

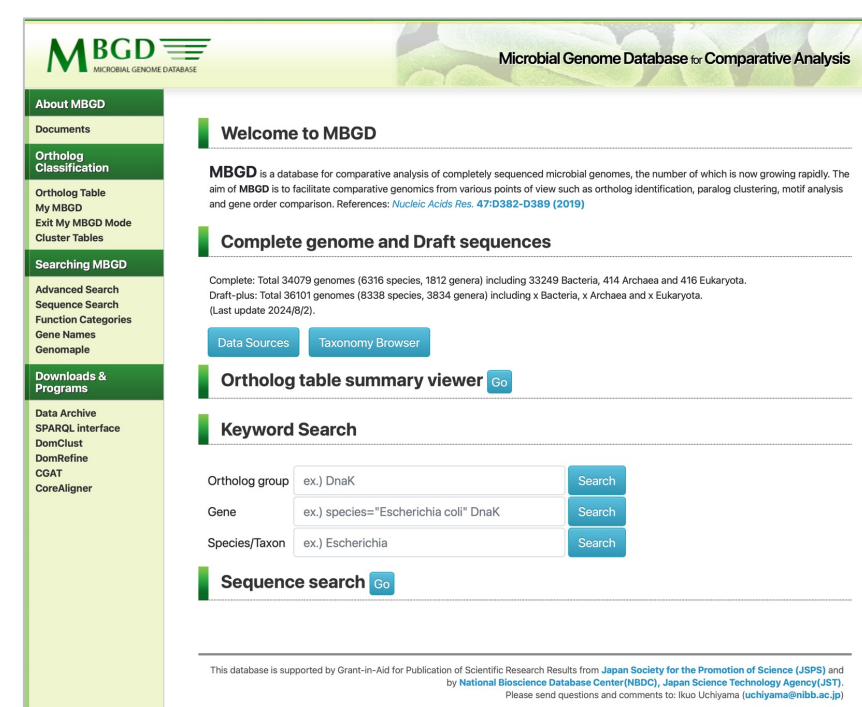


MBGD 2024: ゲノム・メタゲノム機能解析に向けた微生物比較ゲノムデータベースの改良

○内山郁夫（基生研）、三原基広（ダイナコム）、西出浩世（基生研）、千葉啓和（DBCLS）、高柳正彦（ウェブブレイン）、河合幹彦（基生研）、高見英人（JAMSTEC）

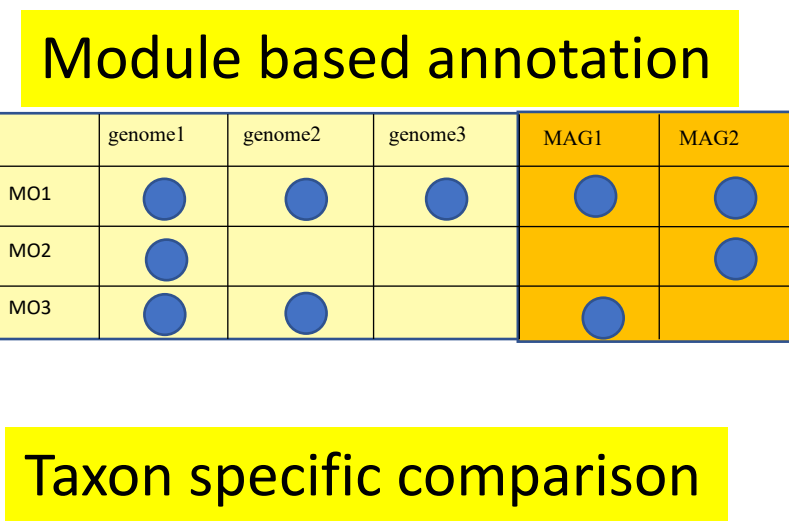
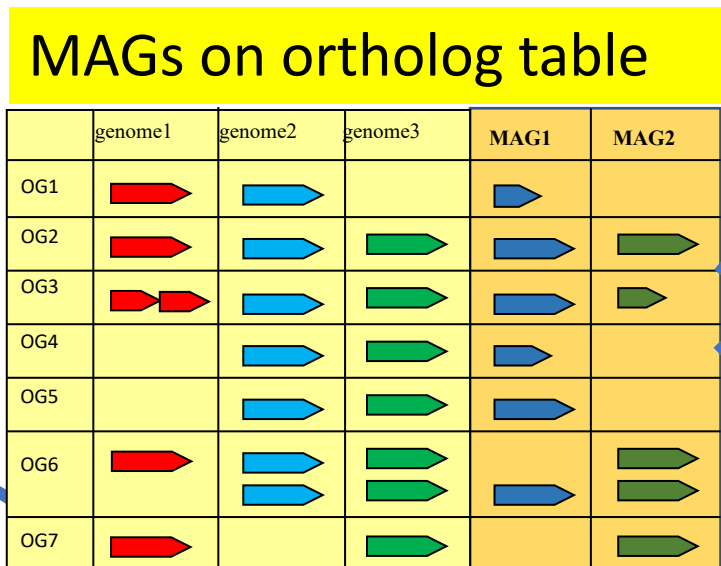
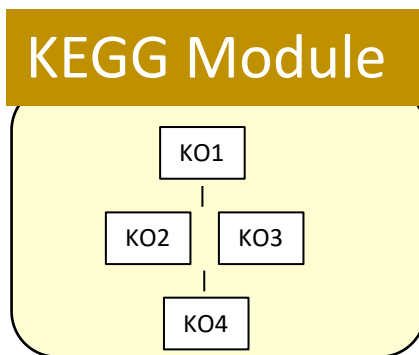
オーソログ解析に基づくゲノムアノテーション

近年、メタゲノムからアセンブルされたゲノム (MAG) の情報が様々な環境から得られるようになり、ゲノム情報から環境中の微生物の生態を明らかにすることが重要になりつつある。我々は、公開されたゲノム情報をオーソログ関係に基づいて整理したデータベースMBGDを構築しており、MAGを含む新規のゲノムのアノテーションを行う目的で改良を行っている。



