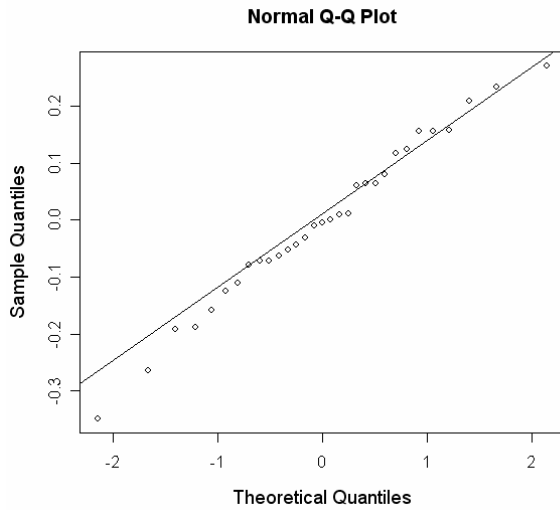


**Additional Figure 5: Error distribution for EGS prediction on real reads, using the bacterial-specific version of the prediction formula.** Relative errors are approximately normally distributed (a) (Shapiro-Wilks test:  $P=0.97$ ), and are independent of genome size (b). The red circles in (b) represent two data sets suffering from strong experimental biases (see main text).

(a)



(b)

