

1

ロングリードデータを使用した菌叢解析

フィルジェン株式会社 バイオインフォマティクス部 (biosupport@filgen.jp)

アンプリコン解析





菌叢解析に使用される16S遺伝子の全長が約1,500bpなのに対し、 一般的なNGSのリード長は数百bpであるため、一部の領域のみを利用している。

 \downarrow

情報量が限られるため、種レベルの分類には不十分な場合がある







一方でロングリードシーケンサーでは、数kbのリードも得られるため、全長を使用可能できる。

しかしながら、ロングリードではエラー率が比較的高いという別の問題が存在する。

 \downarrow

既存のアルゴリズムをそのままロングリードデータに適用することができない。





Curry, K.D., Wang, Q., Nute, M.G. *et al.* **Emu: species-level microbial community profiling of full-length 16S rRNA Oxford Nanopore sequencing data.** *Nat Methods* **19**, 845–853 (2022). https://doi.org/10.1038/s41592-022-01520-4

"Emu"というロングリード用の分類群識別アルゴリズムが2022年に公開された。

CLC Genomic WorkbenchのプラグインCLC Microbial Genomics Moduleにも、 Emuに基づくロングリード用の分類群識別ツールが搭載された(ver.24以降)。

解析の準備



File Edit View Connections Utilities Download Toolbox Help **路路 峰 🖉 つ** 〇 🗵 \rightarrow i ć +Show New Save Import Export Graphics Print Launch Undo Redo Delete Navigation Area 🖄 Standard Import... Ctrl+I ▶協告♡ Tracks... Q[™] <enter search **6** Illumina... 🕀 🕞 CLC_Data 🗄 🔚 CLC_Data2 🚔 PacBio > E CLC_Reference General Other NGS Reads **6** 🗄 🕞 CLC_Training Element Bio... **6 6** Fasta Read Files... Ion Torrent... 62 **6** MGI/BGI... Sanger... 🚰 Singular... e SAM/BAM/CRAM Mapping Files... 💕 Ultima... Import Metadata... RNA Spike-ins... Primers > ₽. Import Gene-pseudogene Table... ŏ₽ Import Immune Reference Segments... E² Import Known Fusion Information Track... 💰 Single Cell Data > Toolbox

デフォルトの状態では、ロングリード(Nanopore)のインポーターが利用できないので、 プラグインLong Read Support(無償)をインストールする必要がある。

解析の準備



File Edit View O	onnections Utilities Download Toolbox Help
$\vdash \vdash \vdash$	$\stackrel{{}_{\scriptstyle \bullet}}{\boxminus} \stackrel{{}_{\scriptstyle \bullet}}{\boxminus} \stackrel{{}_{\scriptstyle \bullet}}{\boxminus} \stackrel{{}_{\scriptstyle \bullet}}{\boxminus} \stackrel{{}_{\scriptstyle \bullet}}{=} \stackrel{{}_{\scriptstyle \bullet}}{\swarrow} \stackrel{{}_{\scriptstyle \bullet}}{\circ} \stackrel{{}_{\scriptstyle \bullet}}{\boxtimes} \stackrel{{}_{\scriptstyle \bullet}}{\to} \stackrel{{}_{\scriptstyle \bullet}}{\circ} \stackrel{{}_{\scriptstyle \bullet}}{\boxtimes} \stackrel{{}_{\scriptstyle \bullet}}{\to} \stackrel{{}_{\scriptstyle \bullet}}$
Show New Save	mport Export Graphics Print Launch Undo Redo Delete
Navigation Area	Standard Import Ctrl+I
 ✓ Mar End Control ✓ Control	Tracks
	🚰 Illumina
	Gamma Oxford Nanopore
	PacBio >
	Import sequences from Oxford Nanopore files Other NGS Reads
	🚔 Fasta Read Files
	🚔 Sanger
	SAM/BAM/CRAM Mapping Files
	Import Metadata
	RNA Spike-ins
	💁 Primers >
	Import Gene-pseudogene Table
	Import Immune Reference Segments
Toolbox	Import Known Fusion Information Track
Processes Toolbox	Single Cell Data >
< enter tool names	29

プラグインをインストールすると、Nanoporeのインポーターが使用可能になる。

Long Read Supportについて





プラグインのインストール後は、ロングリード用のツールがいくつか使用可能になる。 ロングリードのマッピング、トランスクリプトーム解析、ハイブリッドアセンブリ等が可能。

解析の準備





事前にリファレンスのデータベースをダウンロードする必要がある。 (ショートリードの場合と同様)

