



## Geneious Prime でシーケンス解析

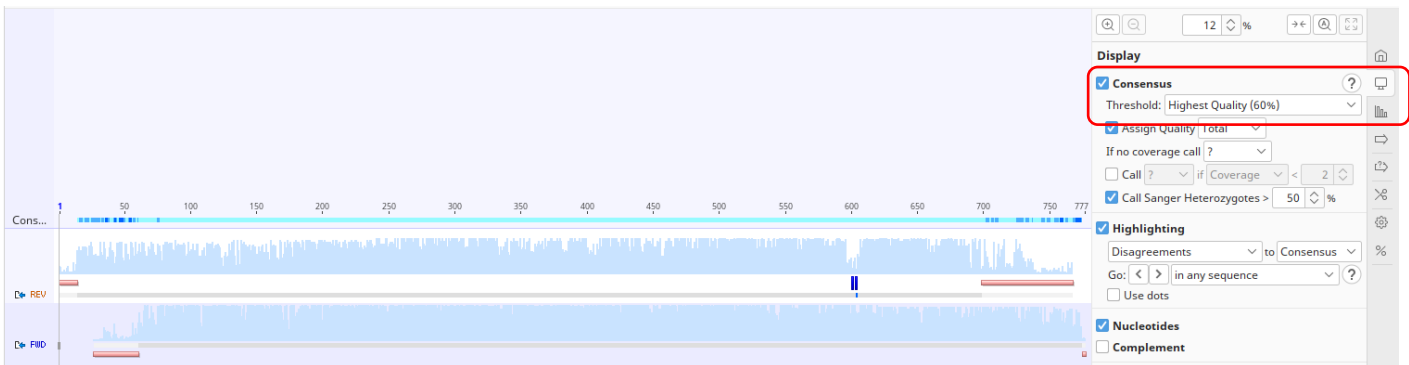
### 第 8 回 クロマトグラムのアセンブル(その 3)



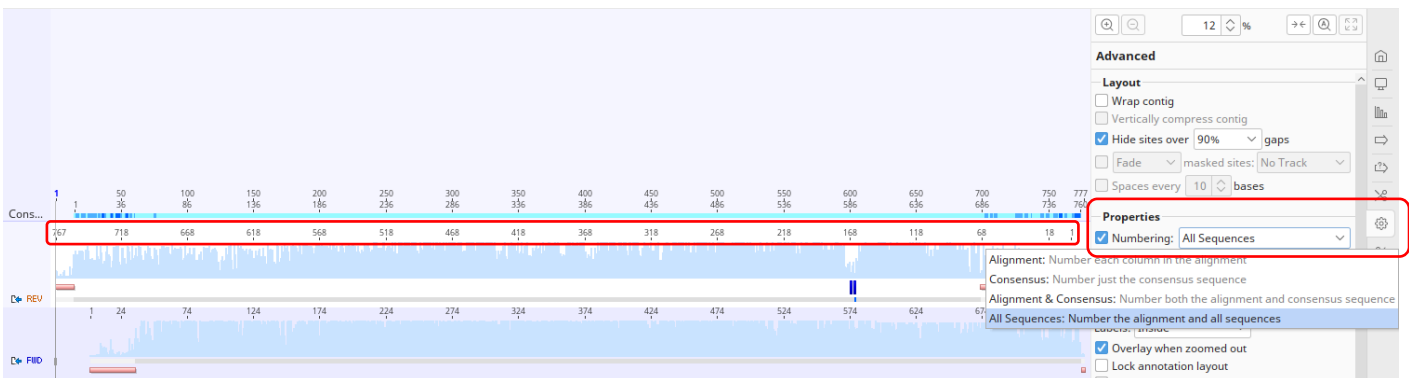
前回(第 7 回 [クロマトグラムのアセンブル\(その2\)](#))のサンガーシーケンスによる双方向核遺伝子シーケンスデータの取り扱い方法に引き続き、今回はアセンブリ/コンセンサスシーケンスの確認やリファレンスシーケンスへのアセンブルなどについてご紹介します。前回から例として使用している *Acrocephalus* シーケンスのアセンブルからの続きになりますので、もしまだご覧になっていない方は[こちら](#)を先にご参照ください。

