

SAG summary report manual

1. 用語説明	2
2. Html レポートの全体像	3
2.1. Html レポートの全体構造	3
2.2. 全体サマリ	4
2.3. サンプル別個別レポート	5
3. サイドバー	6
4. 全体サマリ	7
4.1. 統計表	7
4.2. 系統情報プロット	8
4.3. 系統情報テーブル	9
5. サンプル別個別レポート	10
5.1. 統計情報	10
5.2. リード品質	11
5.3. ゲノム品質	12
5.4. 系統情報	13
5.5. ゲノム情報	14
6. Information	15

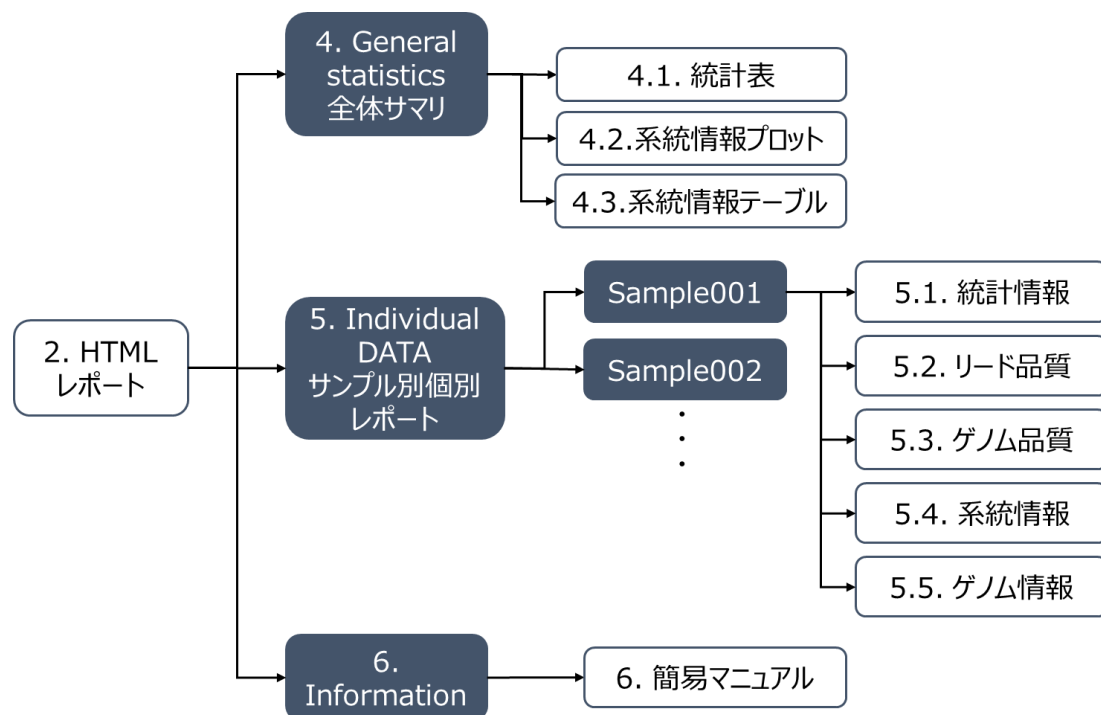
1. 用語説明

SAG	Single-cell Amplified Genome の略。
ゲノム完全性	予め設定されたシングルコピーマーカー遺伝子のカバー率
汚染率	シングルコピーマーカー遺伝子の重複率
High quality	ゲノム完全性が90%超、汚染率5%未満、5S 16S 23S を検出、tRNAを18 種以上検出 参考文献： Bowers R., et. al., Nat Biotechnol, 2017
Medium quality	ゲノム完全性が50%以上、汚染率10%未満
Low quality	ゲノム完全性が50%未満、汚染率10%未満
ANI	Average Nucleotide Identity の略。参照配列に対してアライメントされた配列中の参照配列との一致率。
AF	Alignment Fraction の略。参照配列に対してアライメントされた配列の割合を示す。
Novelty score	ANIとAF から算出され、SAG が新規種である確率を示す。 参考文献： Neha J. Varghese, et. al., Nucleic Acids Research, 2015

