

**Additional data file 1.** Evaluation of genome sequence fragment clustering based on fixed point kernel densities. Shown is the ESOM map presented in Fig. 3, without the U-matrix background displayed. Data points of known identity (from assembly information) are enlarged and colored; others are unassigned. Contour lines are delineated so that 90% of the data points for each genome are included within the bin boundaries.

