

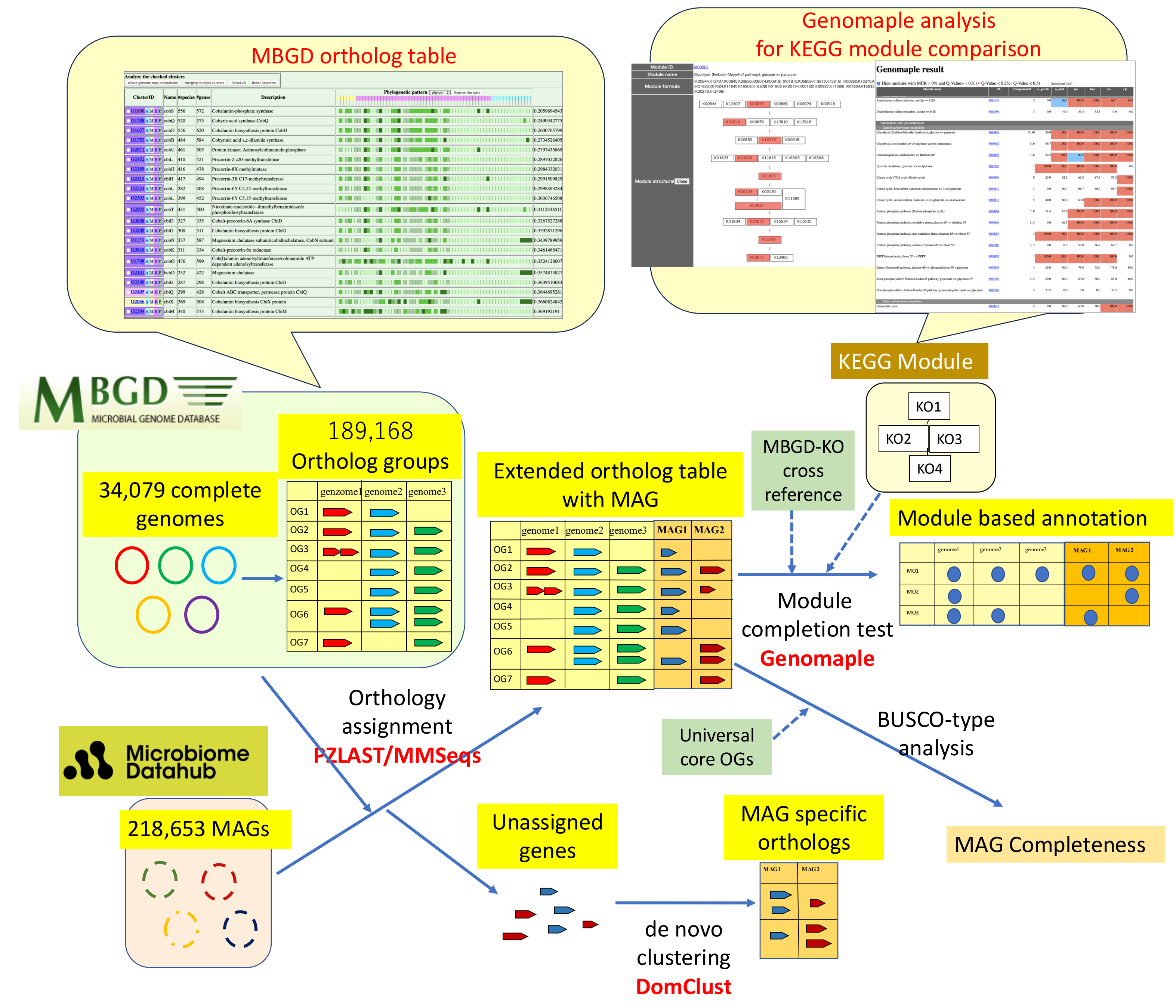


微生物比較ゲノムデータベースMBGDの大規模MAG解析への応用

内山郁夫（基生研）、西出浩世（基生研）、河合幹彦（基生研）、三原基広（ダイナコム）、千葉啓和（DBCLS）、高柳正彦（ウェブブレイン）、森宙史（遺伝研）

オーソログ解析に基づくMAGアノテーション

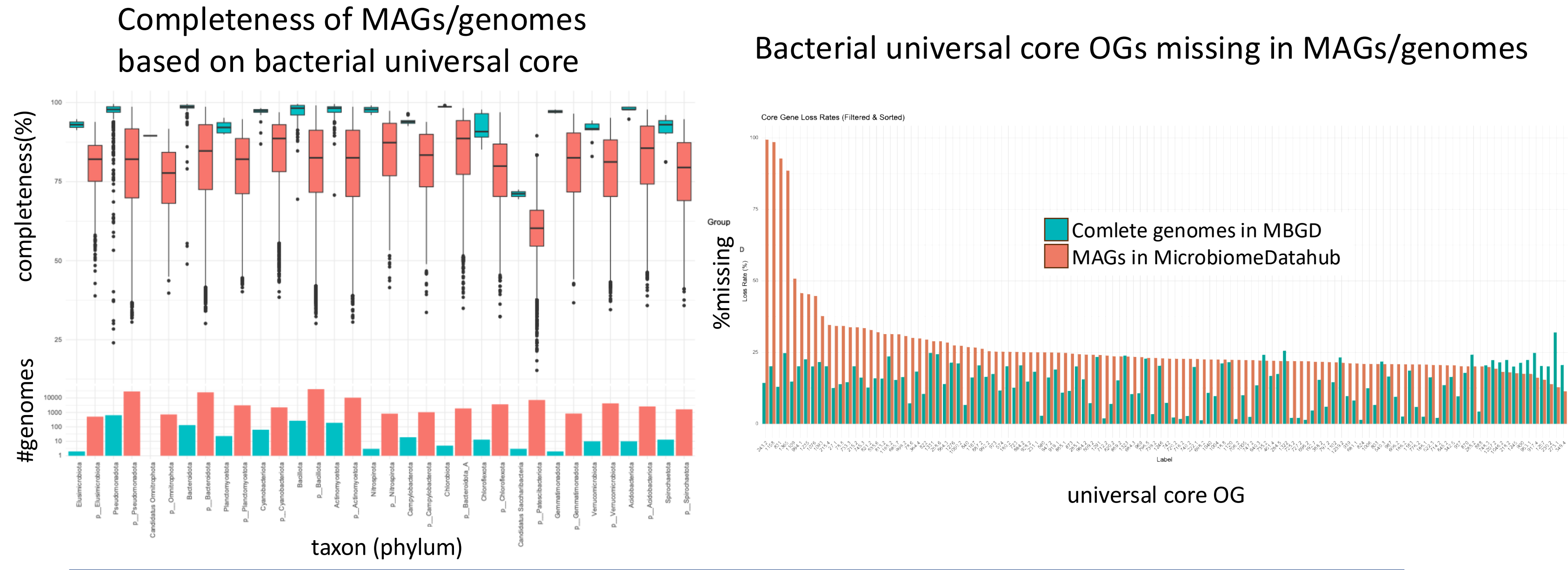
MBGD (http://mbgd.nibb.ac.jp/) はオーソログ解析に基づいて公開された微生物のゲノム情報を整理したデータベースである。MBGDでは、DomClust, DomRefineを用いてオーソロググループを定義した後、オーソロググループごとにMMSeqs検索用の配列プロファイルを作成している。MyMBGDは、このプロファイルを用いてユーザーの新規ゲノム配列を解析する機能で、MBGDオーソロググループに分類した後、Genomableを用いてKEGG module単位のアノテーションを行う。現在、この解析パイプラインを用いて、Microbiome Datahubに登録された21万個のMetagenome Assembled Genome (MAG) データの解析を進めている。



