MRGM : Mouse Reference Gut Microbiome including 1,689 representative prokaryotic species

NETBIOLAB.org
YONSEI UNIVERSITY

*Nayeon Kim[1], Chan Yeong Kim[1], and Insuk Lee[1]**

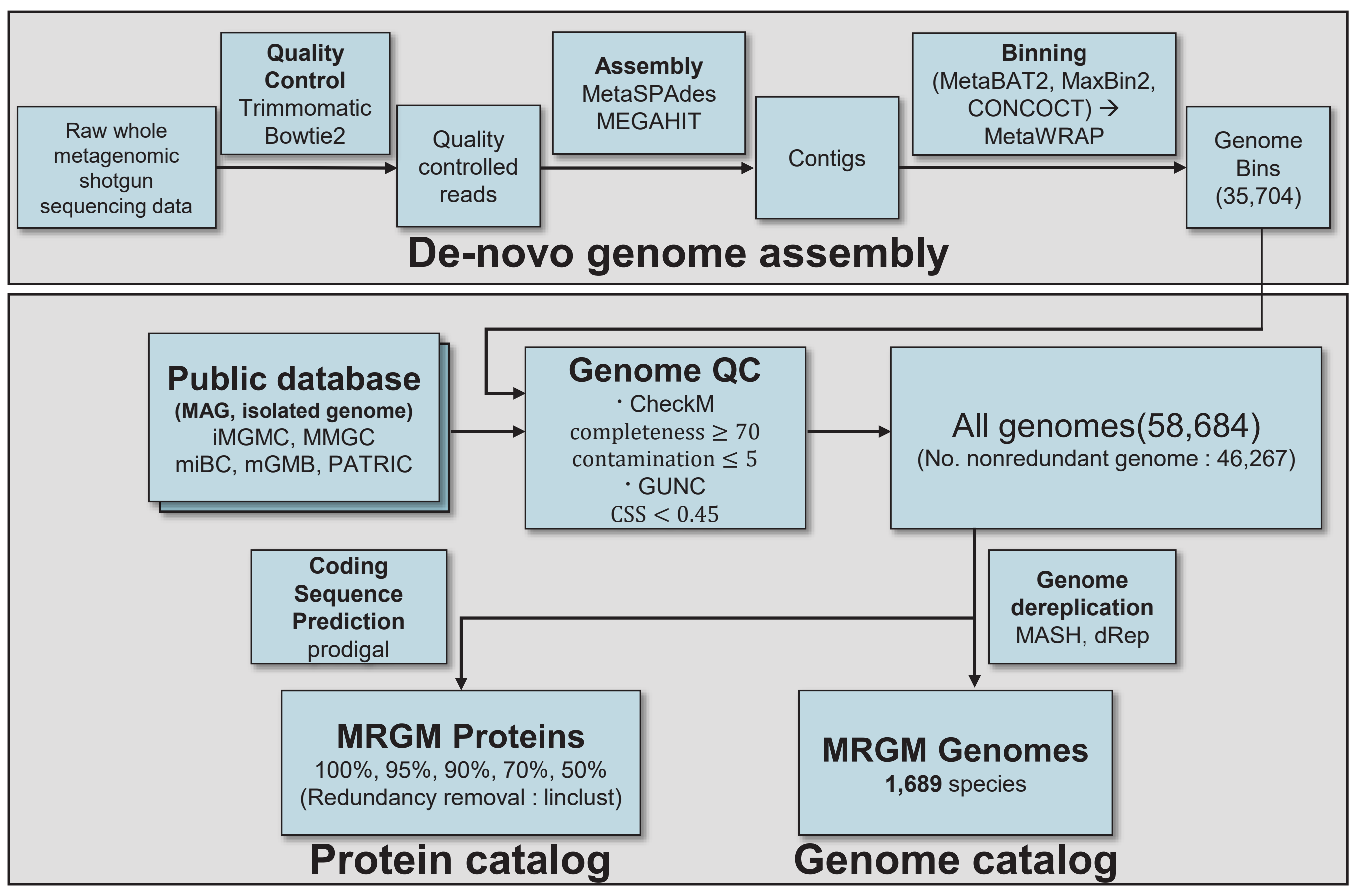
[1] Department of Biotechnology, College of Life Science & Biotechnology, Yonsei University, Seoul 03722, Korea

