

Selected Papers of the First International Conference on Algorithms for Computational Biology (AlCoB 2014)

Adrian-Horia Dediu and Carlos Martín-Vide

Abstract—This special section of the *IEEE/ACM Transactions on Computational Biology and Bioinformatics* contains extended versions of the best papers presented at the First International Conference on Algorithms for Computational Biology (AlCoB 2014). Out of 39 submissions to the conference, only four papers representing the current state-of-the-art in their respective domains were accepted to this special section.

Index Terms—Analysis of algorithms and problem complexity, cutting planes, endoplasmic reticulum, heuristics design, machine learning, network dynamics, next generation sequencing, signal processing systems

1 FOREWORD

THIS special section of *IEEE/ACM Transactions on Computational Biology and Bioinformatics* contains extended versions of some of the best papers presented at the First International Conference on Algorithms for Computational Biology (AlCoB 2014),¹ held in Tarragona, Spain, on July 1-3, 2014, under the organization of Rovira i Virgili University.

During the last decades, there has been an explosion of available data in computational biology allowing new insights into genetic and infectious diseases, cancer, basic biology, and migration patterns, even for human communities. The researchers focus on new methods to analyze the existing data aiming to discover correlations capable to (partially) explain the underlying biological processes. In this context, AlCoB 2014 covered several of the current challenges in computational biology by encouraging, discussing, and publishing the most recent algorithms, capabilities, complexity aspects, and limitations, as well as their implications for several domains like:

- Molecular sequence analysis,
- Recognition of genes and regulatory elements,
- Molecular evolution,
- Biological networks,
- Sequencing and genotyping technologies,
- Computational proteomics,
- Molecular structural biology,
- Computational genomics,
- Evolutionary phylogeny,
- etc.

Out of 39 submissions to the conference, the program committee decided to accept 20 papers (which represents an

acceptance rate of approximately 51%). Among them, there were eight papers for which the authors submitted extended versions to this special section. Each submission was reviewed by two or three referees. After several revisions, four papers were accepted (yielding an acceptance rate of approximately 10% of the submissions to the conference). Let us introduce the papers and briefly present the main achievements of our special section.

2 BRIEF CONTENT

Thiago da Silva Arruda, Ulisses Dias, and Zanoni Dias study inversion events which occur when a segment of DNA sequence in the genome is reversed. Based on a meta-heuristic method called Greedy Randomized Adaptive Search Procedure (GRASP), the authors propose an algorithm to find the optimal sequence of events producing the minimum number of global operations that transform a given genome into another. The length-weighted inversion distance that gives the number of inversions needed to sort a permutation into another is an optimization problem. The algorithm minimizing the length-weighted inversion distance proposed by the authors is capable to infer phylogenies from genomic changes. A comparative analysis is followed by an example of phylogenetic trees for a subset of species in the *Yersinia* genus. The new algorithm shows better results than existing distances or other algorithms minimizing the length-weighted inversion distance. We find more details in the paper entitled “A GRASP-Based Heuristic for the Sorting by Length-Weighted Inversions Problem.”

Ivo Hedtke, Ioana Lemnian, Ivo Grosse, and Matthias Müller-Hannemann study the problem of read trimming for the next generation sequencing (NGS) data, in the paper “Optimal Block-Based Trimming for Next Generation Sequencing.” The goal of block-based trimming is to truncate low-quality positions at both ends of the reads and to remove low-quality reads. The authors formulate three constrained optimization problems corresponding to popular variants of block-based trimming and show that the three problems are

1. For more information, please check the site of the conference: <http://grammars.grlmc.com/AlCoB2014/>

• A.-H. Dediu is with Superdata, Bucharest, Romania.
E-mail: adrian-horia.dediu@superdata.ro.
• C. Martín-Vide is with Rovira i Virgili University, Tarragona, Spain.

For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TCBB.2018.2804398

NP-hard. They also propose three relaxed problems by omitting one set of constraints from each of the three hard problems, providing polynomial time approximations that show improved performance to the traditionally used heuristics for reads trimming. An extensive validation concludes the paper.

The paper entitled “Modeling the Geometry and Dynamics of the Endoplasmic Reticulum Network” by Congping Lin, Laurent Lemarchand, Reinhardt Euler, and Imogen Sparkes proposes a mathematical model to reconstruct geometric network dynamics for Endoplasmic Reticulum (ER) network formation. The authors consider the very interesting problem of learning a generative model for the planar graph structure of 2D ER tubule networks. This problem is fundamental, and having such a generative model would be useful in a variety of applications ranging from image analysis (Bayesian prior) to quantification of ER perturbations. The network topology in the model is determined by a modified optimization procedure which minimizes the total length taking into account both degree and angle constraints, beyond the conditions of connectedness and planarity. A novel feature for solving this optimization problem is the use of “lifted” angle constraints, which allows considerably reduced solution runtimes.

Amina Noor, Aitzaz Ahmad, and Erchin Serpedin propose a variant of the well known Network Component Algorithm (NCA) for matrix factorization exploiting sparsity in the connectivity network, in their paper entitled “SparseNCA: Sparse Network Component Analysis for Recovering Transcription Factor Activities with Incomplete Prior Information.” NCA is an important method for inferring transcriptional regulatory networks (TRNs) and recovering transcription factor activities (TFAs) using gene expression data, and the prior information about the connectivity matrix. Although modeling the transcriptional gene regulation is a very complex task, which depends on a large number of unknown variables of the system, certain characteristics can be estimated by relatively simpler models. The proposed algorithm is rigorously compared to the existing state-of-the-art algorithms both in terms of estimation accuracy and consistency using synthetic data as well as real data. The new algorithm having a lower computational complexity, it outperforms the results of other previous NCA algorithms.

Adrian-Horia Dediu
Carlos Martín-Vide
Guest Editors

ACKNOWLEDGMENTS

The guest editors acknowledge the conference system provided by EasyChair: they managed with it the submissions, the reviewing process, and the preparation of both the conference papers and the extended versions presented in this special issue. They also thank all the authors for their contributions as well as the editorial team of *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, IEEE Publishing Operations, for their excellent work. They address special thanks to all of the reviewers who helped us in producing this special section.



Adrian-Horia Dediu received the first PhD degree in computer science from the University Politehnica of Bucharest, Romania, in 2001, with the thesis entitled “Genetic Algorithms for Behaviour Simulation,” and also received the second PhD degree from Rovira i Virgili University, Tarragona, Spain, in 2015, with the thesis “Learning Automata with Help.” Currently, he is CEO of SuperData Bucharest, with overall responsibility for application design and intelligent data analysis. His research interests include: algorithms and programming languages, learning theory, adaptive and self-organizing systems, natural language processing, knowledge representation, and reasoning.



Carlos Martín-Vide is a professor with Rovira i Virgili University, Tarragona, Spain. His areas of interest are automata and language theory, mathematical linguistics, and molecular computing, where he has published more than 300 papers. He was principal investigator of a number of research projects funded by European and Spanish sources. He has been involved in numerous activities of research evaluation in the European Commission and other national and international bodies. He has been the promotor of several series of conferences and training schools in theoretical computer science, natural computing, natural language processing, bioinformatics, data science, and machine learning. He is a member of the editorial boards of *Information and Computation*, and *Soft Computing*.

▷ For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.