We compared the profiling results to those generated by several taxonomic binners using a simple coverage-approximation conversion algorithm for deriving profiles from taxonomic bins (Supplementary Methods). In taxon identification (recall), most binning methods were indeed superior to profiling methods, particularly at the genus level and below (Supplementary Figs P21-P24). Averaging over all samples at the genus level, the binners had a mean recall and precision of 0.519 and 0.459, respectively, while the profilers showed corresponding metric values of 0.393 and 0.262.

Abundance estimation at higher ranks was more of an issue for the binners, as the L1 norm error at the order level was 0.944 when averaged over all samples, while the profiler average was just 0.681. Overall, though, the binners delivered more accurate abundance estimates, as the binning average Unifrac metric was 6.31, while the profiling average was 7.21. However, for each metric aside from Unifrac, there was at least one profiler that outperformed all binning methods for every taxonomic rank higher than the genus level (though the particular profiler(s) varied depending on metric and taxonomic rank). Notably, similar to the taxon binning results, PhyloPythiaS+ showed a superior performance (averaged over all samples and taxonomic ranks) in all metrics compared to the other taxonomic binners in taxonomic profiling. Note that this in part might be due to its use of contigs, here the gold standard contigs, as input, while most profilers used the raw reads. Using longer correct sequenced for assignment could lead to improved taxon bin assignments.