2. Html レポートの全体像

2.1. Html レポートの全体構造

Html レポートは全体サマリ、サンプル別個別レポート、Information で構成されています。HTML レポートのサイドバーに下図の塗りつぶしの項目があり、クリックすると枠のみの項目が表示されます。項目中の数字は本マニュアルの章番号です。

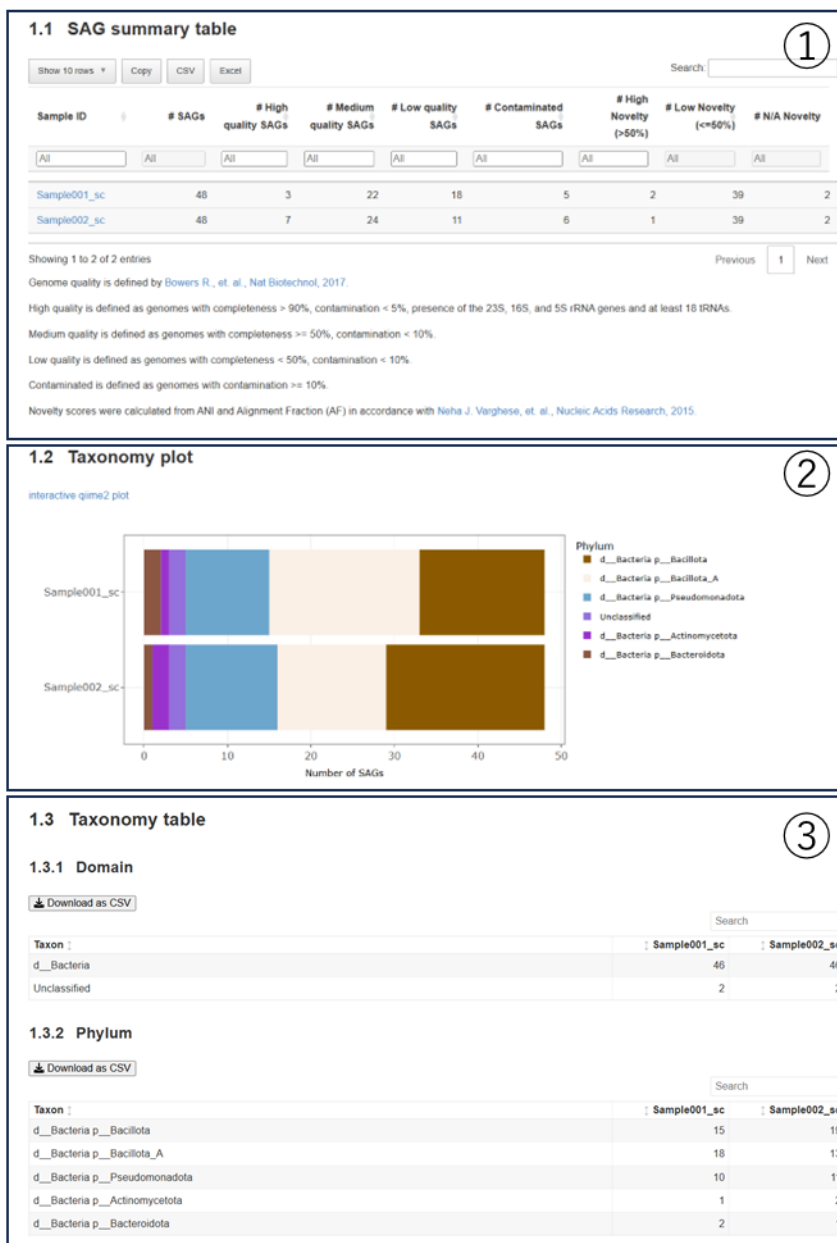


2.2. 全体サマリ

全体サマリは①統計表、②系統情報プロット、③系統情報テーブルで構成されています。

SAG-Report

1 General statistics



① 統計表	SAG の統計情報が表示されます。
② 系統情報プロット	生物系統の門別に色分した積み上げ棒グラフです。
③ 系統情報テーブル	生物系統の各階層の系統別にゲノム数を集計した表です。

