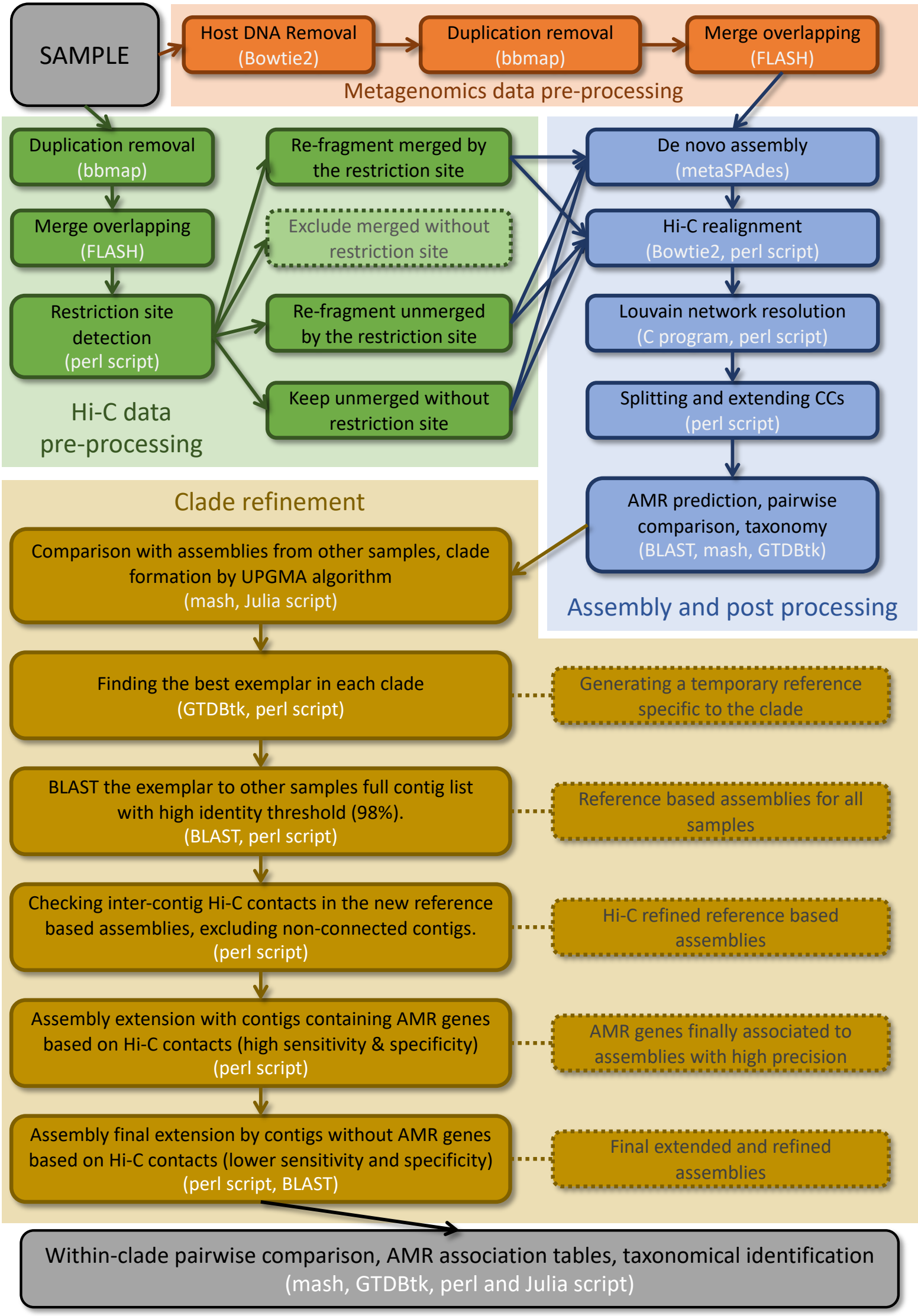
SAMPLE
Host DNA Removal (Bowtie2)
Duplication removal (bbmap)
Merge overlapping (FLASH)
Metagenomics data pre-processing
Duplication removal (bbmap)
Merge overlapping (FLASH)
Restriction site detection (perl script)
Re-fragment merged by the restriction site
Exclude merged without restriction site
Re-fragment unmerged by the restriction site
Keep unmerged without restriction site
Hi-C data pre-processing
De novo assembly (metaSPAdes)
Hi-C realignment (Bowtie2, perl script)
Louvain network resolution (C program, perl script)
Splitting and extending CCs (perl script)
AMR prediction, pairwise comparison, taxonomy (BLAST, mash, GTDBtk)
Assembly and post processing
Clade refinement
Comparison with assemblies from other samples, clade formation by UPGMA algorithm (mash, Julia script)
Finding the best exemplar in each clade (GTDBtk, perl script)
Generating a temporary reference specific to the clade
BLAST the exemplar to other samples full contig list with high identity threshold (98%). (BLAST, perl script)
Reference based assemblies for all samples
Checking inter-contig Hi-C contacts in the new reference based assemblies, excluding non-connected contigs. (perl script)
Hi-C refined reference based assemblies
Assembly extension with contigs containing AMR genes based on Hi-C contacts (high sensitivity & specificity) (perl script)
AMR genes finally associated to assemblies with high precision
Assembly final extension by contigs without AMR genes based on Hi-C contacts (lower sensitivity and specificity) (perl script, BLAST)
Final extended and refined assemblies
Within-clade pairwise comparison, AMR association tables, taxonomical identification (mash, GTDBtk, perl and Julia script)