



**S4 Fig. Effect of Redundans reduction.** (A) Identity distribution of discarded sequences to target sequences. (B) Proportion of overlap between discarded and target sequences. (C) Length distribution of discarded sequences.