

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

Ion Mandoiu, Yi Pan, Raj Sunderraman, and Alexander Zelikovsky



THIS special section includes a selection of papers presented at the Fourth International Symposium on Bioinformatics Research and Application (ISBRA), which was held at Georgia State University in Atlanta, Georgia, on 6-9 May 2008. 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 2008, 94 papers were submitted in response to the call for papers, out of which 35 papers appeared in the ISBRA proceedings published as volume 4983 of Springer Verlag's Lecture Notes in Bioinformatics series.

A small number of authors were invited to submit extended versions of their symposium papers to this special section. Following a rigorous review process, five papers were selected for publication. The selected papers cover a broad range of bioinformatics topics, including multiple local sequence alignment methods, computational prediction of siRNA silencing efficacy, gene network models, microarray data analysis and inference, and reconstruction and analysis of phylogenetic trees.

The first paper by Treangen et al. presents a novel approach to identify interspersed repeats in genome sequences. Existing methods perform pairwise local sequence alignments to identify homologues, but these methods are not scalable and have limited accuracy. The method proposed in the paper uses a clever combination of a gapped extension heuristic and an efficient filtration technique to achieve greater accuracy in the identification of interspersed repeats. The proposed method is implemented and made available for download.

In the second paper, Qiu and Lane adapt the Support Vector Regression approach by considering multiple kernel functions to effectively predict siRNA silencing efficacy. Computational prediction of the initiator siRNA molecules can be of tremendous assistance to the scientist in the screening process before using them in biological experiments. The authors formulate the multiple kernel

learning function into a quadratically constrained quadratic programming problem, provide several heuristics, and empirically establish the superiority of their approach over current methods in accuracy, model complexity, and computational speed.

In the third paper, Park et al. employ gene network models in a novel manner to analyze microarray data to infer cancer progression. This approach considerably improves the estimates of evolutionary distance between tumors over methods that employ only gene expression profiles. They also present three variants of the gene network model approach: one that uses optimized best-fit networks, the second that uses sampling to infer high confidence subnetworks, and the third that uses modular networks inferred from clusters of similarly expressed genes. The three variants show excellent results on lung cancer and breast cancer microarray data.

The last two papers are devoted to advanced methods for the reconstruction and analysis of phylogenetic trees. The paper by Zhu et al. proposes a new way to define and analyze gene clusters and gene order. They show that the bandwidth parameter of a graph is tightly connected with the proposed parameterized definition of gene clusters and affects the number, size, and extent of preservation of identified clusters along phylogenetic trees. The latter property is computed using a new dynamic programming algorithm. The advantages of the proposed analysis methods are illustrated by application to a set of genomes drawn from the Yeast Gene Order Browser.

The paper by Bansal et al. is devoted to the problem of inferring a species supertree by reconciling gene trees, including those constructed for large families of duplicated genes, based on the duplication optimality criterion. The resulting optimization problem (commonly referred to as the gene-duplication problem) is NP-hard and practical solutions are frequently based on local search heuristics. In each step, these heuristics must find a phylogenetic tree that is optimal under the duplication optimality criterion in the *neighborhood* of the current tree, i.e., the set of trees that can be obtained from it by applying a variety of tree edit operations. The authors propose near-linear time algorithms for searching optimal trees within neighborhoods defined by the  $k$ -NNI (Nearest Neighbor Interchange) tree edit operation for  $k \in \{1, 2, 3\}$ . They validate their algorithms using sets of large randomly generated gene trees.

We would like to thank the Program Committee members and external reviewers for volunteering their

- I. Mandoiu is with the Computer Science and Engineering Department, University of Connecticut, 371 Fairfield Rd. Unit 2155, Storrs, CT 06269-2155. E-mail: ion@engr.uconn.edu.
- Y. Pan, R. Sunderraman, and A. Zelikovsky are with the Computer Science Department, Georgia State University, Atlanta, GA 30303-4110. E-mail: {pan, raj, alexz}@cs.gsu.edu.

