

海洋環境解析に向けたメタゲノムおよび 1 細胞配列データ解析用パイプラインの開発

小林 健太^{1,a)} 加藤 有己¹ 谷口 丈晃² 丸山 徹³ 伊藤 通浩³ 五斗 進⁴ 竹山 春子³ 藤渕 航¹

概要: 多数の難培養微生物で構成される海洋環境の理解を促進するために、メタゲノム解析に注目が集まっている。しかしながら、メタゲノムのリード群を完全にアセンブルすることは困難を極めることが知られている。近年、次世代シーケンシング技術の発展とともに、1細胞ゲノムデータが利用できるようになってきた。本稿では、近年の利用可能な技術の動向を踏まえ、メタゲノムおよび1細胞配列データが与えられたとき、アセンブリ、遺伝子の構造および機能推定を行うパイプラインを発表する。具体的に、リードの前処理、アセンブリとアノテーションに特化したパイプラインを提供し、遺伝子アノテーションのマップの可視化を可能とするものである。

1. はじめに

海洋環境は多くの難培養微生物で構成されており、その理解のための方法論の1つとして、難培養微生物のメタゲノム解析が挙げられる。近年では目覚ましい次世代シーケンサー技術の発展があり、メタゲノム解析のためのリード配列を得ることが容易となってきた。しかしながら、一般にメタゲノム解析を行う上で、リード群のアセンブリが困難であることを考慮する必要がある。一方、次世代シーケンサーを利用した1細胞ゲノムデータの解析が盛んに行われていることも注目すべき点である。ここで、我々は難培養微生物の1細胞ゲノムおよびメタゲノム解析を容易に行えるよう、Web サイトから遺伝子の構造および機能推定ができるシステム (Single cell/meta genome analysis system) を開発した^{*1}。具体的に、プラットフォーム環境と解析データがあれば遺伝子アノテーションのマップの可視化作業ができる。以下に解析パイプラインの手法とwebサーバーについて述べる。

2. 手法

解析を円滑に行うためのWebアプリケーションとして、以下の2つを組み込む。

2.1 アセンブラ

配列アセンブリを行うツールとして、メタゲノムと1細胞の両方にアセンブリが可能なSPAdes [1], [4] と、次世代シーケンサーによるリードの深さが不均一でも正しく長いコンティグを構築することができるIDBA-UD [5] をシステムに組み込んだ。

2.2 遺伝子アノテーター

原核生物ゲノムは ab initio 遺伝子予測でゲノム上での遺伝子の位置を予測する。一方、メタゲノムに関しては、メタゲノム用のツール MetaGeneAnnotator [3] を使用し、遺伝子予測のためのリボソーム結合部位の特異的なパターンを検出する。

3. Web サーバー

Webアプリケーションとして実装したパイプラインは大別すると3つの工程に分かれる。まず第1に、リードに対する前処理を行い、既存ツールによって解析に用いるのに不適切なリードを除去し誤り訂正を行う。第2に、前処理済みのリードを用いてアセンブラを実行し、評価に必要な統計量、すなわち、コンティング、スキップフォールド数、N50、長さ (最長、最低、平均、総長) を計算する。最

¹ 京都大学 iPS 細胞研究所
Kyoto University, 53 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan

² 三菱総合研究所
Mitsubishi Research Institute, 2 Nagata-cho, Chiyoda-ku, Tokyo 100-8141, Japan

³ 早稲田大学 理工学術院
Waseda University, 2-2 Wakamatsu-cho, Shinjuku, Tokyo 162-8480, Japan

⁴ 京都大学 化学研究所
Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan

a) k.kobayashi@cira.kyoto-u.ac.jp

*1 近日公開予定である。

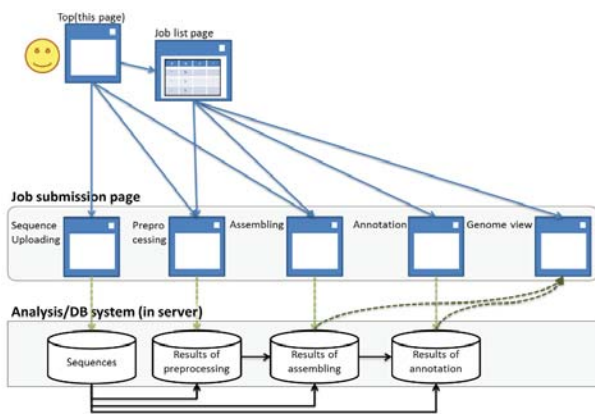


図 1 ユーザーインターフェース

後に、第 2 工程でアセンブラを実行した出力を受け取り、遺伝子構造アノテーションおよび遺伝子機能アノテーションを行う。また、アセンブリでは、前処理リードの他に、ディレクトリに存在する配列リードとアップロードした配列リードを使用することができる。図 1 にユーザーインターフェースの概観を示す。

