



9th Workshop on Biomedical and Bioinformatics Challenges for Computer Science – BBC2016

Stefano Beretta¹, Mario Cannataro², and Mauro Castelli³

¹ Dept. of Informatics, Systems and Communication,
University of Milano - Bicocca, Milano - Italy
stefano.beretta@disco.unimib.it

² Bioinformatics Laboratory, Department of Medical and Surgical Sciences,
University Magna Græcia of Catanzaro, Catanzaro - Italy
cannataro@unicz.it

³ NOVA IMS, Universidade Nova de Lisboa, Lisboa - Portugal
mcastelli@novaims.unl.pt

1 Introduction

Computer science plays a fundamental role in many biological fields, mainly due to the advent of new technologies that produce a vast amount of data which requires a continuous development of new technologies and methods for their analysis. Research areas such as bioinformatics and computational biology have been deeply influenced by this revolution, and several challenges, which require the adoption of computer science methods for their solution, have emerged. More precisely, bioinformatics has started to play a central role in several fields like genomics, transcriptomics, interactomics, and metagenomics, mainly thanks to the introduction of new sequencing technologies which have, on the one hand changed the nature of the data to be analyzed and, on the other hand, opened new directions of research. In order to tackle these new needs, novel methods and approaches, requiring the knowledge of both theoretical and applied aspects of computer science, have been developed. Moreover, although the main focus of computational biology was on modelling and simulating biological processes, the advent of the developed technologies, contributed to make these two disciplines closer and related each other.

Several areas of computer science are employed to tackle the emerging challenges in the life science domain, ranging from high-performance computing to theoretical aspects, or from artificial intelligence to data structures. In particular, these new approaches are likely to involve computational methods, technologies, and infrastructures, such as:

- high-performance architectures and systems (e.g., multicore, GPU);
- distributed computing (e.g. grid, cloud, peer-to-peer);
- computational simulation (mechanistic, stochastic, multi-model);

- algorithms (theoretical and experimental aspects);
- applied bioinformatics (analysis pipelines, tools, applications);
- artificial and computational intelligence (machine learning, agents, evolutionary techniques, bio-inspired methods).

The ninth edition of the *Workshop on Biomedical and Bioinformatics Challenges for Computer Science* (BBC) took place between 6-8 June 2016. San Diego, California (USA), was the setting, and the Catamaran Resort Hotel and Spa was the venue.

The main aim of this renowned international workshop was to highlight recent approaches to face new emerging problems in fields of computer science, bioinformatics and computational biology, as well as to discuss future research directions in these fields. As in the past, also this year the workshop has been organized in conjunction with the *International Conference on Computational Science* (ICCS).

The first session, which is dedicated to *Biomedical* works, comprised four full papers and one abstract. In the first paper *Supermodeling in simulation of melanoma progression*, Witold Dzwiniel et al. propose the use of a supermodeling strategy for simulating cancer proliferation. The approach creates supermodels by coupling models of single cells to model the growth of a melanoma, which was used as a case study also to check the adaptive flexibility of the system. Moreover, thanks to the synchronization of sub-models, the supermodel was shown to be able to simulate qualitatively different modes of cancer growth than those observed for a single model. In the paper *Modeling cumulative biological phenomena with Suppes-Bayes causal networks*, Daniele Ramazzotti et al. present a theoretical framework underlying the Suppes-Bayes Casual Network (SBCN), a model that has been previously applied by the same group to infer cancer progression. The model is shown to be effective for cumulative phenomena. An investigation on the effects of the priors based on Suppes' theory and maximum likelihood regularization parameters on the performance of the CAPRI method has also been discussed. A case study on simulated data is also presented in order to validate the proposed approach and also to assess its performances. In the paper *Targeting deep brain regions in transcranial electrical neuromodulation using the reciprocity principle*, Mariano Fernandez-Corazza et al. explore the problem of targeting deep regions in the brain, by means of computer simulations within a finite element head model constructed from high resolution MRI and CT volumes. In the paper *CFD investigation of human tidal breathing through human airway geometry*, Jamasp Azarnoosh et al. provides a very interesting analysis of the airflow through the human respiratory tract. It presents an analysis of the effect of the extra-thoracic airways on the flow field through the lower airways by means of CFD simulations of the airflow through the human respiratory tract. In the paper *Forward Error Correction for DNA Data Storage*, Meinolf Blawat et al. presents a way to encode information in a synthetic DNA sequence. The authors argue that their forward error correction scheme is able to cope with all error types of today's DNA synthesis, amplification and sequencing processes. This approach has been also tested on a real experiment in which a binary data of 22 MB has been encoded in a DNA sequence and then decoded.

The second session, which is dedicated to *Bioinformatics* works, comprised four full papers and one poster. In the first paper *Generating a 3D Normative Infant Cranial Model*, Binhang Yuan et al. describes an algorithm to generate a normative model of infant's head shape from CT scans of normal infant skulls. A mesh creation, registration, and a normal model which is obtained from CT scans, is discussed with the goal of helping the evaluation of the efficacy of plastic surgeries in a more objective way. More precisely, a correspondence map between meshes based on a registration algorithm was generated, and this averaging algorithm was applied to construct the normative model. In the paper *Denormalize and Delimit: How not to Make Data*

Extraction for Analysis More Complex than Necessary, by Alex Bokov et al. a method to help bio-statisticians in speeding up the data extraction process is described. In particular, this method is for denormalizing data from the approximations of 2NF and 3NF to a single table in 1NF for support to statistical analysis. Finally, the authors describe a tool called DataFinisher, which is part of the i2b2 Query and analysis system, and it is able to reduce complex data to the simple 1NF format, which can be easily read by multiple statistical tools. In the paper *Computationally characterizing genomic pipelines and benchmarking results using GATK best practices on the high performance computing cluster at the University of Kentucky*, Xiaofei Zhang and Sally Ellingson present some experiments in variant calling from NGS data using GATK which is the most widely used pipeline so far. The results were obtained by executing the experiments on the high-performance computing cluster at the University of Kentucky, and analyze different aspects. In terms of accuracy, two variations of the GATK pipeline (UG vs. HC) are assessed, while in terms of computational efficiency, the two pipelines in Single Threading and Multiple Threading are compared. Finally, the incidence of the choice of the evaluation tool and of the assumed gold standard variant calls is discussed. In the paper *Partitioning of arterial tree for parallel decomposition of hemodynamic calculations*, Andrew Svitenkov et al. describes the use of graph partitioning to decompose the human arterial network. More precisely, the detailed model of full human arterial tree and some simpler geometries are discussed. Some graph partitioning strategies are described and applied, and their results are compared. Finally, in the short paper *Novel Allosteric Sites of Insulin-Degrading Enzyme Identified through Applied Structural Bioinformatics Analysis*, Suryani Lukman describes the application of several bioinformatics techniques to analyze the structural dynamics of the Insulin-degrading enzyme (IDE). Novel druggable sites that are specific to particular conformational states of IDE, which can potentially be explored for designing investigative probes or therapeutic agents for specific spatiotemporal contexts, have also been identified.

This workshop proceedings includes the final revised versions of the BBC papers, taking the feedback from reviewers and workshop audience into account. The program chair sincerely thanks the Program Committee members for the time and expertise they put into the reviewing work, the ICCS organization, for providing the opportunity to arrange the BBC workshop in conjunction with the ICCS 2016 conference, and all the workshop attendees.

June 2016

Stefano Beretta
Mario Cannataro
Mauro Castelli