



生体試料と環境試料の菌叢解析

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)

メタゲノム解析のアプローチ







アンプリコン シーケンス (16S/ITS)

- ・ シーケンスコスト低
- ゲノム内に複数の rRNA 遺伝子コピーを持っているため、種の定 量結果が不鮮明

全ゲノム シーケンス (WGS)

- シーケンスコスト高
- アンプリコンシーケンスより高い精度、菌叢のもつ機能推定も可能

解析ワークフロー



QC・トリミング	 NGSより出力された生データが良好か、下流分析に影響する問題がないか確認。 de novo アセンブリに有用な高品質なリードデータが得られる。 FastQCとTrimmomaticツールを統合
宿主ゲノムの除去	 宿主ゲノムがデータのノイズとなる実験の場合は必要に応じて行う宿主ゲノムを除去してクリーンなデータを作成する。 Bowie2、Samtools、および Qualimapツールを統合
Taxonomic Classification	 DNA ショートリードデータに対しに分類学的ラベルを割り当てることで菌叢解析を行う。 WGS および 16S/ITS アンプリコンデータに対応。 Kraken2ツールを統合
Differential Abundance Analysis	• 統計的にサンプルまたは条件間の菌叢の違いを検出。

16S,全ゲノムを同じ手順で解析できます。

Omicsbox



- 解析環境の構築や高スペックPCや導入が必要
- コマンドライン型のツール:パラメータ設定の操作が煩雑





OmicsBoxの菌叢解析

- ロ メーカーのサーバーで高速計算 高価なPCの購入は不要
- ロ マウス操作で簡単に解析
- ロ 16S,全ゲノムを同じ手順で解析できます。





・データが良好か、下流分析に影響する問題がないか確認

Short-Read Quality Assessment with FastQC × Short-Read Quality Assessment with FastQC × File Help View Configuration Input \bigcirc 8 You must select files or a directory. general genome analysis genetic transcript functional meta variation tools omics analysis genomics workflo The FASTQ Quality Check tool provides an easy way to perform a quality control check on sequence data coming Browse... 🕜 Additional Adapter Sequences from high throughput sequencing pipelines. The analysis is performed by nine modules which provide a quick overview of whether the data looks good and there are no problems or biases which may affect downstream FASTA Tools > analysis. Results and evaluations are returned in the form of charts and tables. This tool is based on the popular FastQC software. Short-Read Quality Assessment with FastQC FASTQ Tools > Additional Contaminant Sequences Browse... 6 Short-Read Preprocessing with Trimmomatic BAM Tools > 0 Files Clear Add Files 👩 Input Reads Long-Read Quality Assessment with LongQC Genome Browser 5 \checkmark Chart Read Length Binning 0 Venn Diagram Convert FastO to FastA Version Details: Merge FastQ/FastA Files - FastQC 0.11.8 Demultiplexing with CutAdapt Please Cite: Andrews S. (2010). FastQC: A Quality Control tool for High Thoughput Sequence Data. Retrieved 2018, from http://www.bioinformatics.babraham.ac.uk/projects/fastqc. Default Cancel Default < Back Run Cancel

画面上部のアイコンからQCツールを起動

全てのサンプルをこちらに指定





🐨 Welcome Message 🔀 FASTQ Quality Check (Dataset) 🕸 🔀 FASTQ Quality Check (ERR1948631_1.fastq) 🔀 FASTQ Quality Check (clean_ERR1948631_1.fq) 💿 *Chart: Adapter Conten

FASTQ Quality Check

Name: Dataset

Overall Results

Name	Per Base Sequence Quality	Per Sequence Qualit	y Scores	Per Base Sequence	Content	Per Sequence GC Co	ntent	Per Base N Content
ERR1948631_1.fastq	PASS	PASS		FAIL PASS		PASS		PASS
clean_ERR1948631_1.fq	PASS	PASS	FAIL			PASS		PASS
Name	Sequence Length Distribution	Adapter Content	Overrepr	resented Sequences	Sequenc	e Duplication Levels	Report	
ERR1948631_1.fastq	PASS	FAIL	WARNIN	G	FAIL		0	
clean_ERR1948631_1.fq	WARNING	PASS	WARNIN	G	FAIL		•	

The FASTQ quality check task is performed by nine analysis modules. The table above provides a quick evaluation of whether the results of each module seem entirely normal (pass), sightly abnormal (warning) or very unusual (fail). Note that these evaluations must be taken in the context of what is expected from the library. For example, some experiments may be expected to produce libraries which are biased in particular ways. Therefore, the summary evaluations should be treated as pointers that guide the preprocessing of the libraries.



