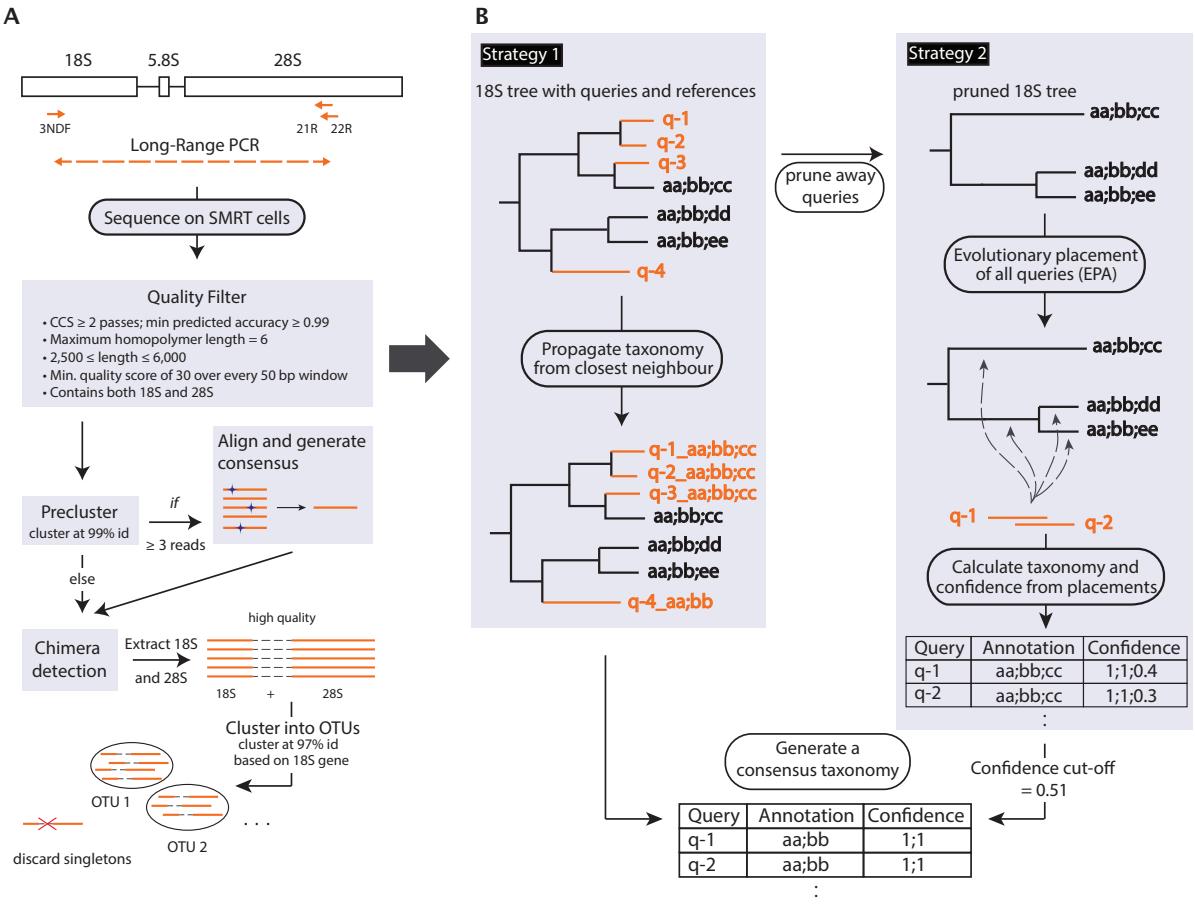

Figure 1. Workflow for generating long rDNA amplicons, quality filtering and taxonomic annotation. (A) General eukaryotic primers (3NDF, 21R and 22R) were used to amplify a ca. 4500 bp fragment of the rDNA operon (including 18S, ITS1, 5.8S, ITS2 and 28S) from environmental soil samples. The amplicons were sequenced on SMRT cells and subject to quality filtering before being pre-clustered at 99% similarity. Preclusters with at least 3 reads were de-noised further by aligning them and generating majority-rule consensus sequences. These were combined with the remaining reads and subject to de-novo chimera detection. The 18S and 28S gene regions extracted from these regions in the final step were physically linked as they originated from the same amplicon. (B) Overview of the taxonomic annotation pipeline. Reference sequences are shown in black with taxonomic ranks separated by ';' (e.g. aa;bb;cc denote three taxonomic rank labels: aa, bb, and cc). Queries are shown in orange. See text for details on the pipeline.