HiBB 2012: 3^{rd} Workshop on High Performance Bioinformatics and Biomedicine

Mario Cannataro

Bioinformatics Laboratory,
Department of Medical and Surgical Sciences,
University Magna Græcia of Catanzaro,
88100 Catanzaro, Italy
cannataro@unicz.it

Foreword

High-throughput technologies (e.g. microarray and mass spectrometry) and clinical diagnostic tools (e.g. medical imaging) are producing an increasing amount of experimental and clinical data, yielding to the so called age of Big Data in Biosciences. In such a scenario, large scale databases and bioinformatics tools are key tools for organizing and exploring biological and biomedical data with the aim to discover new knowledge in biology and medicine. However the storage, preprocessing and analysis of experimental data is becoming the main bottleneck of the biomedical analysis pipeline.

Parallel computing and high-performance infrastructures are more and more used in all phases of life sciences research, e.g. for storing and preprocessing large experimental data, for the simulation of biological systems, for data exploration and visualization, for data integration, and for knowledge discovery.

The current bioinformatics scenario is characterized by the application of well established techniques, such as parallel computing on multicore architectures and grid computing, as well as by the application of emerging computational models such as graphics processing and cloud computing.

Large scale infrastructures such as grids or clouds are mainly used to store in an efficient manner and to share in an easy way, the huge experimental data produced in life sciences, while parallel computing allows the efficient analysis of huge data. In particular, novel parallel architectures such as GPUs and emerging programming models such as MapReduce, may overcome the limits posed by conventional computers to the analysis of large amounts of data.

The third edition of the Workshop on *High Performance Bioinformatics and Biomedicine* (HiBB) aimed to bring together scientists in the fields of high performance computing, computational biology and medicine to discuss the parallel implementation of bioinformatics algorithms, the deployment of biomedical applications on high performance infrastructures, and the organization of large scale databases in biology and medicine. As in the past, also this year the workshop has been organized in conjunction with Euro-Par, the main European (but international) conference on all aspects of parallel processing.

I. Caragiannis et al. (Eds.): Euro-Par 2012 Workshops, LNCS 7640, pp. 217–219, 2013.

The workshop included the invited talk *Using Clouds for Scalable Knowledge Discovery Applications* that was held by Prof. Domenico Talia (University of Calabria and ICAR-CNR, Italy), and presentations were organized in three sessions: *High Performance Bioinformatics and Systems Biology*, *Software Platforms for Bioinformatics and Biomedicine*, and *Grid and Cloud Computing for Life Sciences*.

The invited talk *Using Clouds for Scalable Knowledge Discovery Applications* discussed the use of clouds for the development of scalable distributed knowledge discovery applications for life sciences, and presented a software framework for implementing high-performance data mining applications modelled through workflows on clouds.

The first session, *High Performance Bioinformatics and Systems Biology*, comprised three papers discussing the parallel implementation of bioinformatics and systems biology algorithms.

In the first paper, *P3S: Protein Structure Similarity Search*, Galgonek et al. presented the P3S web-based tool for similarity search in protein structure databases, that uses an efficient indexing technique and the SProt specialized similarity measure.

A second paper by Galgonek et al., *On the Parallelization of the SProt Measure and the TM-score Algorithm*, discussed the parallel implementation of the SProt similarity measure.

In the third paper, *Stochastic Simulation of the Coagulation Cascade: A Petri Net Based Approach*, Castaldi et al. presented a Stochastic Petri Net (SPN) based model used to introduce uncertainty to capture the variability of biological systems. The authors described the SPN model for the Tissue Factor induced coagulation cascade and simulated it using Gillespieś Tau-Leaping Stochastic Simulation, proving that a stochastic approach allows to detect important features that deterministic models cannot discover.

The second session, *Software Platforms for Bioinformatics and Biomedicine*, comprised three papers describing experiences on using Grid and Cloud computing as well as multicore and GPUs architectures, to implement biomedical data analysis.

In the first paper, *Processing the biomedical data on the grid using the UNICORE workflow system*, Borcz et al. described the use of the UNICORE middleware to automate the analysis pipeline in biomedical applications, with focus on the determination of the spectrum of mutations in mitochondrial genomes of normal and colorectal cancer cells, using the PL-Grid Polish National Grid Infrastructure.

In the second paper, *Multicore and Cloud-based Solutions for Genomic Variant Analysis*, Gonzalez et al. introduced the High Performance Genomics (HPG) project that is developing a collection of open-source tools for genomics data, and presented the HPG Variant, a suite for genomic variant analysis that allows to conduct genomic-wide and family-based analysis.

In the third paper, *A Novel Implementation of Double Precision and Real Valued ICA Algorithm for Bioinformatics Applications on GPUs*, Foshati et al.

presented the parallel implementation on a GPU using the CUDA programming toolkit, of an ICA (Independent Component Analysis) algorithm named Joint Approximate Diagonalization of Eigen-matrices (JADE), for the analysis of EEG signals.

The third session, *Grid and Cloud Computing for Life Sciences*, included the aforementioned invited talk, a paper describing an environment for scientific workflows on heterogeneous grids, and a panel on *High Performance Bioinformatics and Biomedicine*.

In particular, in the paper *PROGENIA: an approach for Grid Interoperability at Workflow level*, Mirto et al. presented the PROGENIA Workflow Management System that supports interoperability at workflow level. The presented system allows the deployment of scientific workflows on different grids based on Globus, gLite, and Unicore middlewares, and has been tested on several bioinformatics case studies.

The workshop was concluded by a panel on *High Performance Bioinformatics and Biomedicine* chaired by Mario Cannataro. The goal of the panel was to discuss emerging topics and problems regarding the application of High Performance Computing techniques, such as Grid and Cloud Computing, to biomedical, bioinformatics, and systems biology applications.

The three panelists, Francesco Archetti, Piotr Bala, and Domenico Talia, were asked to answer the following questions:

- The role of systems biology to speed up the development of new drugs.
- Role of grids and clouds in high performance bioinformatics and biomedicine.
- Integration and analysis of omics and clinical data.

This post-workshop proceedings includes the final revised versions of the HiBB papers and invited talk, taking the feedback from reviewers and workshop audience into account.

The program chair sincerely thanks the Program Committee members and the following additional reviewers: Giuseppe Agapito, Marco Aldinucci, Barbara Calabrese, Maria Mirto, and Simona Rombo, for the time and expertise they put into the reviewing work, the Euro-Par organization, for providing the opportunity to arrange the HiBB workshop in conjunction with the Euro-Par 2012 conference, and all the workshop attendees who contributed to a lively day.

October 2012

Mario Cannataro