猫も杓子も

Geneious Prime でシークエンス解析

第 23 回 De novo アセンブリ



(その 3: トリム済みペアリードのアセンブル)

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](#)

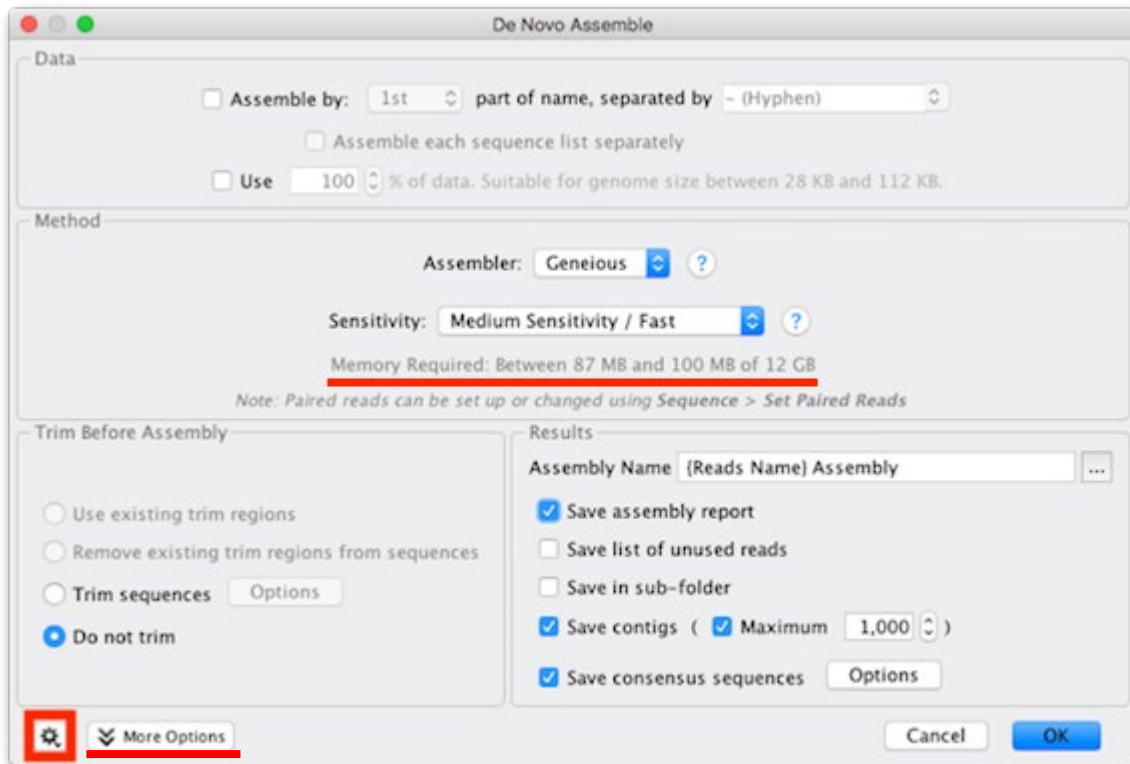
3 回目となる今回は、トリミングされたペアリードをアセンブルする実際の流れについてです。

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。

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、ツールバーの **Align/Assemble** ボタンをクリックして **De novo Assemble** を選択すると、De novo Assembler Settings ウィンドウが開きます。

この設定画面は **Data, Method, Trim, Results** のセクションに分かれており、**More Options** ボタンをクリックすると、**Advanced** 設定セクションが表示されます。各種設定の詳細については、[オンラインマニュアル](#)または[日本語マニュアル](#)をご参照ください。

標準で選択される Geneious de novo アセンブラは、インプットデータを解析し、適切な **Sensitivity** を自動的に判別しますので、ほとんどの場合は Sensitivity 設定を調整する必要はありません。また、アセンブリの実行に必要と予想される RAM 量を推定し、選択したデータセットをアセンブルするのに十分な RAM がないと判断した場合には警告が表示されます。



Sensitivity を変更すると、**Advanced** の様々な設定が調整されます。これらの詳細設定がどのように変化するかを確認したい場合は、設定ウィンドウの左下にある **More Options** ボタンをクリックしてから、**Sensitivity** の設定を変更することで確認することができます。また、各詳細設定にマウスオーバーすることで、その設定を説明するツールチップが表示されます。ご自身で特定の設定を変更したい場合は、**Sensitivity** を **Custom Sensitivity** に設定します。

