

Figure S1: Summary of all the unique entities per phylum for each of the four entity types (in log10 scale) that appear in PREGO coming from the *Literature* channel.

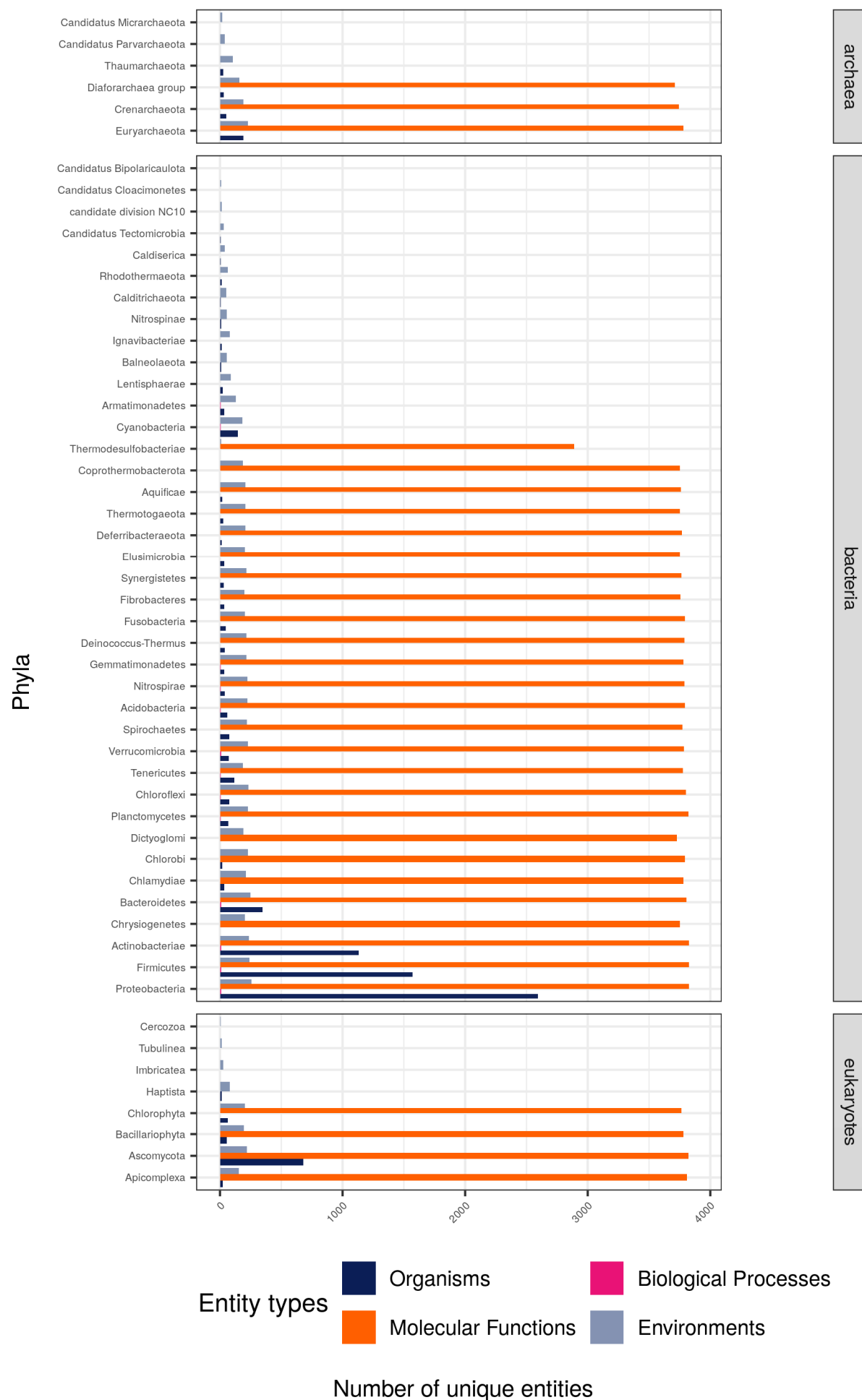


Figure S2: Summary of all the unique entities per phylum for each of the four entity types (in log10 scale) that appear in PREGO coming from the *Environmental samples* channel.

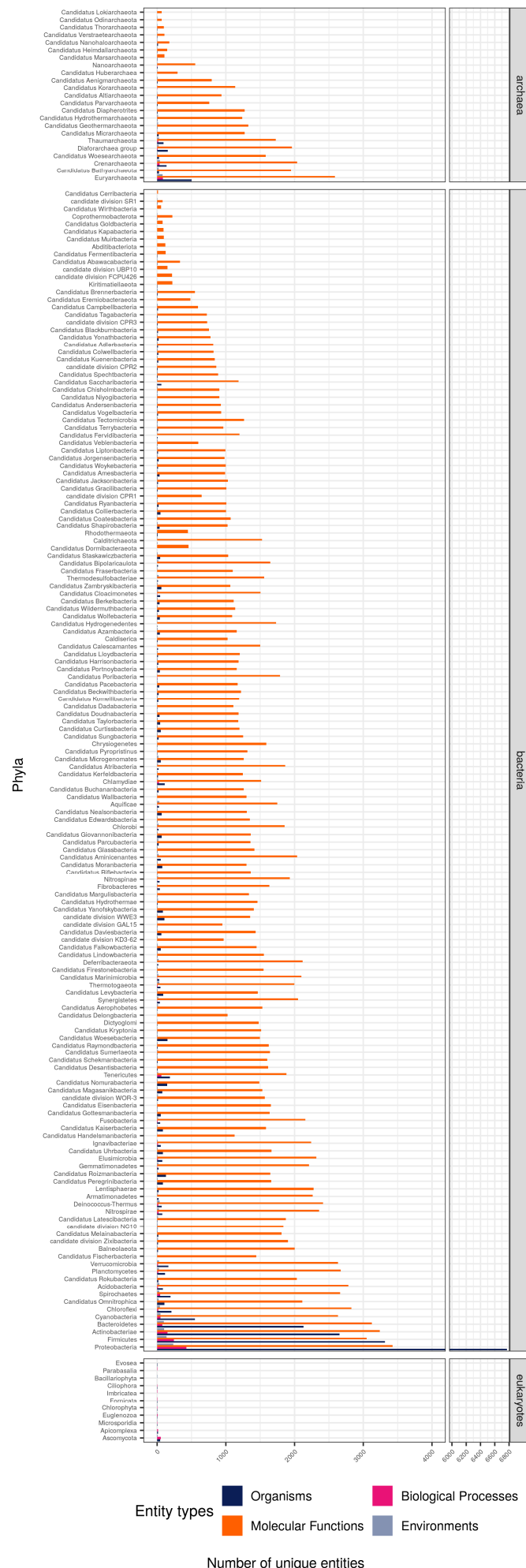


Figure S3: Summary of all the unique entities per phylum for each of the four entity types (in log10 scale) that appear in PREGO coming from the *Annotated genomes and Isolates* channel.