解析の準備



	1	Nar
ERR3363792	2	ERF
ERR3363793	3	ERF
ERR3363794	4	ERF
ERR3363781	5	FRE
ERR3363782	6	FRF
ERR3363/83	7	FRF
ERR3363804	0	
ERR3363805	0	
:	9	
	10	E K I

- A	A	В
1	Name	Month
2	ERR336781	April
3	ERR336782	April
4	ERR336783	April
5	ERR336792	June
6	ERR336793	June
- 7	ERR336794	June
8	ERR336804	August
9	ERR336805	August
10	ERR336806	August
1.1		

インポートするリード名と、それぞれが属するグループの情報を含んだエクセルファイルを用意しておく。

今回使用するツール





今回主に使用するツールは、Microbial Genomics Moduleのフォルダに入っている "Classify Long Read Amplicons"ツールです。 出力フォーマットの共通性











ショートリード用のツールと同様のフォーマットで結果が出力される。

この後は、これまでと同様に図表の作成(サンバースト図・ヒートマップ)が可能

本日の内容





ロングリードのインポート



File Edit View (Connec	tions Utilities Download Toolbox Help			
$\exists \exists f \in f$	ß	$ \stackrel{{}_{\frown}}{\boxminus} \stackrel{{}_{\frown}}{\boxminus} \stackrel{{}_{\frown}}{\boxminus} \stackrel{{}_{\frown}}{=} \stackrel{{}_{\frown}}{\swarrow} \stackrel{{}_{\frown}}{\circ} \stackrel{{}_{\bullet}}{\circ} \stackrel{{}_{\frown}}{\circ} \stackrel{{}_{\bullet}}{\circ} $	X		
Show New Save	Import	Export Graphics Print Launch Undo Redo D	elete		
Navigation Area	ഷ	Standard Import Ctrl+			
◎ 箇 巴 い Qv Center search t	E\$	Tracks			
Enter Search of	i	Illumina			
CLC_Data2	1	Oxford Nanopore			
	1	PacBio	>		
	F	Other NGS Reads	>		
	F	Fasta Read Files			
	1	Sanger			
	Ê	SAM/BAM/CRAM Mapping Files			
		Import Metadata			
	1	RNA Spike-ins			
	3	Primers	>		
	₽.	Import Gene-pseudogene Table			
	¥\$	Import Immune Reference Segments			
Toolbox		Import Known Fusion Information Track			
Processes Toolbox	ൽ	Single Cell Data	>		
<enter name="" tool=""></enter>		£2			

ImportメニューのNanoporeのインポーターを利用してインポート

サンプルごとの組成解析





Classify Long Read Ampliconツールを起動し、インポートしたロングリードデータを選択 (各リードが別々のサンプルに由来する場合はbatchをチェック)

サンプルごとの組成解析



G. Classify Long Read Ampl	icons X	6. Select sequence list	×
1. Choose where to run	Settings	Navigation Area Reference Data	Selected elements (1)
2. Select sequencing reads		Qv <enter search="" term=""></enter>	Greengenes2 (2022.10)
3. Batch overview	Select reference	CLC_Data2 → → CLC_Data2	
4. Settings	Amplicon database := Greengenes2 (2022.10)	Arrow Metadata Arrow	
5. Result handling	Read alignment		
	Automatic Manual	rattus_norvegicus test	
	Match score 2	Break Barrier Bar	
	Mismatch cost 4	NCB Download C Results C Results	
	Gap open cost 4	Reseq_demo Reseq_demo Reseq_demo	
	Gap extend cost 2	Amplicon-Based Analysis Greengenes	<u> </u>
	Long gap extend cost 1	Greengenes 97% (v13_8)	
	Score bonus for global alignment 0	Read_data Data QC and OTU Clustering	
	Abundance estimation	E Caracteria Constraints and Assign Taxonomies E Caracteria Constraints and Assign Taxonomies E Caracteria Constraints	
	Minimum relative abundance 0.0001	Comparison of the second	
		B-C Resistance Analysis Resistance Analysis	
		GRNA-GEMO DeNovo_demo	
Help Reset	Previous Next Finish Cancel		OK Cancel

Amplicon databaseでリファレンスとして使用するデータベースを選択

サンプルごとの組成解析



· IIII ERR3363783 (abundance table)

