
 猫も杓子も

Geneious Prime でシーケンス解析

第 7 回 クロマトグラムのアSEMBL(その 2)



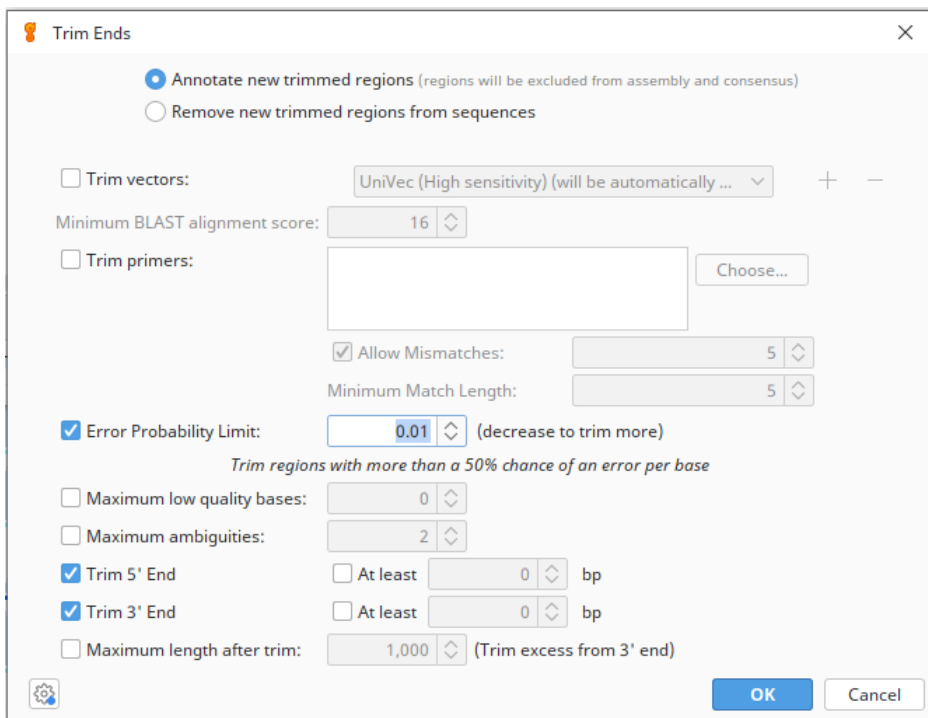
前回([第 6 回 クロマトグラムのアSEMBL\(その 1\)](#))のサンガーシーケンスのトリミング、アラインメントやアSEMBLなどの方法に引き続き、今回は双方向核遺伝子シーケンスデータの取り扱いについてご紹介します。

今回例として使用する *Acrocephalus* のシーケンスリストは、Geneious の Local フォルダ → Tutorials フォルダ → Assembling_Chromatograms フォルダに含まれています。リストには 3 種の *Acrocephalus reed warbler* (ヨシキリ)の核遺伝子の順方向と逆方向のシーケンスが含まれています。シーケンスには、種を示す 3 文字のコード(aru = *A. arundinaceus*, ニシオオヨシキリ; dum = *A. dumetorum*, シベリアヨシキリ; ort = *A. orientalis*, オオヨシキリ)があり、フォワードプライマーで配列決定したかリバースプライマーで配列決定したか示す「F」「R」が示されています。



Acrocephalus のシーケンスリストをダブルクリックすると、新規ウィンドウで開くことができます。下にスクロールすると、シーケンスの概要が表示されます。いくつかのシーケンス(dum2 や dum4 など)では、途中でシーケンスの品質が落ちていることに気付くと思います。

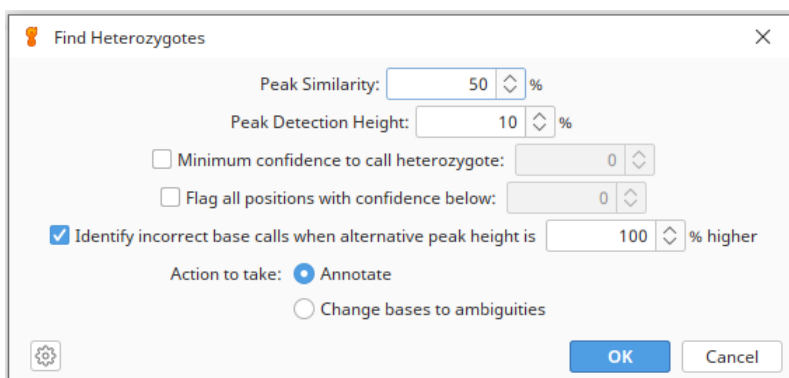
前回の手順と同じ用に **Annotate and Predict** メニュー → **Trim Ends** をクリックして、品質の悪いシーケンスを末端からトリムします。今回はトリムした領域を完全に削除するのではなく、アノテーションをつけるため、「**Annotate new trimmed regions**」を選択します。**Error probability limit** を 0.01 に設定し、**OK** をクリックします。トリミングが終了したら、シーケンスリストを保存し、ウィンドウを閉じます。



