



Supplementary Figure 6. Results of Simka on low covered samples from the Global Ocean Sampling project (GOS) [1] Comparison of the clustering obtained by Simka using its quantitative Bray-Curtis distance without k -mer solid filter (top) and when filtering out k -mers seen only one time (bottom). This experiments show two important points 1) Simka is able to retrieve the structure of the oceans as in the original GOS study even if those data are very low covered 2) the solid filter has almost no impact on the clustering of the samples. This is confirmed by the Spearman correlation value of 0.981 obtained between both distance matrices.

[1] Rusch *et al.* The sorcerer II global ocean sampling expedition: northwest atlantic through eastern tropical pacific. PLoS Biol, 5(3):e77, 2007.