正常(PASS) わずかに異常(WARNING) 異常(FAIL)

シーケンスデータの品質をすばやく評価



Filgen

biosciences & nanoscience

Per Base Sequence Quality (Sanger / Illumina 1.9 encoding) [clean_paired_SRR3233859_1. fastq.gz]



レポートのアイコンをクリック→さらに詳細な結果を見ることが可能

トリミング



gener tool	ral	genome analysis	• (gi va	enetic riation		ript	functional analysis	✓ meta genomics	workflo
F	FASTA To	ols	>	1					
F	ASTQ To	ools	>	5	hort-Read	Quality	y Assessment	t with FastQC	1
E	ВАМ Тоо	ls	>	9	Short-Read	Prepro	cessing with	Trimmomatic	
C	Genome	Browser	>	l	ong-Read	Quality	/ Assessment	with LongQC	
N	Venn Dia	gram		C	Convert Fas	tQ to F	astA		
				1	Verge Fast(Q/FastA	A Files		
				[Demultiple>	cing wi	th CutAdapt		

QC後必要に応じてトリミングを行う



様々なトリミング方法 トリミング後再度QCを行う

【補足】必ずしもすべてのQC項目を正常(PASS)にする必要はありません。



FalseやWarningと判定されていても、それを説明できる理由が あれば、適切なリードであると判断することが多いです。

例)

左の結果ではGC ContentがFalseですが、

メタゲノムでは多様な種が含まれるため正規分布から外れた形状は想定されます。

宿主ゲノムの除去





宿主DNAを含むリードは Kraken2では分類できず、データセットのノイズが増加する原因となることがあります。 リードデータを宿主ゲノムにマッピングすることで、未分類のリードの数を減らすことができます。 このプロセスを繰り返して、(系統発生的に近いターゲットゲノムなどを使用して) データを改良することができます。



①Contaminant Removalを選択

	Contaminant Removal		
	Contaminant Removal in Please add input files.		\Diamond
	Remove sequences that are consic may contain a considerable amou tool you can map the reads agains is not recommendable for long-re	lered contaminants with Bowtie2. E.g. Human gut m nt of human DNA although the libraries were already t a target genome to only keep the unaligned and co ad sequence data in this configuration.	etagenomics NGS read data c cleaned in vitro. With this ontaminant free data. Bowtie2
	Note: This tool makes use of free of future release depending on the o	loud computation resources. This is an introductory verall resource consumption of this feature.	offer and may change in a
	Sequencing Data	Single-End Reads	~ ?
②リードデータを入力 🔸	Reads Paired-End Configuration Define the pattern to distinguish extension, and the start of the nam Upstream Files Pattern	0 upstream files from downstream files. The pattern is s me should be the same for both files of each sample. R1	Files Clear Add Files
	Downstream Files Pattern	_R2	0
③宿主ゲノムの生物種を指定 リファレンスゲノムデータを 入力することも可能	Database Index Target Genome	Homo sapiens (grch38)	BrowseØ
④保存先の指定	Default	< Back Next > Rui	Cancel

宿主ゲノムの除去



Alignment Overview

Global

Sample	Total Alignments	Mapped / Contaminants	Unmapped / Contaminant-Free	Duplicated Reads (estimated)	Duplication Rate
control2	185,179	35 / 0.019%	185,144 / 99.981%	1 / 0.001%	2.94
control1	237,711	47 / 0.02%	237,664 / 99.98%	3 / 0.001%	6.82
control4	237,480	16 / 0.007%	237,464 / 99.993%	0	0
control3	132,728	23 / 0.017%	132,705 / 99.983%	0	0
smoking5	33,077	1 / 0.003%	33,076 / 99.997%	0	0
smoking4	9,810	20 / 0.204%	9,790 / 99.796%	3 / 0.031%	17.65
smoking1	29,645	41 / 0.138%	29,604 / 99.862%	14 / 0.047%	29.63
smoking3	25,220	10 / 0.04%	25,210 / 99.96%	0	0
smoking2	8,297	42 / 0.506%	8,255 / 99.494%	3 / 0.036%	2.56
control5	234,046	1 / 0%	234,045 / 100%	0	0

