Team Member	Initial Research & Proposal	First Presentation Preparation	Presentation	Final Project Phase	Final Report
Kevin He	Developed core ideas, primary editor of proposal	Dealt with minHash and minimizers	Participated in creating slides and presenting	Collaborated on simple-diamond implementation	Contributed to writing and editing
Chaocheng Chuang	Sourced references, refined proposal	Implemented ORF filter, assisted with DIAMOND	Participated in creating slides and presenting	Collaborated on simple-diamond implementation	Contributed to writing and editing
Jingkai Guo	Participated in literature review	Created utils for processing data	Participated in creating slides and presenting, edited recording	Implemented multithreading, conducted experimental runs	Contributed to writing and editing
Peijia Ye	Identified and acquired datasets	Ran the filter, conducted BLASTx and DIAMOND experiments	Participated in creating slides and presenting	Managed datasets, conducted experimental runs	Contributed to writing and editing

Comment (if you have):

Kevin created the ORF filter and the 6 reading frames and made the double indexing scheme to match together seeds between the protein index and the query index. Kevin also implemented the minimizer and minhash sketching scheme.

In our project, Chao Cheng Chuang developed a series of computational techniques for genomic data analysis. He implemented an ORF filter for identifying protein-coding regions and a reduced alphabet approach to simplify data complexity. His adaptation of the Smith-Waterman algorithm used the BLOSUM62 matrix for enhanced alignment accuracy. He also established a double indexing baseline for efficient genomic database searches and a sort-merge join baseline for effective data handling. Lastly, he applied uniform sketching for concise sequence representations, facilitating faster analysis

Jingkai Guo: create utils for processing data like generating random samples, reads to protein, reverse complement six frames, etc…. Experimenting multithreaded sort as parallel computing to increase the computing performance. Synthesis the function to find the best alignment then save a file. Helps kevin to finished the diamond main

Peijia Ye identified and acquired protein datasets of E.coli and the associated DNA sequences, ran the orf filter, conducted BLASTx and DIAMOND experiments, participated in creating slides and presenting, preprocessed datasets, conducted experimental runs with uniform, minHash and minimizer, generating the matching scores, contributed to writing and editing the readme file and abstract & introduction part of final report