リードの方向を設定したり、**Heterozygote Finder** を使用したりするためには、リストから各シーケンスを個々に抽出する必要があります。Acrocephalus のシーケンスリストを選択し、**Sequence → Extract Sequences from List** をクリックします。Acrocephalus Sequences というサブフォルダにシーケンスが保存されるように選択します。

Heterozygotes プラグインがインストールされていない方は、**Tools メニュー → Plugins** で、利用可能なプラグインの一覧から選択して、**Install** をクリックするとインストールすることができます。

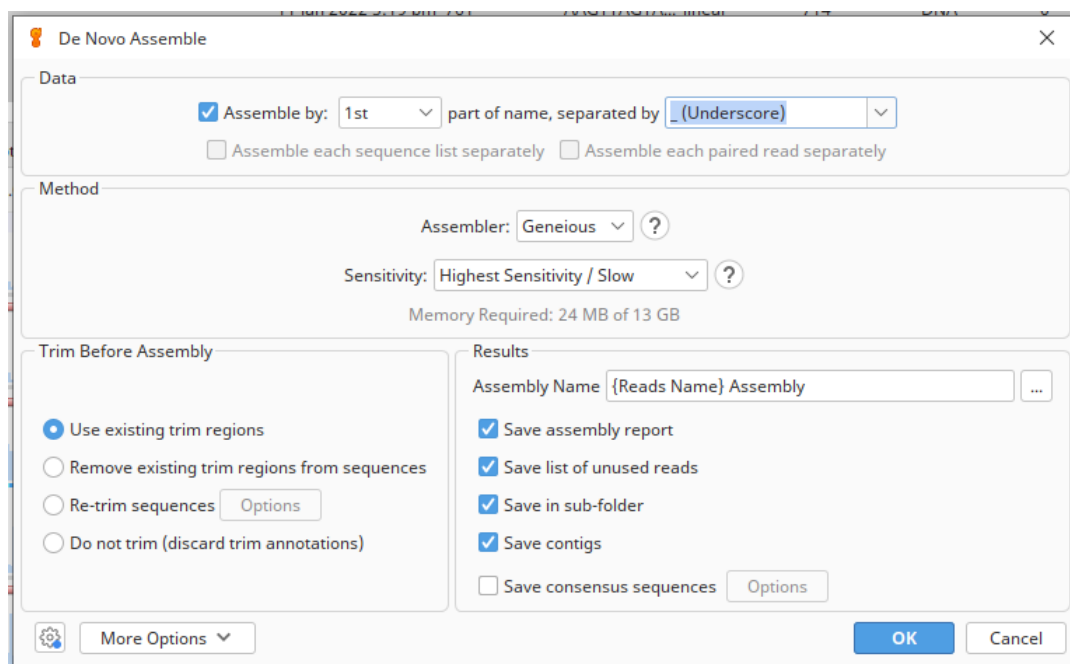
個々のシーケンスファイルに対して **Heterozygote Finder** を実行し、同じ位置で 2 つの異なるヌクレオチドがコールされている塩基を特定し、アノテーションを付けます。これらは核のシーケンスであるため、それぞれが 2 つの対立遺伝子を表し、2 つの対立遺伝子が異なる塩基を持つヘテロ接合位置、2 つのクロマトグラムのピークが存在する可能性があります。Acrocephalus Sequences フォルダ内のファイルをすべて選択し、**Annotate and Predict → Find Heterozygotes** をクリックします。**Peak Similarity** を 50% に設定し、**Action to take: Annotate** を選択します。



解析が終了したら、**OK** をクリックして、保存します。ヘテロ接合体としてアノテーションされた塩基については、次回以降、順方向と逆方向のシーケンスをアSEMBルした後に、再度ご紹介します。

各個体について、順方向と逆方向のシーケンスをアセンブルします。各ペアのシーケンスが同じ向きになるように、まずリードの方向を設定する必要があります。**command/ctrl** キーを押しながら、フォルダ内の全ての順方向シーケンス(最後の文字が F で始まる名前)を選択し、**Sequence** → **Set Read Direction** を選択します。**Forward** にチェックを入れ、**OK** をクリックします。リバースリードの方向は設定する必要はありません。

次に、フォルダ内の全ての配列を選択し、**Align/Assemble** → **De Novo Assemble** を選択します。**Assemble by** にチェックを入れ、「1st」 part of name, separated by 「_(Underscore)」と設定します。順方向と逆方向のシーケンスのペアごとに 1 つのコンティグが作成されます。**Sensitivity** を **Highest Sensitivity/Slow** に設定し、**Save assembly report**, **Save list of unused reads**, **Save in sub-folder**, **Save contigs** にチェックが入っていることを確認します。**Use existing trim regions** を選択します。このオプションを選択すると、アセンブラはトリム済みとしてアノテーションされた領域を無視しますが、シーケンス上ではこれらの領域を確認することができます。設定が完了したら **OK** をクリックします。



Assembly というサブフォルダが作成され、コンティグと Assembly Report が保存されています。また、unused reads というシーケンスリストが表示されますが、これにはアセンブルできなかったシーケンスが含まれています。このシーケンスリストを見てみると、途中から品質の悪かったシーケンス(dum2、dum4)が含まれていることが確認できます。

次回、リファレンスシーケンスへのアセンブル、アセンブリ/コンセンサスシーケンスの解析に続きます(予定)。