

MAG 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. MAG 統計	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. データ品質	11
5.3. 系統情報	12
5.4. ゲノム情報	13
6. Information	14

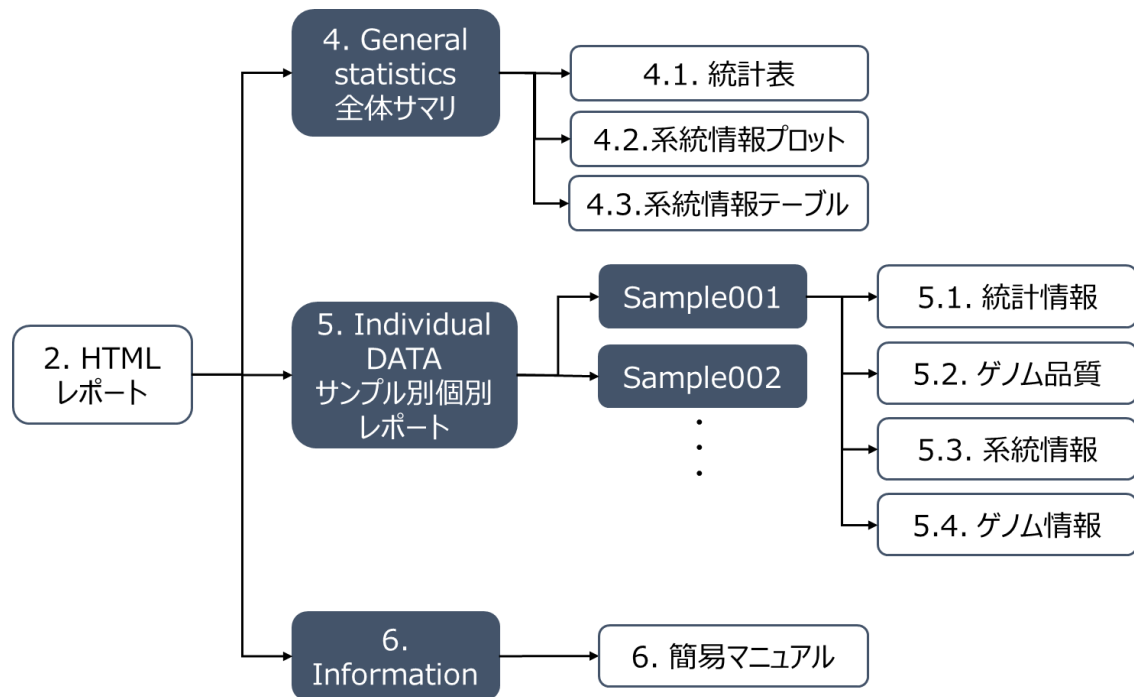
1. 用語説明

MAG	Metagenome-Assembled 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
Complete	下記 High quality の条件に加え、N50 と同一の長さの環状コンティグを保有
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
Novelty score	ANI と AF から算出される、MAG が新規の種である確率 参考文献： Neha J. Varghese, et. al., Nucleic Acids Research, 2015
N50	配列長の加重平均。配列を長い順に並べて加算したとき、全体の半分の長さに達したときの配列の長さ。

