



not identify a set of protein clusters that characterize the totality of Antarctic CBS.

We repeated the analysis on proteins for which we could infer a functional annotation from sequence databases. KEGG [52] functional ortholog (KO) profiles were inferred for each genome by EggNOG-mapper and compared to the GTDB representative genomes from the same order. Also in this case, we could not identify a set of functions that characterize the Antarctic CBS across the whole dataset. This is also evident in the t-SNE analysis (Fig. 3c), where Antarctic CBS and reference genomes invariably pooled together according to taxonomy down to the order level, indicating similar functional potential within each order, regardless of provenience, when compared to the functional differences between different orders.

However, within each order, the pairwise Jaccard distances between the KO profiles of the Antarctic MAGs were in most cases lower than between MAGs and GTDB representatives (Fig. S9). These data indicated the existence of differences in the functional potential of Antarctic MAGs and taxonomically related reference genomes. In particular, there was evidence that specific functions were overrepresented in distinct phylogenetic clades. For instance, in the Acetobacteraceae family (order Acetobacteriales) (Fig. 4a) the t-SNE based on the KO profiles shows a separation of the Antarctic MAGs from the available genomes included in GTDB (Fig. 4b). In particular, the Antarctic clade is enriched (Fisher's exact test,  $p < 0.01$ , Bonferroni corrected) in genes related to membrane transport, carbohydrate, and amino acid metabolisms by factors from 3 to 31 (Fig. 4c, Table S8).