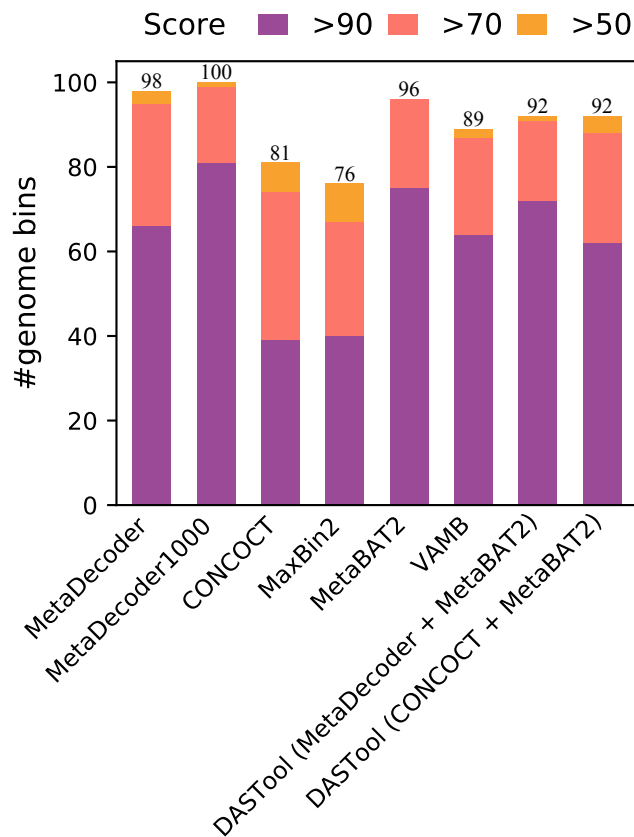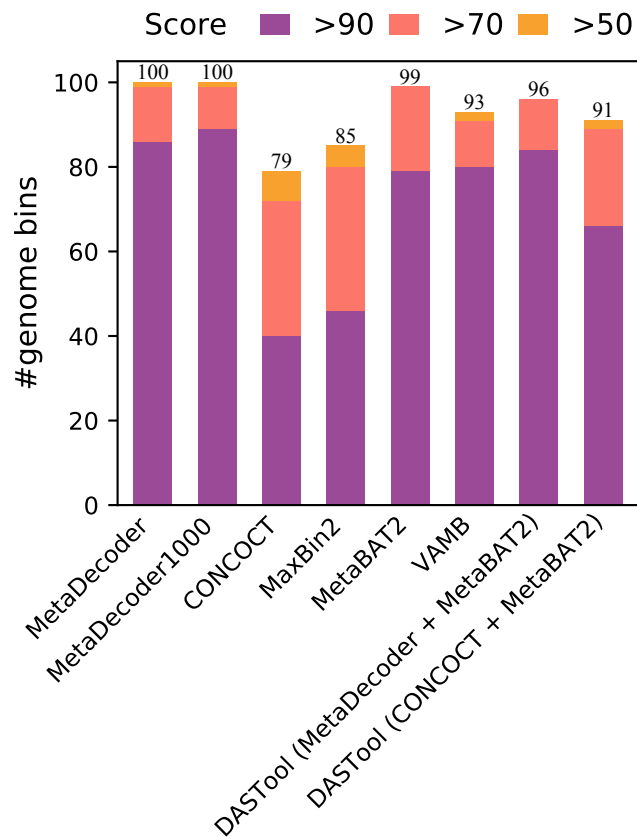**Supplementary Figure S2.** Clustering benchmarks on a simulated dataset. The number of identified bins with different score levels were shown. All programs were run with their default parameters in both single-sample **(A)** and multi-sample **(B)** modes. MetaDecoder with minimum sequence length setting to 1 Kb (MetaDecoder1000) was also added for benchmarking. Assessments were evaluated using AMBER (version 2.0.2).