

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

Ion I. Măndoiu, Yi Pan, and Alexander Zelikovsky

THIS special section includes a selection of papers presented at the Second International Workshop on Bioinformatics Research and Application (IWBRA '06). The workshop was held on 28-31 May 2006 in Reading, United Kingdom, as part of the 2006 International Conference on Computational Science.

Authors were invited to submit papers that demonstrate original unpublished research in all areas of bioinformatics and computational biology and their applications. Ninety-three submissions were received in response to the call for papers. Following a rigorous review process, the program committee selected 35 of the submissions for oral presentation and publication in the workshop proceedings, published as part of the ICCS proceedings in volume 3515 of Springer-Verlag's Lecture Notes in Computer Science series. Extended versions of a small number of workshop papers were invited to this special section, of which six were selected for publication.

- The paper by Zheng et al. addresses an important problem arising in the comparison of genomic maps—constructing a set of conflict-free synteny blocks that contain as much as possible of the original data and can then be directly analyzed with genome rearrangement algorithms. The proposed strategy consists of 1) generating prestrips—common subsequences of two or more markers on single chromosomes in both genomes, 2) extracting mutually compatible prestrips containing a maximum number of markers using a Maximum Weighted Clique (MWC) search algorithm, and 3) adding markers not included in the MWC if they do not increase rearrangement distance between the two genomes. The authors use this strategy to analyze the rice and sorghum comparative map.
- The paper by Blin et al. explores the computational complexity of computing (dis)similarity measures between two genomes when they contain duplicated genes or genomic markers. Recently proposed

methods rely on 1) a model of one-to-one correspondence between genes (in the exemplar model, all but one copy are deleted for each gene family and, in the matching model, a maximal correspondence for each gene family is computed) and 2) computing a correspondence resulting from an optimal permutation for the given (dis)similarity measure. The problem of computing an optimal correspondence is shown to be NP-complete for three (dis)similarity measures on permutations: the number of common intervals, the maximum adjacency disruption (MAD) number, and the summed adjacency disruption (SAD) number. For the MAD and SAD number measures, the problem is further shown to be APX-hard.

- The paper by Bonizzoni et al. investigates the computational and approximation complexity of the Exemplar Longest Common Subsequence (ELCS) problem. This problem generalizes the Longest Common Subsequence problem by allowing symbols from a set of mandatory symbols and a set of optional symbols. Even for two sequences, the problem is shown to be NP-hard and APX-hard. For instances with two sequences, the authors give a polynomial-time algorithm for the case when each mandatory symbol appears at most three times and fixed-parameter algorithms where the parameter is the number of mandatory symbols.
- The paper by Davila et al. considers the planted (l, d) motif search problem which asks for a substring of length l that occurs with up to d errors in each one of a given number of sequences. A series of exact algorithms with reduced space requirements are proposed and implemented. These algorithms take less time on difficult instances that have been previously solved and can solve challenging instances such as $(17, 6)$ and $(19, 7)$, which previously could not be handled by any of the existing methods.
- The paper by Schneider et al. introduces SynPAM, a new method of estimating evolutionary distances between coding DNA sequences based on synonymous codon substitutions. The observed substitution pattern is analyzed using maximum-likelihood based on a 64×64 Markov model with transition rates estimated from empirical data. Comparisons with previous methods show that

• I.I. Măndoiu 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 and A. Zelikovsky are with the Computer Science Department, Georgia State University, Atlanta, GA 30303-4110. E-mail: {pan, alexz}@cs.gsu.edu.

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

SynPAM estimates have consistently lower variance and more often result in correct phylogenetic reconstruction of the vertebrate tree, particularly for longer evolutionary distances.

- The paper by Sridhar et al. addresses the problem of reconstructing phylogenetic trees using binary character states. The authors propose and implement algorithms which are fixed parameter tractable in the case of perfect phylogeny with a constant number of additional mutations, thus providing the first practical phylogenetic tree reconstruction algorithms that find guaranteed optimal solutions for data sets of biologically meaningful size and complexity.