2. Html レポートの全体像

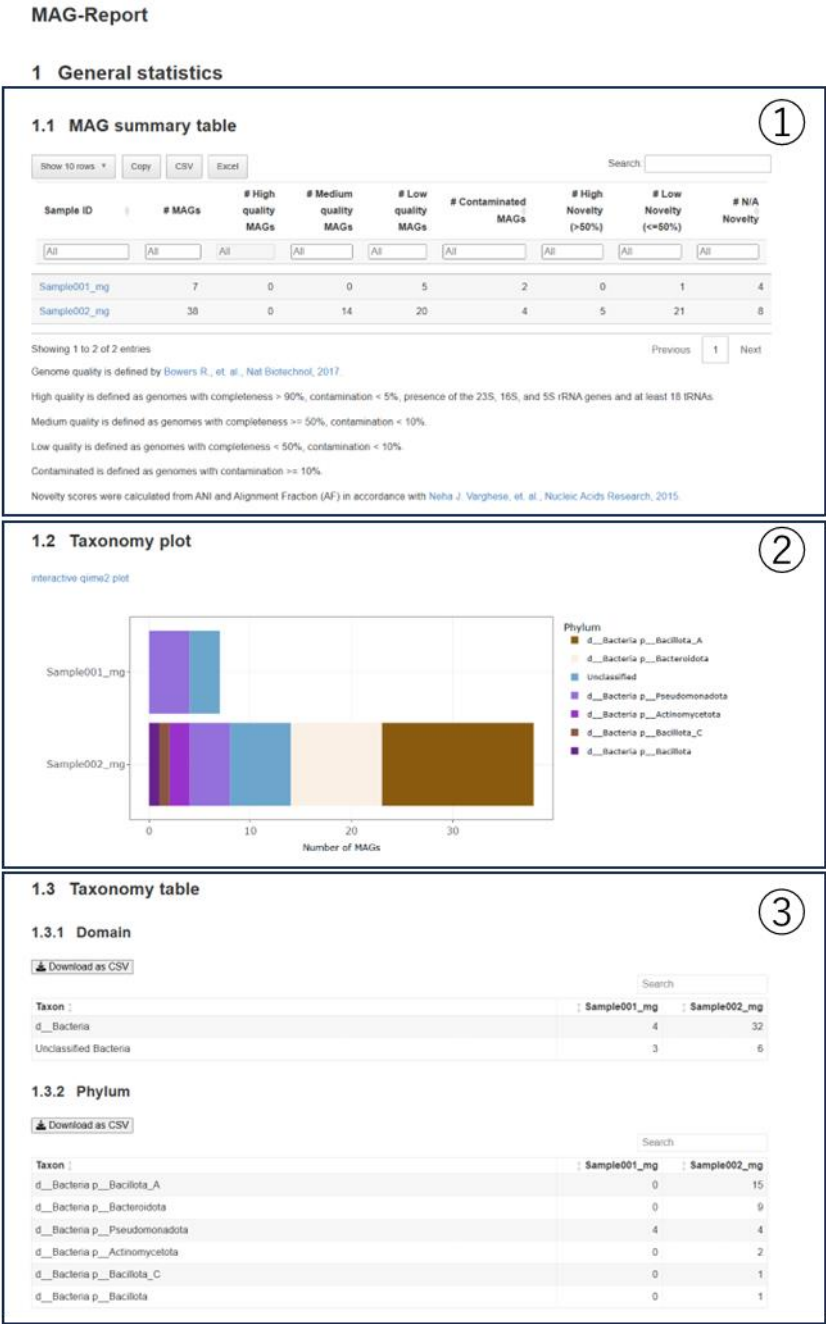
2.1. Html レポートの全体構造

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

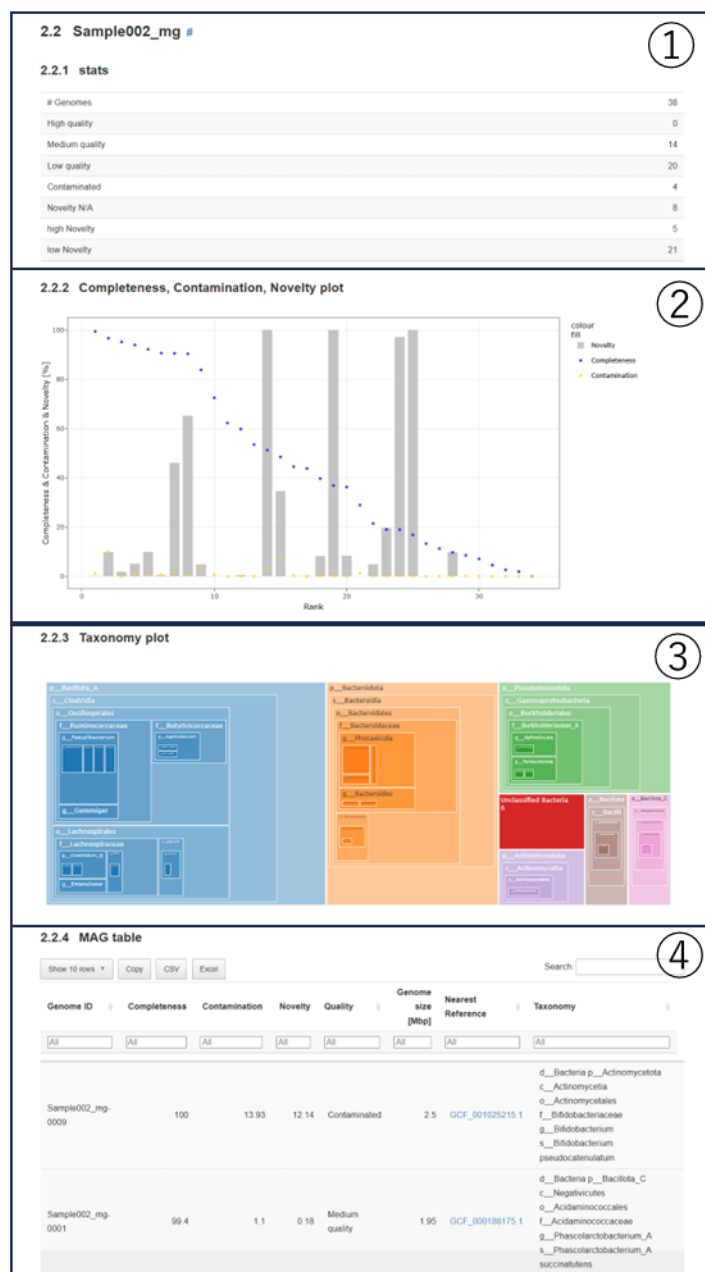
全体サマリは①統計表、②系統情報で構成されています。



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

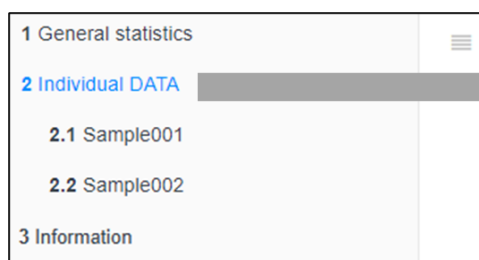
サンプル別個別ページは①統計情報、②データ品質、③系統情報、④ゲノム情報で構成されています。



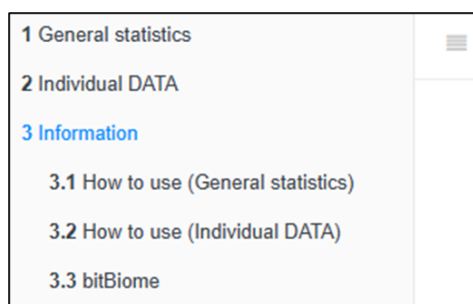
① 統計情報	サンプルごとのデータ品質などが表示されます。
② ゲノム品質	MAG ごとのゲノム完全性、汚染率、新規性プロットです。
③ 系統情報	検出された生物系統情報が Tree map で表示されます。
④ ゲノム情報	MAG の詳細情報がテーブルで表示されます。

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. MAG 統計

MAG 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 MAG summary table								
Show 10 rows ▾		Copy	CSV	Excel	Search:			
Sample ID	# MAGs	# High quality MAGs	# Medium quality MAGs	# Low quality MAGs	# Contaminated MAGs	# High Novelty (>50%)	# Low Novelty (<=50%)	# N/A Novelty
All	All	All	All	All	All	All	All	All
Sample001_mg	7	0	0	5	2	0	1	4
Sample002_mg	38	0	14	20	4	5	21	8
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.								

