FOREWORD

ADVANCES IN COMPUTATIONAL SYSTEMS BIOINFORMATICS

YING XU and PETER MARKSTEIN

This is the eighth special issue devoted to the Computational Systems Bioinformatics Conference series (CSB). The eight papers included in this Special Issue are selected from the 19 extended abstracts published in the *Proceedings of the Eighth Computational Systems Bioinformatics Conference (CSB'09)*. The conference was held in August 2009 at Stanford, California. In selecting papers for this Special Issue, we were most interested in works that present novel solutions to, or in-depth analyses of, difficult biological problems using bioinformatics techniques. We also tried to ensure that the selected papers cover a broad range of topics. All selected papers have been expanded significantly over the versions in the conference proceedings, and have gone through the rigorous review process of the *Journal of Bioinformatics and Computational Biology*. The final result is this collection of eight exciting and high-quality papers:

In “Classification and identification of fungal sequences using characteristic restriction endonuclease cut order,”¹ Sengupta, Bastola and Ali present a novel alignment-free approach to classification and identification of organisms based on a concept “enzyme cut order” that the authors introduced. Their applications to fungal sequences demonstrated the effectiveness of the new method.

Mongiovì *et al.* present an interesting inexact graph-matching algorithm, in their paper “Sigma: A set-cover-based inexact graph matching algorithm,”² by representing and solving the program as a set cover problem, which overcomes the issues often encountered in traditional graph-matching algorithms based on graph isomorphism and variations. The authors applied their algorithm to the compound screening problem, which is typically modeled as finding similar subnetworks in a database of networks, and showed the new algorithm’s superior performance over existing ones.

In their paper “Understanding distal transcriptional regulation from sequence, expression and interactome perspectives,”³ Rao and co-authors present an exciting computational study geared towards more effective ways for identification of regulatory enhancers in eukaryotic genomes, based on two new insights: (a) chromatin

modification signatures can discriminate proximal and distally located regulatory regions, and (b) the notion of promoter–enhancer cross-talk might have implications in the search for regulatory sequences that cooperate with the promoter to yield tissue restricted, gene expression.

The main theme of the paper “Querying metabolism under different physiological constraints”⁴ is about development of a new query language MQL for querying metabolic data collected under different experimental conditions. In this paper, the authors Cakmak, Ozsoyoglu and Hanson demonstrate the usefulness of their newly developed querying language.

Yin and Li present a unified statistical model, based on conditional random fields, for handling data smoothing, segmentation and copy number state decoding of the array comparative genomic hybridization (aCGH) data, in their paper “Detecting copy number variations from array CGH data based on a conditional random field model.”⁵ They demonstrate that their new method outperforms the popular Bayesian Hidden Markov Model-based approach on real aCGH data.

In their paper “Characterizing the space of interatomic distance distribution functions consistent with solution scattering data,”⁶ Kavathekar *et al.* present a computational approach to solving three-dimensional macromolecular structures based on solution scattering data from neutrons or X-rays. A key contribution of the work is that the proposed algorithm avoids the issue of data over-fitting, a general problem with most of the existing methods, through introduction and application of multiple alternative curves. The authors solve the computational problem using a linear programming approach.

Sheikh *et al.* present a formulation of the half-sibling reconstruction problem from microsatellite data in their paper “Combinatorial reconstruction of half-sibling groups from microsatellite data.”⁷ The authors present both exact solution and a heuristic solution to this computationally challenging problem. Application results are presented to show the usefulness of the proposed solutions.

In “Hierarchical classification of gene ontology terms using the GOstruct method,”⁸ Sokolov and Ben-Hur describe a computational method for protein function prediction based on careful analyses of the functional hierarchy of Gene Ontology. The authors demonstrated that their method outperforms functional prediction based on BLAST-based nearest neighbor method on widely used benchmark dataset.

As we conclude this overview, we would like to thank the following colleagues for carefully reviewing the manuscripts. Without their kind and timely help, this special issue would not have been possible. They are Asa Ben Hur, Andreas Beyer, Bhaskar DasGupta, Pan Du, Bertie Gottgens, Tamer Kahveci, Sun Kim, Ion Mandoiu, Sean Mooney, Quaid Morris, Sara Mostafavi, Enrico Pontelli, Johu Rousu, Mudita Sigal, Alex Zelikovsky, and Xuegong Zhang. Also we would like to thank Ms. Joan Yantko of the University of Georgia for her tireless help in communication with the authors and the reviewers during the editing of this special issue.

References

1. Sengupta R, Bastola D, Ali H, Classification and identification of fungal sequences using characteristic restriction endonuclease cut order, *J Bioinform Comput Biol* **8**(2):181–198, 2010.
2. Mongiovì M, di Natale R, Giufno R, Pulvirenti A, Ferro A, Sigma: A set-cover-based inexact graph matching algorithm, *J Bioinform Comput Biol* **8**(2):199–218, 2010.
3. Rao A, States DJ, Hero III AO, Engel JD, Understanding distal transcriptional regulation from sequence, expression and interactome perspectives, *J Bioinform Comput Biol* **8**(2):219–246, 2010.
4. Cakmak A, Ozsoyoglu G, Hanson RW, Querying metabolism under different physiological constraints, *J Bioinform Comput Biol* **8** (2):247–293, 2010.
5. Yin X-L, Li J, Detecting copy number variations from array CGH data based on a conditional random field model, *J Bioinform Comput Biol* **8**(2):295–314, 2010.
6. Kavathekar PA, Craig BA, Friedman AM, Bailey-Kellogg C, Balkcom DJ, Characterizing the space of interatomic distance distribution functions consistent with solution scattering data, *J Bioinform Comput Biol* **8**(2):315–335, 2010.
7. Sheikh SI, Berger-Wolf TY, Khokhar AA, Caballero IC, Ashley MV, Chaovalitwongse W, Chou CA, DasGupta B, Combinatorial reconstruction of half-sibling groups from microsatellite data, *J Bioinform Comput Biol* **8** (2):337–356, 2010.
8. Sokolov A, Ben-Hur A, Hierarchical classification of gene ontology terms using the GOStruct method, *J Bioinform Comput Biol* **8** (2):357–376, 2010.