MBGD 2022
15,397 genomes
4,747 species
1,444 genera
+ 1760 draft/genera

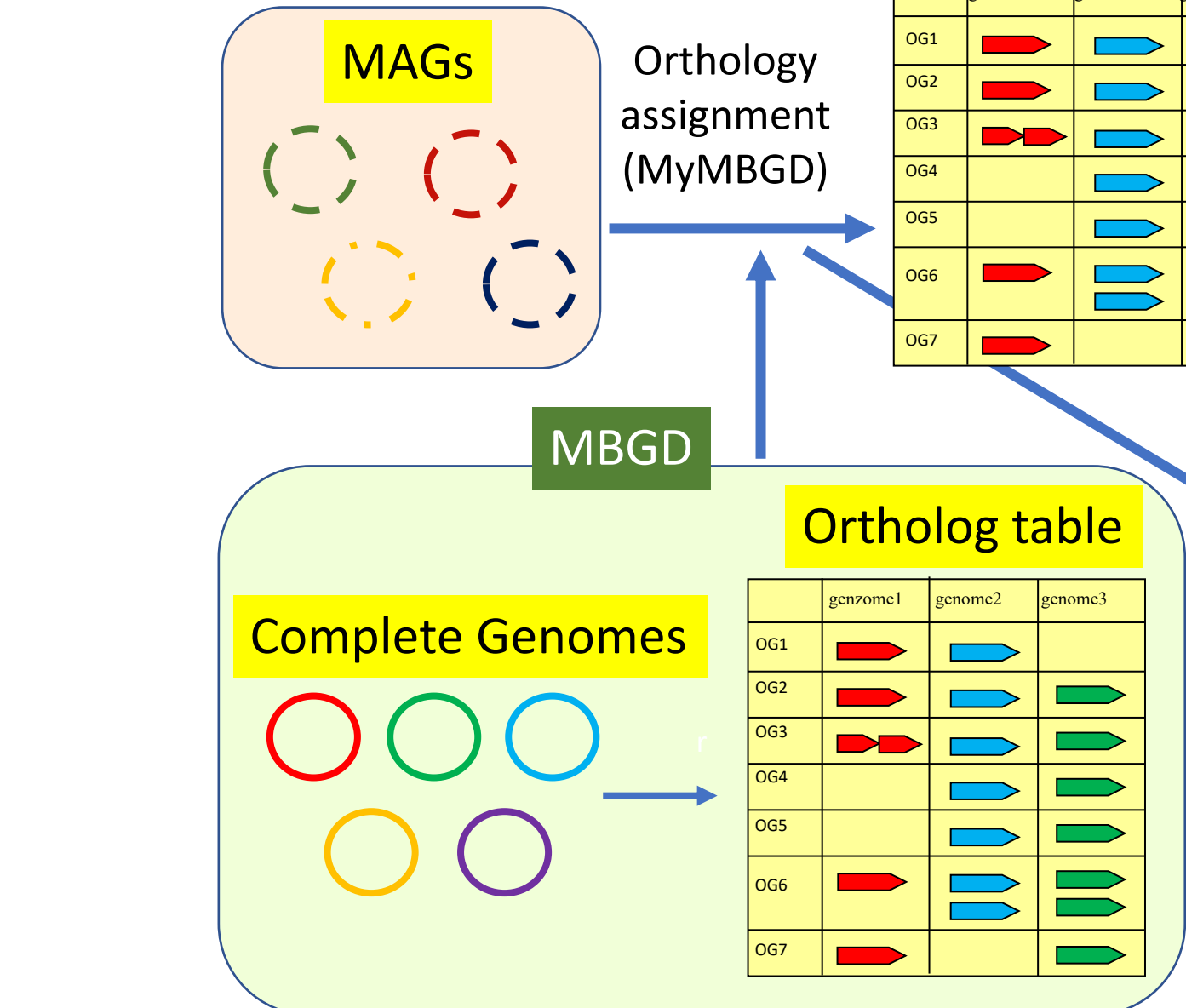
MBGD 2024
34,079 genomes
6,268 species
1,812 genera
+ 2022 draft/genera