Universal core OGを使ったMAGの完全性解析

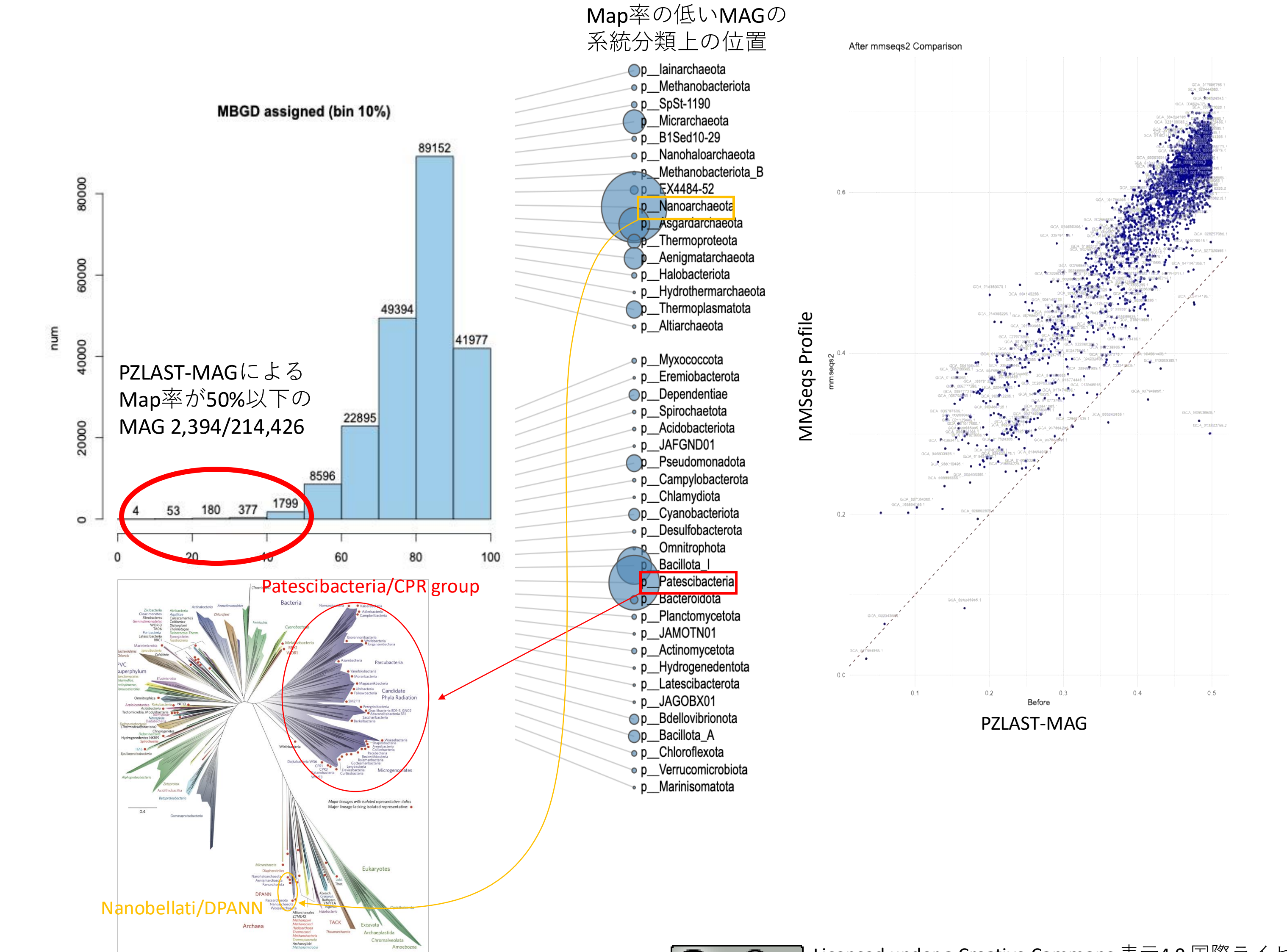
MBGDでは、BacteriaとArchaeaについて、ほとんどの種においてゲノム中にただ一つ存在するオーソロググループ(OG)をUniversal core OGとして定義しており、それを用いてBUSCO/CheckM様の完全性解析を行うことができる。