前回作成した Assembly サブフォルダから aru2 コンティグをクリックして開き、順方向と逆方向のシーケンスがどのようにアセンブルされたかを確認します。

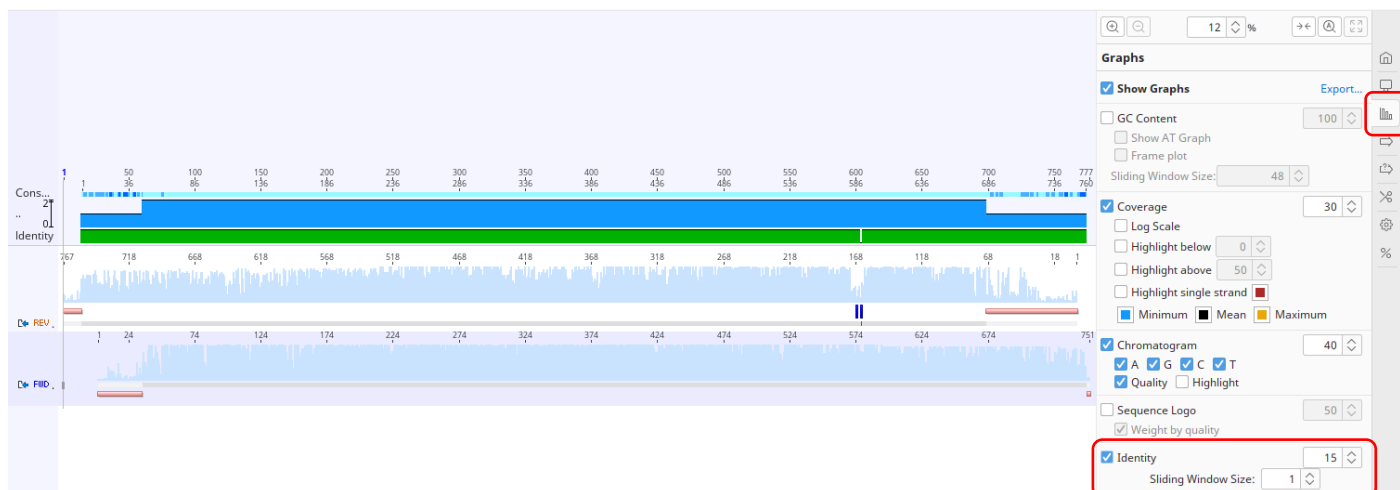
シーケンスビューアの右側にある **Display** タブ(モニタのマーク)で、コンセンサスシーケンスをコールするためのオプションをチェックします。同一遺伝子の順方向と逆方向のシーケンスをアセンブルした場合、通常は各塩基のクオリティが最も高いシーケンスからコンセンサスをコールするため、**Consensus** の項目で **Highest Quality (60%)** を選択します。



さらに **Advanced** タブ(ギアのマーク)で、**Numbering** を **All sequences** に設定します。これにより、各シーケンスに元のリードの塩基番号が表示され、2 つのシーケンスがどのようにアセンブルされたかを確認することができます。R のシーケンスが逆向きになっていることがわかります。