Taxon specific comparison

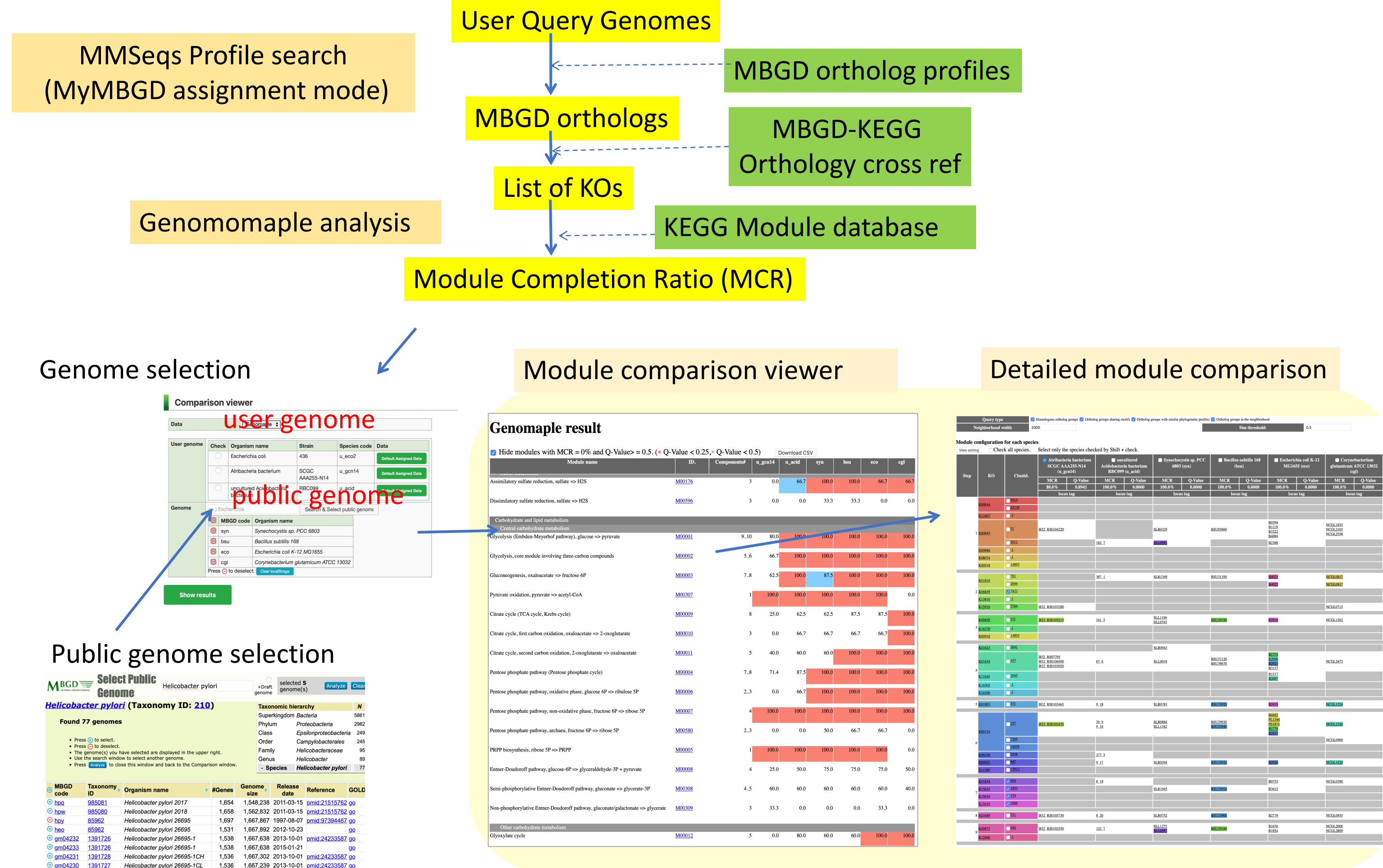
Phylogenetic profile analysis

Unassigned sequences
de novo clustering



Genomableを用いた機能モジュールアノテーション

MBGDには、利用者が登録したゲノムに対してオーソログ解析を行うMyMBGD機能がある。 MyMBGDには assignment modeとclustering modeの2つのモードがあり、ゲノムアノテーションの目的ではassignment modeを用いる。 Assignment mode では、ユーザゲノムの各遺伝子に対して、MMSeqsプロファイル検索で既存のオーソロググループに割り当てた後、Genomableシステム (Takami et al, DNA Res. 23, 467-475) を用いて、KEGG Moduleデータベースで定義されたルールに基づいて機能モジュール単位のアノテーションを行う。



