

猫も杓子も

## Geneious Prime でシーケンス解析

### 第 24回 De novo アセンブリ



#### (その 4: ノーマライズデータのアセンブル)

NGS リードを処理し、de novo アセンブルする一般的な流れについてご紹介しています。

#### De novo アセンブリ

[その 1 : 前処理の概要](#)

[その 2 : 前処理の実例](#)

[その 3 : トリム済みペアリードのアセンブル](#)

その 4 回目となる今回は、ノーマライズしたデータセットのアセンブルについてです。

ご紹介している内容はチュートリアルとしてまとめられていますので、ご自身で試してみたい方は、[こちらからダウンロード](#)し、ZIP ファイルを解凍せずにそのまま Geneious Prime にドラッグ&ドロップしてインストールしてください。

チュートリアルで用意されているデータセットでは問題になりませんが、実際に解析に用いられる大規模なデータセットの de novo アセンブルでは、かなり多くの RAM と計算時間が必要となる場合があります。そのような場合に、アセンブル前にデータセットをノーマライズ(正規化)することにより、アセンブリの精度は落とさずに、必要な RAM と計算時間を大幅に削減することができます。

トリミングされたペアリード(前回からの続きの場合は **SRR513053 subset (trimmed)**)を選択し、**Sequence メニュー** → **Error Correct and Normalize reads** に進みます。今回はデフォルト設定でノーマライズのみを行い、**Error correction** は行いませんので、オプションのチェックを外します。**OK** をクリックして実行すると、**SRR513053 subset (trimmed) (normalized)** というノーマライズ済みリストが作成されます。

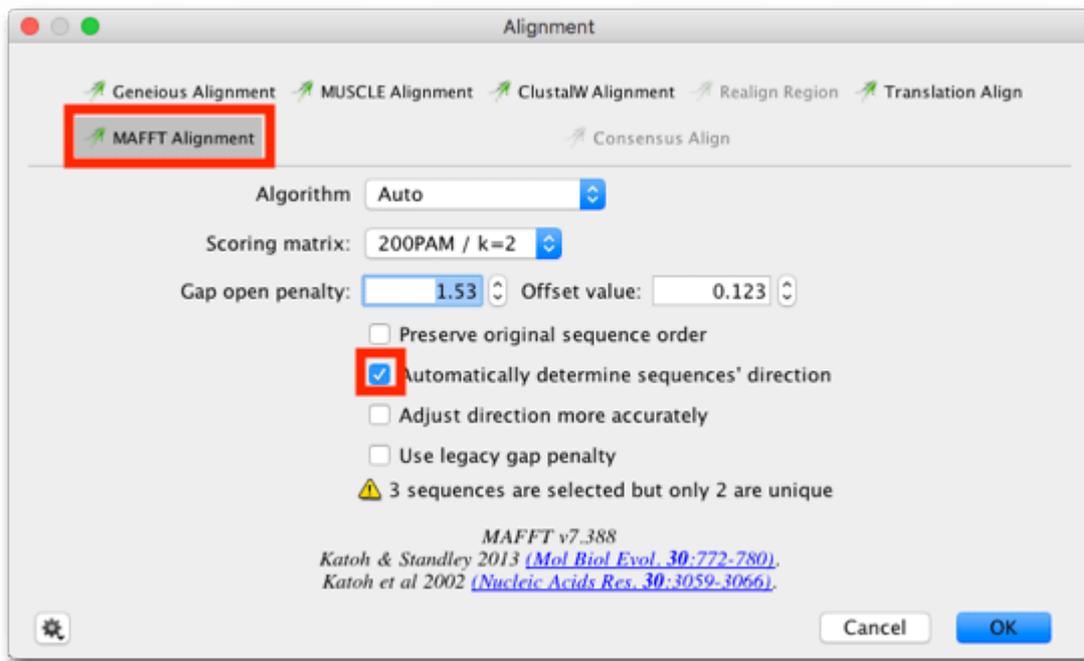
ノーマライズ済みリストをクリックして確認すると、デモデータでは 3,776 リードとなっていて、元のリストの 7,800 リードと比べて半分以下に減少していることがわかります。

さらに、ノーマライズ済みリストを用いて前回と同じ手順でアセンブリを行ってみると、元のデータセットを用いた時の約 1/3 の時間で完了することがわかります。作成されたコンティグを選択し、**Statistics** タブを表示すると、平均カバレッジが 52.3 であることが確認できます。

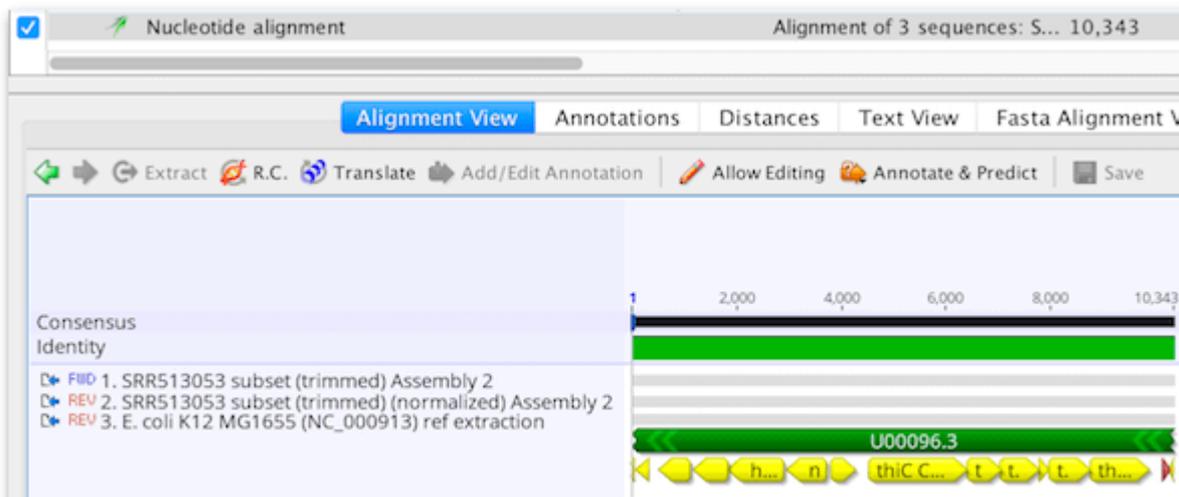
アセンブリの精度を確認するために、各アセンブリから生成された 2 つのコンセンサスファイル

(SRR513053 subset (trimmed) Assembly 2 と SRR513053 subset (trimmed) (normalized) Assembly 2)と、チュートリアルで提供されているリファレンス配列 E. coli K12 MG1655 (NC\_000913) ref extraction のアラインメントを行ってみます

ツールバーの Align/Assemble ボタンをクリックして **Multiple Align** を選択し、MAFFT アライメントツールでアライメントすることを選択して、**Automatically determine sequences' direction** にチェックを入れ、OK をクリックしてアラインメントを行います。



作成された **Nucleotide alignment** を選択して表示すると、アラインメントの上部にある Identity グラフが緑色のバーになっているのがわかると思います。これは 3 つの配列がアラインメント内のすべての位置で同一であることを示しています。



NGS リードの de novo アセンブリについては、次回はひとまずの最終回となりますが、最後に皆様のお役に立つ(かも知れない)Tips をいくつかご紹介する予定です。

Geneious 製品概要・フリートライアルリクエストについては[こちら](#)

『Geneious Prime でシーケンス解析』の過去の記事は[こちらでチェック!](#)

TDB News 10. 2023  
トミーデジタルバイオロジー株式会社  
Phone 03-6240-0843 Fax 03-6240-0461