2.3. サンプル別個別レポート

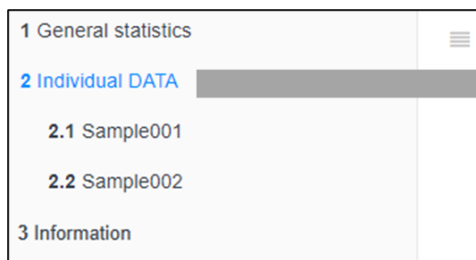
サンプル別個別ページは①統計情報、②データ品質、③ゲノム品質、④系統情報、⑤ゲノム情報で構成されています。



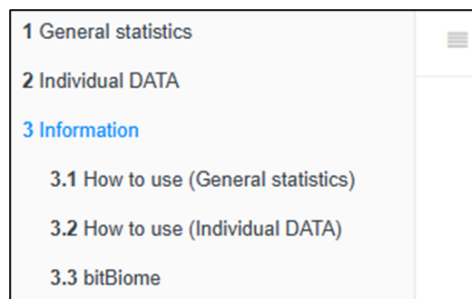
① 統計情報	サンプルごとのデータ品質などが表示されます。
② リード品質	SAG ごとのリードの品質プロットです。
③ ゲノム品質	SAG ごとのゲノム完全性、汚染率、新規性プロットです。
④ 系統情報	検出された生物系統情報が Tree map で表示されます。
⑤ ゲノム情報	SAG の詳細情報がテーブルで表示されます。

3. サイドバー

Html レポートのサイドバー部分は全体サマリ、サンプル別個別レポート、Information で構成されています。



サイドバーの Individual DATA をクリックすることで、表示する Sample を選択できます。



Information をクリックすると、簡易 Help が表示されます。

4. 全体サマリ

4.1. 統計表

SAG の統計情報です。表上部の各ボタンをクリックすると、表示行数の変更 (Show)、クリップボードへのコピー (Copy)、CSV 形式で保存 (CSV)、Excel 形式で保存 (Excel) が可能です。

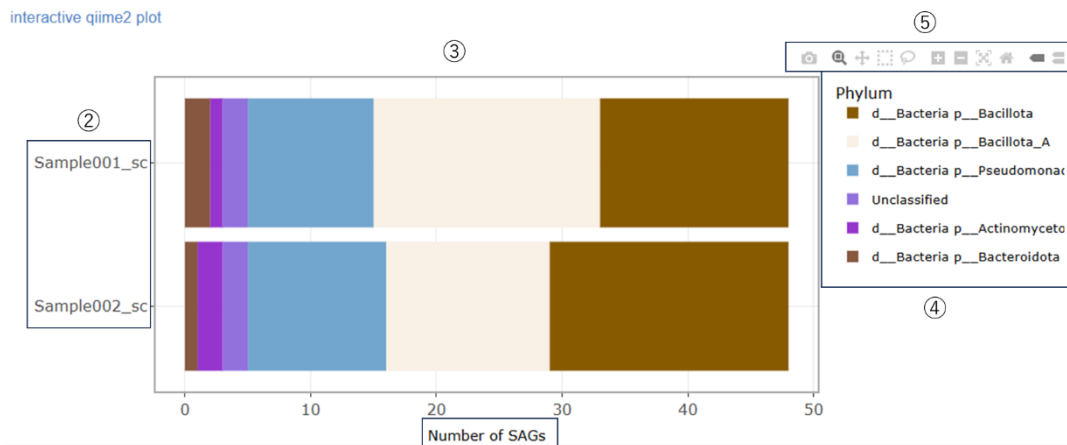
1.1 SAG summary table								
Show 10 rows ▼		Copy	CSV	Excel	Search:			
Sample ID	# SAGs	# High quality SAGs	# Medium quality SAGs	# Low quality SAGs	# Contaminated SAGs	# High Novelty (>50%)	# Low Novelty (<=50%)	# N/A Novelty
All	All	All	All	All	All	All	All	All
Sample001_sc	48	3	22	18	5	2	39	2
Sample002_sc	48	7	24	11	6	1	39	2
Showing 1 to 2 of 2 entries					Previous	1	Next	
Genome quality is defined by Bowers R., et. al., Nat Biotechnol, 2017.								
High quality is defined as genomes with completeness > 90%, contamination < 5%, presence of the 23S, 16S, and 5S rRNA genes and at least 18 tRNAs.								
Medium quality is defined as genomes with completeness >= 50%, contamination < 10%.								
Low quality is defined as genomes with completeness < 50%, contamination < 10%.								
Contaminated is defined as genomes with contamination >= 10%.								
Novelty scores were calculated from ANI and Alignment Fraction (AF) in accordance with Neha J. Varghese, et. al., Nucleic Acids Research, 2015.								

