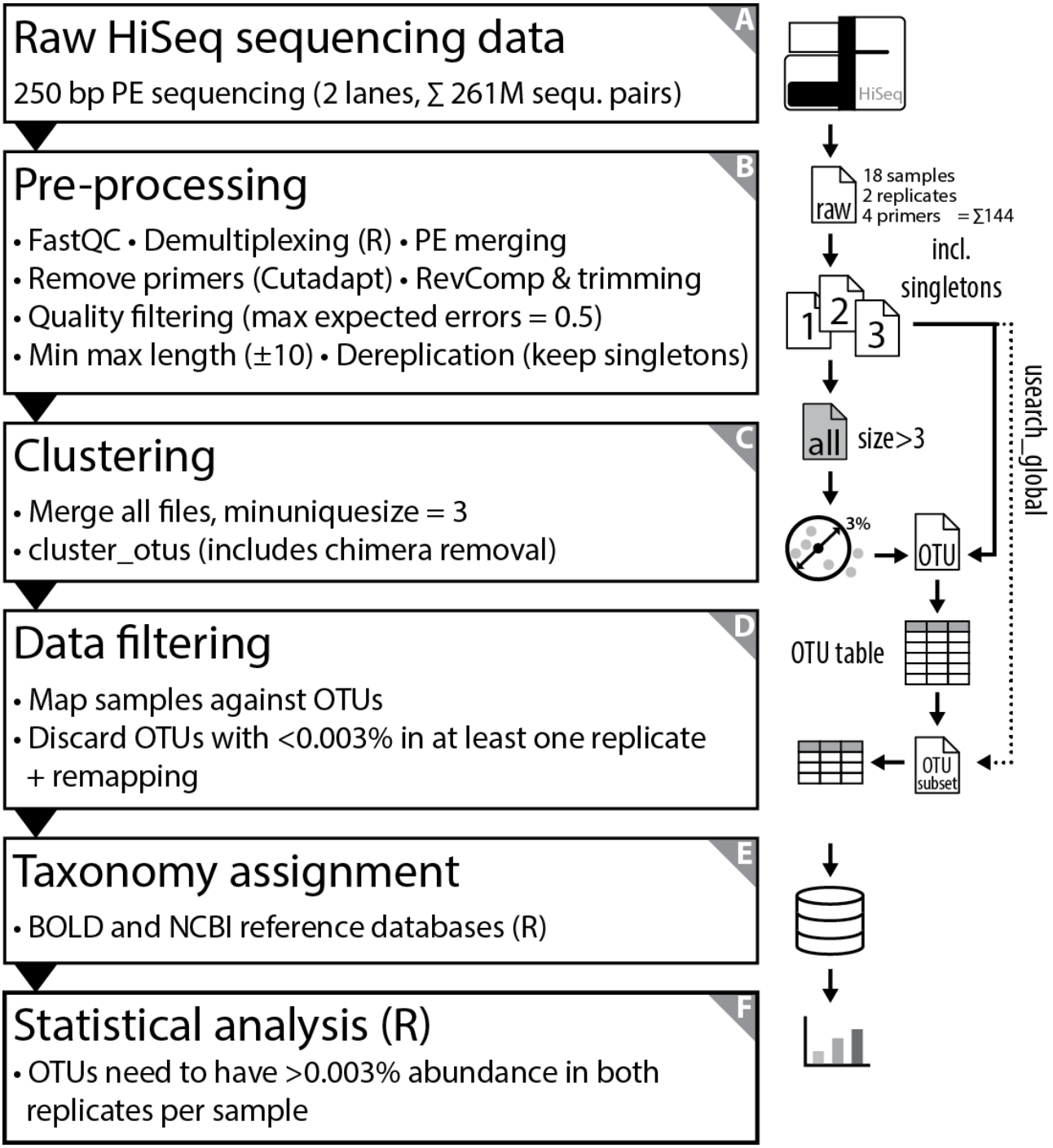

Figure S2: Detailed overview of the bioinformatic processing of the Illumina high throughput sequencing data. Raw sequence data (**A**) is demultiplexed and preprocessed (PE merging, remove primers, trimming, reverse complement, removal of low quality reads) (**B**). The processed sequences are then pooled and demultiplexed with a minimum size of 3, to reduce noise by sequencing errors in clustering (**C**). Reads from all samples are then compared against the generated OTUs and OTUs with a minimum of 0.003% of sequences assigned in at least one replicate, are discarded (**D**). Only OTUs which are present with >0.003% abundance in both replicates are kept for statistical analysis of individual samples. All reads are again mapped against the OTU subset to generate the final OTU table, with taxonomy being assigned to each centroid using NCBI and BOLD (**E**). Only OTUs with >0.003% abundance in both replicates per sample are kept for statistical analysis, OTUs below are set to 0% (**F**).