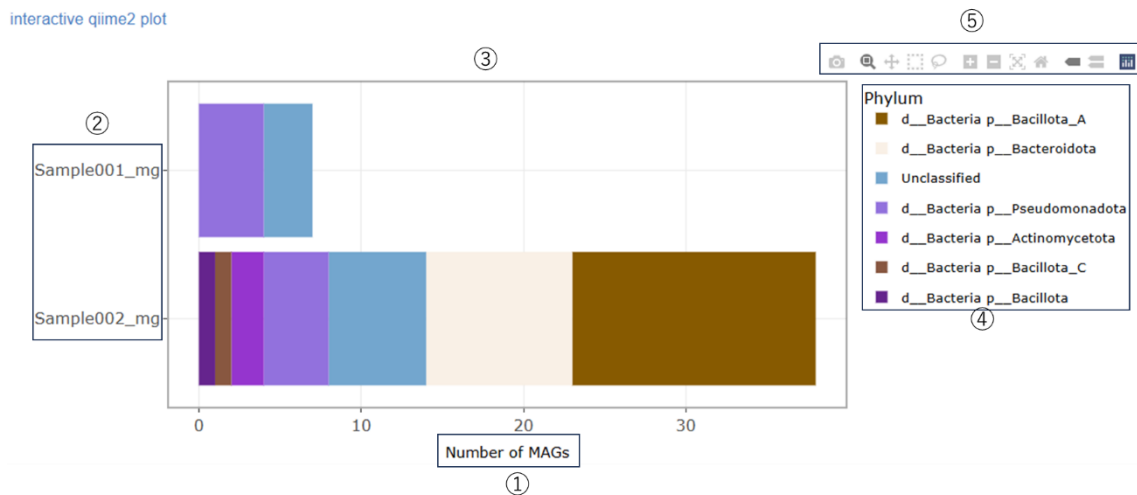
列の説明

Sample ID	サンプル名
# MAGs	MAG の数
# High quality MAGs	High quality MAG の数
# Medium quality MAGs	Medium quality MAG の数
# Low quality MAGs	Low quality MAG の数
# Contaminated MAGs	ゲノム汚染率 $\geq 10\%$ の MAG の数
# High Novelty (> 50%)	新規種である確率が高い (> 50%) MAG の数
# Low Novelty (<= 50%)	新規種である確率が低い (<= 50%) MAG 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) MAG 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。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_mg	↕ Sample002_mg
d__Bacteria p__Bacillota_A	0	15
d__Bacteria p__Bacteroidota	0	9
d__Bacteria p__Pseudomonadota	4	4
d__Bacteria p__Actinomycetota	0	2
d__Bacteria p__Bacillota_C	0	1
d__Bacteria p__Bacillota	0	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。

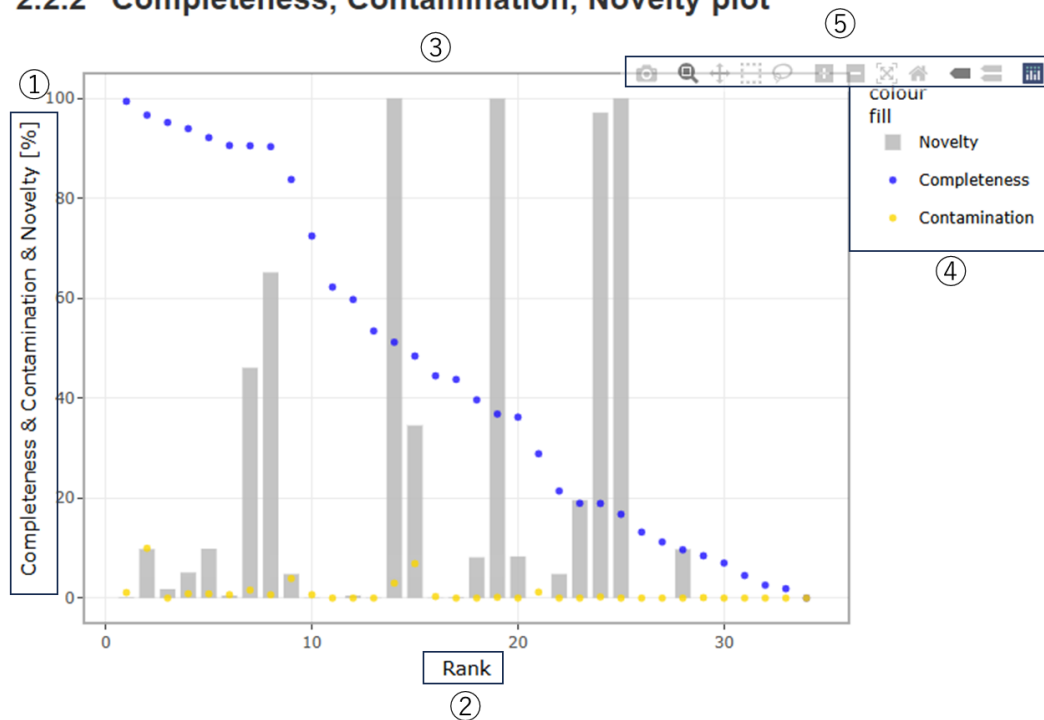
2.2.1 stats

# Genomes	38	検体中のゲノム数
High quality	0	High qualityのゲノム数
Medium quality	14	Medium qualityのゲノム数
Low quality	20	Low qualityのゲノム数
Contaminated	4	Contamination 率が高いゲノム数
Novelty N/A	8	Novelty score が算出されなかったゲノム数
high Novelty	5	Novelty score が50%以上のゲノム数
low Novelty	21	Novelty score が50%以下のゲノム数

