

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 <u>その1:前処理の概要</u> 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、<u>こちらからダウンロード</u>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、 <u>オンラインマニュアル</u>または<u>日本語マニュアル</u>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。

	De Novo Assemble				
Data					
Assemble by: 1st 0 part		t of name, separated by - (Hyphen)			
	Assemble each seq	uence list separately			
	Ura 100 0 % of data Su	itable for second size batteen 20 VD and 112 VD			
	O USE 100 O M OF Data. Su	nable for genome size between 28 KB and 112 KB.			
Method					
	Assembl	er: Geneious 🗧 🤉			
	Sensitivity: Mediu	im Sensitivity / Fast 📀 ?			
	Memory Required: E	Setween 87 MB and 100 MB of 12 GB			
	Note: Paired reads can be set up	o or changed using Sequence > Set Paired Reads			
Trim Before As	sembly	n Results			
		Assembly Name (Reads Name) Assembly			
🔘 Use existin	ng trim regions	Save assembly report			
Remove ex	isting trim regions from sequences	Save list of unused reads			
O Trim seque	ences Options	Save in sub-folder			
		Save contigs ( 🗹 Maximum 1,000 C)			
O not thi					
		Save consensus sequences Options			
X More	Options	Cancel			

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センブルされるため、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、<u>N50</u>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。

<ul> <li>SRR513053 subset (trimmed) Assembly 2</li> </ul>	7,800 reads from SRR51305: 7,800 reads from SRR51305:	3 subset (trimmed) a 3 subset (trimmed) a
SRR513053 subset (trimmed) Assembly Report	-	
		ŝH
L Asst. 113	Andread Lineage Info	
		01 0
R513053 subset (trimmed) A	ssembly Report	
17,800 reads were assembled to produce 1 contin	, ,	
mbled using Genelous assembler. Show Options mbly Duration: 28.10 seconds (3 minutes 3 to 14 s	conds CPU time)	
ber of contigs: 1		
	Options	
Jata		
Assemble by: 1st 0 p	art of name, separated by - (Hyphen)	
Assemble each se	quence list separately	
Use 100 C % of data 5	ultable for genome size between 28 KB and 112 K8.	
Aethod		
Assemt	der: Geneious C ?	
Sensitivity: Medi	um Sensitivity / Fast 0 ?	
Memory Required	Between 87 MB and 100 MB of 13 GB	
Note: Paired reads can be set i	ip or changed using Sequence > Set Paired Reads	
rim Before Assembly	Results	
	Assembly Name (Reads Name) Assembly	
O Use existing trim regions	Save assembly report	
<ul> <li>Use existing trim regions</li> <li>Remove existing trim regions from sequences</li> </ul>	Save assembly report	
Use existing trim regions Remove existing trim regions from sequences Trim sequences Options	<ul> <li>Save assembly report</li> <li>Save list of unused reads</li> <li>Save in sub-folder</li> </ul>	
Use existing trim regions Remove existing trim regions from sequences Trim sequences Options Do not trim	<ul> <li>Save assembly report</li> <li>Save list of unused reads</li> <li>Save in sub-folder</li> <li>Save contigs ( Maximum 1,000 C)</li> </ul>	)

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 ードデータをアセンブルした場合、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。