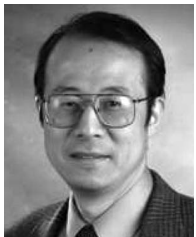
For information on obtaining reprints of this article, please send e-mail to: tccb@computer.org.

time to review the submissions to the symposium and the special section. We would also like to thank former Editor-in-Chief, Professor Dan Gusfield, as well as the current Editor-in-Chief, Dr. Marie-France Sagot, for continuing to provide us with the opportunity to showcase some of the exciting research presented at ISBRA in 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 Mandoiu  
Yi Pan  
Raj Sunderraman  
Alexander Zelikovsky  
*Guest Editors*



**Ion Mandoiu** received the MS degree from Bucharest University in 1992 and the PhD degree from Georgia Institute of Technology in 2000, both in computer science. He is an associate professor in the Computer Science and Engineering Department at the University of Connecticut, Storrs. His main research interests are in the design and analysis of approximation algorithms for NP-hard optimization problems, particularly in the areas of bioinformatics, design automation, and ad hoc wireless networks, areas in which he has authored more than 65 refereed journal and conference proceeding papers. He has also co-edited (with A. Zelikovsky) a book called *Bioinformatics Algorithms: Techniques and Applications* published in the Wiley Book Series on Bioinformatics. Dr. Mandoiu has served as a program committee chair for several international conferences and workshops, including the *International Symposium on Bioinformatics Research and Applications* and the *IEEE International Conference on Bioinformatics and Biomedicine*. He also serves on the editorial board of the *International Journal of Bioinformatics Research and Applications*, and has been a guest editor for the *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, the *IEEE Transactions on Nanobiosciences*, the *International Journal of Wireless and Mobile Computing*, and the *Journal of Universal Computer Science*. Dr. Mandoiu is a 2006 recipient of the US National Science Foundation Faculty Early Career Development Award.



**Yi Pan** received the BEng and MEng degrees in computer engineering from Tsinghua University, China, in 1982 and 1984, respectively, and the PhD degree in computer science from the University of Pittsburgh, Pennsylvania, in 1991. He is the chair and a professor in the Department of Computer Science and a professor in the Department of Computer Information Systems at Georgia State University. Dr. Pan's research interests include parallel and distributed computing, networks, and bioinformatics. Dr. Pan has published more than 100 journal papers with 37 papers published in various IEEE journals. In addition, he has published more than 100 papers in refereed conferences. He has also authored/edited 34 books (including proceedings) and contributed many book chapters. Dr. Pan has served as an editor-in-chief or editorial board member for 15 journals including five IEEE Transactions and a guest editor for 10 journals including the *IEEE/ACM Transactions on Computational Biology and Bioinformatics* and the *IEEE Transactions on NanoBioscience*. He has organized several international conferences and workshops and has also served as a program committee member for several major international conferences such as INFOCOM, GLOBECOM, ICC, IPDPS, and ICPP. Dr. Pan has delivered more than 10 keynote speeches at many international conferences and is a speaker for several distinguished speaker series. He is listed in *Men of Achievement*, *Who's Who in Midwest*, *Who's Who in America*, *Who's Who in American Education*, *Who's Who in Computational Science and Engineering*, and *Who's Who of Asian Americans*.



**Rajshekhar Sunderraman** received the BE (Honors) degree in electronics engineering from the Birla Institute of Technology and Science, Pilani, India, in 1980, the MTech degree in computer technology from the Indian Institute of Technology, Delhi in 1982, and the PhD degree in computer science from Iowa State University in 1988. Dr. Rajshekhar Sunderraman is currently a professor of computer science at Georgia State University in Atlanta. Dr. Sunderraman's research interests are in deductive databases, logic programming, modeling of incomplete and inconsistent information, semistructured data, and data modeling of biological data. He has published more than 100 research papers in leading computer science journals and conferences including the *ACM Transactions on Database Systems*, the *IEEE Transactions on Knowledge and Data Engineering*, and the *IEEE Conference on Data Engineering*. He is the author of a leading textbook called *Oracle 10g Programming: A Primer* published by Addison Wesley.



**Alexander Zelikovsky** received the PhD degree in computer science from the Institute of Mathematics at the Belorussian Academy of Sciences in Minsk, Belarus, in 1989 and worked at the Institute of Mathematics in Kishinev, Moldova, from 1989 to 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 to 1997 and a postdoctoral scholar at the University of California, Los Angeles, from 1997 to 1998. He is an associate professor in the Computer Science Department at Georgia State University, which he joined in 1999. Dr. Zelikovsky's 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 150 refereed publications and coeditor of one book. Dr. Zelikovsky is a 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 five journals and has been a guest editor for five special issues including three in IEEE Transactions. 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.