Universal core OG：90%以上の生物種のゲノムに存在し、ゲノム当たりのパラログ遺伝子数が1.2である。



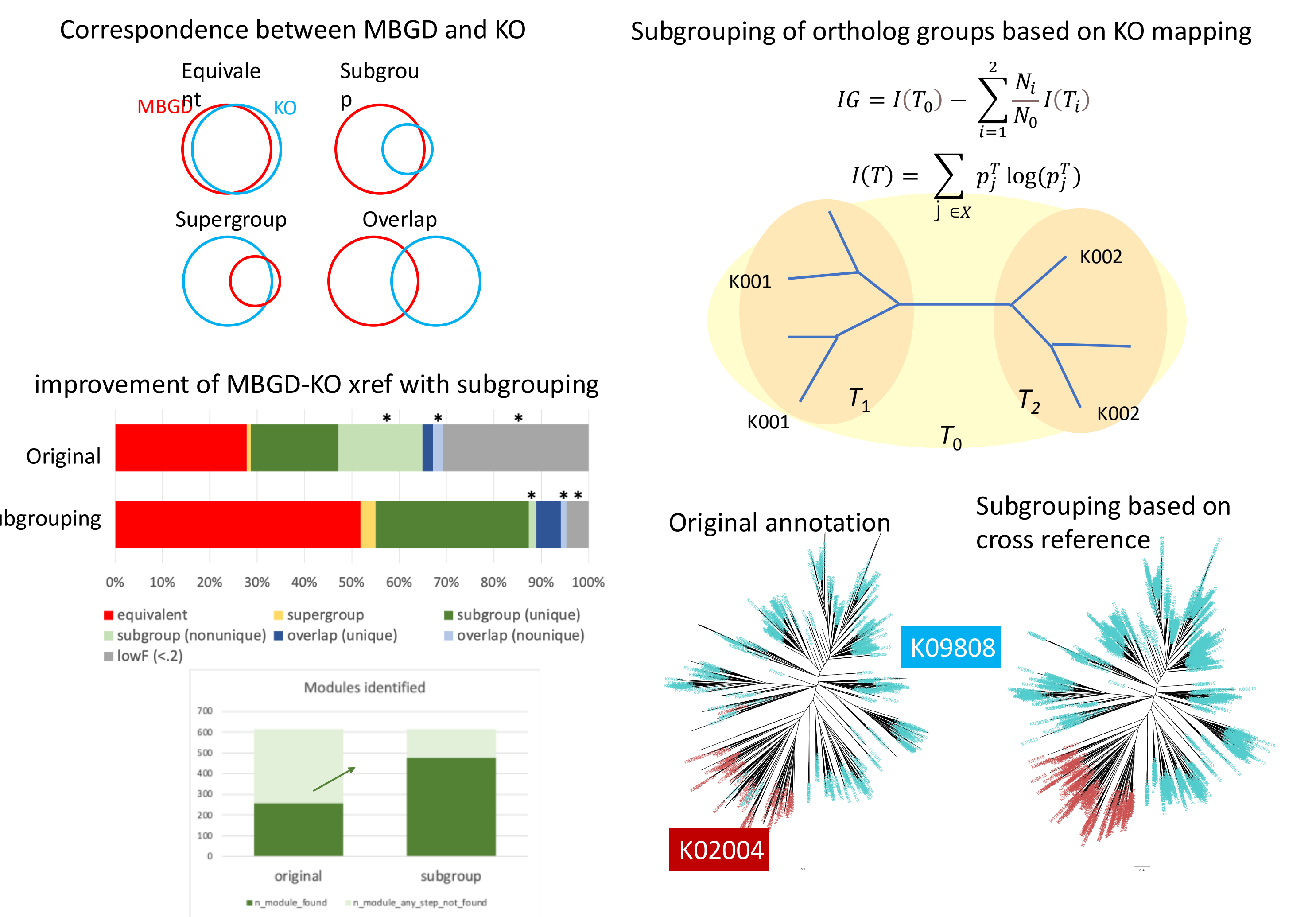
Map率の低いMAGの再解析

多くの場合、PZLASTでMapされた遺伝子の割合は70%を超えるが、一部のMAGではMap率が50%以下と低い値を示した。それらは、PatescibacteriaやNanoarchaeotaなど、一部の系統分類群に偏っている。そのようなMAGについてMMseqs profileで再解析すると、多くの場合MAP率は60-70%程度まで上昇した。



