The University of Melbourne
School of Computing and Information Systems
COMP90025 Parallel and Multicore Computing
Project 2B

Multiple Sequence Alignment

Marco Marasco 834482
Word count: 214

Parallelisation Approach

The approach taken was to use **OpenMP** and **OpenMPI** over the cluster, inspired by [1]. At runtime, the master node **broadcasts** all the genes, and some additional information including the largest gene length and number of genes.

Each worker (master node included) determines which alignment jobs it needs to compute, using the alignment **job number** and the worker's **rank** within the MPI communicator. The worker then allocates a single table to be used by all jobs for the DP alignment algorithm, with the smallest possible dimensions to reduce memory allocation time.
After the worker computes an alignment using the DP algorithm from project 2A, the problem number, penalty, and hash of each alignment are stored in an MPI struct. Once all jobs are completed, the results are sent back to the master node.

Once the master node has computed its own alignment jobs, it determines the number of problem results it will receive from each node, and uses MPI's **gatherv** function to collect the results. The master node then computes the final hash required for output, and returns.

Future work would be to explore load balancing jobs amongst nodes.

References

[1] J. González-Domínguez, Y. Liu, J. Touriño, and B. Schmidt, "MSAProbs-MPI: parallel multiple sequence aligner for distributed-memory systems," *Bioinformatics*, vol. 32, no. 24, pp. 3826–3828, Aug. 2016, ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btw558. eprint: https://academic.oup.com/bioinformatics/article-pdf/32/24/3826/16921328/btw558.pdf. [Online]. Available: https://doi.org/10.1093/bioinformatics/btw558.