. CLC Genomics Workbench 24.0.1 - <New Workspace> - Evaluation 123 days remaining **n** File Edit View Connections Utilities Download Toolbox Help $\Box \in \mathcal{C} \land \mathcal{C} \: \mathcal{C$ 0 \$ 6 6 & A ◀ I III ERR3363783 (abundance table) × Navigation Area ▶協告♡ A- A+ 7 Table Settings Rows: 97 Filter to Selection... Filter 👻 Q* <enter search term Column width - IDI SKA Metadata Table-9 Taxonomy ERR3363783 Abundance Automatic . Name Combined Abundance ERR3363805 Stramenopile Bacteria: Cyanobacteria: Chloroplast: Stramenopiles ERR3363805 (sampled Show column Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae SRA MetadataTable-10 Comamonadaceae DID . imnohabitans Bacteria: Proteobacteria: Betaproteobacteria: Burkholderiales: Comamonadaceae: Limnohabitan ERR3363806 🖂 Name Sphinoobacteriale Bacteria: Bacteroidetes: Sphingobacterija: Sphingobacteriales - ERR3363806 (sampled) Cytophagaceae Bacteria; Bacteroidetes; Cytophagia; Cytophagales; Cytophagaceae ERR3363782 (abundance table Z Taxonomy Arcobacter Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales; Campylobacteraceae; Arcobact Combined Abundance Bacteria; Verrucomicrobia; Opitutae; [Cerasicoccales]; [Cerasicoccaceae] [Cerasicoccaceae] ERR3363793 (sampled) (abundance table) Phodoferay Barteria: Proteobarteria: Betanroteobarteria: Burkholderiales: Comamonadareae: Dhodoferay 🗌 Min ERR3363794 (abundance table) hodobacter Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae; Rhodobacter 🗌 Max ERR3363805 (sampled) (abundance table) Salmonella Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella ERR3363806 (sampled) (abundance table) Mean Enterobacteriaceae Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae - IB ERR3363781 (abundance table) (merged abundance t Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales Median ERR3363781 (abundance table) (merge abundance tat ingomonadales lvdrogenophaga Bacteria: Proteobacteria: Betaproteobacteria: Burkholderiales: Comamonadaceae: Hydrogenophaga Std ERR3363781 (abundance table) (merged abundance ta Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Shigella; Sonnei 🖲 📋 Recycle bin ERR3363783 Abundance vobacterium Bacteria: Bacteroidetes: Flavobacterija: Flavobacteriales: Flavobacteriaceae: Flavobacterium CLC_Data2 Select All Bacteria; Firmicutes; Bacilli; Bacillales; Listeriaceae; Listeria; Seeligeri Seeligeri E CLC References Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudom domonas eae: Pseudomon CLC_TrainingDate Deselect All Curvus Bacteria: Proteobacteria: Betaproteobacteria: Burkholderiales: Comamonadaceae: Limnohabitans: Curvus Bacteria; Proteobacteria; Betaproteobacteria; Rhodocyclales; Rhodocyclaceae Data docvclaceae Aureus Bacteria: Firmicutes: Bacilli: Bacillales: Staphylococcaceae: Staphylococcus: Aureus how abundance values as Processes Toolbox Favorites Bacillus Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus O Raw aromonas Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae; Polaromonas enter tool name> Relative Nitrospira Bacteria: Nitrospirae: Nitrospira: Nitrospirales: Nitrospiraceae: Nitrospira 🗄 🚘 Microbial Workflow Bacteria; Proteobacteria; Betaproteobacteria; Rhodocyclales; Rhodocyclaceae; Dechloromonas Dechloromon Aggregate feature Single Cell Workflows Bacteria: Proteobacteria: Gammaproteobacteria: Enterobacteriales: Enterobacteriaceae: Escherichia: Coli Name Installed Workflows Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales; Helicobacteraceae; Sulfuricurvum; Kujiense Kujiense Installed Workflows Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales; Sphingomonadaceae; Novosphingobium Bacteria; Proteobacteria; Betaproteobacteria; Rhodocyclales; Rhodocyclaceae; Zoogloea Hide incomplete features lovosphingobium Zoogloea Appregate sample Microbial Genomics Module Sulfurospirillum Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales; Campylobacteraceae; Sulfurospirillum Metagenomics Wetagenomics Wetagenome SC-1-84 Bacteria: Proteobacteria: Betaproteobacteria: SC-I-84 Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales; Xanth - Amplicon-Based Analysis Lachnospiraceae Bacteria: Firmicutes: Clostridia: Clostridiales: Lachnospiraceae - 🚯 Normalize OTU Table by Copy Number (beta) Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales ickettsiales K Filter Samples Based on Number of Reads Armatimonadaceae Bacteria: Armatimonadetes: Armatimonadia: Armatimonadales: Armatimonadaceae OTU Clustering Sphingomonadacea Bacteria: Proteobacteria: Alphaproteobacteria: Sphingomonadales: Sphingomonadaceae Remove OTUs with Low Abundance Oxalobacteraceae Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Oxalobacteraceae Align OTUs using MUSCLE Enterica Bacteria: Proteobacteria: Gammaproteobacteria: Enterobacteriales: Enterobacteriaceae: Salmonella: Enterica Cryaerophilus Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales; Campylobacteraceae; Arcobacter; Cryaerophilus Detect Amplicon Sequence Variants Polynucleobacter Bacteria: Proteobacteria: Betaproteobacteria: Burkholderiales: Oxalobacteraceae: Polynucleobacter Taxonomic Analysis Bacillaceae Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae E 🔞 Abundance Analysis Preudomonadacea Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadacea Typing and Epidemiology Functional Analysis Methylophilaceae Bacteria; Proteobacteria; Betaproteobacteria; Methylophilales; Methylophilaceae Alcaligenaceae Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Alcaligenaceae 🖶 📆 Drug Resistance Analysis RS62 Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae; RS62 B-RA Databases Paracoccus Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae; Paracoccus 🗄 🔚 Panel Suppor Enterococcus Bacteria: Firmicutes: Bacilli: Lactobacillales: Enterococcaceae: Enterococcus In Long Read Support actobacillus Bacteria; Firmicutes; Bacilli; Lactobacillales; Lactobacillaceae; Lactobacillus Genome Finishing Module luviicola Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Cryomorphaceae; Fluviicola - 🗟 Classical Sequence Analysis Dok59 Bacteria; Proteobacteria; Betaproteobacteria; Rhodocyclales; Rhodocyclaceae; Dok59 🗑 🙀 Whole Genome Alignmen Bacteria; Bacteroidetes; [Saprospirae]; [Saprospirales]; Saprospiraceae; Haliscomenobacter Haliscomenobacte Molecular Biology Tools BLAST Create Abundance Subtable Create Normalized Abundance Subtable Quality Control 🏥 🛄 🔕 🔛 Help View Settings... S 🗇 Ide.. 1 element(s) are selected