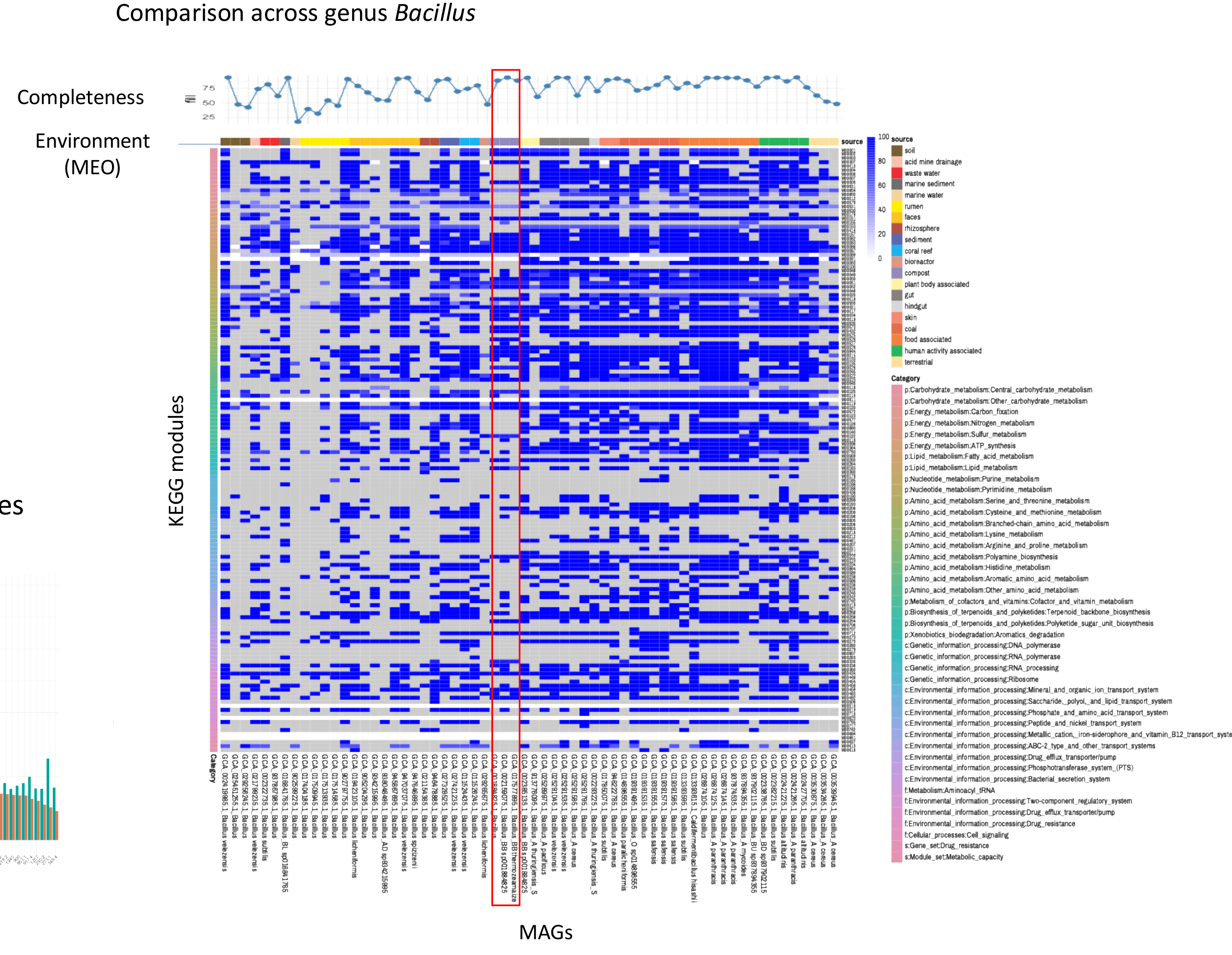
KEGG Ontologyを参照したサブグループ化

MBGDに Genomable解析を適用する際、MBGD Ortholog groupとKEGG Orthology(KO)のクロスリファレンスを参照してKOのリストに変換する。このクロスリファレンスを改善するため、KOを参照したMBGDオーソロググループのサブグループ化を行った。



