

{tag}

{/tag}

IJCA Proceedings on International Conference
on Advances in Computer Engineering and Applications

© 2014 by IJCA Journal

ICACEA - Number 5

Year of Publication: 2014

Authors:

Charu Sharma

Pankaj Agrawal

Preeti Gupta

{bibtex}icacea1455.bib{/bibtex}

Abstract

The growth of bioinformatics and computational biology industry, multiple sequence alignment (MSA) applications have become an important emerging workload. In spite of the large amount of recent attention given to the MSA software design, there has been little quantitative understanding of the performance of such applications on modern microprocessors and systems. In this paper we try to analyze performance and characteristics of MSA software from

the perspective of multicore machines. We use several popular MSA programs employing a wide variety of alignment approaches. The basic workload characteristics and the efficiencies of various multicore machines features are examined . In order to mapping parallelism in multicore machines we try to explore different parallel programming approaches using threads and MPI

Refer

ences

- J. D. Thompson, J. e. Thierry, O. Poch. (2003). RASCAL: rapid scanning and correction of multiple sequence alignments, *Bioinformatics*, Vol. 19, No. 9.
- V Amouda et. al. / *International Journal of Engineering Science and Technology* Vol. 2(11), 2010, 6361-6370.
- Thompson JD, Poch O. (2006). Multiple sequence alignment as a workbench for molecular systems biology, *Curr Bioinformatics* , 1:95-104.
- Boukerche A, Demelo A, Ayalarincon M, Walter M. (2007). Parallel strategies for the local biological sequence alignment in a cluster of workstations, *J Parallel Distrib Comput* , 67:170-185.
- Essoussi N, Boujenfa K, Limam M. (2008). A comparison of MSA tools. *Bioinformation* , 2:452-455.
- Kleinjung J, Douglas N, Heringa J. (2002). Parallelized multiple alignment, *Bioinformatics* , 18:1270-1271.
- Schmollinger M, Nieselt K, Kaufmann M, Morgenstern B. (2004). DIALIGN P: Fast pair-wise and multiple sequence alignment using parallel processors. *BMC Bioinformatics* 2004, 5:128.
- Li K-B. (2003). ClustalW-MPI: ClustalW analysis using distributed and parallel computing, *Bioinformatics* , 19:1585-1586.
- Mikhailov D, Cofer H, Gomperts R. (2001). Performance optimization of Clustal W: parallel Clustal W, HT Clustal, and Multiclustals. In *White papers*. Silicon Graphics, Mountain View, CA.
- Zafalon, F. D. Geraldo. Et al. (2013). "Improvements in the score matrix calculation method using parallel score estimating algorithm", *Journal of Biophysical Chemistry*, Vol. 4, No. 2, 47-51.
- Zhu, X. Li, K. et al. (2011). "A Data Parallel Strategy for Aligning Multiple Biological Sequences on Homogeneous Multiprocessor Platform", *Sixth Annual ChinaGrid Conference*.
- Agarwal, P. Rizvi, S. A. M. (2009). "Solving sequence Alignment Problems using Pipeline Approach", *BIJIT-BVICAM's International Journal of Information Technology*
- Saeed, F. et al. (2009). "A Domain Decomposition Strategy for Alignment of Multiple Biological Sequences on Multiprocessor Platforms", *J. Parallel Distrib. Comput.*
- Trelles O. (2001). "On the parallelisation of bioinformatics applications", *Brief Bioinform*, 2:181-194.
- Naveed , T. Siddiqui, I. S. (2005). "Parallel Needleman-Wunsch Algorithm for

Grid", Biogridpaper.