下のアイコン列から、積み上げ棒グラフやサンバースト図を表示させることが可能



	G. Add Metadata to Abundar	nce Table				×
Tools	 Select an abundance table Select input file and map 	Set parameters				
De Novo Assemble Metagenome	columns to attributes	metasemi.xlsx	x			Browse
Amplicon-Based Analysis Normalize OTU Table by Copy Number (beta) Filter Samples Based on Number of Reads	3. Result handling	Enco	oding UTF-8	~	Start at row 0	•
OTU Clustering		Parsing				
Remove OTUs with Low Abundance Align OTUs using MUSCLE Gassify Long Read Amplicons		Field separat	tor , 🗸	Quote syn	mbol 🐪 🗸	
Detect Amplicon Sequence Variants		Preview and n	nappings			
Taxonomic Analysis Taxonomic Analysis Taxonomes by Sequence Sin Pangenomes by Taxonomy		Named of Named of Named of Named of Named of Named of Named Science (Named Scienc	columns column names will be used	d as metadata categori	es.	
Taxonomic Profiling		Name	Month			
🔄 🚱 Identify Viral Integration Sites		Name	Month			
🖻 🔯 Abundance Analysis		ERR336781	April			
		ERR336782	April			
Assign Taxonomies to Sequences in Abundance Table		ERR336783	April			
Alpha Diversity		ERR336792	June			
Beta Diversity		ERR336794	June			
PERMANOVA Analysis		ERR336804	August			
Differential Abundance Analysis		ERR336805	August			
Create Heat Map for Abundance Table		ERR336806	August			
Add Metadata to Abundance Table						
Ketine Adundance Table		[
	Help Reset		Pr	revious Next	Finish	Cancel

Add Metadata to Abundance Tableツールで、組成表にメタデータの情報を付加





Merge Abundance Tablesを使用し、サンプルごとの組成表をすべて選択





Differential Abundance Analysisツールを起動し、マージした組成表を選択



Choose where to run	Experimental design and comparison	S	
Select abundance table with two or more samples			
Experimental design and comparisons			
Result handling	Experimental design		
Result numaring	Test differential abundance due to	Sample Name	~
	Correct for	(Nothing selected)	÷
	All group pairs Against control group Among comparison groups: Control group	ielected 9 elements. 9	•
	1 Attention: No replicates for grou	ıp: 9, 10, 11, 20, 21, 22, 32, 33, 34. The results may be unreliable. Proceed v	ith care

比較に用いるキーおよび比較方法を選択 (組成表とメタデータが紐づいている場合は、群間比較等が可能)

ヒートマップの作成





Create Heat Map for Abundance Tableツールを起動し、マージした組成表を選択



お問い合わせ先:フィルジェン株式会社 TEL: 052-624-4388 (9:00~17:00) FAX: 052-624-4389 E-mail: biosupport@filgen.jp