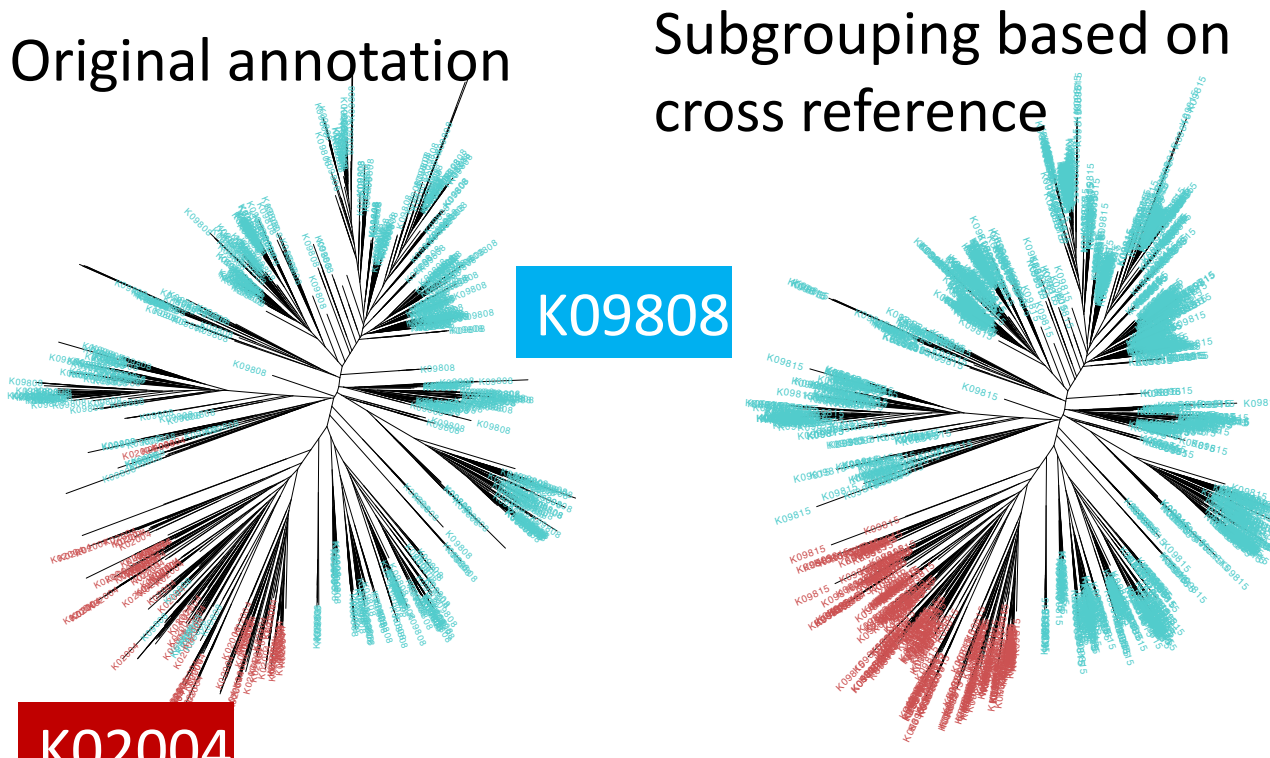
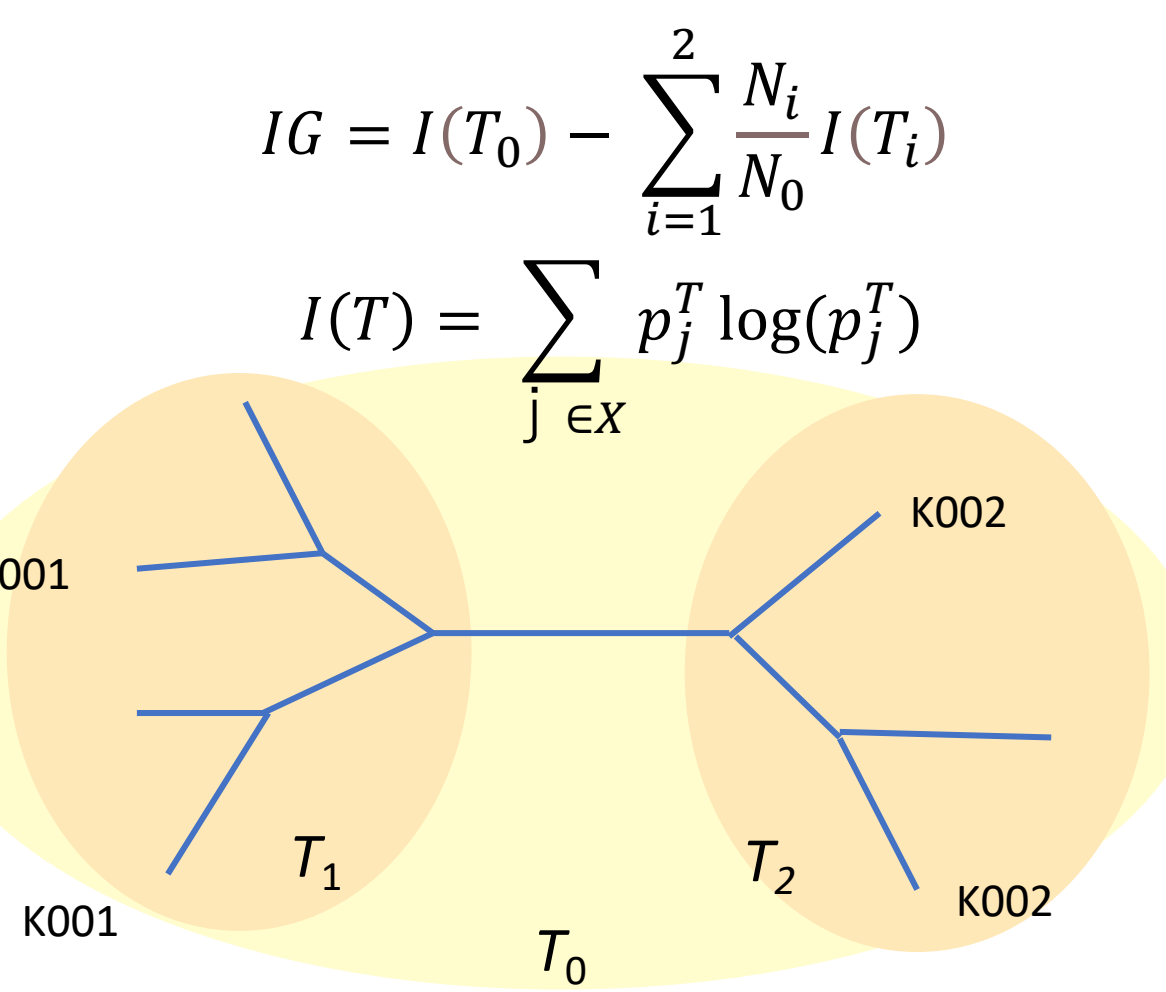
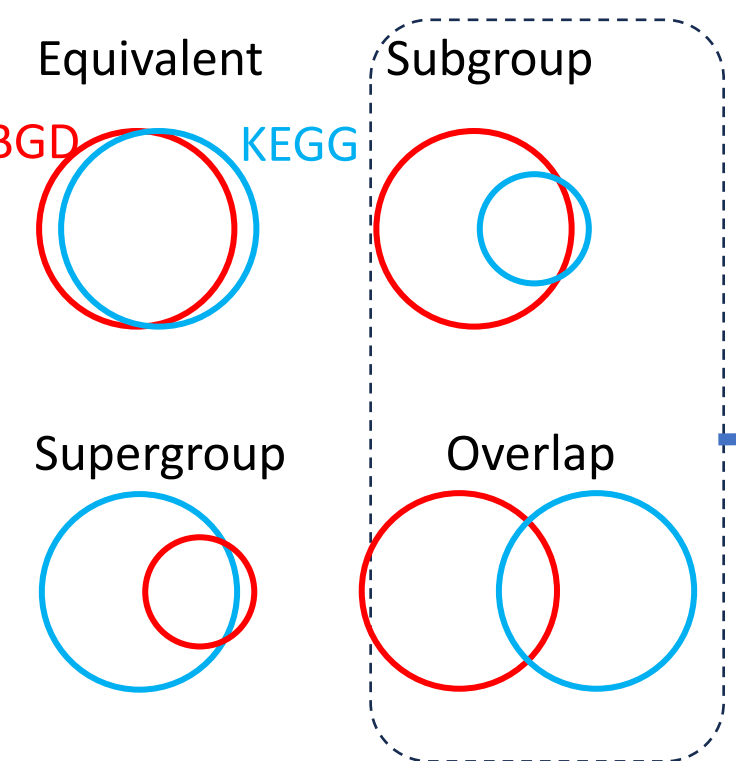
サブグループ化によるMBGD-KEGG クロスリファレンスの改良

Genomable解析において、MBGDのオーソロググループとKEGG Orthology (KO) のグループとの対応づけが必要になる。MBGDのサブグループ化を行うことで、その対応づけの改善を行った。

