1147-92-584 **Mahdi Belcaid*** (mahdi@hawaii.edu), 1680 East-West Road, Room 317, Honolulu, HI 96826. *A Probabilistic Approach for DNA Sequence Partitioning Using Dimensionality Reduction.* Preliminary report.

DNA clustering is an essential computational step in a plethora of Next-Generation sequencing experiments, such as amplicon sequencing, high-throughput immune system characterization, and functional studies. However, clustering of very large datasets, such as those produced by current-generation sequencers, is highly CPU and memory intensive. As such, analysis datasets comprising millions of sequencing reads cannot be tackled on commodity hardware without a significant compromise in sensitivity. To address this limitation, we propose a new probabilistic data partitioning technique that uses random projections to divide large datasets into smaller, non-overlapping subsets which can be subsequently clustered at lower computational costs. This partitioning step, in addition to substantially lowering resources requirements, provides an opportunity for intuitive parallelization of DNA sequence clustering, and for the independent processing of each subset using sensitive clustering algorithms. (Received January 26, 2019)