We would like to thank all of the authors for their high-quality submissions and the IWBRA 2006 program committee and anonymous reviewers for volunteering their time and expertise in evaluating the scientific merits of submitted papers. We would also like to thank the Editor-in-Chief, Professor Dan Gusfield, for providing us with the opportunity to showcase, in this special section, a sampling of the exciting research presented at IWBRA 2006.

Ion I. Măndoiu
Yi Pan
Alexander Zelikovsky
Guest Editors



Ion I. Măndoiu 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. Between 2000 and 2003, Dr. Măndoiu was a postdoctoral researcher and then a research scientist at the University of California at Los Angeles and at the University of California at San Diego. Currently, he is an assistant professor with the Computer Science and Engineering Department at the University of Connecticut, Storrs. He is the author of more than 60 refereed journal and conference articles. 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. Dr. Măndoiu is a founding cochair of the ACIS International Workshop on Self-Assembling Wireless Networks (SAWN) and has served as program committee chair or cochair for several conferences, including the 2007 ACM/IEEE System Level Interconnect Prediction Workshop (SLIP), the 2007 International Symposium on Bioinformatics Research and Applications (ISBRA), and the 2007 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). He also serves on the editorial board of the *International Journal of Bioinformatics Research and Applications* and is a guest editor for the *IEEE Transactions on Nanobiotechnology*, the *International Journal of Wireless and Mobile Computing*, and the *Journal of Universal Computer Science*. He received the US National Science Foundation Faculty Early Career Development Award in 2006.



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 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. His research interests include parallel and distributed computing, optical networks, wireless networks, and bioinformatics. Dr. Pan has published more than 100 journal papers with more than 30 papers published in various IEEE journals. In addition, he has published more than 100 papers in refereed conferences (including IPDPS, ICPP, ICDCS, INFOCOM, and GLOBECOM). He has also coedited more than 30 books (including proceedings) and contributed several book chapters. His pioneer work on computing using reconfigurable optical buses has inspired extensive subsequent work by many researchers and his research results have been cited by more than 100 researchers worldwide in books, theses, journal and conference papers. He is a co-inventor of three US patents (pending) and five provisional patents and has received many awards from agencies such as the NSF, AFOSR, JSPS, IISF, and Mellon Foundation. His recent research has been supported by the NSF, NIH, NSFC, AFOSR, AFRL, JSPS, IISF, and the states of Georgia and Ohio. He has served as a reviewer/panelist for many research foundations/agencies such as the US National Science Foundation, the Natural Sciences and Engineering Research Council of Canada, the Australian Research Council, and the Hong Kong Research Grants Council. 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 Nano-Bioscience*. 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. He was an IEEE Distinguished Speaker (2000-2002), a Yamacraw Distinguished Speaker (2002), a Shell Oil Colloquium Speaker (2002), and a senior member of the IEEE. He is listed in *Men of Achievement*, *Who's Who in the 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*.



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 (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 (1995-1997) and a postdoctoral scholar at the University of California at Los Angeles (1997-1998). He is an associate professor in the Computer Science Department at Georgia State University, which he joined in 1999. He is the author of more than 140 refereed publications. Dr. Zelikovsky's research interests include bioinformatics, discrete and approximation algorithms, combinatorial optimization, VLSI physical layout design, and ad hoc wireless networks. He is a founding cochair of the ACIS International Workshop on Self-Assembling Wireless Networks (SAWN) and the International Workshop on Bioinformatics Research and Applications (IWBRA). He is program committee cochair of the 2007 International Symposium on Bioinformatics Research and Applications (ISBRA). He also serves on the editorial board of the *International Journal of Bioinformatics Research and Applications* and is a guest editor for the *LNCS Transactions on Computational Systems Biology*, the *IEEE Transactions on Nanobiotechnology*, the *International Journal of Wireless and Mobile Computing*, and the *Journal of Universal Computer Science*. Dr. Zelikovsky received the best paper award at the joint Asia-South Pacific Design Automation/VLSI Design Conferences in 2003 and the best poster awards at the Annual BACUS Symposium on Photomask Technology in 2005 and the Fifth Georgia Tech International Conference on Bioinformatics in 2005.