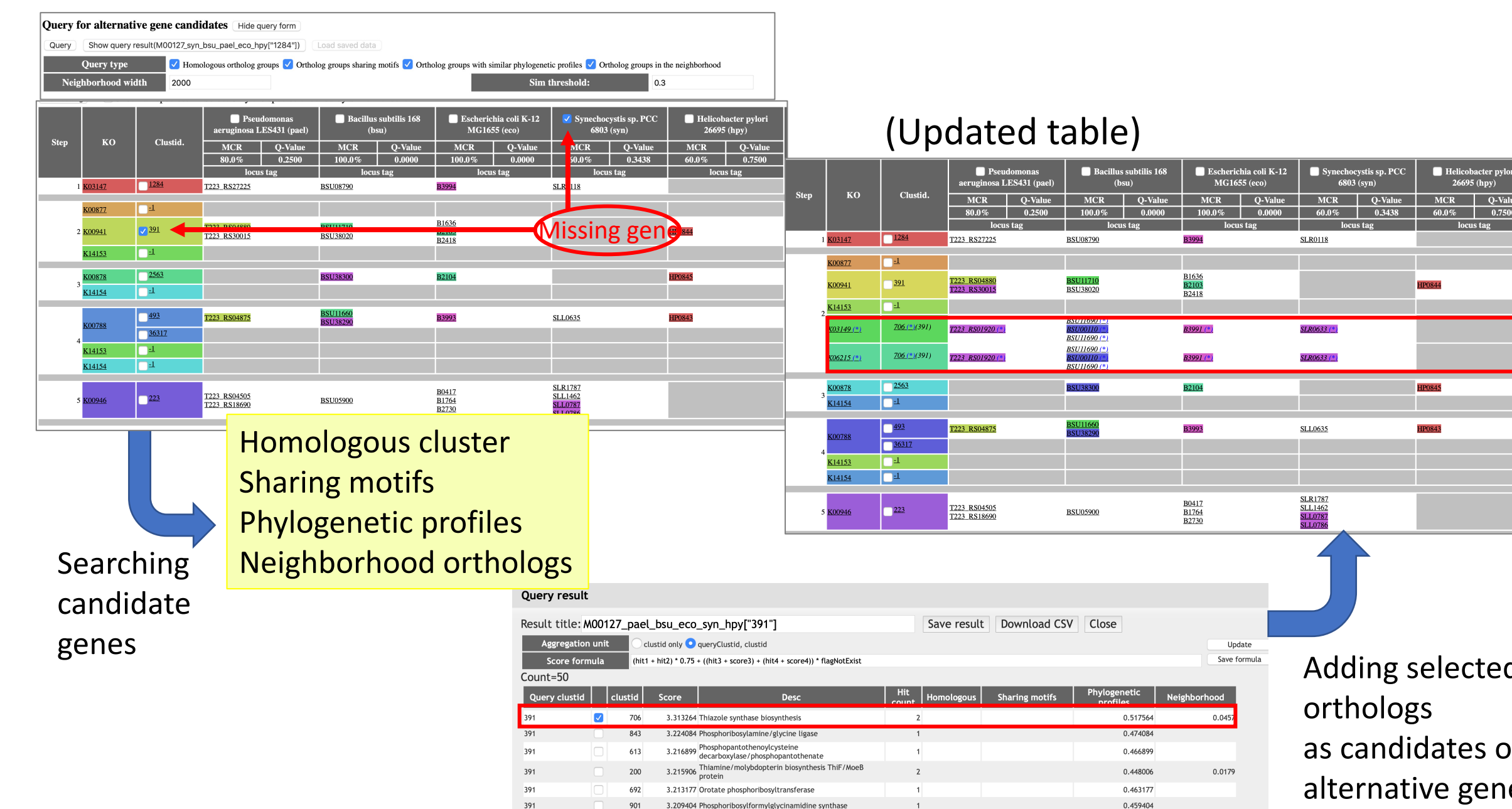
Definition of cross reference

Subgrouping MBGD clusters based on KO

Example: Cluster 606: Lipoprotein-releasing ABC transporter permease subunit LolC

