



**Figure 4.** ML tree showing phylogenetic distribution of eukaryotic ribosomal diversity from soil samples. The tree corresponds to the best ML tree inferred from a concatenated alignment of 18S and 28S genes using the GTR+Gamma model. Bipartition support is derived from 300 bootstrap replicates. Purple circles represent BS  $\geq 75\%$ . The tree includes 589 queries from soil samples (coloured in orange) and 430 reference sequences (coloured in black) and all major eukaryotic supergroups (with the exception of Excavata which were eliminated during removal of long branches; see Methods for details). For details on taxon sampling, see Methods. The lower panel shows a subtree in Rhizaria. Zooming in illustrates how transferring the taxonomic annotation to the queries (coloured in orange) makes the phylogeny more informative, especially in clades (such as Rhizaria) where there are few references (labelled in black) with both the 18S and 28S available.