GenomableによるKEGG Module 解析

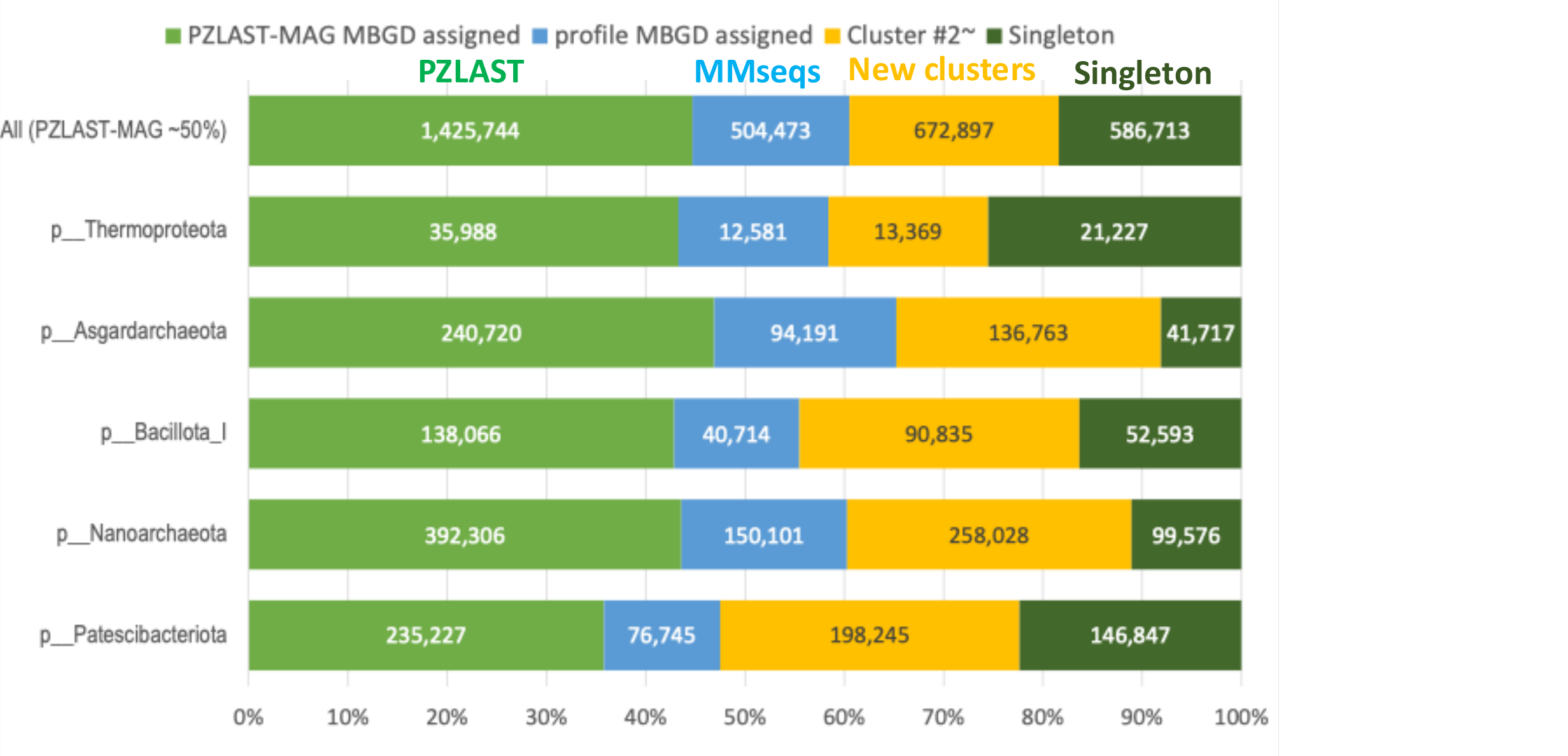
ヒットしたMBGD-OGから、MBGD-KOクロスリファレンスを通してKOをアサインし、Genomableを用いて各MAGのKEGG Module 充足率を計算した。これに基づき、近縁種に分類されたMAGのKEGG module構成を比較するシステムを構築している。



