクスツールが登録されている Galaxy ToolShed 上における既存の MinION 用ツールの登録状況についても調査を行うとともに、Galaxy 環境にコマンドラインツールを実装するために必要な依存関係の解決方法やツール XML ファイルの作成方法について調査を行い、テストツールの Galaxy 環境への実

装を行った。

(2) 最終年度

既存のコマンドライン上で動作する各種解析ツールを Galaxy 環境上に実装し、GUI 解析環境の構築を行った。VirtualBox 上に構築した Galaxy 環境では、仮想ディスクの拡張性に制約があることや環境構築が煩雑であったため、新たにコンテナ型環境の Docker をベースに Galaxy 環境の構築を行った。Docker 版 Galaxy の docker-galaxy-stable (<https://github.com/bgruening/docker-galaxy-stable>) をベースに、Illumina と MinION 用の各種ツールを GUI で使用できるように Galaxy に組み込んだ。現時点で使用可能なツールは次のとおりである。

- リードの QC 用
 - FastQC
 - Trimmomatic
 - NanoPlot
 - Porechop
- FASTQ ファイルの操作
 - Seqtk
- メタゲノム解析用
 - Centrifuge
 - Krona
- マッピング
 - BWA-MEM
- *De novo* アセンブル
 - Unicycler
 - Canu

これらのツールの他にも Galaxy ToolShed に登録されている様々な NGS 解析ツールが Galaxy 環境に追加インストール可能である。

本研究で構築したクラウド型 GUI 解析環境は Docker イメージとして Docker Hub (<https://hub.docker.com>) にイメージ名 “youyuh48/pitanano-galaxy” として公開した。本ツールの利用方法はホームページ (<https://github.com/youyuh48/pitanano-galaxy>) から閲覧可能である。

以上のように本研究では、Web アプリケーションの Galaxy をベースとした MinION 向けのクラウド型解析環境の構築を行った。本解析環境はコンテナ型環境の Docker 上で動作し、ユーザはノートパソコン上で Galaxy サーバを Docker コンテナとして起動することで、Web ブラウザを利用してグラフィカルな操作系でデータ解析が可能となり、バイオインフォマティクス解析の高度化に資すると期待される。

5 . 主な発表論文等

〔雑誌論文〕(計0件)

〔学会発表〕(計2件)

志波優、作ったツール、どうやってユーザーに届ける? ~ Galaxy, Docker, クラウドを活用する: Galaxy にツールを実装・公開するまでの流れ、第7回生命医薬情報学連合大会、2018

志波優、統合データ解析環境 Galaxy を使った再現可能なデータ解析: ワークフロー共有に Galaxy を利用する、第40回日本分子生物学会年会、2017

〔図書〕(計0件)

〔産業財産権〕

出願状況 (計0件)

名称：
発明者：
権利者：
種類：
番号：
出願年：
国内外の別：

取得状況（計 0 件）

名称：
発明者：
権利者：
種類：
番号：
取得年：
国内外の別：

〔その他〕

ホームページ等

<https://github.com/youyuh48/pi-tanano-galaxy>

6 . 研究組織

(1)研究分担者

研究分担者氏名：無し

ローマ字氏名：無し

(2)研究協力者

研究協力者氏名：無し

ローマ字氏名：無し

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。