I. Introduction

- Gut microbiome is associated with many diseases and can be modulated by diet and drugs. Mouse may provide effective *in vivo* model for the study of diet and drug treatment for gut microbiome. For the gut microbiome profiling with high resolution and coverage, a comprehensive reference of gut microbial genomes is required.
- We constructed mouse reference gut microbiome(MRGM) which include newly assembled genomes from more than 800 metagenomes and assembled genomes from previous databases.
- There are several improvements in our mouse reference gut microbiome compared with previous databases. First, we included metagenome-assembled genomes(MAGs) from many samples. Second, we filtered the assembled genomes not only by conventional criteria, completeness and contamination, but also by genome chimerism. Third, we chose genomes with 70% completeness for this database and found that such threshold could achieve an optimal trade-off between accuracy or reference genomes and coverage of sequencing reads with taxonomic classification.
- The data will be available via a web server (www.mbiomenet.org/MRGM/).

II. Results

1) Pipeline



Quality Control
Trimmomatic
Bowtie2

Raw whole metagenomic shotgun sequencing data → Quality controlled reads

Assembly
MetaSPAdes
MEGAHIT

Quality controlled reads → Contigs

Binning
(MetaBAT2, MaxBin2, CONCOCT) → MetaWRAP

Contigs → Genome Bins (35,704)

De-novo genome assembly

Public database
(MAG, isolated genome)
iMGMC, MMGC
miBC, mGMB, PATRIC

Genome QC
· CheckM
completeness ≥ 70
contamination ≤ 5
· GUNC
CSS < 0.45

All genomes(58,684)
(No. nonredundant genome : 46,267)

Coding Sequence Prediction
prodigal

Genome dereplication
MASH, dRep

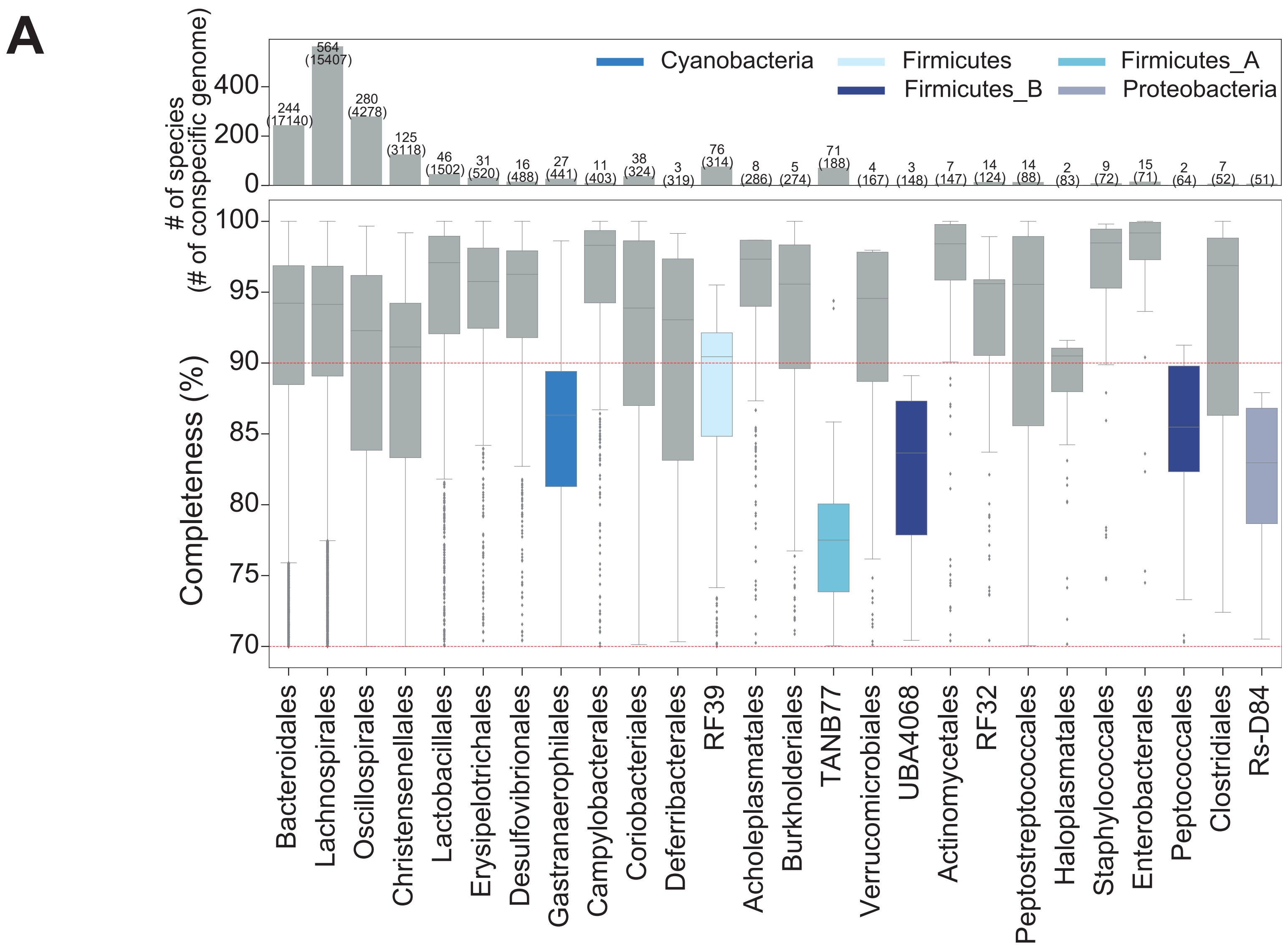
MRGM Proteins
100%, 95%, 90%, 70%, 50%
(Redundancy removal : linclust)