今回は Geneious de novo アセンブリ アルゴリズムを使用して、デフォルトの設定のままアセンブリを行います。設定がデフォルトであることを確認するには、Window の左下隅にある歯車ボタンをクリックして、**Reset to defaults** を選択します。

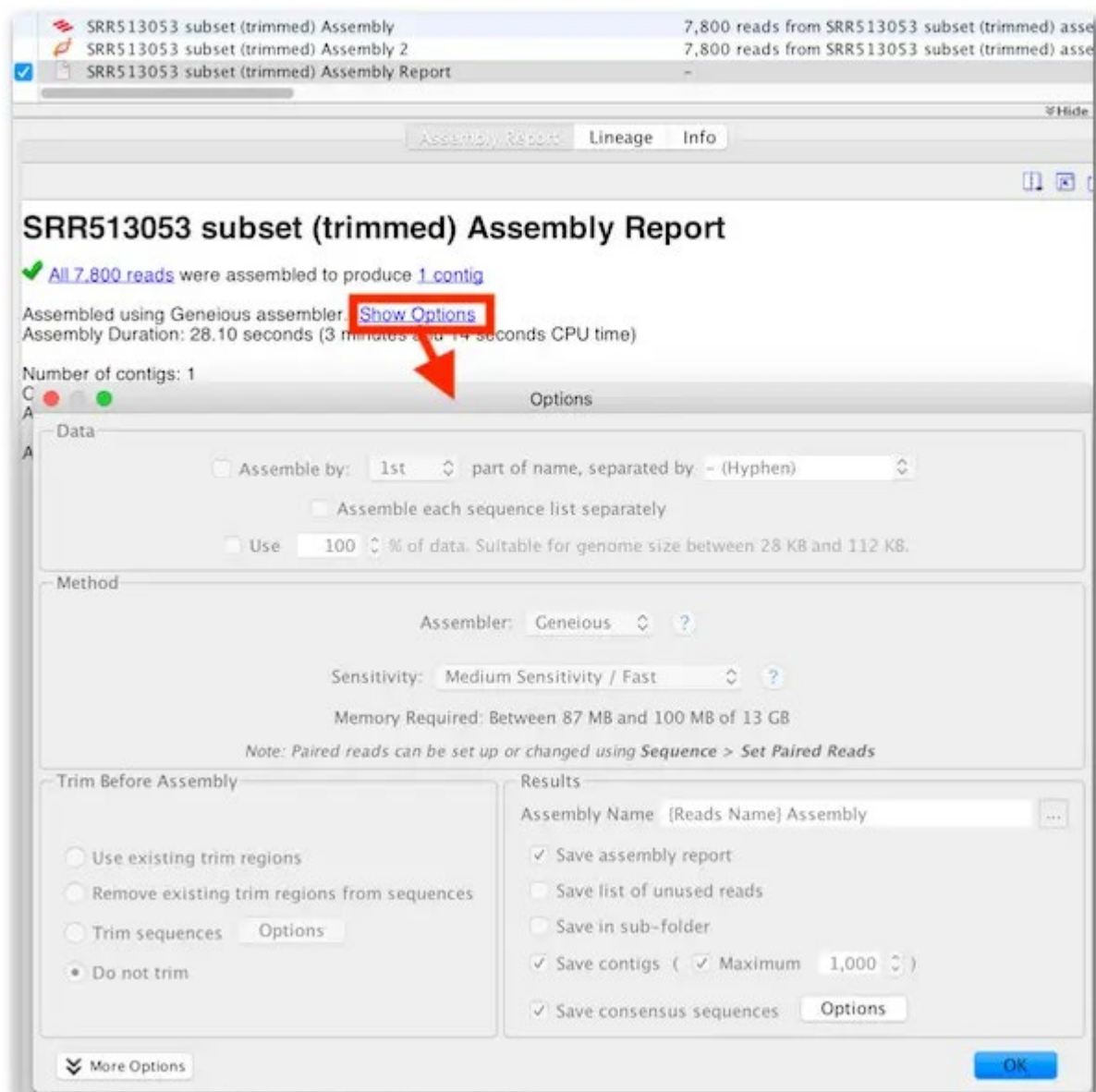
Trim セクションは、主にサンガーリードのためのものです。今回は前処理ステップですでにトリム済みの NGS データを使用しますので、ここでは **Do not trim** に設定されていることを確認します。