ACTG Content

Sample	A's	C's	T's	G's	N's	GC (%)
control2	1,158 / 27.697%	954 / 22.818%	1,097 / 26.238%	972 / 23.248%	0	46.07
control1	1,607 / 25.919%	1,412 / 22.774%	1,678 / 27.065%	1,503 / 24.242%	0	47.02
control4	614 / 28.088%	530 / 24.245%	495 / 22.644%	547 / 25.023%	0	49.27
control3	615 / 25.445%	494 / 20.439%	704 / 29.127%	604 / 24.99%	0	45.43
smoking5	51 / 27.273%	51 / 27.273%	34 / 18.182%	51/27.273%	0	54.55
smoking4	811 / 26.4%	709 / 23.079%	719 / 23.405%	833 / 27.116%	0	50.2
smoking1	1,685 / 25.871%	1,641 / 25.196%	1,532 / 23.522%	1,655 / 25.411%	0	50.61
smoking3	343 / 24.189%	368 / 25.952%	351 / 24.753%	356 / 25.106%	0	51.06
smoking2	1,614 / 23.655%	1,730 / 25.355%	1,716 / 25.15%	1,763 / 25.839%	0	51.19
control5	20/25.974%	19 / 24.675%	11 / 14.286%	27/35.065%	0	59,74

Coverage

Sample	Mean	Standard Deviation
control2	0X	0.001X
control1	02	0.0028



宿主ゲノムが除去されたリードデータの他 リード中に含まれた宿主ゲノムの割合などを含むレポートやグラフが作成される

9



meta genomics workflows		
Load	>	e: f.taxonomic_classification.005.2 $ imes$
Taxonomic Classification	>	Contaminant Removal
Metagenomic Assembly		Kraken 2
Metagenomic Gene Prediction		Database Info

 \times 5

①Kraken2を選択

S Kraken 2	
Input vou must select files or a directory.	

Kraken 2 is a taxonomic sequence classifier that assigns taxonomic labels to short DNA reads. It does this by examining the k-mers within a read and querying a database with those k-mers. This database contains a mapping of every k-mer in Kraken's genomic library to the lowest common ancestor (LCA) in a taxonomic tree of all genomes that contain that k-mer. The set of LCA taxa that correspond to the k-mers in a read are then analyzed to create a single taxonomic label for the read; this label can be any of the nodes in the taxonomic tree. Kraken is designed to be rapid, sensitive, and highly precise.

Note: This tool makes use of free cloud computation resources. This is an introductory offer and may change in a future release depending on the overall resource consumption of this feature. DefCen

Database

equencing Data	Single-End Reads	~ 🕻
	Fasta	
Reads, Contigs or Genes	Single-End Reads Paired-End Reads	
Paired-End Configuration		
Paired-End Configuration Define the pattern to distinguish ups xtension, and the start of the name	stream files from downstream files. The pattern is searched ri should be the same for both files of each sample.	ght before the file
Paired-End Configuration Define the pattern to distinguish upp xtension, and the start of the name Upstream Files Pattern	stream files from downstream files. The pattern is searched ri should be the same for both files of each sample.	ght before the file
Paired-End Configuration Define the pattern to distinguish upp xtension, and the start of the name Upstream Files Pattern Downstream Files Pattern	stream files from downstream files. The pattern is searched ri should be the same for both files of each sample. R1 R2	ght before the file

<u>_اار</u>	κ <i>≕</i> -	_クを フ	+
2.7	1 [•] J	リセノ	∇_{-}

		\overline{O}
Enable Filter		(
Kraken Confidence Filter	0.05	
Minimum Hit Groups	2	÷ (
lease Cite: Wood DE. and Salzberg SL. (2014). lignments. Genome biology, 15(3), Wood DE. Lu J. and Langmead B. (Kraken: ultrafast metagenomic sequence classificatio R46. 2019). Improved metagenomic analysis with Kraken :	on using exact
N 257		
1), 257.		