Image: Distance International Internation Internation Internation Internation Internation Internation       Image: Distance International Internation Internation Internation Internation Internation         Image: Distance International Internation Internation Internation Internation Internation       Image: Distance International Internation Internation       Image: Distance International Internation         Image: Distance International Internation Internation       Image: Distance International Internation Internation Internation Internation Internation Internation       Image: Distance International Internation		Conception Virtual Gel Lengths Graph Insert Sizes	Text View Lineage Info			
Consensus Consensus	🗇 🕸 🕞 Extract 💋 R.C. 😵 Translat	e 🏟 Add/Edit Annotation 🛛 🥜 Allow Editing 🚔 Annotate & Predict	Save .	8080	?	
Coverage of Paired Distance Color Options Coverage of Paired Distance Color Options Paired Distance Color Options Coverage of Paired Distance Color Options Paired Distance Color Options Coverage of Paired Distance Color Options Paired Distance Color Options Coverage of Paired Distance Color Options Coverage of Paired Distance Color Options Paired Distance Color Options Coverage of Paired Distance Color Options Coverage of Paired Distance Color Options Paired Distance Color Option Paired Distance Color Option Paired Distance Color Option Pai	Consensus CAGETCGATATCGEGTTCT	610 620 620 640 640 640 640 640 640 640 640 640 64	о 670 ктовоттттст Р Р 56 С%	++ P 22		
0       Indectant Anticipation Control	Coverage	Paired Distance Color Options	General	General		
0 = 000 - 000       000 - 000<		Wrong Direction	Colors: ACEL Paired Di	stan Options	ŵ	
0       0	De FIO - ERR., EAGETEBATATEREBITET	Under Expected Distance	Them Graphs	Options >		
0 = 00 + 000       Execution       Constant constant         0 = 00 + 000       Execution       Unpaired Forward       Unpaired Forward         0 = 00 + 000       EXECUTION       Whighlighting       Whighlighting         0 = 00 + 000       EXECUTION       Whighlighting       Complement         0 = 00 + 000       EXECUTION       Whighlighting       Wrap         0 = 00 + 000       EXECUTION       Wrap       Wrap       Wrap         0 = 00 + 000       EXECUTION       Wrap       Wrap       Show Names       Show Names         0 = 00 + 000       EXECUTION       Wrap       Wrap       Show Sequence numbers       Show Sequence numbers         0 = 00 + 000       EXECUTION       Wrap       Wrap       Show Sequence numbers       Show Sequence numbers         0 = 00 + 000       EXECUTION       Wrap       Wrap       Show Sequence numbers       Show Sequence numbers         0 = 00 + 000       EXECUTION       Wrap       Wrap	De FRU ERR., EARCTER De FRU ERR., EARCTER De FRU ERR EARCTER	Expected Distance	Terr Annotations	Options >	d	
0       PR0       ERA       CANCECCUARATECCUCENTER       Unpaired Reverse       Impaired Reverse       Im	De REV - ERR., CASCING INCID	Unpaired Forward	Consensus	Options >		
0 H00 - ERR.       EXERCISENTATEGRATISTIC       Preview       Translation       Options >       %         0 H00 - ERR.       EXERCISENTATEGRATISTIC       Wrap       Translation       %         0 H00 - ERR.       EXERCISENTATEGRATISTIC       %       %       %       %         0 H00 - ERR.       EXERCISENTATEGRATISTIC       %       %       %       %       %       %         0 H00 - ERR.       EXERCISENTATEGRATISTIC       %       %       %       %       %       %       %         0 H00 - ERR.       EXERCISENTATEGRATISTIC       %	De FEV ERR CARCTOGATATOGOUTTOT De FEO ERR CARCTOGATATOGOUTTOT De FEV ERR CARCTOGATATOGOUTTOT	Unpaired Reverse	Highlighting		0	
0       FIGURE ERRAL       Biological constraints       Image: State of the s	D FRO ERR., EASCTOGATATUSCOTTOT	Preview	Translation	Options >	%	
0       FRV = ERR.       EXEMPTEDESCRIPTION       1 <td< td=""><td>D FRO ERR. DESCRIPTION</td><td></td><td>I Wrap ✓ Show Names</td><td></td><td>0,</td></td<>	D FRO ERR. DESCRIPTION		I Wrap ✓ Show Names		0,	
PRV     ERK     Educational constraints       C FRV     ERK     Educational constraints       D FRV </td <td>D. FEV - ERR., EASETOMATATESEGTICT</td> <td></td> <td>Show Description 0</td> <td>r  </td> <td>-11</td>	D. FEV - ERR., EASETOMATATESEGTICT		Show Description 0	r	-11	
0     PR0     ERR.     EXACCESSANATEGROCETED     PR0       0     PR0     ERR.     EXACESSANATEGROCETED     PR0       0     FR0     ERR.	D REV ERR. EAGETCEATATOCOTTOT		Show sequence numbers			
D FRO E ERR. EXAMPLEMENTATION CONTENT PRO E ERR. EXAMPLEMENTATION CONTENT PRO E ERR. EXAMPLEMENTATION CONTENT PRO E ERR. EXAMPLEMENTATION CONTENT D FRO E ERR. EXAMPLEMENTED PRO E ERR. EXAMPLEMENT	D FID ERR., EASTEGATATOSCUTTO D FIV ERR., CASETOSATATOSCUTTO D FIV ERR., EASTEGATATOSCUTTO					
ReV = ERR.     Execution structure	De FID - ERR., EAGETCGATATORGITET De FID - ERR., EAGETCGATATORGITET		tron mun			
De FRO - ERR., EXEMPTION CONTRACTOR CON	D REV ERR., ETCLATATION OTTO		COLUMN COLUM			
De FRO - ERR. DANNE MINOR Restore Defaults Close men	De FRO ERR., CONTATUCCOTTOT De FRO ERR., ENGETCONTATUCCOTTOT De SAU ERR.					
	De FRO ERR., EARLET ENDER	Restore Defaults Close	·			

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。



ビューアー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 でシークエンス解析』の過去の記事は<u>こちらでチェック!</u>

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