



SI Fig. 4. The possible effect of annotation errors on gene family clustering. For every organism, the percentage of analyzed families for which we found a tblastn-hit on the raw genome sequence, with a certain sequence coverage (i.e. the fraction of the protein-length found on the raw genome sequence), is shown.