③任意で設定(セミナーではデフォルト設定) Runボタンをクリックすると解析が開始される

複雑な設定はなく解析を実行できます。





\Xi Rank	😇 Taxid	😇 Scientific Name		— f	— f	= f	- + -	f = f	= f	= + =	= f = f	/ = f.VM	B2 = f.VMB3	Â.	Hide Side Panel
unclassified	0	Unknown											144289		\bigtriangledown Actions
species	2654248	Mesorhizobium sp. INR15	サイドノ	パネ	いん	より	様々	な	図を				1		△ Charts
species	2588711	Pelagovum pacificum	留出に	- <i>∦</i> ⊏	сt).	オス	71	ゕ゙゙゙゚゚゚゚゙゙゙゙゙゙	+=	- - त			1		Taxa Pie Chart
species	1916956	Synechococcus sp. SynAce01	同半に	-1 F	עני	9 6		Ŋ. C	C 9	9 0			0		
species	2654249	Mesorhizobium sp. NBSH29											0		Taxa Bar Chart
species	1720344	Psychrobacter sp. AntiMn-1		0	0	0 0	1	0	0	0 0	0	0	0		Rarefaction Curves
superkingdom	2	Bacteria <bacteria></bacteria>		159	13	66 5	6 47.	. 40	13	19 1	4 37059	4098	284656		Diversity Curves
genus	131079	Limnobacter		0	1	11 3	6	2	0	0 0	0	0	0		PCoA Plot
genus	6	Azorhizobium		0	1	7 2	6	5	0	2 0	0	0	0		
species	7	Azorhizobium caulinodans		0	1	7 2	6	5	0	2 0	0	0	0		✓ Export
species	9	Buchnera aphidicola		0	0	1 0	0	2	0	1 0	0	0	1		
species	1179672	Flavohacterium sp. KBS0721		0	0	1 1	1	1	0	1 0	0	0	1		

菌種同定に関する結果の表が作成されます。





Kronaチャート このグラフを使用すると、サンプルを相互に比較しやすくなります。



積み上げ棒グラフ

各サンプルの界から種までの菌種同定を棒グラフで見やすく表示できます。

菌叢解析







12

菌叢解析



【Diversity Curve】 シーケンスのカバレッジが十分に深いかどうかを判断することができます。

【Rarefaction Curves】 追加のサンプルをデータセットに含めることの 微生物の多様性における利点を評価することが可能です。









14

Differential Abundance Analysis of Tax	a (f.taxonomic_classification.005.2)	– 🗆 X
Filtering and Normalization		\bigcirc
Differential Abundance Analysis of Taxa is differ between two microbial communitie	s a tool to identify Operational Taxonomic Units (OTUs es.	i) that significantly
This feature is based on edgeR, which is p	part of the Bioconductor project.	
Filter OTUs with low counts		
Minimum Sample Filter	1	÷ ?
Counts per Million	1	0
Calculate normalization factors		
Normalization Method	TMMwsp	~ 8
Default	< Back Next > Run	Cancel

①先ほどの結果のサイドパネルから
 Differential Abundance Analysisを選択

②任意でフィルターを調整



比較解析





③実験デザインファイルを入力後

コントロール群、テスト群を指定



Statistical Test		
Exact Test		0
GLM Test	GLM Likelihood Ratio Test \sim	0
Robust		0

Please Cite:

Robinson MD., McCarthy DJ. and Smyth GK. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics (Oxford, England), 26(1), 139-40.

