

Guest Editors' Introduction to the Special Section on Bioinformatics Research and Applications

Ion I. Mandoiu, Jianxin Wang, and Alexander Zelikovsky



THIS special section includes a selection of papers presented at the Eighth International Symposium on Bioinformatics Research and Application (ISBRA), which was held in Dallas, Texas, on 21-23 May 2012. The ISBRA symposium provides a forum for the exchange of ideas and results among researchers, developers, and practitioners working on all aspects of bioinformatics and computational biology and their applications. In 2012, 66 papers were submitted in response to the call for papers, out of which 26 papers appeared in the ISBRA proceedings published as volume 7292 of Springer Verlag's Lecture Notes in Bioinformatics series. Extended versions of nine symposium papers were invited and accepted for publication in this special section following a rigorous review process. The selected papers cover a broad range of bioinformatics topics, including biological networks, computational complexity of problems in structural biology and genomics, and phylogenetic inference and analysis. Below, we briefly introduce each of them.

The first paper by Xi Chen, Jianhua Xuan, Chen Wang, Ayesha N. Shajahan, Rebecca B. Riggins, and Robert Clarke describes a stability-based network component analysis (sNCA) to uncover regulatory networks by integrating gene expression data and binding motif information. The proposed sNCA approach overcomes a major limitation in many existing methods as caused by the inconsistency between gene expression data and TF-gene binding information. Experimental results on simulated and biological data show that sNCA achieves an improved and robust performance in TF identification as compared to the previously published NCA method and can help identify biologically meaningful regulatory networks associated with the development and progression of breast cancer.

In the second paper, Matteo Re and Giorgio Valentini address an important problem in medical informatics, namely, identifying novel effects of existing therapeutic drugs by integrating heterogeneous types of network data on drug structure and effects. The authors explore various strategies for integrating such heterogeneous networks into

a unified model and propose a novel scheme based on kernelized scoring functions for measuring similarity in these networks. Extensive cross-validation experiments are conducted on a data set of more than 1,200 FDA approved drugs to assess the effectiveness of various network integration approaches and scoring functions.

The protein chain pair simplification problem (CPS-3F) was proposed in 2008 to simultaneously simplify both chains with respect to each other under the discrete Fréchet distance. It is unknown whether CPS-3F is NP-complete. In the third paper of the special section, Timothy Wylie and Binhai Zhu define a constrained CPS-3F problem (CPS-3F+), and prove that it is polynomially solvable by presenting a dynamic programming solution. The authors also compare the CPS-3F+ solutions with previous empirical results and demonstrate some of the benefits of the simplified comparisons.

In the fourth paper, Laurent Bulteau and Minghui Jiang explore the approximation complexity (1, 2)-exemplar distance between genomes. Computing this distance from a genome without duplicate genes to the genome with at most one duplicate for each gene requires keeping just one copy of each gene in the second genome so as to minimize the distance between the two reduced genomes according to some measure. It is shown that this problem is hard to approximate for popular genome rearrangement measures (such as breakpoint distance, signed reversals, and signed double-cut-and-joins) and edit distance measures (such as Levenshtein and Hamming distances). They notice that no constant-factor approximation is known for any of these measures.

The fifth paper by Daniele Catanzaro, Martine Labbé, and Bjarni Halldórsson addresses the Parsimonious Loss of Heterozygosity Problem (PLOHP), which seeks to detect deletion polymorphisms based on unexpected patterns of loss of heterozygosity in a population. More specifically, PLOHP asks for partitioning putative polymorphisms of a set of individuals into the minimum number of deletion areas. The inferred deletion areas could provide fundamental insights for the development of new diagnostics and treatments. The paper presents a generalized PLOHP formulation which incorporates different recombination models and prior knowledge about the locations of deletions and reduces it to the clique partition problem in undirected catch-point interval graphs. Although the problem is shown to be NP-hard, a proposed integer programming formulation can exactly solve real instances in practical time.

The last four papers address problems in evolutionary biology. The paper by Yann Christinat and Bernard Moret is

- I. Mandoiu is with the Computer Science and Engineering Department, University of Connecticut, 371 Fairfield Rd., Unit 4155, Storrs, CT 06269-4155. E-mail: ion@engr.uconn.edu.
- J. Wang is with the School of Information Science and Engineering, Central South University, Changsha, 410083, Hunan, China. E-mail: jxwang@mail.csu.edu.cn.
- A. Zelikovsky is with the Computer Science Department, Georgia State University, Atlanta, GA 30303-4110. E-mail: alexz@cs.gsu.edu.

For information on obtaining reprints of this article, please send e-mail to: tcbb@computer.org.

devoted to modeling transcript evolution. They present a new model and a tool, called TrEvoR, for reconstructing transcript evolution viewed as an additional layer above a well-known exon/intron evolution model. The proposed two-level model is shown to perform better than the standard gene-centric one. It is also argued that transcript phylogenies may be used in studies on function evolution since conserved transcripts are more likely to share protein domains than functional sites.