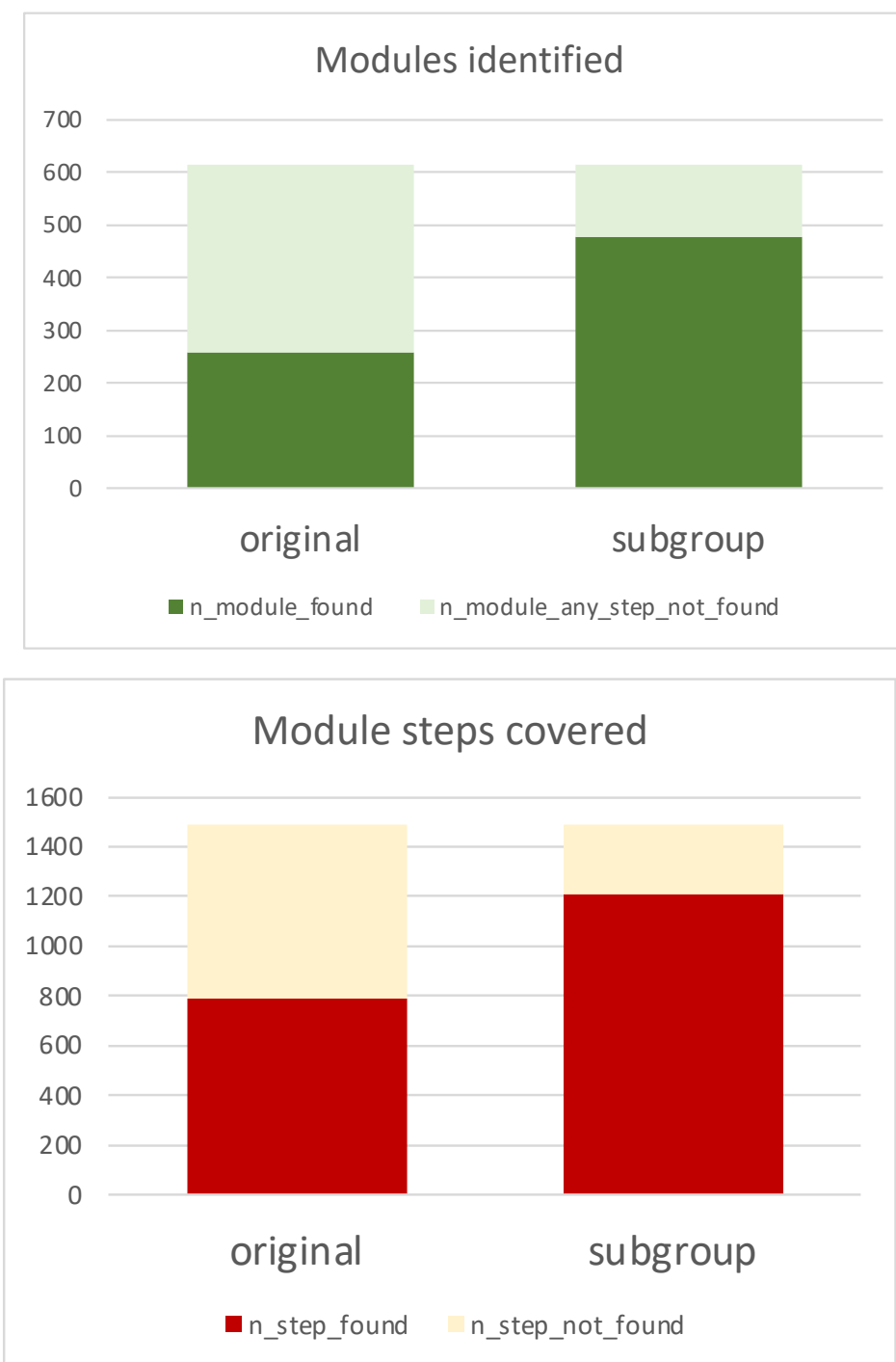
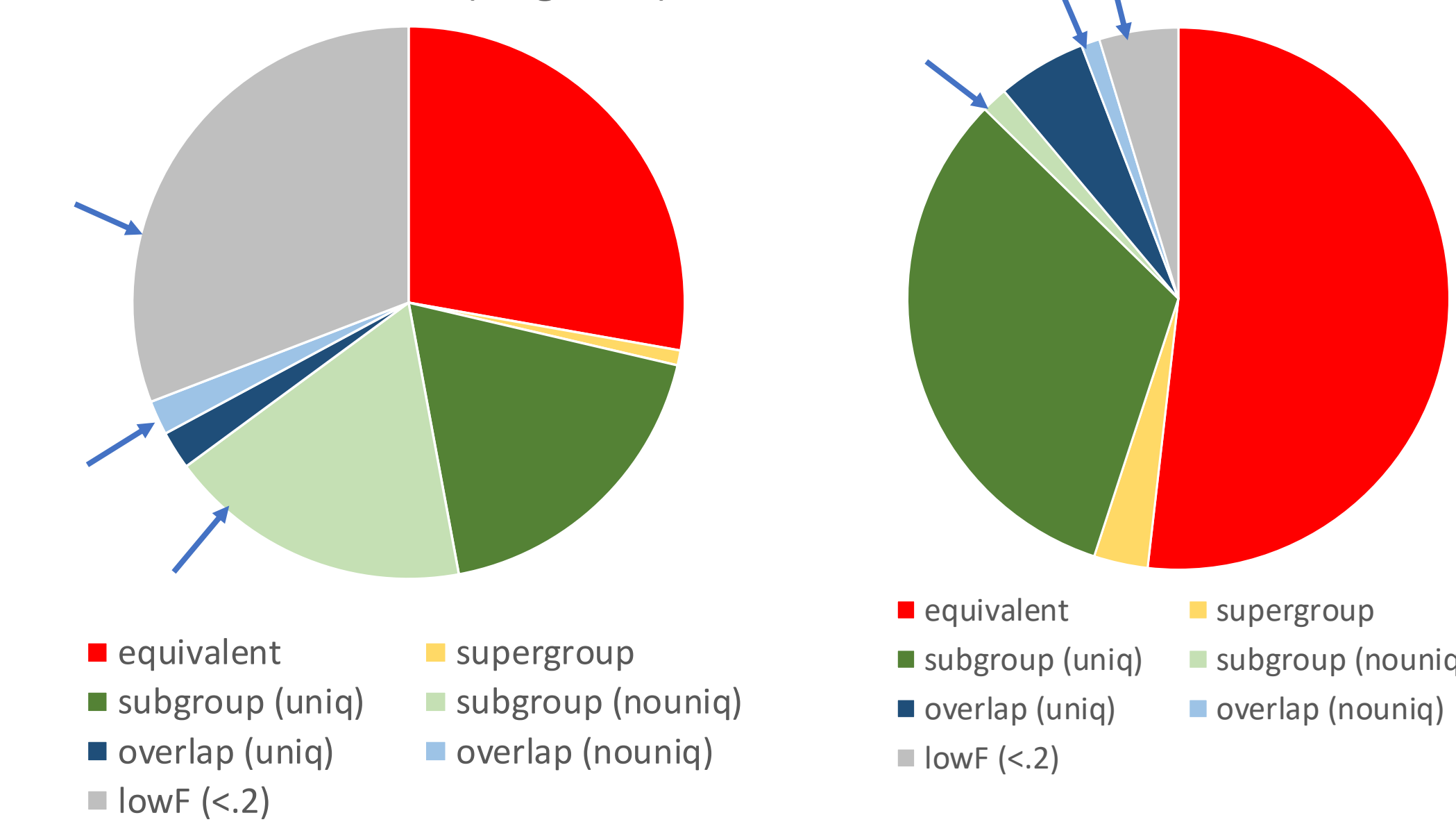


Query interface to search for alternative genes for missing orthologs in a KEGG module



