Isolate 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
3. サイドバー 5
4. 全体サマリ 6
   4.1. 統計表 6
   4.2. リード品質 7
   4.3. ゲノム品質 8
   4.4. 系統情報 9
   4.5. ゲノム情報 10
5. Information 11

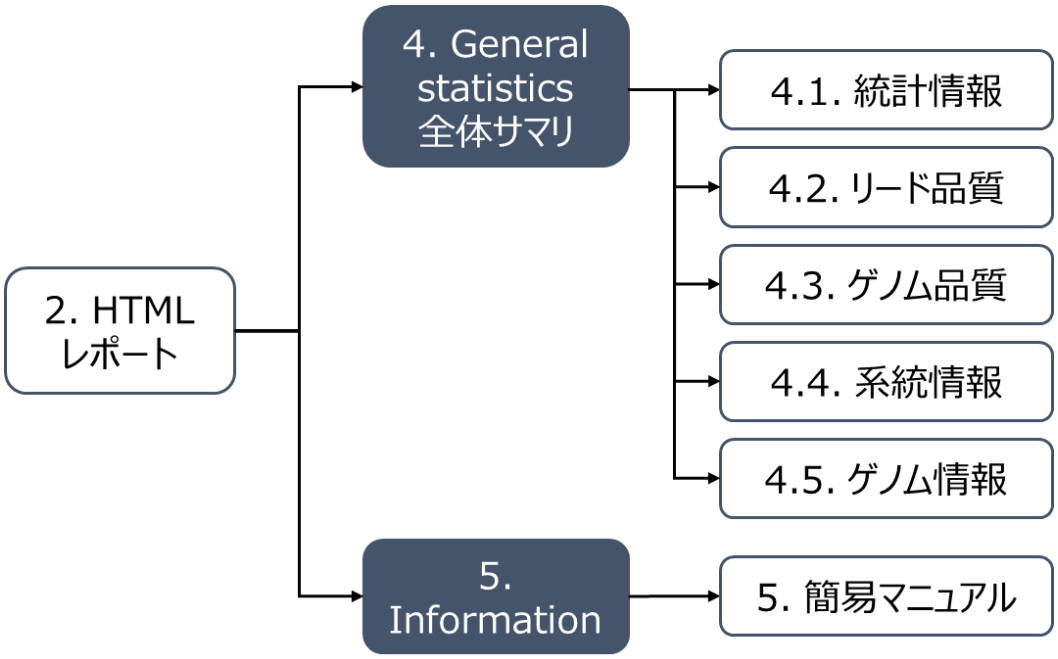
1．用語説明

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を示す。
Novelty score	ANI と AF から算出され、菌株ゲノムが新規種である確率を示す。
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、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。



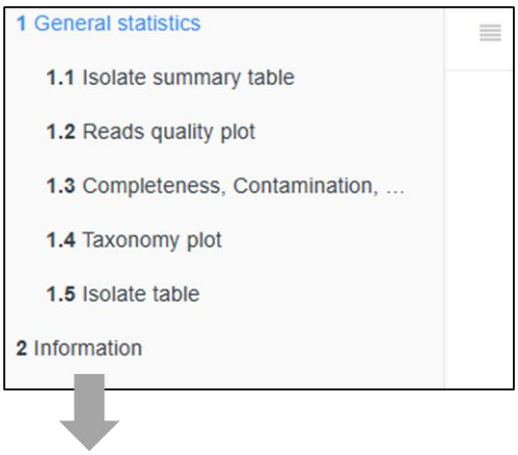
2.2. 全体サマリ

全体サマリは①統計情報、②リード品質、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	株ゲノムごとのリード情報のプロットです。
③ ゲノム品質	株ゲノム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	株ゲノム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、Information で構成されています。