Sample ID	サンプル名
# SAGs	SAG の数
# High quality SAGs	High quality SAG の数
# Medium quality SAGs	Medium quality SAG の数
# Low quality SAGs	Low quality SAG の数
# Contaminated SAGs	ゲノム汚染率≥10%の SAG の数
# High Novelty (> 50%)	新規種である確率が高い (> 50%) SAG の数
# Low Novelty (<= 50%)	新規種である確率が低い (<= 50%) SAG の数
# N/A Novelty	参照配列を発見できず、新規性を算出できなかった (N/A) SAG の数

4.2. 系統情報プロット

検出された生物系統の門レベルで色分けした積み上げ棒グラフです。カーソルをプロット上に合わせると、カーソル位置のデータ情報が表示されます。左上にある「interactive qiime2 plot」をクリックすると、よりインタラクティブな系統情報プロットが表示されます。

1.2 Taxonomy plot



- ① ゲノム数
- ② サンプル名
- ③ 門ごとのゲノム数が棒グラフで表示されます。
- ④ 凡例:門ごとに色分けされています。
- ⑤ プロットを操作するパネルです。

4.3. 系統情報テーブル

生物系統の階層別にテーブルが表示されます。各系統のゲノム数がサンプル別に集計されています。
「Download as CSV」をクリックすることで、表示されているテーブルをダウンロードできます。Seach 内に
キーワードを入力することで、注目する系統のみを表示させることができます。

1.3.2 Phylum		
Download as CSV Search		
Taxon ↑	Sample001_sc ↑	Sample002_sc ↑
d__Bacteria p__Bacillota	15	19
d__Bacteria p__Bacillota_A	18	13
d__Bacteria p__Pseudomonadota	10	11
d__Bacteria p__Actinomycetota	1	2
d__Bacteria p__Bacteroidota	2	1

列の説明

Taxon	生物系統名
検体名	各検体におけるゲノム数

5. サンプル別個別レポート

5.1. 統計情報

サンプルごとのデータ品質などが表示されます。

2.1.1 stats

# Genomes	48	検体中のゲノム数
Total Reads	14.4 Mreads	検体中の全リード数
Total Length	2.17 Gbp	検体中の全リード長
Clean Reads	14.2 Mreads	高品質のリード数
Clean Length	2.1 Gbp	高品質のリード長
High quality	3	High qualityのゲノム数
Medium quality	22	Medium qualityのゲノム数
Low quality	18	Low qualityのゲノム数
Contaminated	5	Contamination 率が高いゲノム数
Novelty N/A	2	Novelty score が算出されなかったゲノム数
high Novelty	2	Novelty score が50%以上のゲノム数
low Novelty	39	Novelty score が50%以下のゲノム数