Results セクションでは、**Save an Assembly report**、**Save contigs**、**Save consensus sequences** の各オプションをチェックします。

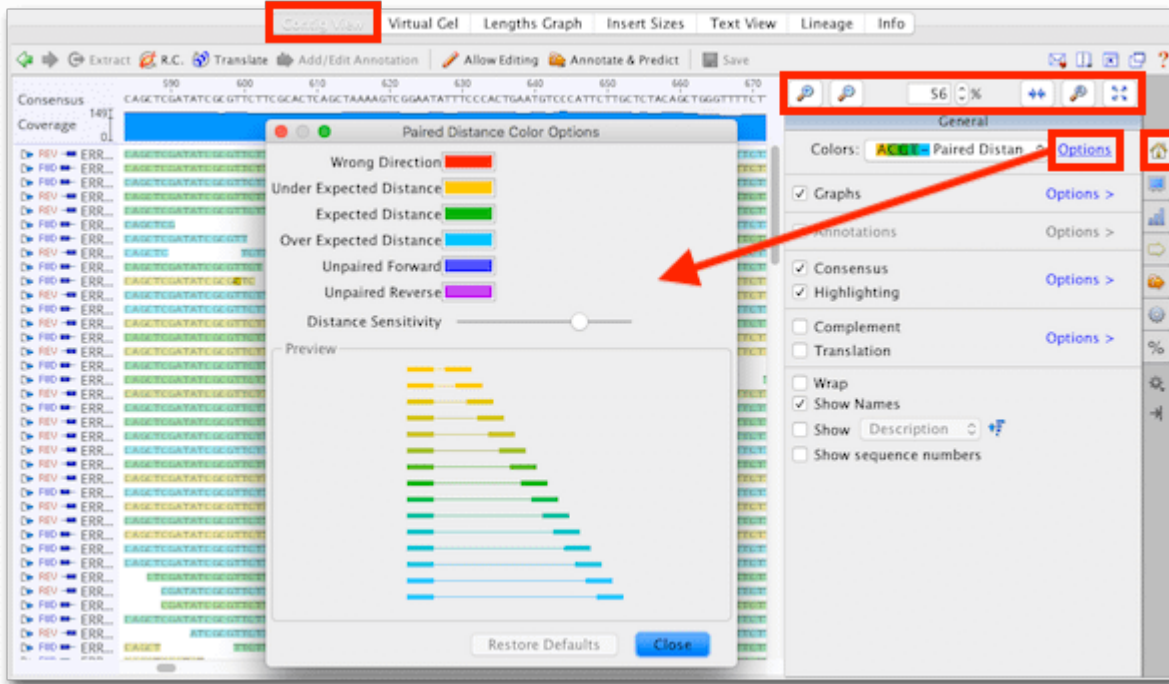
OK をクリックするとアセンブリが開始されます。アセンブリが完了すると、アセンブリレポート、アセンブリ(コンティグ)ファイル、アセンブリから生成されたコンセンサス配列の 3 つの新しいファイルが作成されます。

アセンブリレポートをクリックすると、レポートを確認することができます。チュートリアルデータの場合は、すべてのリードが 1 つのコンティグとしてアSEMBルされるため、レポートはシンプルなものになりますが、実データの場合など、複数のコンティグからなる、より複雑なアセンブリの場合は、[N50](#)を含む様々な統計情報がレポートされます。