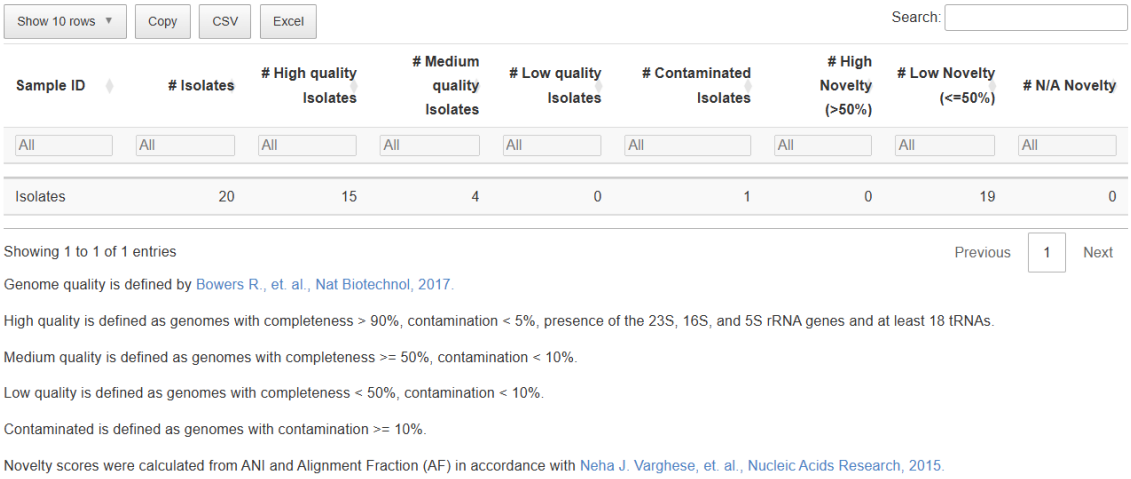
Information をクリックすると、簡易 Help が表示されます。

4. 全体サマリ

4.1. 統計表

株ゲノムの統計情報です。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 Isolate summary table


Show 10 rows Copy CSV Excel Search:

Sample ID	# Isolates	# High quality Isolates	# Medium quality Isolates	# Low quality Isolates	# Contaminated Isolates	# High Novelty (>50%)	# Low Novelty (<=50%)	# N/A Novelty
All	All	All	All	All	All	All	All	All
Isolates	20	15	4	0	1	0	19	0

Showing 1 to 1 of 1 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%.

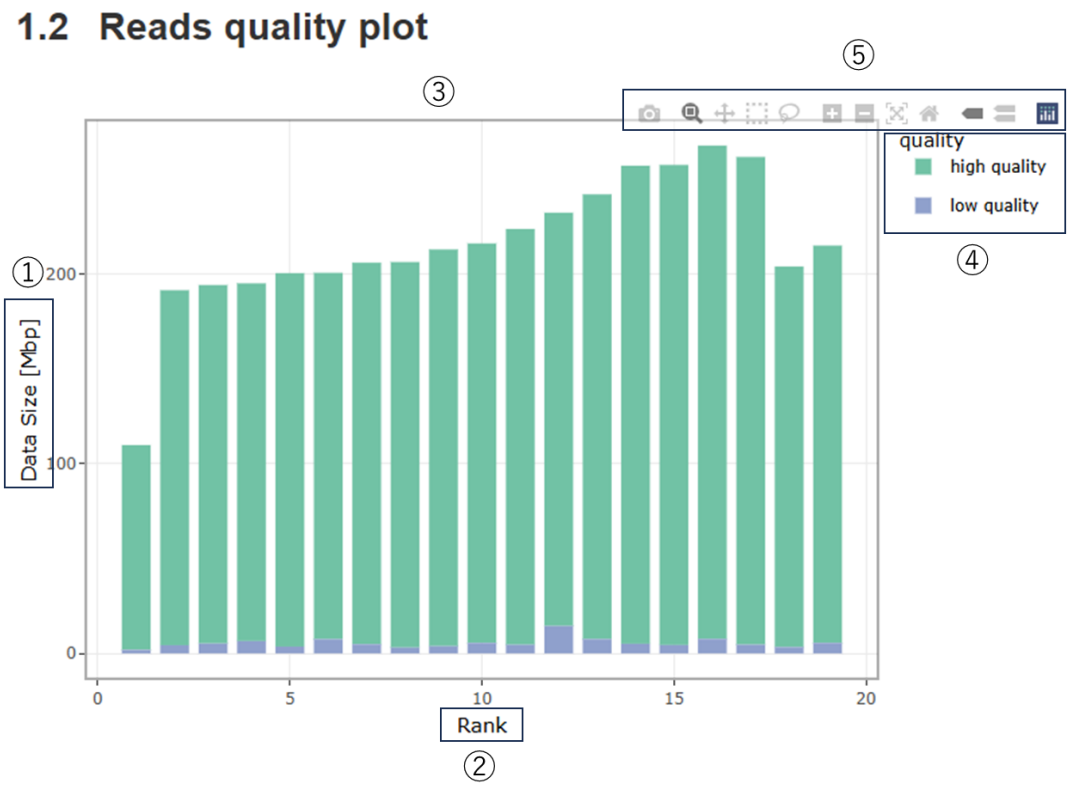
Contaminated is defined as genomes with contamination >= 10%.