5.2. リード品質

SAG ごとのリード品質のプロットです。カーソルをプロット上に合わせると、カーソル位置のデータ情報が表示されます。

2.1.2 Reads quality plot

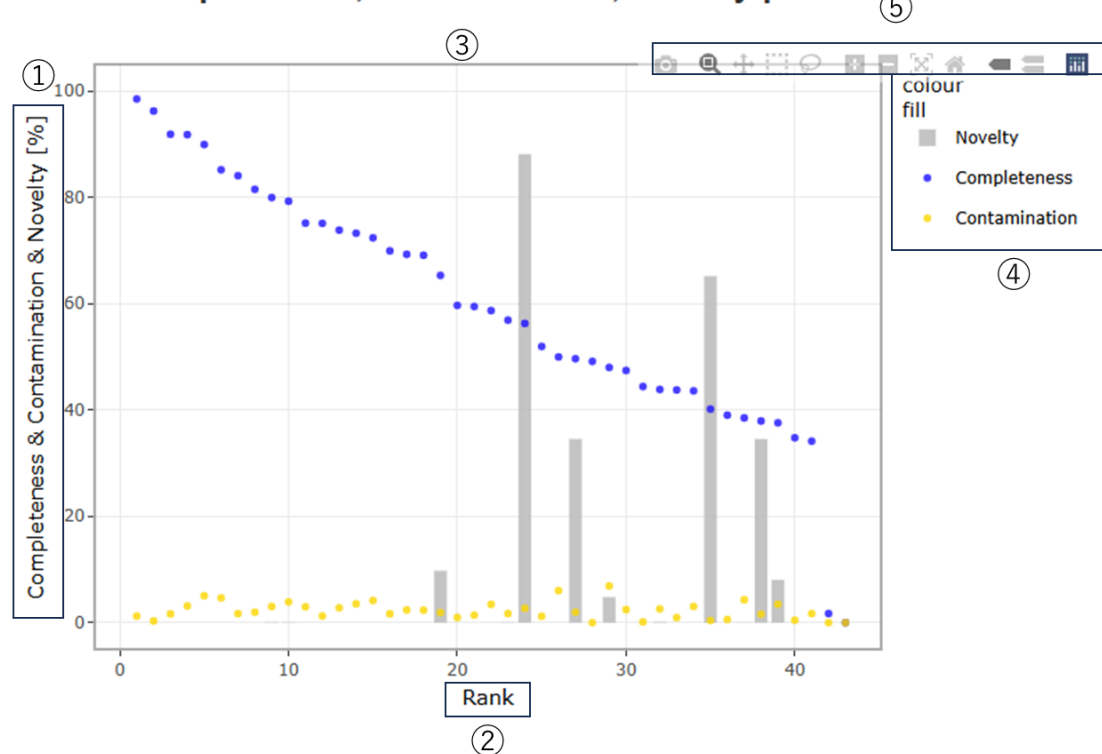


- ① SAG リードの塩基数 (単位: Mbp)
- ② Rank (ゲノム完全性の降順) で並べ替えた SAG
- ③ SAG ごとのリードの合計塩基数が棒グラフで表示されます。
- ④ 凡例: 高品質塩基(high quality)、低品質塩基(low quality)で色分けされています。
- ⑤ インタラクティブに図を操作するパネルです。