また、アセンブリレポートにはアセンブリ時に使用された設定を表示する **Show options** リンクがあります。

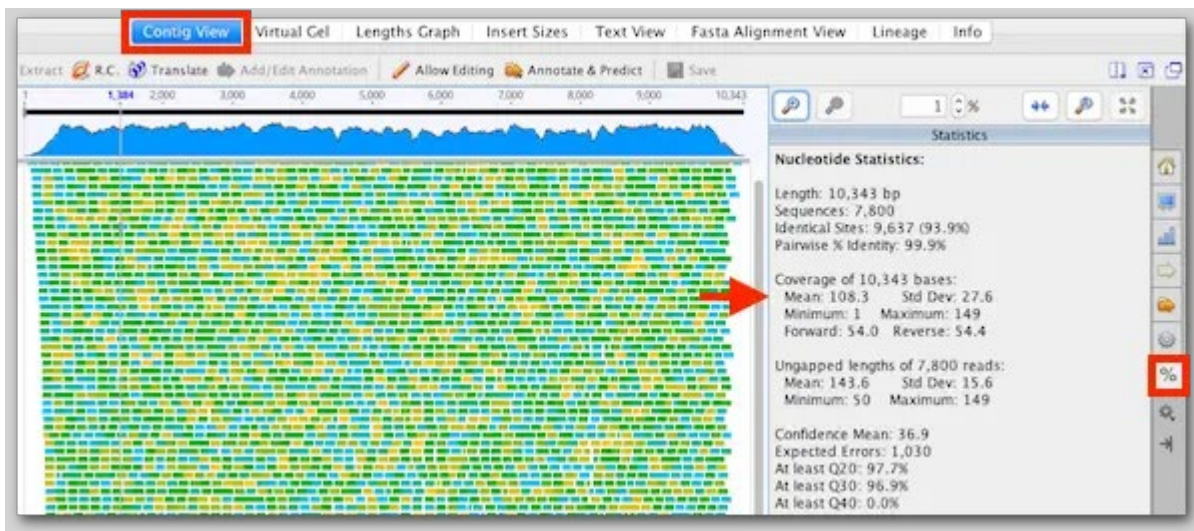


アセンブリ(コンティグ)ファイルをクリックするとアセンブリされたコンティグが表示されます。ペアリードデータをアSEMBルした場合、デフォルトの設定では、ペアリードの距離がペアリング時に設定した **Expected distance** とどれぐらい異なるかによって色付けされています。サイドパネルにある Home タブを選択し、**Options** リンクをクリックすると、カラースキームを表示することができます。