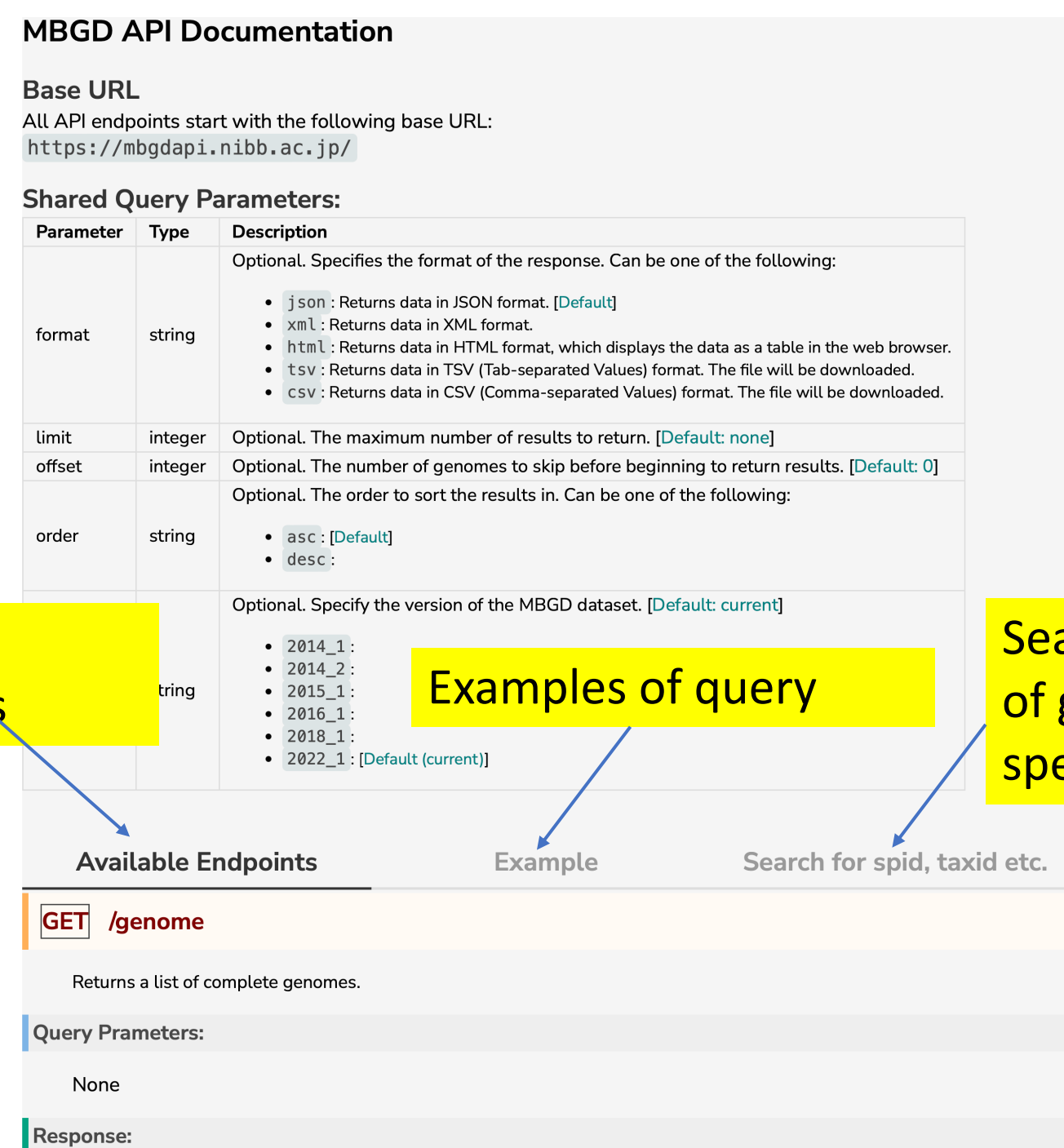
Criss reference (original)

Cross reference (after subgrouping)



MBGD API

Web上のユーザインターフェイスに加えて、MBGDの検索機能をアプリケーションから使うためのWeb APIも公開した。
URL: <https://mbgdapi.nibb.ac.jp/>