5.3. ゲノム品質

SAG ごとのゲノム完全性、汚染率、新規性のプロットです。カーソルをプロット上に合わせると、カーソル位置のデータ情報が表示されます。

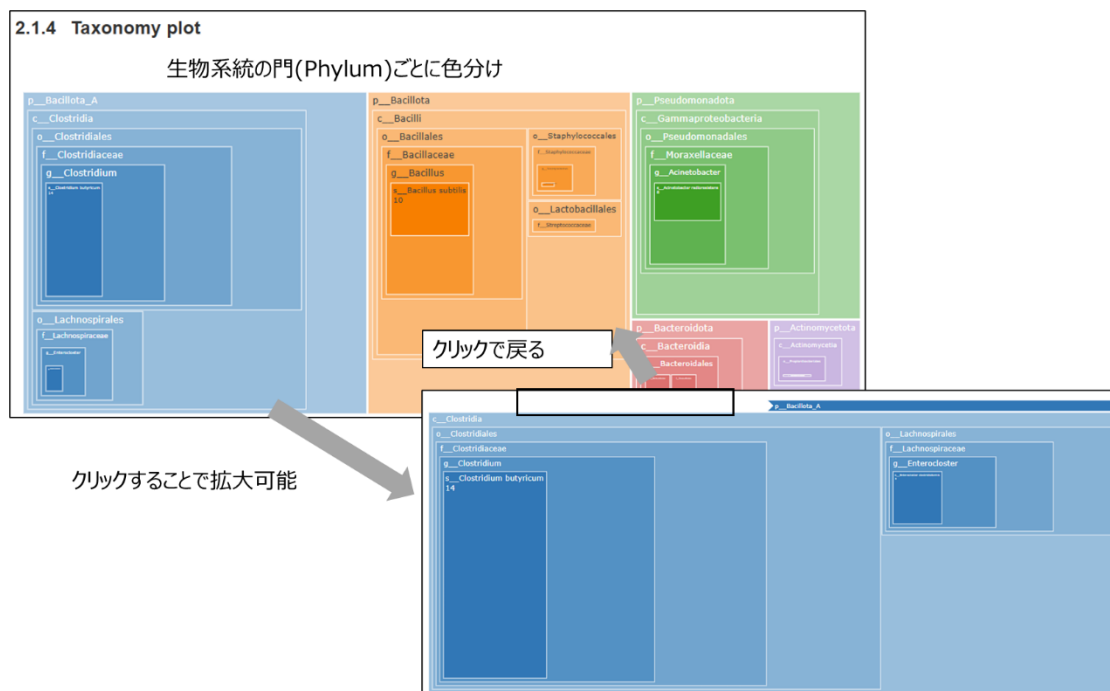
2.1.3 Completeness, Contamination, Novelty plot



- ① ゲノム完全性、汚染率、新規性（単位：%）
- ② Rank（ゲノム完全性の降順）で並べ替えた SAG
- ③ 新規性が棒グラフ、ゲノム完全性（青）と汚染率（黄）がドットプロットで表示されます。
- ④ 凡例：棒グラフが新規性 (Novelty)、ゲノム完全性 (Completeness) と汚染率 (Contamination) で色分けされています。
- ⑤ 図をインタラクティブに操作するパネルです。

5.4. 系統情報

検出された生物系統情報が Tree map で表示されます。面積は SAG の数に比例します。系統情報解析に参照配列を使用して SAG の系統情報を同定しています。



5.5. ゲノム情報

SAG のゲノム完全性、汚染率、新規性、系統情報などの表です。

2.1.5 SAG table

Show 10 rows *CopyCSVExcel

Search:

Genome ID	Completeness	Contamination	Novelty	Quality	Genome size [Mbp]	Nearest Reference	Taxonomy
All	All	All	All	All	All	All	All
Sample001_sc-00035	98.49	1.24	0	High quality	3.3	GCF_006757745.1	d__Bacteria p__Pseudomonadota c__Gammaproteobacteria o__Pseudomonadales f__Moraxellaceae g__Acinetobacter s__Acinetobacter radioresistens
Sample001_sc-00003	96.23	0.33	0	High quality	2.32	GCF_006742205.1	d__Bacteria p__Bacillota c__Bacilli o__Staphylococcales f__Staphylococcaceae g__Staphylococcus s__Staphylococcus epidermidis

NCBI のゲノム情報に移動

NIH National Library of Medicine
National Center for Biotechnology Information

Search NCBI

NCBI Datasets Taxonomy Genome Gene Command-line tools Documentation

Genome assembly ASM674220v1

Download datasets URL FTP Actions

NCBI RefSeq assembly

GCF_006742205.1

:

Submitted GenBank assembly

GCA_006742205.1

:

Taxon

Staphylococcus epidermidis

:

Strain

NRRC 100911

:

Relation to type material

assembly from type material

:

Submitter

Biological Resource Center, National Institute of Technology and Evaluation (NBRC)

:

Date

Jun 11, 2019

:

列の説明

Genome ID	SAG 名
Completeness	ゲノム完全性 (%)
Contamination	汚染率 (%)
Novelty	ゲノム新規性 (%)
Quality	SAG の品質 (High, Medium, Low quality, Contamination)
Genome size [Mbp]	SAG の合計塩基数 (Mbp)
Nearest Reference	SAG に最も近い参照ゲノム配列 (リンク先 NCBI Assembly)
Taxonomy	生物系統名

6. Information

簡易 Help 画面が表示されます。本マニュアルに記載されている使用方法を確認できます。