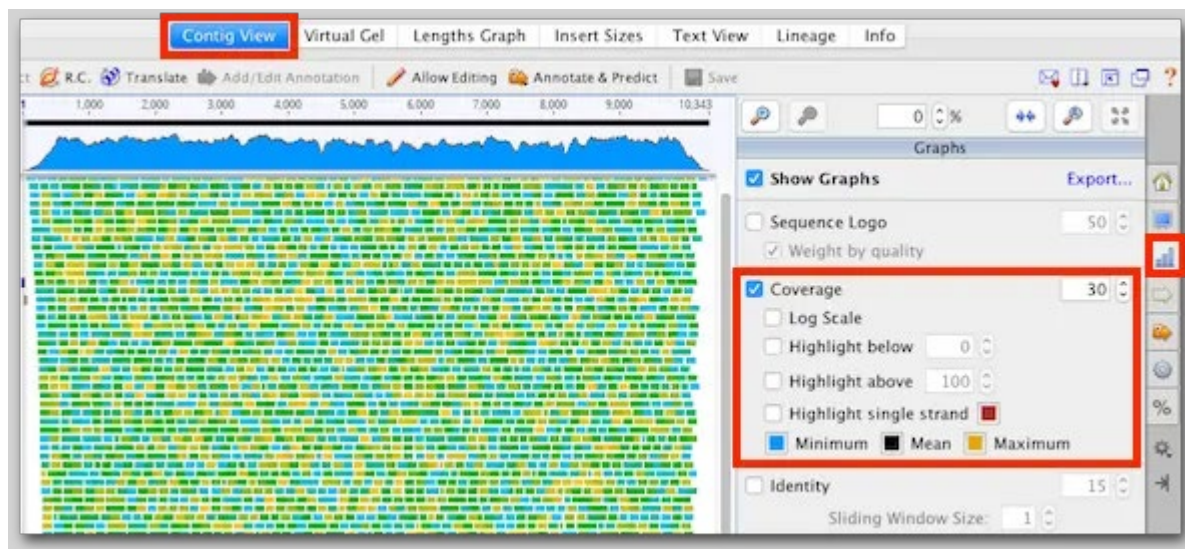


その上に表示されているズームコントロールを使ってズームインすることで、塩基レベルでの配列を確認することができます。

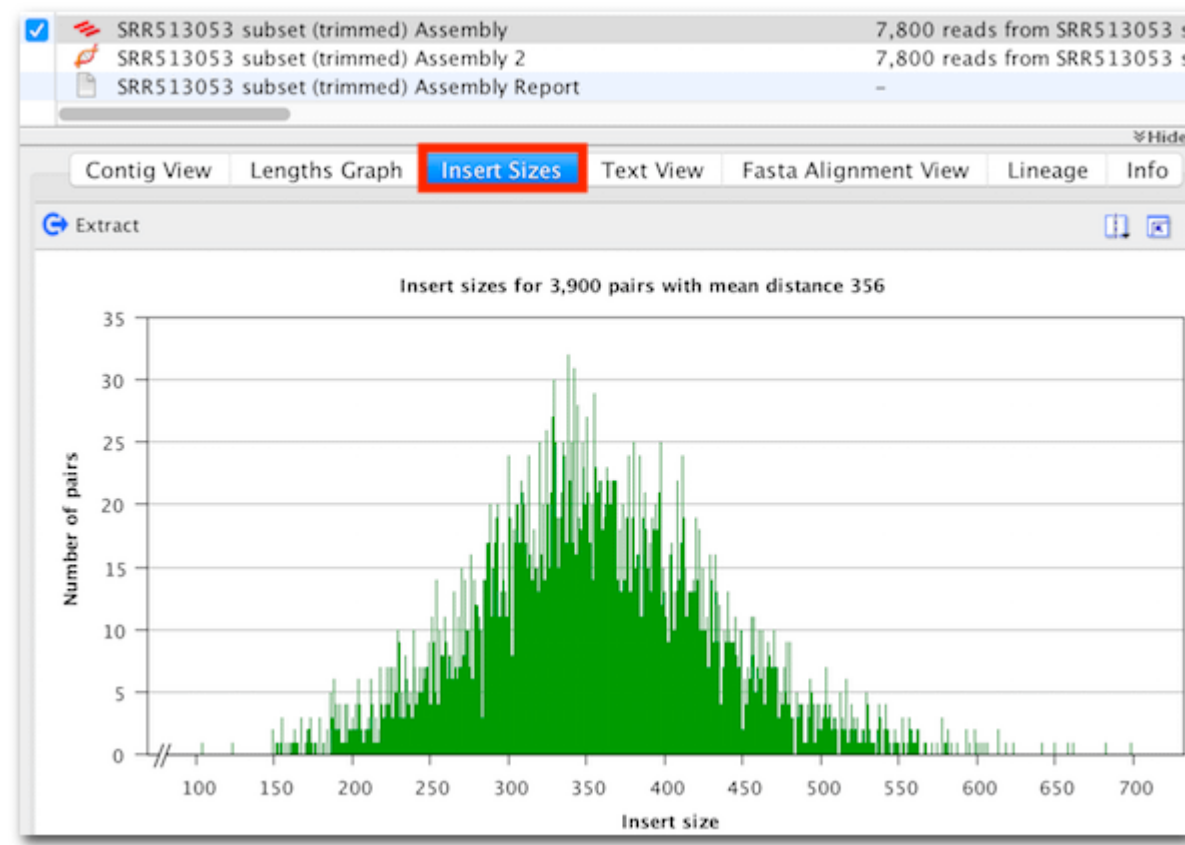
また **Statistics** タブ(%)をクリックすると、アセンブリの平均カバレッジなど、様々な統計情報を表示することができます。



サイドパネルにある **Graphs** タブをクリックすると、コンティグの上で青く表示されているカバレッジグラフの設定を変更することができます。カバレッジが高い/低い、または一本鎖である領域を特定するのに便利です。



ビューアパネルの上部にある **Insert Sizes** タブをクリックすると、ペアリードのアセンブリに基づいて計算されたインサートサイズの分布を示すヒストグラムが表示されます。チュートリアルデータの例では、平均ペア距離が想定インサートサイズの 350 bp に近いことがわかります。



次回はノーマライズ(正規化)したデータセットのアセンブルについてご紹介する予定です。アセンブル前にデータセットをノーマライズすることによって、アセンブリの精度は落さずに、必要な時間を大幅に短縮することができます。

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 9. 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