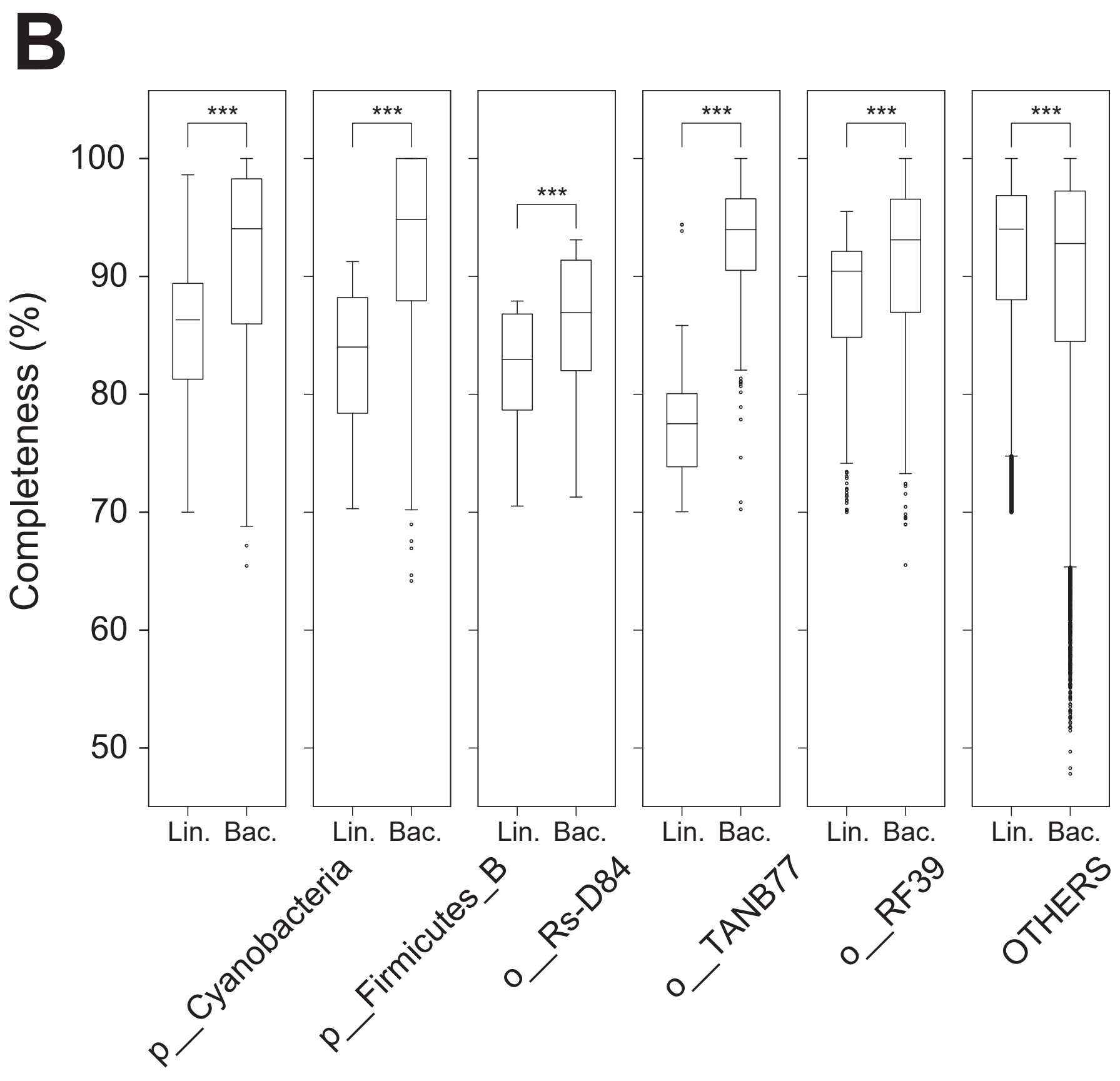
MRGM Genomes
1,689 species

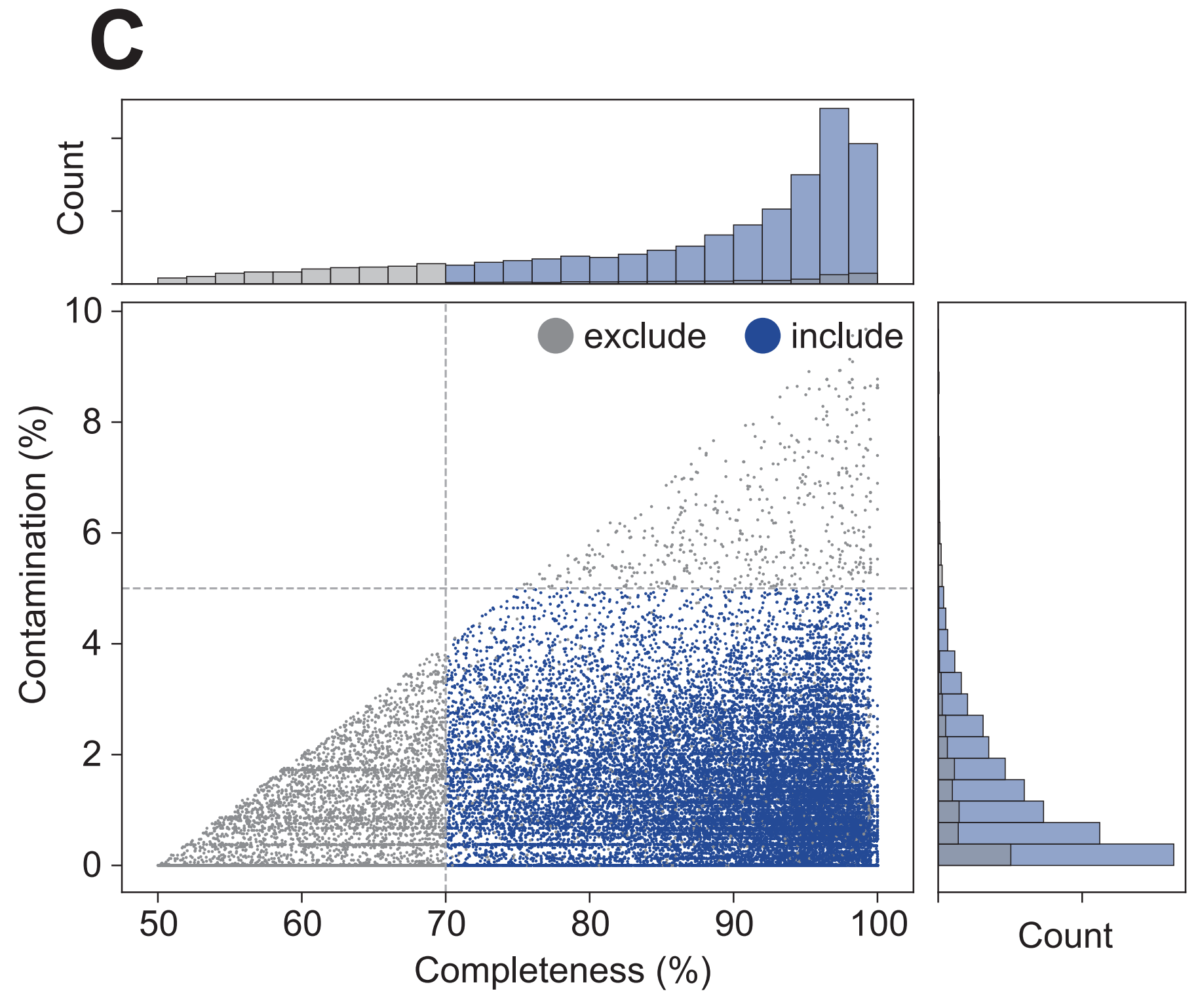
Protein catalog

Genome catalog

2) Underestimated lineage by checkM completeness