Default	< Back	Next >	Run	Cancel

④どの分類階級に対して比較するか指定 Runボタンをクリックすると解析が開始される





Ir	= Tags	🐨 Taxald	≡ Sr	cientific Name	⊤ FC	🐨 logFC	😇 PValue	= FDR	⊤ LR			Hide Toolbar
		698776	Cellulosilyticum		9.5944	3.26219	0.08771	0.23337	2.91587			Actions
		630749	Spongiibacter		-6.21314	-2.63532	0.30332	0.47851	1.05953			Summary Report
		929812	Gibbsiella		1.59187	0.67072	0.57597	0.6815	0.31279			Set Over/Under Tags
		1204360	Siansivirga		-14.03284	-3.81074	0.10363	0.25584	2.64875			Summary Chart 🔬
		295595	Candidatus Hepatoplasma		-3.60657	-1.85063	0.4667	0.57957	0.52977			Heatmap
		2282742	Desulfofarcimen		-3.60657	-1.85063	0.4667	0.57957	0.52977			
UNDER		2004797	Luteitalea		-7.3956	-2.88667	1.2033E-6	1.8724E-5	23.57194			
		2282740	Desulfallas		-1.46261	-0.54854	0.69094	0.76309	0.15807			T
OVER		76023	Halothece		92.90584	6.5377	1.6794E-21	8.6240E-20	90.69112			
OVER		400634	Lysinibacillus		9.80931	3.29415	0.00734	0.04512	7.18874			
	L	1696	Brevibacterium		-2.1776	-1.12274	0.15785	0.33407	1.99473			
		$\mathcal{N} = \mathbf{F}$:R(緑)(n々ガがん	ト与さわ	ます		912			
rress 😨 File Manaç V XXXX OmicsBoxWorkspac	ager 🖾 🌚 Application Message	over (* **	赤)UNDE	R(緑)(● Welcom ● OTU Dataset	Dタグが付 e Message ⁽) OTU DA I J Differential : taxonomic_classificatio	ト与され Report (taxonomic_ Abundan n ODAT	ます。 ^{Classification ODAT)} ce Testing	¤ Report	912	サ- ヒ-	イドパネルから -トマップの作成	も可能(次スラ
gress Tile Manaç 4 ~ OmicsBoxWorkspac	ager ☆ @ Application Message	°=√ ⊅	赤)UNDE	R(緑)(で Welcom OTL Dataset Over	Dタグが付 e Message @ OTU DA i J Differential L taxonomic_classificatio view Total features: 1,027	ト与され Report (taxonomic_ Abundan n ODAT	ます。 ^{classification ODAT)}	≌ Report	912 102	サ- ヒ-	イドパネルから -トマップの作成	も可能(次スライ
ress File Manage ComicsBoxWorkspac Fストにん 食出され	C ager ☆ ② Application Message たe 使用されたバ れたOTUの数	OVER ([*] [■] ××× 、 、 、 ラメーター ななどのレ ⁷	赤)UNDE -、 ポート	R(緑)(「」 「」 のTL Dataset	Dタグが付 Message のTU DA I D Differential : taxonomic_classificatio view Total features: 1,027 Contrast Group: VAB, VA Reference Group: PAB, I	ト与され Report (taxonomic_ Abundan n ODAT MB PMB	ます。 classification ODAT) ce Testing	¤ Report	912 102	サ- ヒ-	イドパネルから -トマップの作成	も可能(次スラ
es @ File Manac white Box Workspac こ ストにん 食出され	c ager ⊠ @ Application Message kce 使用されたパ れたOTUの数	OVER (え [■] ××× 、 、 、 、 、 、 、 、 、 、 、 、 、	赤)UNDE -、 ポート	R(緑)(Welcom Yy OTL Dataset Over 	Dタグが付 Message ① OTU DA I D Differential に taxonomic_classificatio view Total features: 1,027 Contrast Group: VAB, VA Beference Group: PAB, I Its	ト与され Report (taxonomic_ Abundan n ODAT ив PMB	ます。 classification ODAT) ce Testing	¤ Report	912 102	サ- ヒ-	イドパネルから -トマップの作成	も可能(次スラ
gress @ File Manac ↓ OmicsBoxWorkspac テスト(こ(食出され	c ager ☆ @ Application Message kce	OVER (え [■] [■] 、 、 、 、 、 、 、 、 、 、 、 、 、	赤)UNDE -、 ポート	R(緑)(で で で で で で で で で で で で で	Dタグが付 Message ① OTU DA I D Differential C taxonomic_classificatio view Total features: 1,027 Contrast Group: VAB, VA Beference Group: VAB, VA Beference Group: PAB, I uits tially abundant (DA) feat presented (logFC < -1).7 epresented (logFC < -1).7 epresented (logFC < -1).7	ト与され Report (taxonomic_ Abundan n ODAT //B PMB ures (FDR < 0.05): 7 : 90 n	ます。 classification ODAT) ce Testing	x Report	912	サ- ヒ-	イドパネルから -トマップの作成	も可能(次スラ

菌叢解析







サイドパネルからヒートマップを作成できます。

全ゲノムメタゲノム解析の機能推定





Omicsboxでは菌叢解析だけでなく、ショットガンメタゲノムシーケンス (全ゲノム)を使用した機能推定解析もサポート





OmicsBox DBEAF

- 16S,全ゲノムを同じ手順で解析できます。
- 初心者でも解析できるインターフェース
- 7日間無料のデモライセンス→ <u>詳細(PDF)</u>









お問い合わせ先:フィルジェン株式会社

TEL 052-624-4388 (9:00 \sim 17 : 00)

FAX 052-624-4389

E-mail: biosupport@filgen.jp