- C. Notredame, D. G. Higgins, and J. Heringa. (2000). T-COFFEE: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* , 392:205–217.
- C. Notredame, L. Holm, and D. G. Higgin. (1998) COFFEE: An objective function for multiple sequence alignment. *Bioinformatics*, 14(5):407–422.
- J. D. Thompson, D. G. Higgins, and T. J. Gibson. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Research*,22(22):4673–4680.
- C. Notredame and D. G. Higgins. (1996). SAGA: Sequence alignment by genetic algorithm, *Nucleic Acids Research*, 24(8):1515–1524.
- Robert C. Edgar. (2004). " MUSCLE: multiple sequence alignment with high accuracy and high throughput", *Journal,Nucleic Acid Research*, Vol. 32(5).
- J. Blazewicz et al. (2013) "G-MSA — A GPU-based, fast and accurate algorithm for multiple sequence alignment", *J. Parallel Distrib. Comput.* 73: 32–41.
- JF. J. Esteban et al. (2013). "Direct approaches to exploit many-core architecture in bioinformatics", *Future Generation Computer Systems* Vol. 29:15–26.
- Rezaei. S et al. (2006). "DIVIDE-AND-CONQUER ALGORITHM FOR CLUSTALW-MPI",*IEEE CCECE/CCGEI*, Ottawa, May 2006.
- Xiangyuan Zhu. Et al. (2013) "A data parallel strategy for aligning multiple biological sequences on multi-core computers", *Computers in Biology and Medicine* 43: 350–361.
- Dohi. K, K. Benkrid,C. Ling et al. (2010). "Highly Efficient Mapping of the Smith-Waterman Algorithm on CUDA-compatible GPUs" *IEEE,ASAP*.
- Thompson,J. D. , Higgins,D. G. and Gibson,T. J. (1994). "ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting,position-specific gap penalties and weight matrix choice," *Nucleic Acids Research*, vol. 22, pp. 4673–4680.
- O. Miquel, G. Fernando, N. Cedric, C. Fernando, Exploiting parallelism on progressive alignment methods, *J. Supercomput.* (2009) 79–87.
- B. Azzedine, C. Alba, A. -R. Mauricio, E. Maria, Parallel strategies for the local biological sequence alignment in a cluster of workstations, *J. Parallel Distrib. Comput.* 67 (2007) 170–185.
- K. Taecho, j. Hyun, Clustalxseed: a GUI-based grid computation version for high performance and terabyte size multiple sequence alignment, *BMC Bioinformatics* 11 (2010) 467. [
- H. M. Wong, V. Bharadwaj, Aligning biological sequences on distributed bus networks: a divisible load scheduling approach, *IEEE Trans. Inf. Technol. Biomed.* 9 (4) (2005) 489–501.
- V. Bharadwaj, H. M. Wong, Handling biological sequence alignments on networked computing systems: a divide-and-conquer approach, *J. Parallel Distrib. Comput.* 69 (2009) 854–865.
-
- H. P. L. Diana, V. Bharadwaj, A. B. David, On the design of high-performance algorithms for aligning multiple protein sequences on mesh-based multi- processor architectures, *J. Parallel Distrib. Comput.* 67 (2007) 1007–1017.
- D. P. Tommaso, M. Orobítg, F. Guirado, F. Cores, T. Espinosa, C. Notredame,

Cloud-coffee: implementation of a parallel consistency-based multiple alignment algorithm in the t-coffee package and its benchmarking on the amazon elastic-cloud, *Bioinformatics* 26 (15) (2010) 1903–1904.

- H. F. B. Vicente, L. M. David, P. Sylvain, S. Johannes, Parallel geometric algorithms for multi-core computers, *Comput. Geometry* 43 (2010) 663–677.
- L. Kuo-Bin, CLUSTALW-MPI: CLUSTALW analysis using distributed and parallel computing, *Bioinformatics* 19 (12) (2003) 1585–1586.
- Z. Jaroslaw, Y. Xiao, R. Adrain, A. Srinivas, Parallel-tcoffee: a parallel multiple sequence aligner, in: *Proc. ISCA PDCS*, 2007, pp. 248–253.
- S. Fahad, K. Ashfaq, A domain decomposition strategy for alignment of multiple biological sequences on multiprocessor platforms, *J. Parallel Distrib. Comput.* 69 (2009) 666–677.
- K. Kazutaka, T. Hiroyuki, Parallelization of the MAFFT multiple sequence alignment program, *Bioinformatics* 26 (15) (2010) 1899–1900.

Index Terms

Computer Science

Distributed Computing

Keywords

Multiple Sequence Alignment Parallelism Multicore Machines Parallel Strategies.