The next paper, by Si Li, Kwok Pui Choi, Taoyang Wu, and Louxin Zhang, presents a likelihood-based approach to reconstruct the evolutionary history of protein-protein interaction (PPI) networks, incorporating in their model both the topology of the network and information about the duplication relationships among paralogs. The authors propose fast algorithms extending over previous work by exploiting known duplication histories given as duplication forests. Empirical experiments on simulated data show that the proposed approach is superior to a previous maximum likelihood algorithm, NetArch. The authors also provide results obtained by applying the proposed method on real PPI networks of yeast, fruitfly, and worm.

The paper by Ayshwarya Subramanian, Stanley Shackney, and Russell Schwartz addresses the challenging problem of inferring cancer markers that could be used as signatures of cancer progression. The authors propose an HMM-based approach for inferring such phylogenetic markers. Experiments on both synthetic and real data show that the proposed method improves upon previous methods and is highly effective in identifying phylogenetically informative markers.

Finally, the paper by Andre Wehe, J. Gordon Burleigh, and Oliver Eulenstein introduces knowledge-enhanced phylogenetic problems for both supertree and supermatrix phylogenetic analyses. These problems seek an optimal phylogenetic tree that can only be assembled from a user-supplied set of, possibly incompatible, phylogenetic relationships. A knowledge-enhanced search heuristic approach is proposed to be applied to discrete character data sets using the maximum parsimony (MP) objective. Although the proposed approach is not guaranteed to find exact solutions, it can improve upon parsimony solutions from commonly used MP heuristics.

We would like to thank the program committee members and external reviewers for volunteering their time to review the submissions to the symposium and the special section. We would also like to thank Editor-in-Chief Ying Xu for continuing to provide us with the opportunity to disseminate a selection of the exciting research presented at ISBRA to the wide readership of the *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. Last, but not least, we would like to thank all ISBRA authors—the symposium could not continue to thrive without their high-quality contributions.

Ion I. Mandoiu
Jianxin Wang
Alexander Zelikovsky
Guest Editors



Ion I. Mandoiu received the MS degree from Bucharest University in 1992 and the PhD degree from the Georgia Institute of Technology in 2000, both in computer science. He holds the United Technologies Corporation Associate Professorship in Engineering Innovation in the Computer Science and Engineering Department at the University of Connecticut. His main research interests are in the areas of bioinformatics and computational genomics, with a

special focus on the development of high-performance computational methods for the analysis of high-throughput sequencing data. He has published more than 90 refereed articles in journals and conference proceedings and eight book chapters. He has also coedited eight conference proceedings and the book *Bioinformatics Algorithms: Techniques and Applications*, published in the Wiley Book Series on Bioinformatics. His research program has been funded by federal grants from the US National Science Foundation (NSF), NIH, and USDA, including a prestigious US NSF Faculty Early Career Development (CAREER) award. Dr. Mandoiu serves as an associate editor for *BMC Bioinformatics* and is on the editorial board of *In Silico Biology* and the *International Journal of Bioinformatics Research and Applications*. He has served as a guest editor for more than 12 special issues in *BMC Bioinformatics*, *BMC Genomics*, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, and *IEEE Transactions on Nano-Bioscience*. He has also served as program committee chair for numerous international conferences.



Jianxin Wang received the BS, MS, and PhD degrees in computer science from Central South University of Technology, P.R. China. Currently, he is the vice dean and a professor in School of Information Science and Engineering, Central South University, Changsha, Hunan, P.R. China. He is currently serving as executive editor of the *International Journal of Bioinformatics Research and Applications* and is serving on the editorial boards of the *International Journal of Data Mining*

and *Bioinformatics*. He has also served as a program committee member for many international conferences. He was a program committee cochair for the Seventh and Eighth International Symposium on Bioinformatics Research and Applications (ISBRA'11 and ISBRA'12), will be a program committee cochair for the Eighth International Frontiers of Algorithmics Workshop (FAW'14) and 10th International Symposium on Bioinformatics Research and Applications (ISBRA'14). His current research interests include algorithm analysis and optimization, parameterized algorithm, bioinformatics, and computer network. He has published more than 200 papers in various International journals and refereed conferences. He is a senior member of the IEEE.



Alexander Zelikovsky received the PhD degree in computer science from the Institute of Mathematics of the Belorussian Academy of Sciences in Minsk, Belarus, in 1989 and worked at the Institute of Mathematics in Kishinev, Moldova, from 1989-1995. Between 1992 and 1995, he visited Bonn University and the Institut für Informatik in Saarbrücken, Germany. Dr. Zelikovsky was a research scientist at the University of Virginia from 1995-1997 and a

postdoctoral scholar at the University of California, Los Angeles from 1997-1998. He is a professor in the Computer Science Department at Georgia State University, which he joined in 1999. His research interests include bioinformatics, discrete and approximation algorithms, combinatorial optimization, VLSI physical layout design, and ad-hoc wireless networks. He is the author of more than 190 refereed publications and coeditor of four books. Dr. Zelikovsky received the SIAM Outstanding Paper Prize, the best paper award at the joint Asia-South Pacific Design Automation/VLSI Design Conferences, and the best poster awards at the Annual BACUS Symposium on Photomask Technology and the Fifth Georgia Tech International Conference on Bioinformatics. He is founding cochair of the ACIS International Workshop on Self-Assembling Wireless Networks (SAWN) and the International Symposium on Bioinformatics Research and Applications (ISBRA). He has also served on the editorial boards of six journals and as a guest editor for six special issues including four in IEEE transactions.