



Parallel Programming in Bioinformatics: Some Interesting Approaches

Miguel A. Vega-Rodríguez¹ · José M. Granado-Criado¹

Published online: 29 October 2018

© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract

Bioinformatics allows and encourages the application of many different parallel programming approaches. This special issue brings together high-quality state-of-the-art contributions about parallel programming in bioinformatics, from some interesting points of view or perspectives. The special issue collects considerably extended and improved versions of the best papers, accepted and presented in PBio 2017 (5th International Workshop on Parallelism in Bioinformatics, and part of ICA3PP 2017). The domains and topics covered in these 2 papers are timely and important, and the authors have done an excellent job of presenting the material.

Keywords Parallel programming · Bioinformatics · Interesting approaches

1 Introduction

In bioinformatics, we can find a variety of problems which are affected by huge processing times and memory/storage consumption, due to the large size of biological data sets and the inherent complexity of biological problems. In fact, bioinformatics is among the most exciting research areas in which parallel programming finds application. Successful examples are mpiBLAST, RAxML-HPC or ClustalW-MPI, among many others. Therefore, bioinformatics allows and encourages the application of many different parallel programming approaches: multicore computing, cluster computing, supercomputing, cloud computing, grid computing, green computing, hardware accelerators such as GPUs, FPGAs, etc.

✉ Miguel A. Vega-Rodríguez
mavega@unex.es

José M. Granado-Criado
granado@unex.es

¹ Department of Computer and Communications Technologies, University of Extremadura, Escuela Politécnica, Campus Universitario s/n, 10003 Cáceres, Spain

This special issue brings together high-quality state-of-the-art contributions about parallel programming in bioinformatics, from some interesting points of view or perspectives. This special issue collects the best papers, accepted and presented in [1] (5th International Workshop on Parallelism in Bioinformatics, and part of ICA3PP 2017). These papers have been considerably extended and improved by the authors from their original conference versions.

2 Important Data

At present, the application of parallel programming in bioinformatics is a very popular research topic. As an example of the current interest in this field, it is worth mentioning that, for this special issue, we have managed a total of 17 high-quality submissions from different countries, such as USA, Argentina, Germany, Switzerland, Poland, Sweden, Portugal, or Spain. After a first review round, 4 submissions were selected as the best papers for this special issue. All the papers included in this special issue were reviewed by at least four expert reviewers. Furthermore, all the papers in the special issue received a minimum of three review rounds. Finally, only 2 papers of high quality in emerging research areas were accepted for inclusion in the special issue (therefore, from the initial point, an acceptance rate = $2/17 = 11.76\%$). In conclusion, we think these papers bring us an international sampling of significant work.

The domains and topics covered in these 2 papers are timely and important, and the authors have done an excellent job of presenting the material. We are confident that this special issue will be useful for all the readers who are engaged in the many issues surrounding the application of parallel programming in the bioinformatics domain.

3 Detailed Content

The title of our first paper is “SWIMM 2.0: Enhanced Smith–Waterman on Intel’s Multicore and Manycore Architectures based on AVX-512 Vector Extensions”, by E. Rucci, C. Garcia-Sanchez, G. Botella-Juan, A. De Giusti, M. Naiouf, and M. Prieto-Matias (see [2]). The Smith–Waterman algorithm is the most commonly used method for local sequence alignments. However, it has very high computational requirements for large protein databases. Although the acceleration of this algorithm has already been studied on many parallel platforms, there are hardly any studies which take advantage of the latest Intel architectures based on AVX-512 vector extensions. This paper fills this gap. The novelty of this vector instruction set requires the revision of previous programming and optimization techniques. The software detailed in this paper is competitive in terms of performance compared with other state-of-the-art implementations. Moreover, these successful performance rates also make this software a very efficient energy footprint implementation.

The second paper, “A History-based Resource Manager for Genome Analysis Workflows Applications on Clusters with Heterogeneous Nodes” by F. Badosa, A. Espinosa, C. Acevedo, G. Vera, and A. Ripoll (see [3]), is focused on the bioinformatics workflows and how to manage the computational resources required by them when they

are executed. Bioinformatics workflows require large amounts of resources and are commonly executed in clusters. Determining the adequate amount of resources for bioinformatics applications is a tricky matter, since the resource usage of a single application might vary substantially from one execution to the next one. Resource management systems in clusters do not consider these variations and subsequent needs. As a result, the computing power offered by clusters is not harnessed properly, compromising both application performance and resource efficiency. This paper proposes a history-based resource manager for bioinformatics workflows applications running on clusters with heterogeneous nodes. The proposed resource manager features a prediction model that generates multiple performance predictions for each job under different combinations of cluster resources. The proposal shows an average workflow makespan improvement of 32%, an average workflow efficiency improvement of 79%, and an average resource usage improvement of 99%.

4 Conclusion

We sincerely hope that you enjoy this special issue. We also have hope that the paper collection can pleasantly introduce the readers to the composite and challenging area of the application of parallel programming in bioinformatics, giving a fresh view of some state-of-the-art solutions from interesting points of view. Before concluding we want to express our sincere gratitude to some people who have helped us in this challenge. First of all, we would like to thank Prof. Dr. Alexandru Nicolau (Editor-in-Chief of the International Journal of Parallel Programming), for trusting us. We also extend our sincere thanks to all the authors who submitted papers for this special issue and the many reviewers, whose dedicated efforts made this special issue possible.

Acknowledgements This work was partially funded by the AEI (State Research Agency, Spain) and the ERDF (European Regional Development Fund, EU), under the Contract TIN2016-76259-P (PROTEIN Project).

References

1. PBio: 5th International Workshop on Parallelism in Bioinformatics. <http://arco.unex.es/mavega/pbio/2017/> (2017). Accessed 6 Sept 2018
2. Rucci, E., Garcia-Sanchez, C., Botella-Juan, G., De Giusti, A., Naiouf, M., Prieto-Matias, M.: SWIMM 2.0: enhanced Smith–Waterman on Intel’s multicore and manycore architectures based on AVX-512 vector extensions. *Int. J. Parallel Program.* (2018). <https://doi.org/10.1007/s10766-018-0585-7>
3. Badosa, F., Espinosa, A., Acevedo, C., Vera, G., Ripoll, A.: A history-based resource manager for genome analysis workflows applications on clusters with heterogeneous nodes. *Int. J. Parallel Program.* (2018). <https://doi.org/10.1007/s10766-018-0600-z>

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.