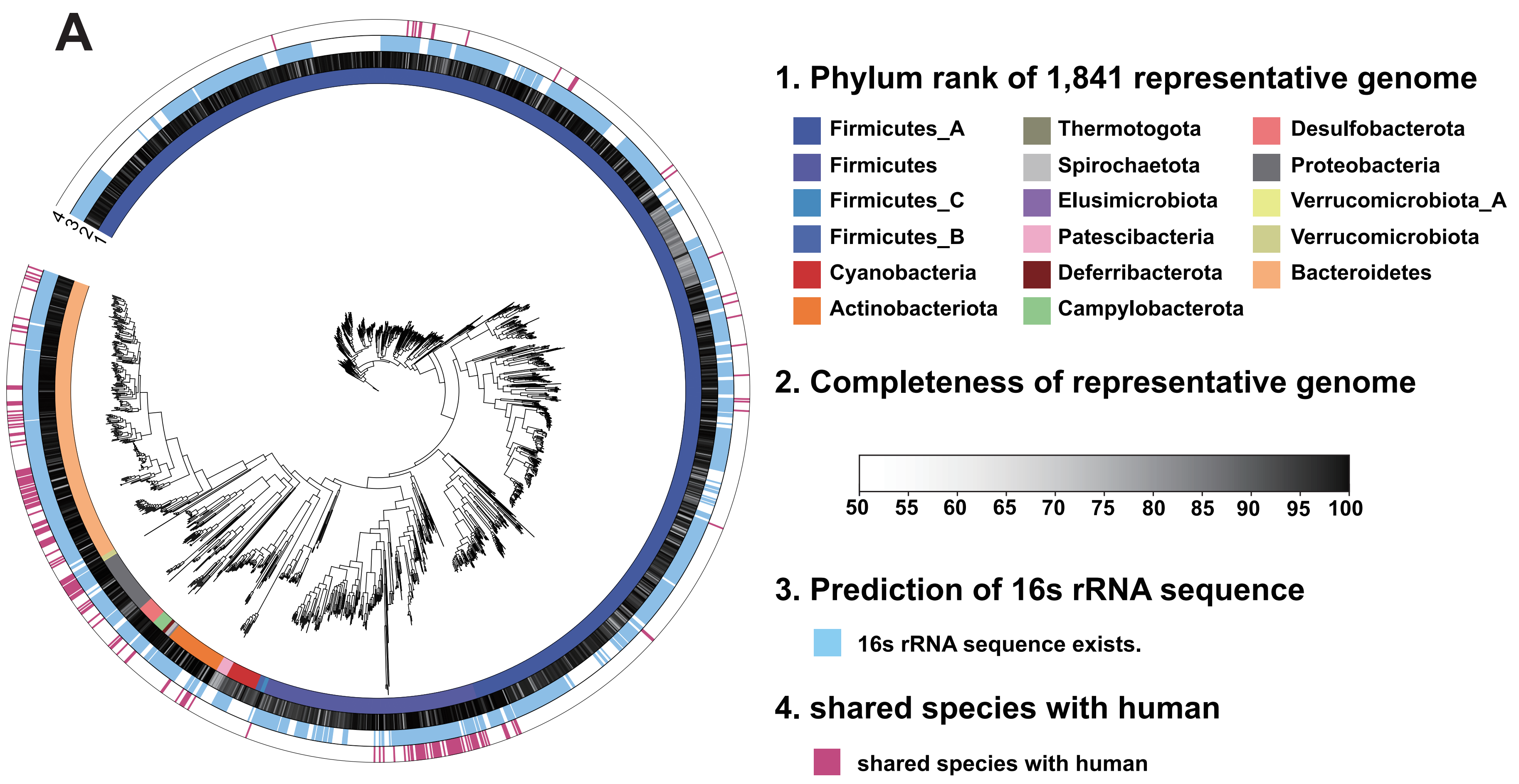






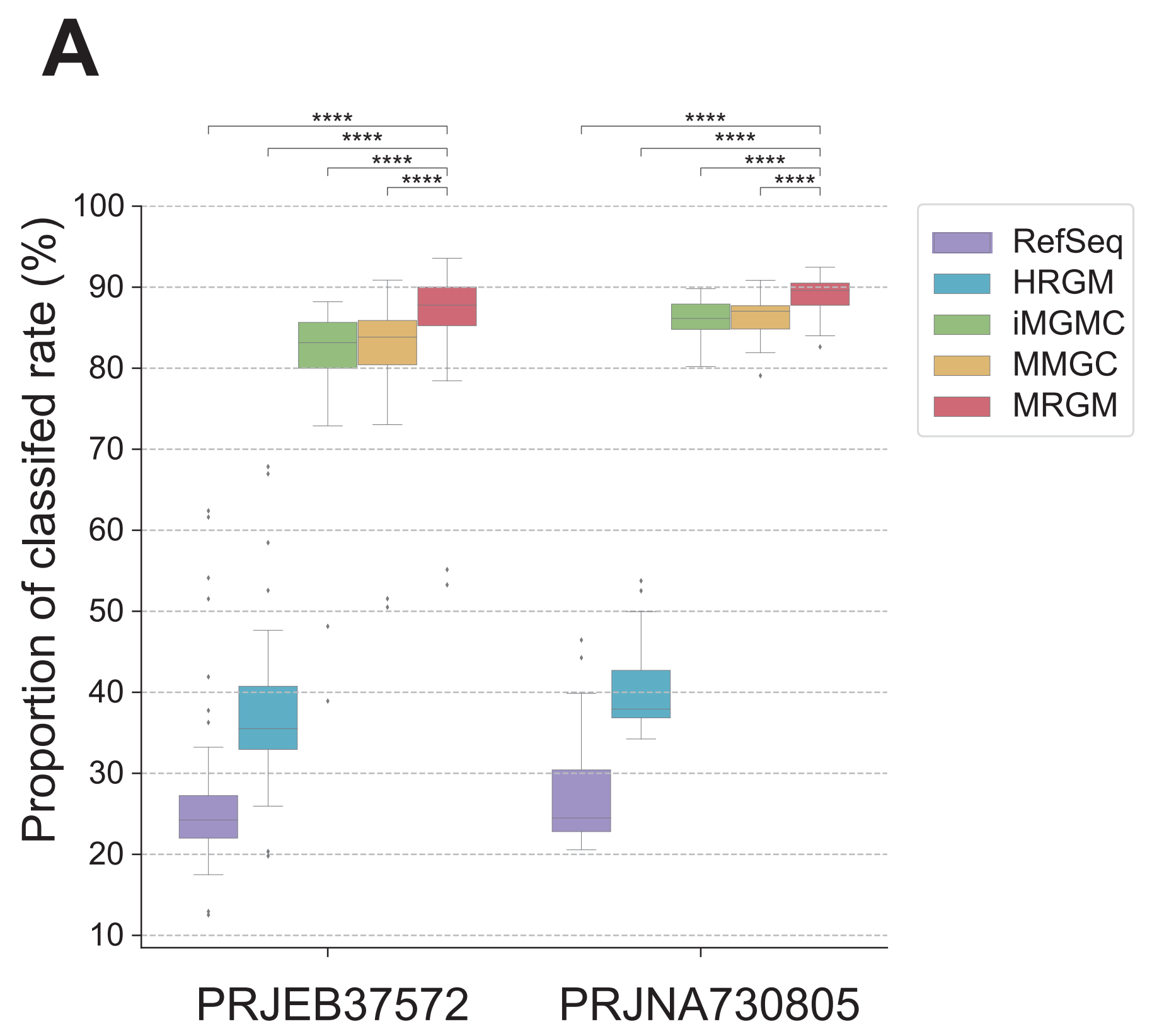
- Some lineages have overall low completeness measured with checkM lineage markers (Fig. A).
- Completeness of underestimated lineages was significantly improved when measured with bactial marker genes (Fig. B).
- Therefore, we used genomes with completeness ≥ 70.
- checkM results of all 35,704 newly assembled MAGs. MAGs with completeness ≥ 70, contamination ≤5, and passing GUNC[4] filtering are included (Fig. C).