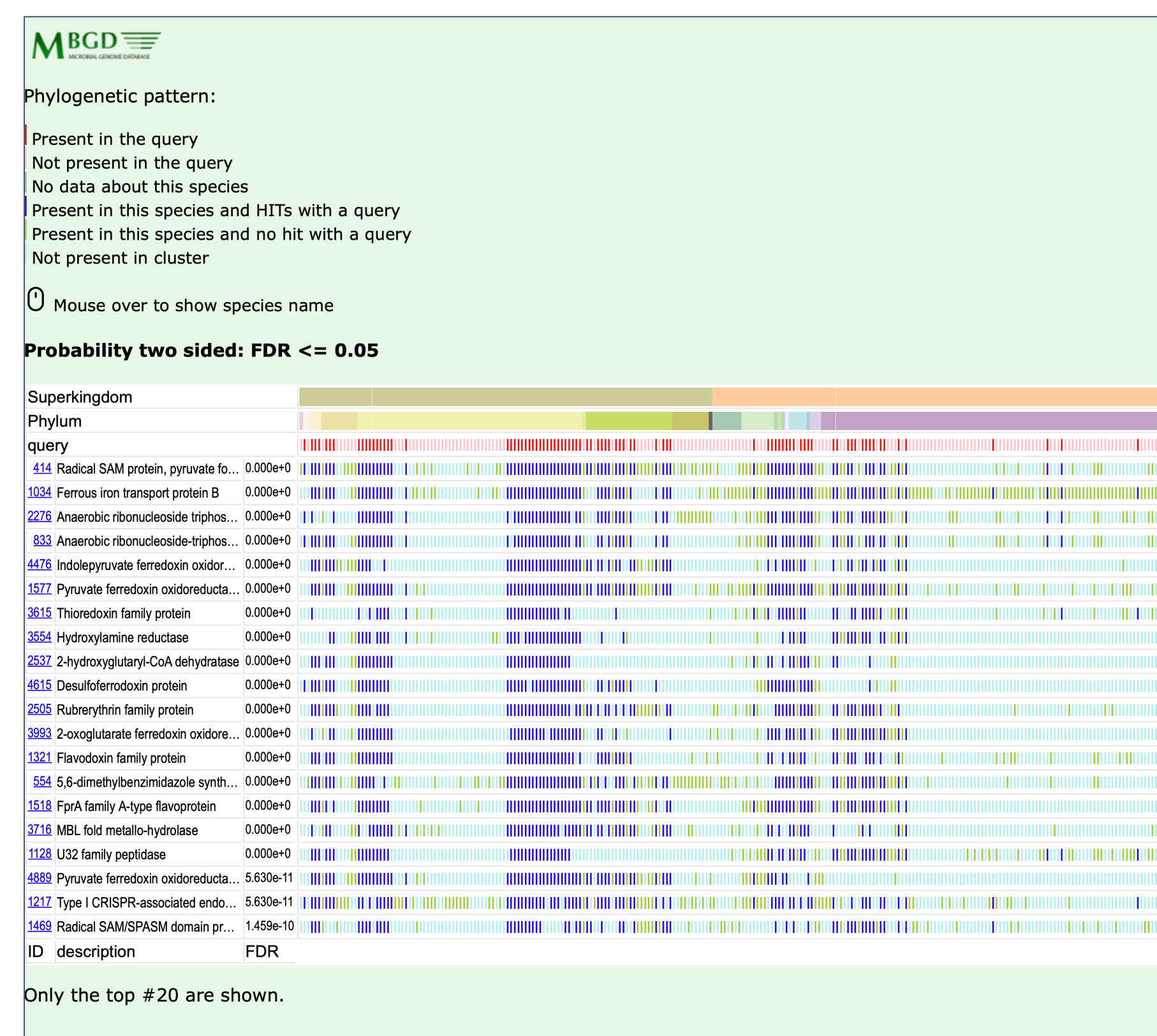
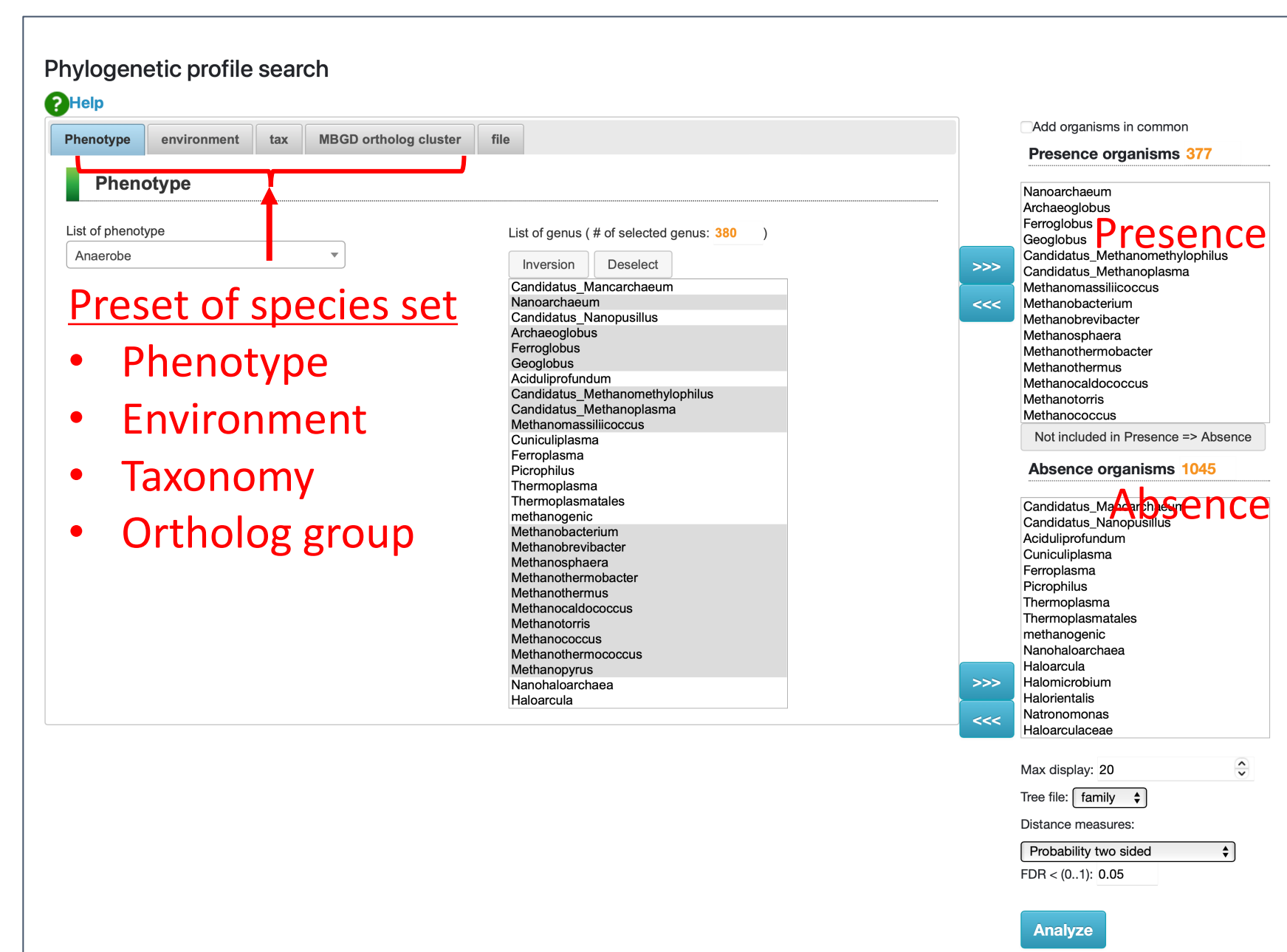
Show usages for available queries

Examples of query

Search for identifiers of genomes etc. specified in a query

系統プロファイル検索機能

系統プロファイル検索機能は、利用者が指定した系統プロファイル(その遺伝子を持つ生物種と持たない生物種の集合)に対して類似した系統プロファイルを持つオーソロググループを検索する機能で、表現型、系統分類、環境などを指定して生物集合を検索し、系統プロファイルを組み立てる機能を有している。



まとめ

- MBGD2024では、34,079ゲノムが登録されている。前回バージョン (MBGD2022)と比べてゲノム数で2.2倍、genus数では1.3倍に増加し、オーソロググループ数も1.3倍に増加した。
- MBGDのオーソロググループを新規ゲノム解析に活用するため、MyMBGD機能を拡張し、各遺伝子をオーソロググループと対応づけた後に、KEGG Moduleの機能モジュールを同定するGenomable機能を実装した。
- この機能を効果的にするため、MBGDのオーソロググループをKEGG Orthologyに合わせてサブグループ化する手続きを作成し、KEGG Moduleへの割り当てを改善した。
- 表現型、系統分類、環境などを指定して生物種集合を検索し、系統プロファイルを組み立てて類似の系統プロファイルを持つオーソロググループを検索する系統プロファイル検索機能を実装した。
- MBGDの検索機能をアプリケーションから使うためのWeb APIを公開した。



Licensed under a Creative Commons 表示4.0 国際ライセンス
© 2024 内山郁夫（基礎生物学研究所）