Available online at www.sciencedirect.com
ScienceDirect
J. Parallel Distrib. Comput. 66 (2006) 1481
Journal of Parallel and Distributed Computing
www.elsevier.com/locate/jpdc

Guest editorial

Grids in bioinformatics and computational biology

Bioinformatics is fast emerging as an important discipline for academic research and industrial application. Research and development in bioinformatics and computational biology create and develop advanced information and computational techniques to manage and extract useful information from the DNA/RNA/protein sequence being generated by high-throughput technologies (e.g., DNA microarrays, DNA sequencers). It is the comprehensive application of mathematics (e.g., probability and graph theory), statistics, science (e.g., biochemistry), and computer science (e.g., computer algorithms and machine learning) to the understanding of living systems. Those techniques are extremely computationally or data intensive, providing motivation for using Grid technology [1].

Grids are an enabling technology that permits the transparent coupling of geographically dispersed resources (machines, networks, data storage, visualization devices, and scientific instruments) for large-scale distributed applications. Grids provide several important benefits for users and applications to share: computing and data storage, knowledge, instruments.

This special issue of the *Journal Parallel and Distributed Computing* (JPDC) is composed of original research papers that describe recent advances in using grids in bioinformatics and computational biology.

The first article addresses high-performance cDNA sequence analysis using Grid technology. The capability of handling high-throughput sequencing is becoming increasingly important in bioinformatics. This study concerns the development of a high-performance pipeline for analyzing cDNA sequences produced by a high-throughput pyrosequencer. The results are stored in an output database directly from the Grid sites using the web services technology.

The second paper presents a parallel evolution strategy for the well-known protein threading problem. The protein threading problem is the problem of determining the three-dimensional structure of a given but arbitrary protein sequence from a set of known structures of other proteins. Three parallel algorithms based on evolution strategies have been proposed. The experiments produce encouraging preliminary results in terms of threading energy as well as significant reduction in threading time.

The third paper directly focuses on designing a parallel framework for searching in large biological databases. Biological databases storing DNA sequences, protein sequences,or mass spectra are growing exponentially. The proposed frame work runs as a persistent service, processing all submitted queries. Superlinear speedups have been achieved using real biological databases and an actual searching algorithm for mass spectrometry.

The fourth paper presents an approach to construct large suffix trees on a computational grid. The suffix tree is a key data structure for biological sequence analysis, since it permits efficient solutions to many string based problems. The authors show that the distributed grid implementation leads to a significant run-time save.

The next paper deals with an application that involves magnetic resonance scanners for near real-time medical image processing. The objective is to improve patient care by enabling real-time, computational intensive medical image processing, directly at an MR scanner. The results indicate that real-time processing of medical imaging data on a shared HPC resource is reliable and possible in a clinically acceptable time.

The sixth paper also targets a widely researched topic that of large-scale multiple assignment. Indeed, multiple sequence alignment and phylogenetic tree construction are important in computational biology. A parallel method which is based on the distributed caching of intermediate results has been proposed. Preliminary results using actual biological data have been reported.

The final paper deals with rRNA probe design. It is a significant computational challenge task in view of the continuing rapid increase in the number of available 16S ribosomal RNA (RRNA). A fast software tool, named ProkProbePicker, has been presented. A parallel version of this algorithm is described. A linear speedup has been obtained which revealed the outstanding scalability of the parallelized version of the algorithm.

Reference

[1] Zomaya, A.Y. (ed.), *Parallel Computing for Bioinformatics and Computational Biology*, Wiley, New York.

Co-Guest Editors
El-Ghazali Talbi
LIFL-INRIA Futurs-University of Lille, France
E-mail address: talbi@lifl.fr
*URL:*http://www.lifl.fr/∼ talbi

Albert Zomaya
University of Sydney, Australia
*E-mail address:*zomaya@it.usyd.edu.au
*URL:*http://www.cs.usyd.edu.au/∼ zomaya

0743-7315/$ - see front matter
doi:10.1016/j.jpdc.2006.09.001