This special section includes a selection of papers presented at the Sixth International Symposium on Bioinformatics Research and Application (ISBRA), which was held at the University of Connecticut in Storrs on 23-26 May, 2010. 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 2010, 57 papers were submitted in response to the call for papers, out of which 20 papers appeared in the ISBRA proceedings published as volume 6053 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, six papers were selected for publication. The selected papers cover a broad range of bioinformatics topics, including protein structure alignment methods, comparison of biological networks, inference, reconstruction, and analysis of the evolutionary history.

The first paper, “A Spectral Approach to Protein Structure Alignment,” by Yosi Shibberu and Allen Holder gives a novel contact geometry description of protein folds that allows for scaling the distance matrix so that its eigenvalues are positive. This mathematical structure is used to develop two new protein fold alignment algorithms. The first algorithm is fast enough to identify folds in large data sets, while the second additionally incorporates the intrinsic geometry and the 3D geometry of a fold to make high-quality alignments.

In the second paper “Asymmetric Comparison and Querying of Biological Networks,” Nicola Ferraro, Luigi Palopoli, Simona Panni, and Simona E. Rombo propose a novel method for biological network alignment. It exploits differences in the characterization of species—the protein-protein interaction network of a better characterized species guides the alignment to the less characterized one via a finite automaton. The method allows for computing most meaningful protein pairings between networks as well as finding matching subgraphs. Experimental results show that all found matchings are biologically relevant and the technique is robust with respect to the misdefined or misplaced interactions occurred in databases.

The remaining four papers describe new methods in the area of reconstruction of biomolecular evolution. “The Plexus Model for the Inference of Ancestral Multidomain Proteins” by John Wiedenhoeft, Roland Krause, and Oliver Eulenstein introduces a graph-theoretic concept of a novel network-like structure, called plexus, which represents the evolution of protein domains and their combinations. They formulate an optimization problem modeling multidomain protein evolution and describe its solution. The application to empirical data sets results in inferring credible scenarios for the evolution of multidomain proteins and reveals inconsistencies in the initial phylogenetic trees.

“Uncovering Hidden Phylogenetic Consensus in Large Data Sets” by Nicholas D. Pattengale, Andre J. Aberer, Krister M. Swenson, Alexandros Stamatakis, and Bernard M.E. Moret presents a novel framework for defining rogue taxa (i.e., data with no well-defined place in the tree) that maximizes the relative information present in a consensus tree computed upon removing these rogue taxa. This framework leads to a bicriterion optimization problem balancing the loss of taxa with the gain in resolution. A greedy heuristic for finding an appropriate set of rogue data is proposed and tested on pathological cases selected both from the literature as well as from new biological data. Removing rogue data allows to obtain better bootstrap scores for multiple remaining edges. The proposed algorithm can handle large practical data sets.

“Extensions and Improvements to the Chordal Graph Approach to the Multistate Perfect Phylogeny Problem” by Ron Gysel and Dan Gusfield describes an improved chordal graph approach to the classic Multistate Perfect Phylogeny Problem. Triangulations of the partition intersection graph help to formulate integer linear program for finding minimum number characters whose removal will result in perfect phylogeny for the remaining data. This allows, for the first time, the solving of the missing data character removal problem for an arbitrary number of states. A novel speeding-up preprocessing is given for computing of the minimal graph separators also used for solving perfect phylogeny problem with missing data.

“A Consensus Tree Approach for Reconstructing Human Evolutionary History and Detecting Population Substructure” by Ming-Chi Tsai, Gyu Blloch, R. Ravi, and Russell Schwartz deals with the reconstruction of
human evolutionary history and detection of population substructure. The proposed method is based on the general consensus tree concept summarizing multiple likely phylogenies on millions of small regions spanning the human genome. Although each of them represent distorted versions of the “global” evolutionary history and population structure, the method can identify major splits or subdivisions between population groups supported by multiple trees. The optimization method based on the minimum description length model assembles a true evolutionary history model resistant to overfitting and to noise in the SNP data and provides a de novo inference of population subgroups comparable in quality to that provided by current best methods. 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 the Editor-in-Chief, Dr. Marie-France Sagot, for continuing to provide us with the opportunity for wider dissemination 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.