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（Isolates）
# Isolates	株ゲノムの数
# High quality Isolates	High quality ゲノムの数
# Medium quality Isolates	Medium quality ゲノムの数
# Low quality Isolates	Low quality ゲノムの数
# Contaminated Isolates	ゲノム汚染率≧10%のゲノムの数
# High Novelty (> 50%)	新規種である確率が高い（> 50%）ゲノムの数
# Low Novelty (<= 50%)	新規種である確率が低い（<= 50%）ゲノムの数
# 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）ゲノムの数

4.2. リード品質

株ゲノム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。



① 株ゲノムリードの塩基数（単位: Mbp）

② Rank (ゲノム完全性の降順) で並べ替えた株ゲノム

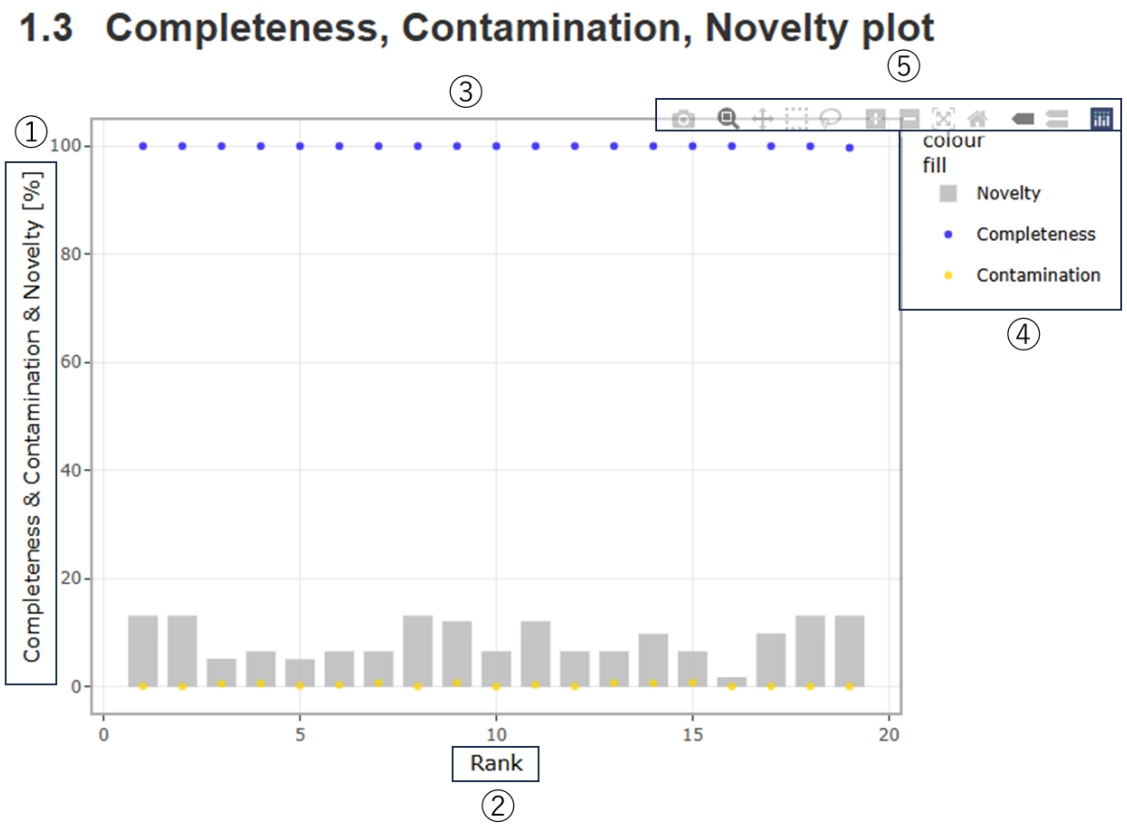
③ 株ゲノムごとのリードの合計塩基数が棒グラフで表示されます。

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。

4.3. ゲノム品質

株ゲノムのゲノム完全性、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。



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) で並べ替えた株ゲノム

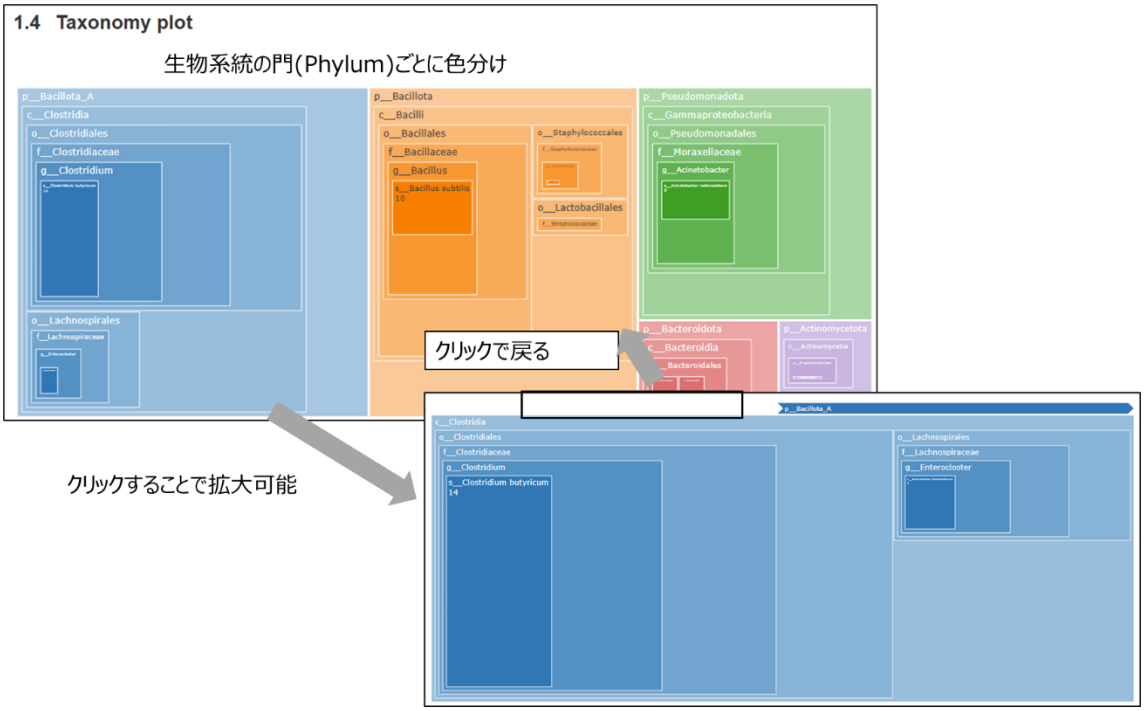
③ 新規性が棒グラフ、ゲノム完全性（青）と汚染率（黄）がドットプロットで表示されます。

