

Table 2. Summary of datasets used in this study. RS = reference sequences, QS = query sequences.

	Eukaryote 18S	Eukaryote 18S+28S	Apicomplexa
Usage	Taxonomic annotation of environmental queries	Phylogenetic reconstruction of environmental samples	Assessing the effect of query length on inferring phylogeny of environmental diversity
Alignment size (# taxa:# positions in alignment)	a) 2311:1588 ¹ b) 2311:RS=1588, QS=V4 region ²	1019:4304	a) 107:1509 (18S) b) 107:3250 (18S+28S) c) 107:RS=1588, QS=V4 region (18S-V4)

¹ This dataset is referred to as “MSA” in main text.

² This dataset is referred to as “MSA-V4” in main text