3.1 各手法の操作

以下、各計算パイプラインについて Web 上での操作方法を説明する。

(1) 前処理パイプライン

前処理を行う Preprocessing ページは、データのアップロードで配列データを指定し、サブミットボタンをクリックするとジョブ一覧ページにジャンプする。リードに対する前処理では、クオリティが低いリードの除去、3' 端の低クオリティ領域の除去、N が多いリードの除去、短いリードの除去の各設定を行い、それらを除外した処理済みのリードを作成することができる。

(2) アセンブリパイプライン

アセンブラを実行する Assembling ページでは、前処理ページ同様に、Assembling ページにあるデータのアップロードで配列データを指定し、サブミットボタンをクリックするとジョブ一覧ページにジャンプする。また、オプションとして前処理済み配列データのほかに、配列データのアップロード、アップロードした配列の使用、所定のディレクトリに存在する配列の使用が指定できる。

(3) ジョブ一覧ページ

サブミットボタンクリック後の解析結果一覧ページでは、解析ジョブに関する情報（ジョブ名、統計値）が一覧表示される。また、ゲノムビューページでは、アノテーションを可視化したマップを表示することが可能である（図 2 参照）。さらに、ゲノムビューページ下部にある KEGG map をクリックすることで、リンク先の KEGG マップ [2] を閲覧することができる。

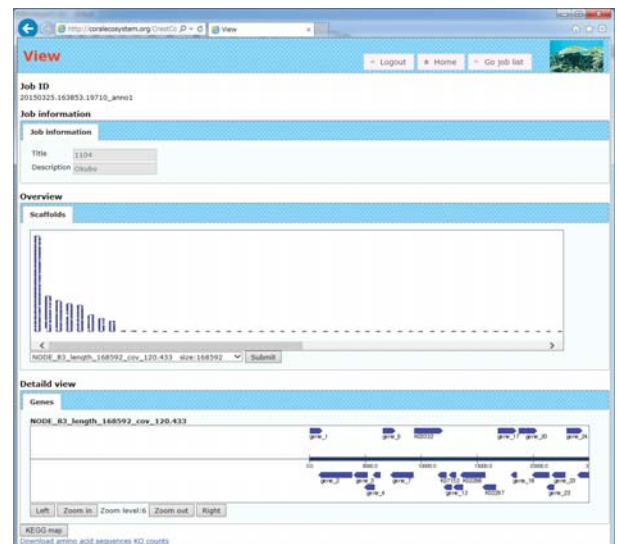


図 2 ゲノムビューページ

4. おわりに

本研究で開発した Single cell/meta genome analysis system により、Web ブラウザを用いて容易に難培養微生物のメタゲノムおよび 1 細胞配列データ解析を行うことができる。また、解析結果の可視化をブラウザで簡単に閲覧できるほか、KEGG 代謝マップ等へのリンク機能を可能にしている。今後、Single cell/meta genome analysis system で解析され、公開、更新される海洋環境データが有益な情報を提供することが期待される。

謝辞 本研究は一部、CREST（海洋生物多様性および生態系の保全・再生に資する基盤技術の創出）からの助成金を受けている。

参考文献

- [1] Bankevich A. *et al.*: SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing, *J. Comput. Biol.*, 19(5):455-477 (2012).
- [2] Kanehisa M., Goto S., Sato Y., Kawashima M., Furumichi M. and Tanabe M.: Data, information, knowledge and principle: back to metabolism in KEGG, *Nucleic Acids Res.*, 42(Database issue):D199-205 (2014).
- [3] Noguchi H., Taniguchi T. and Itoh T.: MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes, *DNA Res.*, 15(6):387-396 (2008).
- [4] Nurk S. *et al.*: Assembling single-cell genomes and mini-metagenomes from chimeric MDA products, *J. Comput. Biol.*, 20(10):714-737 (2013).
- [5] Peng Y., Leung H.C., Yiu S.M. and Chin F.Y.: IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth, *Bioinformatics*, 28(11):1420-1428 (2012).