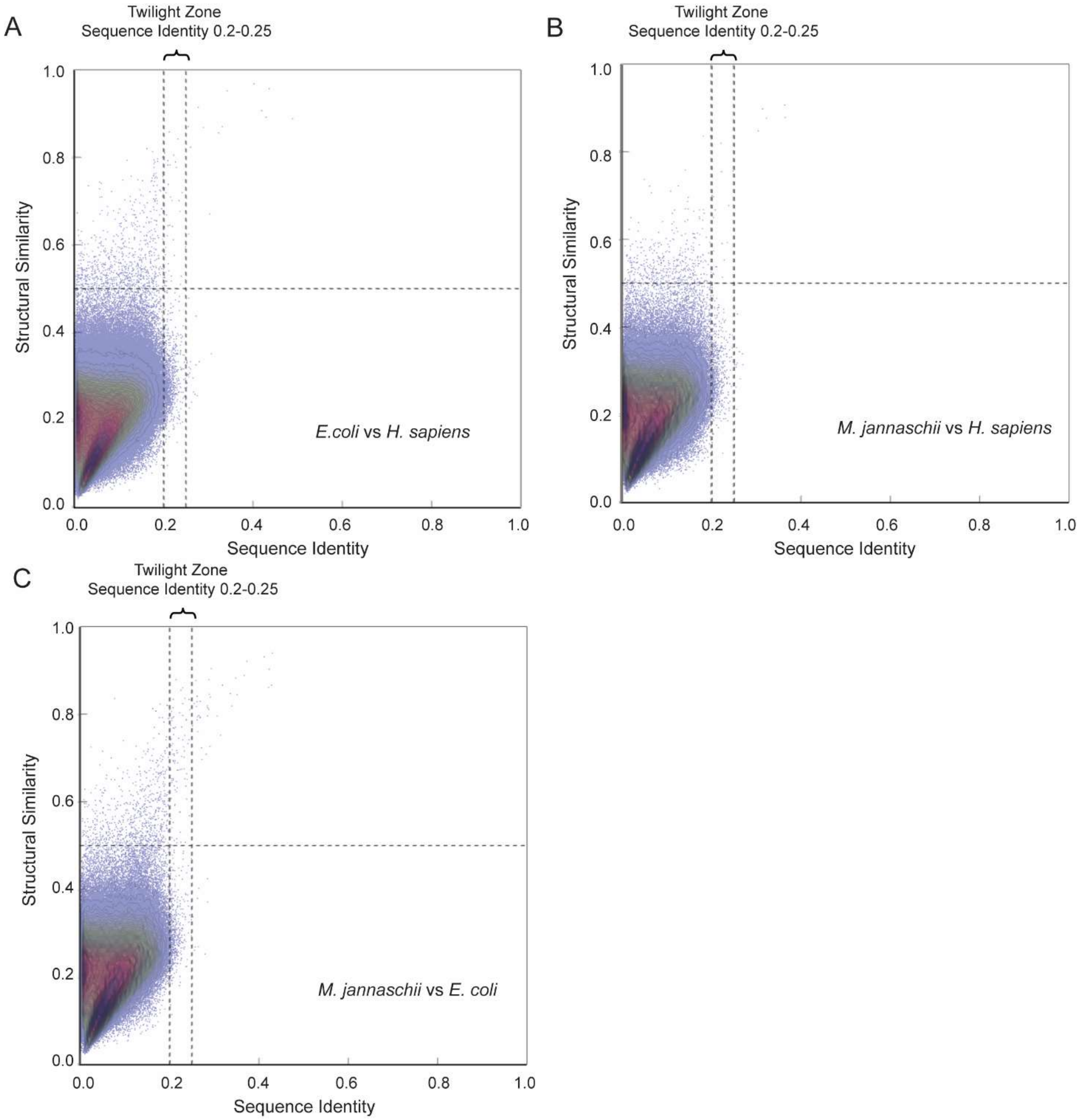

Supplemental Figure S2. Characterization of the Distribution of structural similarity and sequence identity through random sampling

Overlay of 2D scatter plot and 2D contour plot depicting the distribution of structural similarity and sequence identity of the proteins. The subfigures compare the proteomes of *E. coli* vs *H. sapiens* (A), *M. jannaschii* vs *H. sapiens* (B) and *M. jannaschii* vs. *E. coli* (C). A total number of 1,000,000 unique pairs of proteins were randomly selected to compare *E. coli* and *H. sapiens*, and 500,000 unique pairs were selected for comparing *M. jannaschii* with either *H. sapiens* or *E. coli*. The vertical dashed lines indicate the sequence identity of 0.2 and 0.25, and the horizontal dashed line indicates the structural similarity of 0.5.