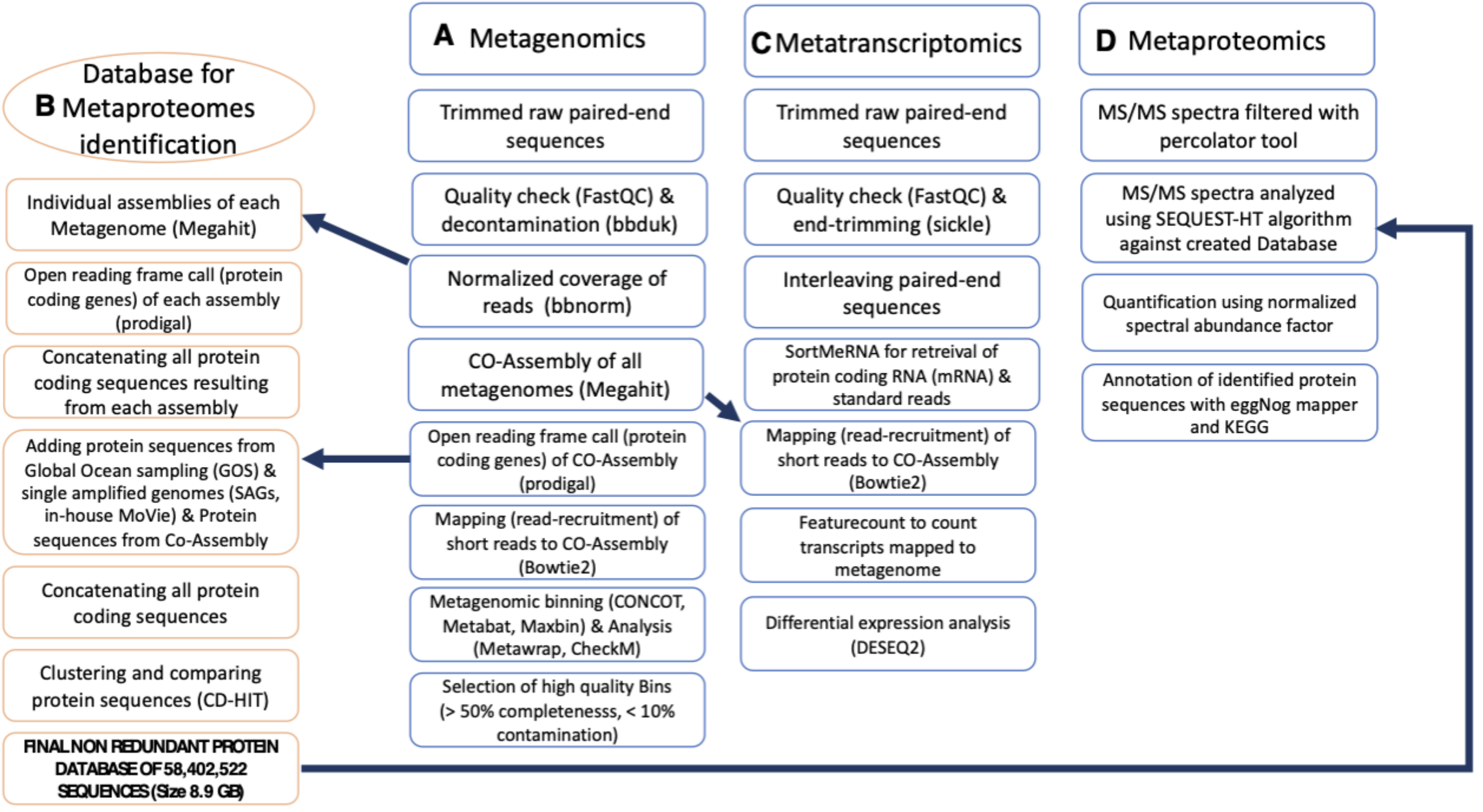
A Metagenomics
Trimmed raw paired-end sequences
Quality check (FastQC) & decontamination (bbduk)
Normalized coverage of reads (bbnorm)
CO-Assembly of all metagenomes (Megahit)
Open reading frame call (protein coding genes) of CO-Assembly (prodigal)
Mapping (read-recruitment) of short reads to CO-Assembly (Bowtie2)
Metagenomic binning (CONCOT, Metabat, Maxbin) & Analysis (Metawrap, CheckM)
Selection of high quality Bins (> 50% completenesss, < 10% contamination)
C Metatranscriptomics
Trimmed raw paired-end sequences
Quality check (FastQC) & end-trimming (sickle)
Interleaving paired-end sequences
SortMeRNA for retreival of protein coding RNA (mRNA) & standard reads
Mapping (read-recruitment) of short reads to CO-Assembly (Bowtie2)
Featurecount to count transcripts mapped to metagenome
Differential expression analysis (DESEQ2)
D Metaproteomics
MS/MS spectra filtered with percolator tool
MS/MS spectra analyzed using SEQUEST-HT algorithm against created Database
Quantification using normalized spectral abundance factor
Annotation of identified protein sequences with eggNog mapper and KEGG
Database for B Metaproteomes identification
Individual assemblies of each Metagenome (Megahit)
Open reading frame call (protein coding genes) of each assembly (prodigal)
Concatenating all protein coding sequences resulting from each assembly
Adding protein sequences from Global Ocean sampling (GOS) & single amplified genomes (SAGs, in-house MoVie) & Protein sequences from Co-Assembly
Concatenating all protein coding sequences
Clustering and comparing protein sequences (CD-HIT)
FINAL NON REDUNDANT PROTEIN DATABASE OF 58,402,522 SEQUENCES (Size 8.9 GB)