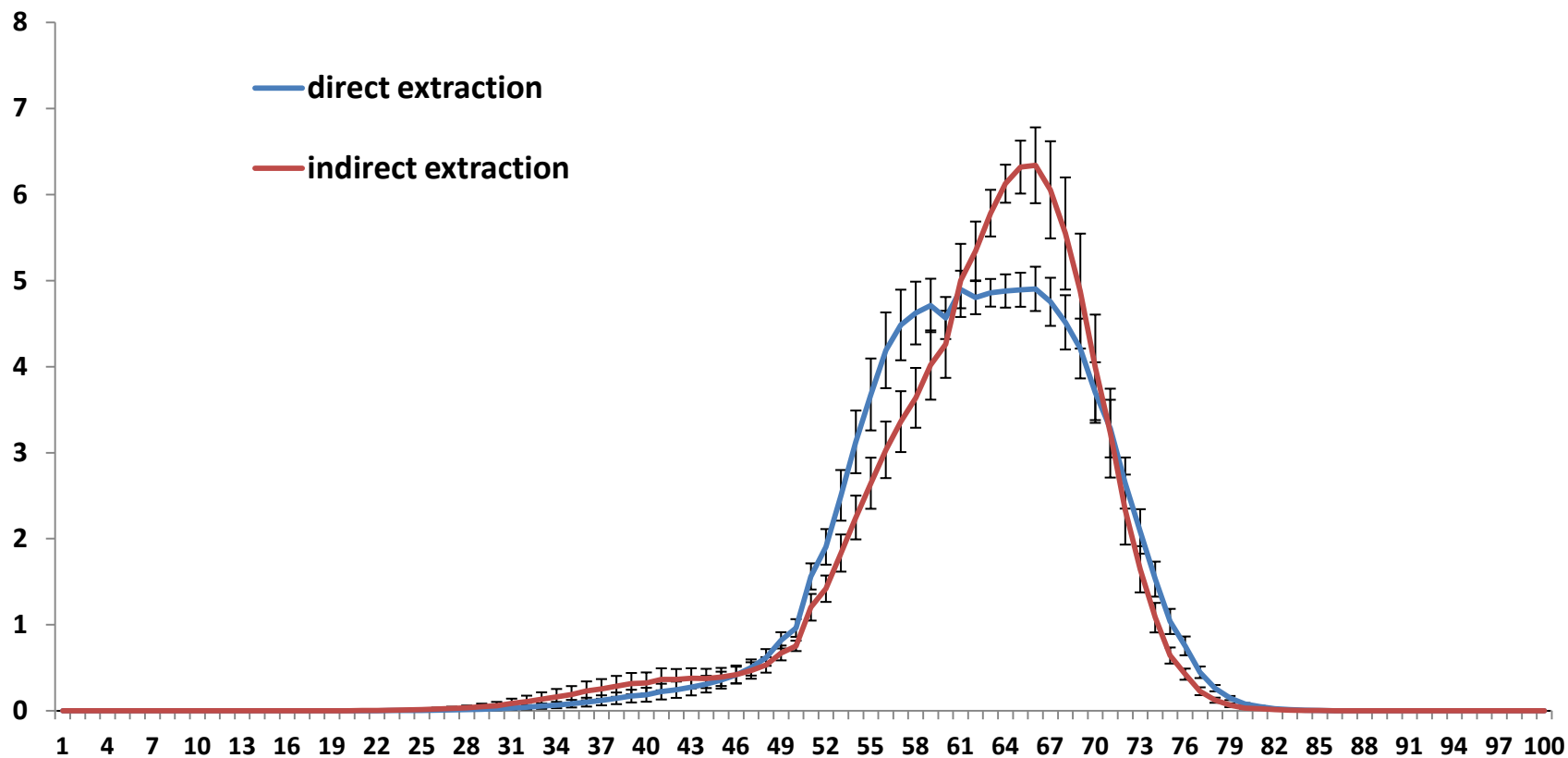


Figure S1. The average and standard deviation of the GC percentage profiles of the pyrosequence reads from all the direct and indirect DNA extractions of the Rothamsted Park Grass soil. There are 6 direct and 7 indirect DNA extractions and associated pyrosequencing runs used.



**Figure S2.** The principal component analysis of three ecosystems using the relative distribution of reads in the different phylogenetic data based on the SEED annotation for all the metagenomic sequences available in the public database in addition to those produced here.

