

Additional Figure 2: EGS predictions work well when analyzing mixtures of different species or read lengths. The figure shows the distributions of relative errors (predicted EGS – known EGS) / (known EGS) from genome size predictions with Eq.(1) of the main text, when mixing simulated reads from different species (a) or reads with different read lengths (b). Mixing reads from different genomes or with different read lengths averages out prediction errors due to biological variation and inaccuracies of read assignment to marker gene OGs. Accordingly, the actual standard deviations of the error distributions are substantially lower than the errors predicted for single genomes or readlengths, SD=3.7% (a) and SD=6.4% (b), respectively. See Supplementary Methods for details.