4) Phylogenetic tree of MRGM genomes



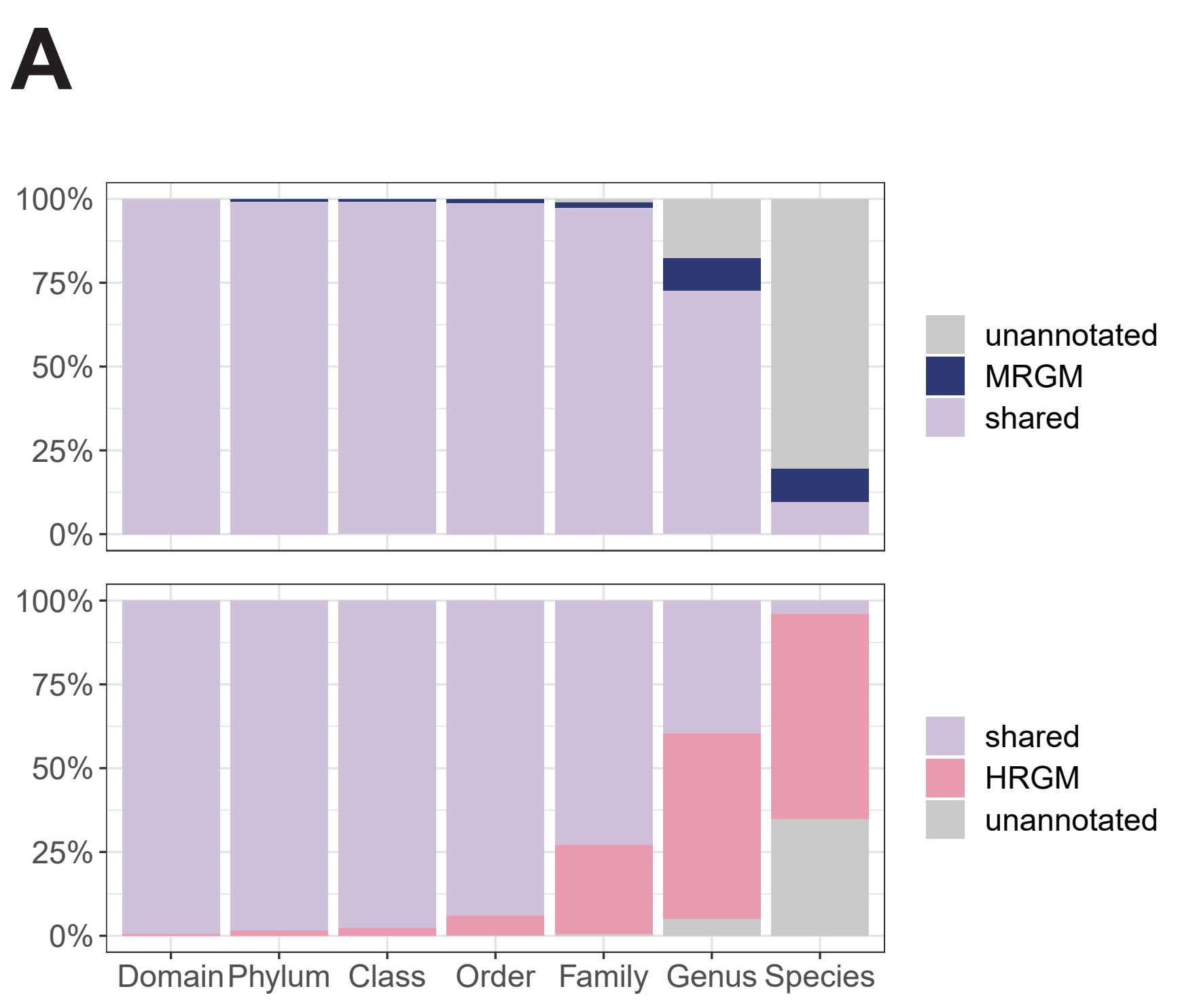
- The phylogenetic tree of 1,689 MRGM representative genomes (Fig. A).
- More than half was Firmicutes_A. The most common phylum after Firmicutes_A was Bacteroidetes.
- We identified that some phyla, such as Bacteroidetes, have relatively more shared species. However, for Firmicutes_A, only a small portion of species is shared. In Firmicutes, species found only in mice and species shared with humans(HRGM) were clearly distinguished.

