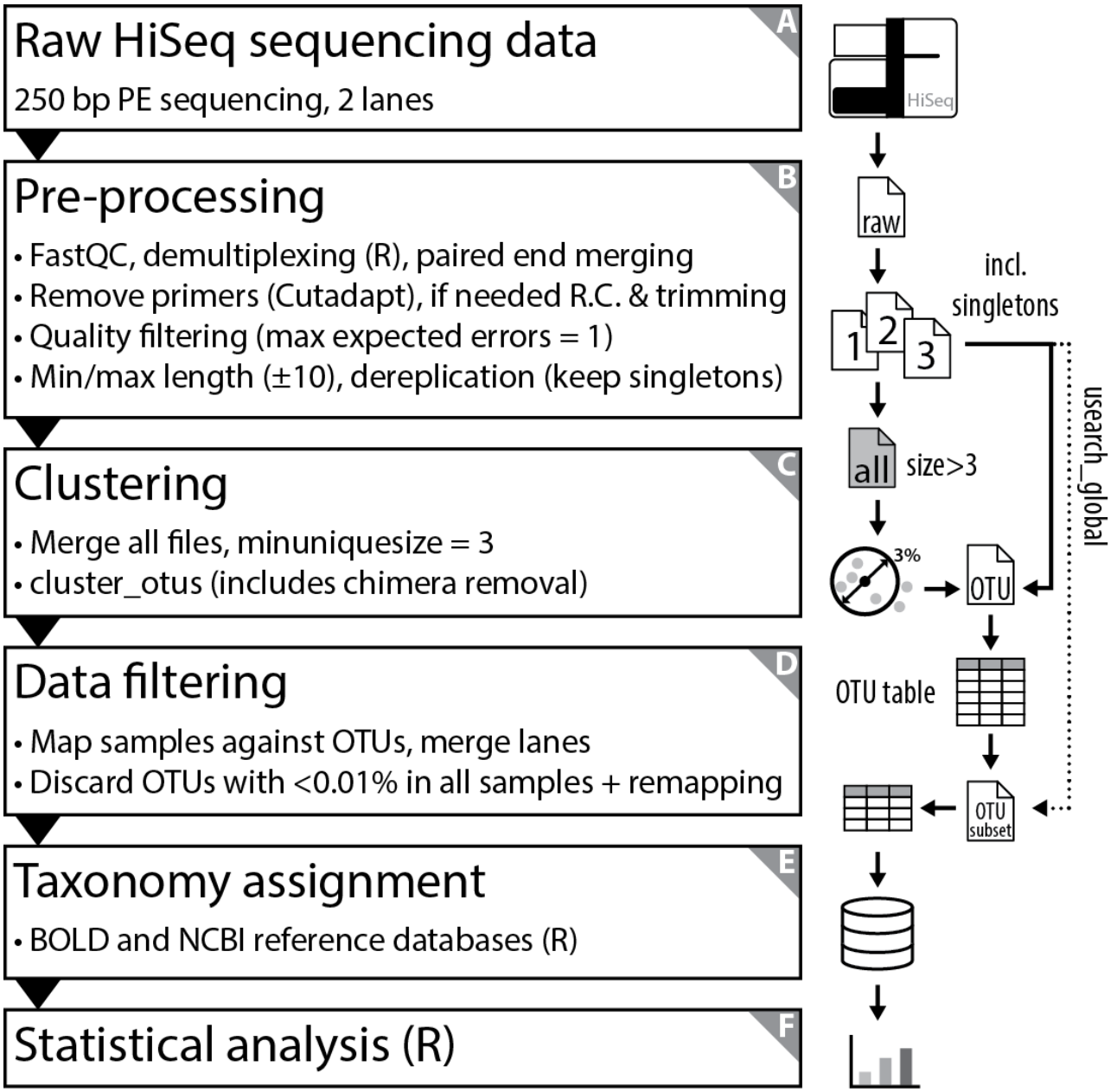

Figure S5: Detailed overview of the bioinformatic processing of the Illumina hight throughput sequencing data. Raw sequence data (**A**) is demultiplexed and pre-processed (paired end merging, remove primers, trimming, reverse complement, removal of low quality reads) (**B**). The processed sequences 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 0.01% of sequences assigned in at least one sample are discarded (data from both lanes is merged at this point) (**D**). 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**).