



**Supplementary Figure 2. Correlation between Simka and BLAT results.** Each row corresponds to a BLAT identity threshold and each column to a  $k$ -mer size used by Simka. The value  $R$  is the Spearman correlation between Simka and BLAT results. Correlation values in bold are the greatest for each identity threshold.