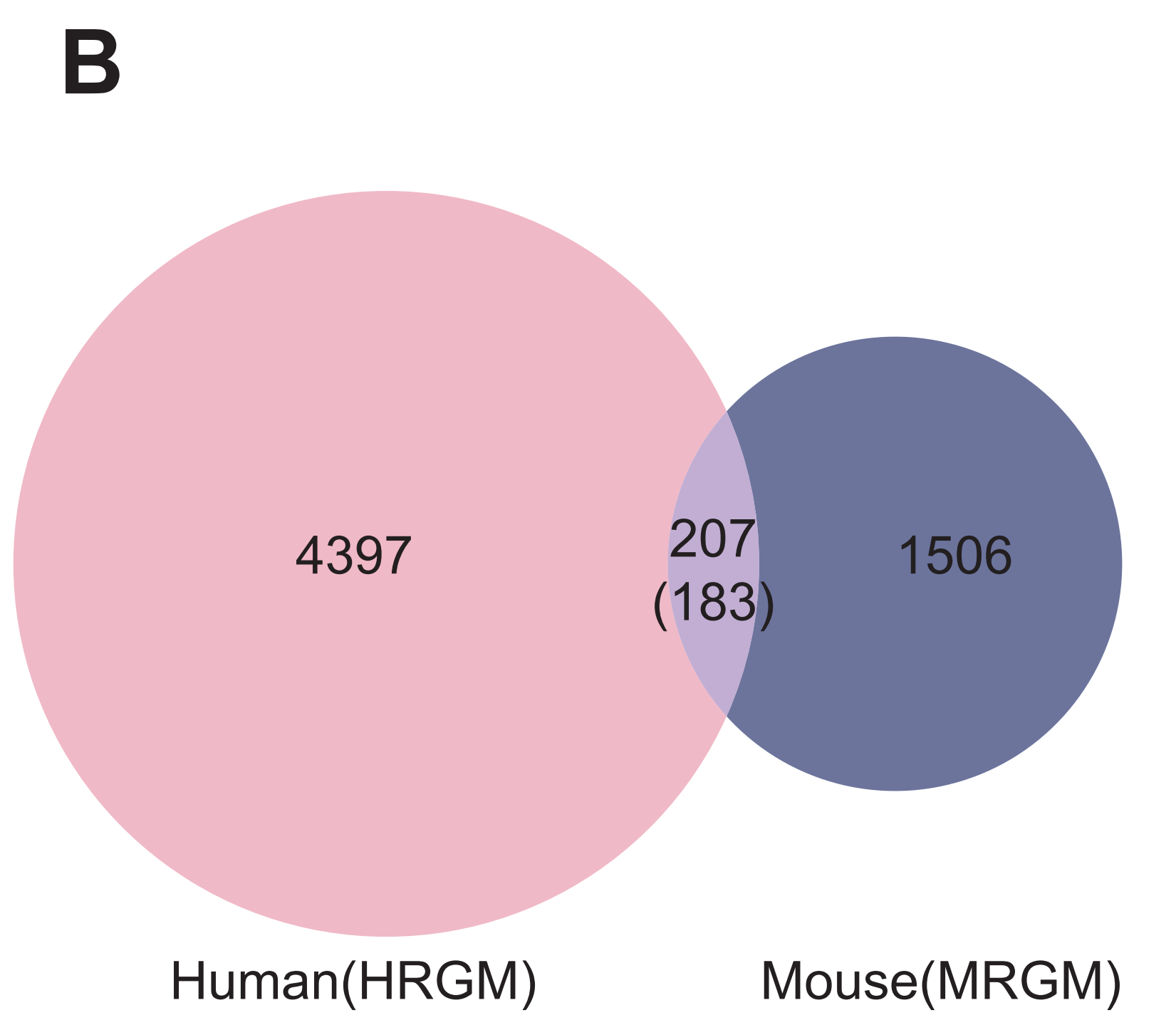
5) Taxonomic profiling result using MRGM



- Taxonomic classified rate of independent validation set that was excluded in every databases. The classified rate was significantly improved compared with other mouse genome catalogs, iMGMC and MMGC [1,2](Fig A).

6) Comparison mouse human gut microbiome(HRGM)[3] and mouse gut microbiome(MRGM)





- Top : MRGM, Bottom : HRGM(Human Reference Gut Microbiome)[4]. The lower the taxonomic level, the lower the shared percentage. In genus level, most mouse gut genomes were shared with humans (Fig. A).
- Only 10% of MRGM overlapped with HRGM in species (Fig. B).

Reference

1) Till R. Lesker et al., An Integrated Metagenome Catalog Reveals New Insights into the Murine Gut Microbiome. Cell Reports (2020)
2) Benjamin S. Beresford-Jones et al., Functional and taxonomic comparison of mouse and human gut microbiotas using extensive culturing and metagenomics. BioRxiv (2021)
3) Chan Yeong Kim et al., Human reference gut microbiome comprising 5,414 prokaryotic species, including newly assembled genomes from under-represented Asian metagenomes. Genome Med (2021)
4) Orakov A et al., GUNC: Detection of Chimerism and Contamination in Prokaryotic Genomes. Benome Biol (2021)