



**Figure 3.** Detailed comparison of taxonomic assignments for the 10 most abundant protist lineages in our data by VSEARCH (similarity-based approach) and the phylogeny-based approach for long (phylo. long) and truncated queries (phylo. V4). **A.** Stacked bars indicate the proportion of queries assigned to each lineage by different methods. Green bars: same taxonomic assignment as VSEARCH; yellow bars: queries assigned to higher ranks in the same taxonomic path; blue bars: queries assigned to the lineage by phylogeny but not by similarity (i.e. false negatives); pink bars: sequences assigned to a different lineage by

phylogeny (i.e. false positives). The number of taxonomic ranks in SILVA (from “Eukaryota” to species level) for each lineage is indicated in brackets, followed by the number of queries (n) assigned to the lineage by VSEARCH. **B.** Violin plots showing the taxonomic resolution of the assignments to each lineage. For the phylogeny-based approach these correspond to the green, yellow and blue bars.