Additional Figure 4: Error distribution for EGS prediction on simulated reads, using the bacterial-specific version of the prediction formula. On a test set of species not used for parameter estimation, we defined relative errors as (prediction by Eq.(1) – known genome size) / (known genome size). Errors are approximately normally distributed (a) (Shapiro-Wilks test: *P*=0.074). Errors are approximately independent of read length (b) and genome size (c). Red dashed lines (b-c) are estimated standard deviations.