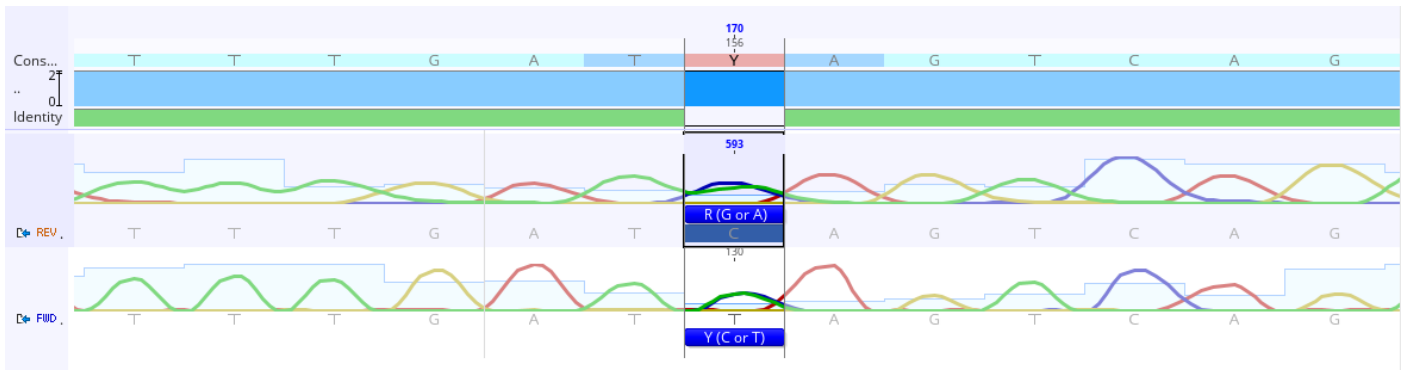
**Graphs** タブ(バーチャートのマーク)で、**Coverage** と **Identity** のボックスにチェックを入れます。**Coverage Graph** はコンセンサスシーケンスがいくつのシーケンスに基づいているかを示し、**Identity Graph** は寄与しているシーケンスが同一かどうかを示します。**Trimmed** とアノテーションされたクオリティの低い領域(ピンク色のバー)が見えますが、アセンブラはこの領域をコンセンサスのコールやカバレッジの算出には用いておらず、この領域ではクオリティの高いシーケンスだけが用いられていることがわかります。



また aru2 には順方向と逆方向のシーケンスの間で不一致な塩基が 1 つだけ存在します。ズームインしてこの塩基を探してみてください。ctrl/command+D のキーボードショートカットを使うと、不一致な塩基に素早くジャンプすることができます。この位置では逆方向のシーケンスの塩基が(A であるべきなのに C と)誤ってコールされています。

この位置の誤ったシーケンスのコールを編集して修正することもできますが、今回の手順ではクオリティが最も高いシーケンスからコンセンサスをコールしているため、コンセンサスシーケンスの塩基は正しいものになっています。下流の解析に使用するのがコンセンサスシーケンスである場合は、コンセンサスが正しければ、個々のリードの不一致をすべて編集する必要はありません。コンセンサスシーケンスを選択し、**Extract** をクリックして、抽出したコンセンサスシーケンスに名前を付け(例:aru2 consensus)、**OK** をクリックします。

次に ort1 アセンブリをクリックして開いてください。このシーケンスにはいくつかのヘテロ接合塩基のアノテーションがあるため、それらが正しくコールされているかどうかを確認する必要があります。ctrl/command+D でヘテロ接合塩基にジャンプします。この塩基(コンセンサスシーケンスの 170 塩基目)には、順方向と逆方向のリードで C と T のピークが重複したダブルピークがあり、これが本当のヘテロ接合塩基であることを示しています。そのため、コンセンサスシーケンスでコールされる塩基は、この位置では C と T の両方のヌクレオチドを含むことを示す「Y」となります([IUPAC による表記法](#)をご参照ください)。



他のコンティグについてもフォワードリードとリバースリード、ヘテロ接合体の塩基の間に不一致がないかをチェックし、必要に応じて [IUPAC の曖昧塩基表記](#) を参照してコンセンサスシーケンスを編集します。シーケンスの編集前には **Allow Editing** をクリックする必要があります。変更を保存し、元のシーケンスに変更を適用するかどうかを確認されたら **Yes** を選択し、コンセンサスシーケンスを選択して Extract します。

また、前回アセンブルがうまくいかなかった 2 つの *A. dumetorum* のシーケンス (dum2 と dum4、シーケンスのクオリティが低いためトリムされていた) のようなシーケンスをアセンブルするため、リファレンスを用いて部分配列をアセンブルする方法をとることができます。Assembly サブフォルダから Unused Reads シーケンスリストをクリックし、ctrl/command キーを押しながら、リファレンスとして使用する dum3 コンセンサスシーケンスをクリックします。**Align/Assemble** → **Map to Reference** をクリックします。dum3 コンセンサスシーケンスがリファレンスとして設定されていることを確認し、**Assemble by** にチェックを入れ、「1st」 part of name, separated by 「\_(Underscore)」と設定します。その他のオプションは、以下のスクリーンショットのように設定します。

Map to Reference
✕

**Data**

Reference Sequence: dum3 consensus - 18 documents Assembly Choose... ?

