Optimizing Traceback in the Smith-Waterman Algorithm for GPU architectures
LeAnn M. Lindsey¹, Hari Sundar¹(Advisor), Muaaz Awan² (Advisor)
¹University of Utah, ²Lawrence Berkeley National Laboratory

Abstract
The traceback phase of the Smith-Waterman algorithm require a significant amount of memory and introduce an irregular memory access pattern which makes it challenging to scale performance on GPU architectures. We developed an efficient traceback algorithm and implementation was integrated into the ADAPT sequence alignment library. To demonstrate its efficacy on high-performance software, we integrated this traceback implementation into Metahipmer, which is a widely used metagenome assembler. Our proposed implementation is 3.6x faster than traceback in GASAL2 implementation into the ADEPT sequence alignment library. To

Challenges of Implementing on GPUs

Sequence Alignment Algorithms

Smith-Waterman Algorithm

Performance Evaluation against Comparison Libraries

Metahipmer2 Integration

References

Acknowledgements

The research was supported by the Exascale Computing Project (17-SC-02-SC), a collaborative effort of two U.S. Department of Energy organizations, Office of Science and National Nuclear Security Administration. This research used the National Center for Computational Sciences’ TC2000 supercomputer, Titan, located in Oak Ridge, Tennessee. It was also supported by the National Energy Research Scientific Computing Center (NERSC). NERSC is a high-performance computing center and is a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under contract DE-AC02-05CH11231. DOE also funds the National Institute of Biomedical Imaging and Bioengineering via Grant 1R01EB020996-01A1. This work is a contribution to DOE Exascale Data Infrastructure (EDIC) project.

ACKNOWLEDGEMENTS

We would like to thank the authors on June 2, 2022. (GASAL2 does not align Protein sequences)

SOLUTION #1

SOLUTION #2