



Supplementary Figure 3. Correlation between k -mer-based distances and taxonomic distances by body sites. Only tools able to scale-up all the datasets were kept in this experiment. Only body sites generating high quality taxonomic distances are shown. The quality requirements are the following: (1) the dataset must contain more than 10 samples to produce a reliable Spearman correlation (2) The dataset must have more than 40% of mapped reads to not generate too many outliers during the comparison (3) The taxonomic distances must be well balanced in the range [0, 1]. In particular, the nasal dataset was discarded because it has only 33% of assigned reads to references. The urogenital dataset was discarded because it has an over representation of taxonomic distances in the range [0.99-1] (median=0.971), meaning that the correlation would have been measured only on a very small range of distances.