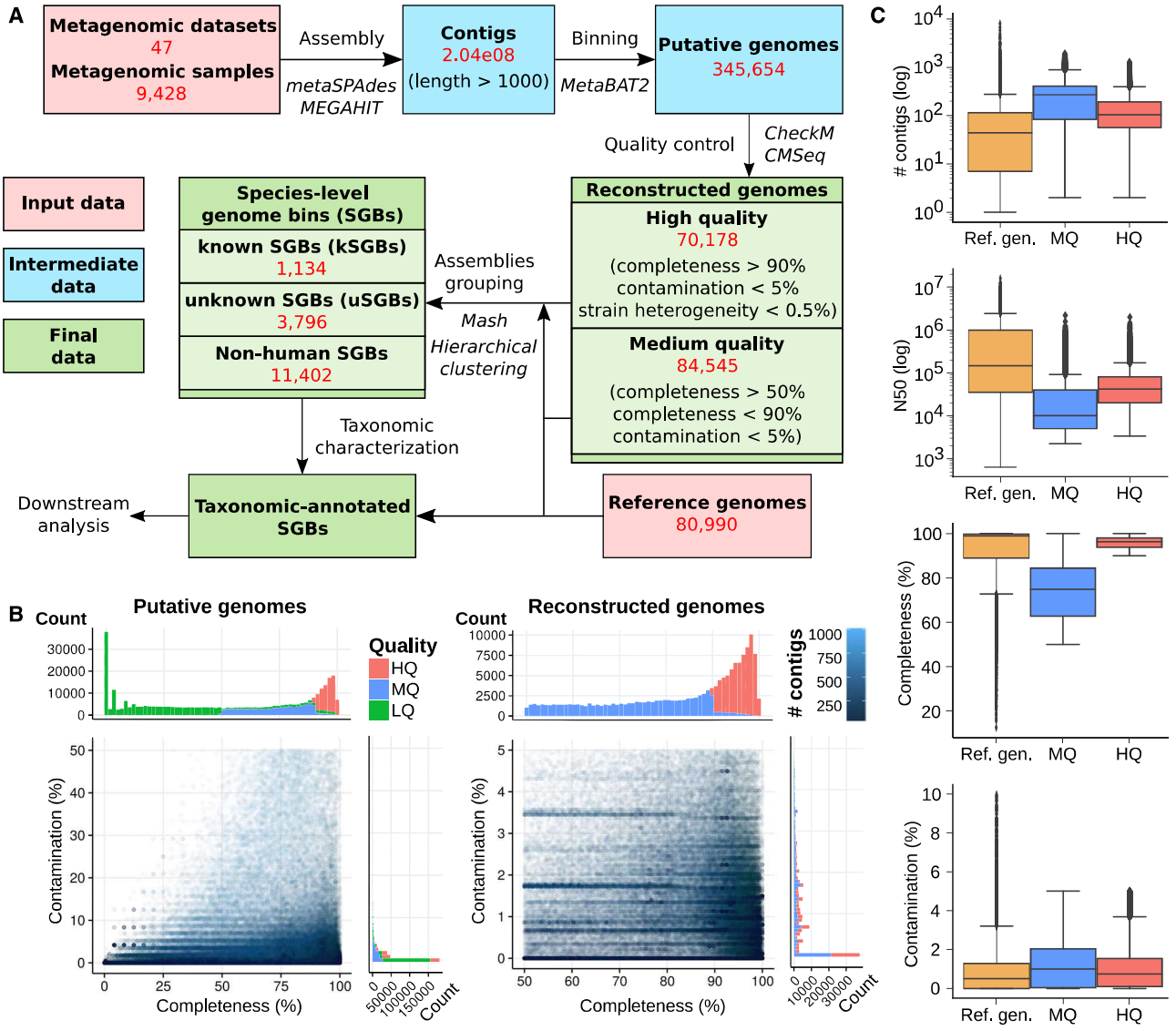

Figure 6. Methodology Overview and Quality Characteristics for the 154,723 Reconstructed Genomes

(A) Overview of the overall strategy and datasets employed for the reconstruction of microbial genomes and their organizations in SGBs.

(B) Completeness and contamination values estimated by CheckM are reported for LQ (low quality, completeness <50% or contamination >5%), MQ (completeness in the range [50%, 90%] and contamination <5%), and HQ (completeness >90%, contamination <5%, CMSeq strain heterogeneity <0.5%) genomes. LQ genomes are excluded from the rest of the analysis.

(C) Comparisons between the genomes from UniRef/NCBI used as references and our reconstructed genomes.

current collections with over 150,000 newly reconstructed genomes, in the process recovering hidden functional and phylogenetic diversity associated with global populations (particularly those that are undersampled from non-Western lifestyles and non-gut areas, Figure 1E). More than 94% of metagenomic reads can now be mapped to the expanded genome catalog for half of the gut microbiomes, enabling a much more comprehensive profiling of these communities. The metagenomic-assembly strategies employed here (Li et al., 2015; Nurk et al., 2017) represent a scalable methodology for very large-scale integration of metagenomes (Figure 6) that we extensively validated (STAR Methods; Figures 7 and S7) and could be fruitfully applied to additional or non-human-associated metagenomes. The methods are also compatible with emerging technologies such as synthetic (Kuleshov et al., 2016) or single-molecule (Brown et al., 2017) long-read sequencing, which will further add to the diversity of microbial genomes. Finally, the study's results themselves emphasize the phylogenetic and functional diversity that remains to be captured from rare organisms, especially for sample types other than stool, global human populations, and varied lifestyles for the human microbiome.