18 documents will be mapped to dum3 consensus

Assemble by: 1st part of name, separated by \_(Underscore)

Assemble each sequence list separately

**Method**

Mapper: Geneious ?

Sensitivity: Highest Sensitivity / Slow ?

Find structural variants, short insertions, and deletions of any size ?

Find short insertions and large deletions up to 1,000 bp

Fine Tuning: Iterate up to 5 times ?

Memory Required: 24 MB of 13 GB

**Trim Before Mapping**

Use existing trim regions

Remove existing trim regions from sequences

Re-trim sequences Options

Do not trim (discard trim annotations)

**Results**

Assembly Name {Reads Name} assembled to {Reference Name} ...

Save assembly report

Save list of unused reads

Save list of used reads  Include mates

Save in sub-folder

Save contigs

Save consensus sequences Options

More Options
OK
Cancel

これで dum2 と dum4 についてもコンティグ・アセンブリが作成できたと思います。dum2 アセンブリを開いて見ると、なぜ de novo でアセンブルできなかったのかがわかります。これは F と R のシーケンスの間で重なる高クオリティなシーケンスがない領域が 4 bp あるため、両方のシーケンスでトリムされたダブルピークの領域がここから始まっていますが、これは 2 つの対立遺伝子のうち 1 つが欠失したインデルを表していると考えられます。



そこでコンセンサスシーケンスの 4 bp のギャップ部分を選択し、**Add Annotation** をクリックし

て、インデルを強調するアノテーションを追加します。アノテーションのタイプを「Polymorphism」、名前を「Indel」に設定します。OK をクリックすると、コンセンサスシーケンスにこのアノテーションが追加されます。Save をクリックし、dum2 コンセンサスシーケンスを新しいファイルとして保存します。dum4 についても同様の手順で処理することができます。

これで前回と合わせて 9 つのサンプルすべてのコンセンサスシーケンスができあがりました。これらのシーケンスは、集団遺伝学的または系統学的な解析に使用できるように、アラインメントすることができます。全てのコンセンサスシーケンスを選択し、Align/Assemble → Multiple Align をクリックします。Geneious Aligner をデフォルトの設定で使用します。

Alignment View **Distances** Text View My god, it's full of stars Geneious BaseMan Lineage Info

Export Matrix

Matrix: % Identity Decimal Places: 2 Style: Heatmap and Numbers

	aru2 conse...	aru3 conse...	aru4 conse...	dum2 cons...	dum3 cons...	dum4 cons...	ort1 conse...	ort2 conse...	ort3 conse...
aru2 consensus		100%	100%	96.76%	96.84%	96.58%	99.01%	99.21%	99.01%
aru3 consensus	100%		100%	96.76%	96.85%	96.58%	99.02%	99.21%	99.01%
aru4 consensus	100%	100%		96.76%	96.85%	96.58%	99.02%	99.21%	99.01%
dum2 consensus	96.76%	96.76%	96.76%		100%	100%	96.68%	96.92%	96.84%
dum3 consensus	96.84%	96.85%	96.85%	100%		99.92%	96.92%	97.10%	96.91%
dum4 consensus	96.58%	96.58%	96.58%	100%	99.92%		96.50%	96.75%	96.67%
ort1 consensus	99.01%	99.02%	99.02%	96.68%	96.92%	96.50%		99.80%	99.61%
ort2 consensus	99.21%	99.21%	99.21%	96.92%	97.10%	96.75%	99.80%		99.80%
ort3 consensus	99.01%	99.01%	99.01%	96.84%	96.91%	96.67%	99.61%	99.80%	

アラインメントを開き、Distances タブをクリックすると、種内/間の塩基多様性の概要が表示されます。予想されるように、シーケンスは種間よりも種内でより類似しています。実際、*A. arundinaceus* (aru)のシーケンスは同一です。ここから Geneious の Tree 構築ツールでシーケンスの系統解析を行うこともできます。また、より高度な集団遺伝学的解析のために、アライメントを Fasta または Nexus フォーマットでエクスポートし、DNAsp などのプログラムで解析することができます。

Geneious 製品概要については[こちら](#)