Mark Borodovsky
Teresa M. Przytycka
Sanguthevar Rajasekaran
Alexander Zelikovsky

Guest Editors

ACKNOWLEDGMENTS

The work of Teresa M. Przytycka is fully supported by intramural program at the National Library of Medicine, National Institutes of Health.

Mark Borodovsky received the MSc degree in physics and operation research as well as the PhD degree in applied mathematics from the Moscow Institute of Physics and Technology, Russia. He is director of the Center for Bioinformatics and Computational Genomics as well as a Regents’ Professor in the Wallace H. Coulter Department of Biomedical Engineering at the Georgia Institute of Technology (Georgia Tech) and Emory University and the School of Computational Science and Engineering at Georgia Tech. Dr. Borodovsky is also a founder of the Georgia Tech graduate program in bioinformatics and a chair of the Georgia Tech PhD program in bioinformatics. His research interests include bioinformatics and machine learning, with a focus on computational analysis of biological sequences, particularly, developing gene finding algorithms. He has published about 100 research papers, many of them in high impact factor journals such as Nature, Science, Proceedings of the National Academy of Sciences (PNAS), Genome Research, Nucleic Acids Research, and Bioinformatics, with more than 6,000 citations in total. His book Problems and Solutions in Biological Sequence Analysis (written together with Svetlana Ekisheva) was published by Cambridge University Press. He has served as a cochair for nine international conferences, was an invited and keynote speaker at a number of conferences and symposiums, and he also serves as a member of editorial boards of Advances in Bioinformatics, Bioinformatics Research and Applications, and the Journal of Biomedicine and Biotechnology. He is a member of Educational Committee of the International Society of Computational Biology.

Teresa M. Przytycka received the master’s degree from Warsaw University, Poland, and the PhD degree from the University of British Columbia, Vancouver. She is currently a senior investigator at the National Center for Biotechnology Information, National Institutes of Health, and is heading a research group focusing on developing algorithmic and graph theoretical approaches to study problems arising in computational and systems biology. She is particularly interested in the dynamical properties of biological systems, including spatial, temporal, and/or contextual variation and exploring how such variations are impacting gene expression, how they affect network topology, function, and the phenotype of the organism. She has published about 100 scientific papers, reviews, and book chapters in the areas of computational biology, computer science, and graph theory. She is a coeditor of a popular book on protein-protein interactions. She has served as a chair and/or program committee member of numerous prestigious computational and systems biology conferences. Currently, she serves on the editorial board of BMC Bioinformatics and the IEEE/ACM Transactions on Computational Biology and Bioinformatics.

Sanguthevar Rajasekaran received the ME degree in automation from the Indian Institute of Science, Bangalore, in 1983 and the PhD degree in computer science from Harvard University in 1988. Currently, he is the UTC Chair Professor of computer science and engineering and the director of the Booth Engineering Center for Advanced Technologies at the University of Connecticut (UConn). Before joining UConn, he served as a faculty member in the CISE Department of the University of Florida and in the CIS Department of the University of Pennsylvania. During 2000-2002, he was the chief scientist for Aroc Systems. His research interests include bioinformatics, parallel algorithms, data mining, randomized computing, computer simulations, and combinatorial optimization. He has published more than 200 articles in journals and conferences. He has coauthored two texts on algorithms and coedited four books on algorithms and related topics. He is an elected member of the Connecticut Academy of Science and Engineering. He is a fellow of the IEEE and a fellow of the AAAS.

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 professor 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 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 160 refereed papers 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 has been a guest editor for six special issues, including four in IEEE transactions.