





第7回 クロマトグラムのアセンブル(その2)

前回(<u>第6回 クロマトグラムのアセンブル(その1)</u>)のサンガーシークエンスのトリミング、アラインメントやアセンブリなどの方法に引き続き、今回は双方向核遺伝子シークエ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 のシークエンスリストは、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クエンスが 含まれています。シークエンスには、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 のシークエ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、シークエンスの概要が表示されます。いくつかのシークエンス(dum2 や dum4 など)では、途中でシークエ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 悪いシークエ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、シークエン スリストを保存し、ウィンドウを閉じます。

🚦 Trim Ends			×	
 Annotate new trimmed regions (regions will be excluded from assembly and consensus) Remove new trimmed regions from sequences 				
Trim vectors: Minimum BLAST alignment score:	UniVec (High sensitivity) (will be automatically > + -			
Trim primers:		Choose.		
	Allow Mismatches:	5	\Diamond	
	Minimum Match Length:	5	\Diamond	
Error Probability Limit:	0.01 🗘 (decrease to trim more)			
Trim regions with more than a 50% chance of an error per base				
Maximum low quality bases:	0			
Maximum ambiguities:	2 🗘			
🗹 Trim 5' End	At least 0 🗘	bp		
🗹 Trim 3' End	At least 0 🗘	bp		
Maximum length after trim:	1,000 🗘 (Trim excess	s from 3' end)		
		ОК	Cancel	

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 クエンスを個々に抽出する必要があります。Acrocephalus のシークエンスリストを選択し、 Sequence \rightarrow 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クエ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。

個々のシークエ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。これらは核のシークエンス 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 \rightarrow 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。

Find Heterozygotes X		
Peak Similarity: 50 🗇 %		
Peak Detection Height: 10 🗇 %		
Minimum confidence to call heterozygote: 0		
Flag all positions with confidence below: 0		
☑ Identify incorrect base calls when alternative peak height is 100 ♦ % higher		
Action to take: 💿 Annotate		
Change bases to ambiguities		
OK Cancel		

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クエンスをアセンブルした後に、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クエンスをアセンブルします。各ペアのシークエンスが同じ向 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クエンス(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 \rightarrow 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クエンスのペアごとに 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、シークエンス上では 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。

Ce Novo Assemble	×		
- Data			
Assemble by: 1st v part of name, separated by (Underscore) v			
Assemble each sequence list separately Assemble each paired read separately			
Method			
Assembler: Geneious \vee ?			
Sensitivity: Highest Sensitivity / Slow 🗸 ?			
Memory Required: 24 MB of 13 GB			
Trim Before Assembly	Results		
	Assembly Name {Reads Name} Assembly		
• Use existing trim regions	✓ Save assembly report		
O Remove existing trim regions from sequences	✓ Save list of unused reads		
ORe-trim sequences Options	Save in sub-folder		
O Do not trim (discard trim annotations)	✓ Save contigs		
	Save consensus sequences Options		
More Options 🖌	OK Cancel		

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 というシークエ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 たシークエンスが含まれています。このシークエンスリストを見てみると、途中から品質の悪かったシ ークエンス(dum2、dum4)が含まれていることが確認できます。

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

Geneious 製品概要については<u>こちら</u>