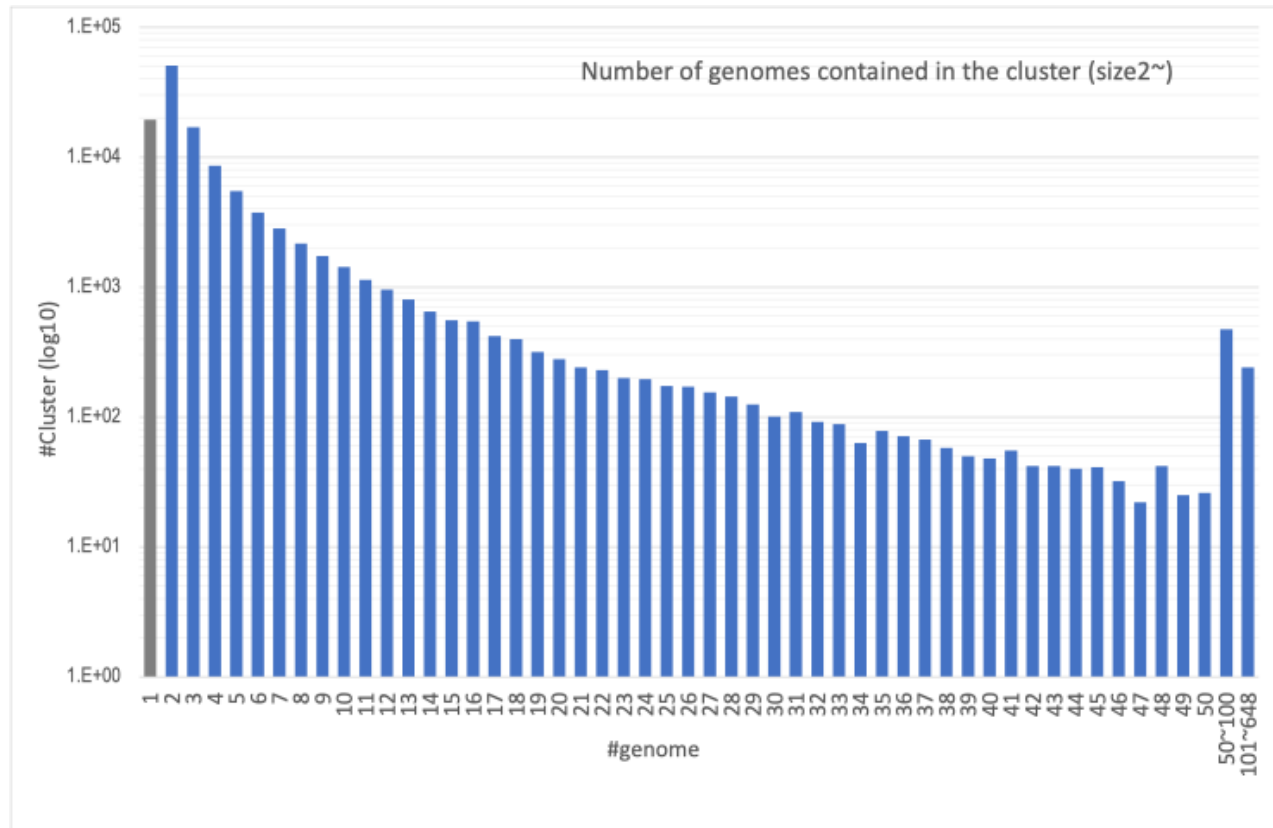
MBGD-OGにマップされない遺伝子の新規クラスタリング

PZLASTによるMap率が50%以下である2,394MAGのうち、MMseqs profileでもヒットしなかった遺伝子に対して、総当たりの相同性検索およびDomClustを用いて新規のオーソロググループを作成した。

低マップ率MAG再解析の結果



新規クラスターのサイズ(MAG数)



新規クラスターのサイズ(phyllum数)

