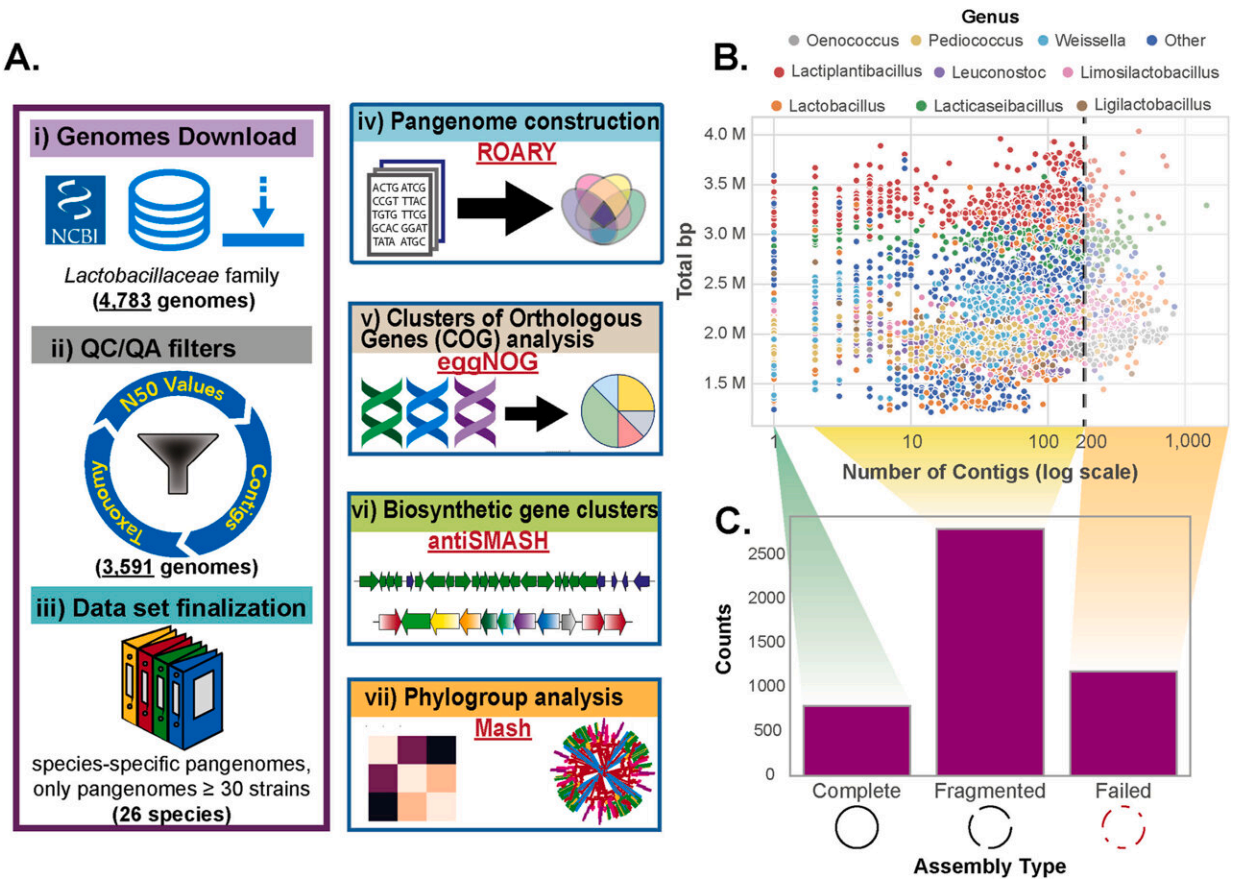
Comprehensive pan-genome analysis of lactic acid bacteria unveils new avenues for food industry and health care

October 10 2023, by Omkar S. Mohite



Credit: Workflow for data curation and subsequent pangenome analysis. A) Graphical summary of workflow and methodologies used. Genomes were downloaded from NCBI and filtered using N50 values, contigs, and GTDB taxonomy as quality control metrics. 26 species had ≥30 genomes available and were then selected for pangenome generation and analysis. The pangenome

generation step outputs a gene presence/absence matrix, which we denote with P. This binary matrix lists gene as rows and strains as columns, with an entry of 1 if a gene is present in a given strain, and 0 if a gene is absent in a given strain. Information about the pangenome, phylogeny, secondary metabolites, clusters of orthologous genes, and Mash clustering analysis was used as a basis for characterizing the Lactobacillaceae family. B) Scatter plot showing the relationship between the number of contigs (log scale) and total base pair length of the 4783 unfiltered Lactobacillaceae genomes, which were initially downloaded. The color of the dots represents various genera. Genomes that passed the quality control step had