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。

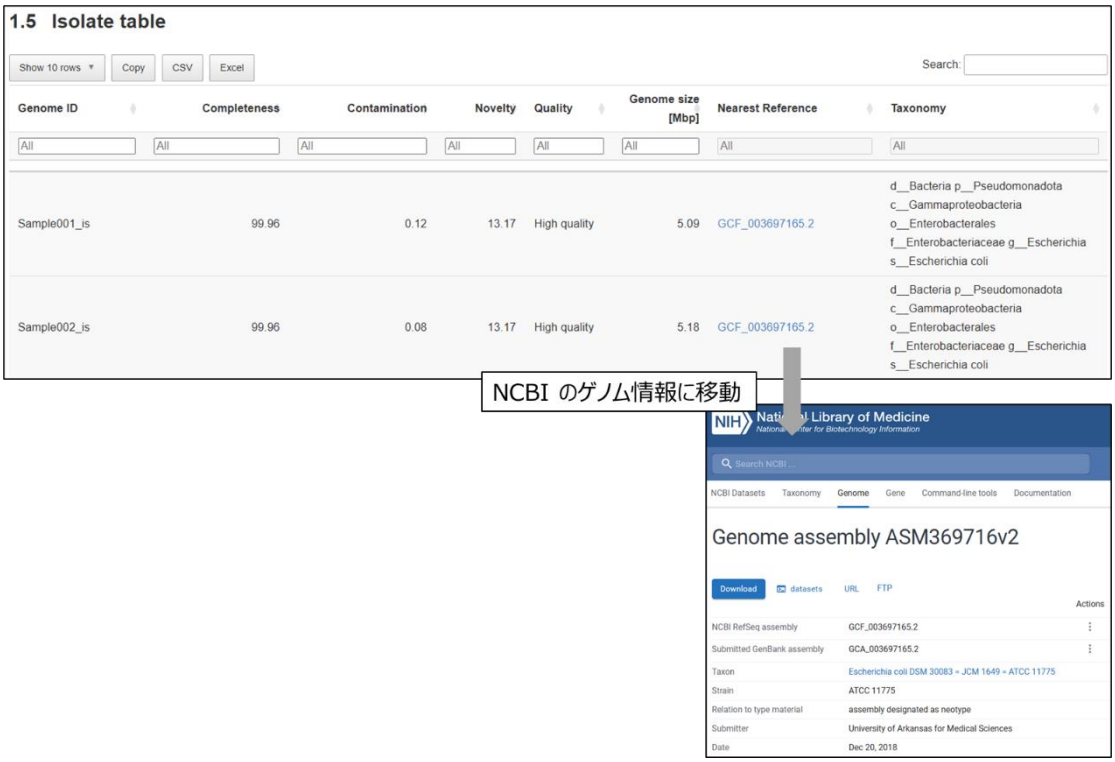
4.4. 系統情報

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



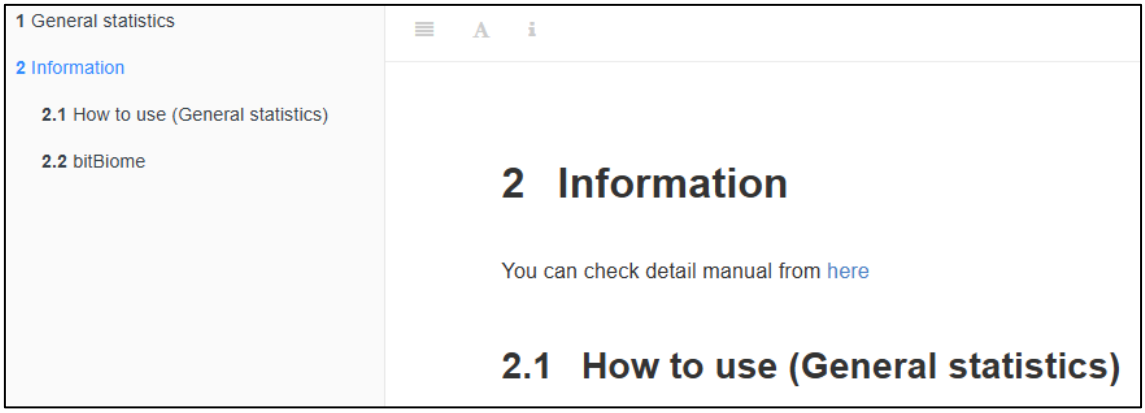
4.5. ゲノム情報

株ゲノムのゲノム完全性、汚染率、新規性、系統情報などの表です。



列の説明

Genome ID	株ゲノム名
Completeness	ゲノム完全性（%）
Contamination	汚染率（%）
Novelty	ゲノム新規性（%）
Quality	株ゲノムの品質
Genome size [Mbp]	株ゲノムの合計塩基数（Mbp）
Nearest Reference	株ゲノムに最も近い参照ゲノム配列（リンク先 NCBI Assembly）
Full taxonomy	生物系統名

5. 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。