5.2. データ品質

MAG 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。

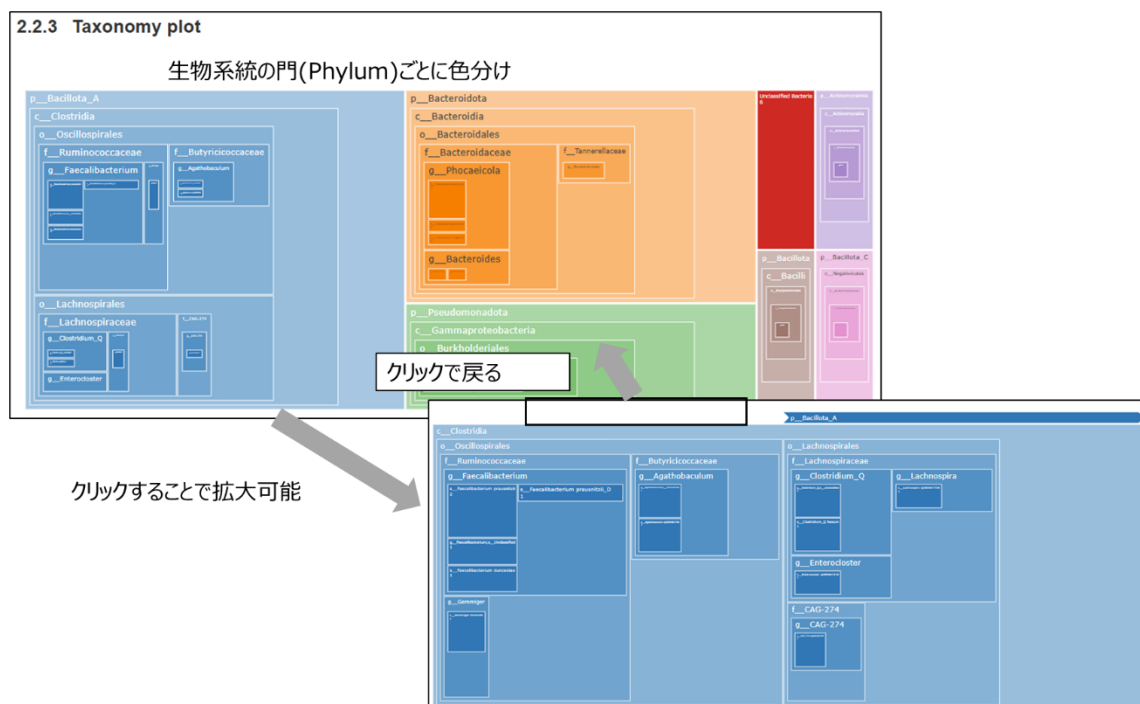
2.2.2 Completeness, Contamination, Novelty plot



- 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 MAG
- 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：新規性(Novelty)、ゲノム完全性(Completeness)、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。

5.3. 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 MAG の数に比例します。系統情報解析に参照配列を使用して MAG の系統情報を同定しています。



5.4. ゲノム情報

MAG のゲノム完全性、汚染率、新規性、系統情報などの表です。

2.2.4 MAG table

Show 10 rows ▾ Copy CSV Excel Search:

Genome ID	Completeness	Contamination	Novelty	Quality	Genome size [Mbp]	Nearest Reference	Taxonomy
Sample002_mg-0009	100	13.93	12.14	Contaminated	2.5	GCF_001025215.1	d__Bacteria p__Actinomycetota c__Actinomycetia o__Actinomycetales f__Bifidobacteriaceae g__Bifidobacterium s__Bifidobacterium pseudocatenulatum
Sample002_mg-0001	99.4	1.1	0.18	Medium quality	1.95	GCF_000188175.1	d__Bacteria p__Bacillota_C c__Negativicutes o__Acidaminococcales f__Acidaminococcaceae g__Phascolarctobacterium_A s__Phascolarctobacterium_A succinatutens

NCBI のゲノム情報に移動

列の説明

Genome ID	MAG 名
Completeness	ゲノム完全性 (%)
Contamination	汚染率 (%)
Novelty	ゲノム新規性 (%)。 近い参照配列が存在しない場合、空欄となります。
Quality	MAG の品質
Genome size [Mbp]	全コンティグ長
Nearest Reference	MAGに最も近い参照ゲノム配列 (リンク先 NCBI Assembly)。 参照配列が存在しない場合、No reference と表示されます。
Taxonomy	系統情報名

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。

