

Series in
Mathematical Biology
and Medicine

Vol. 8

ADVANCES IN BIOINFORMATICS AND ITS APPLICATIONS



editors

Matthew He

Giri Narasimhan

Sergei Petoukhov



Series in
Mathematical Biology
and Medicine

Vol. 8

Proceedings of the International Conference
**ADVANCES IN BIOINFORMATICS
AND ITS APPLICATIONS**

Nova Southeastern University, Fort Lauderdale,
Florida, USA 16 – 19 December 2004

editors

Matthew He

Nova Southeastern University, Florida, USA

Giri Narasimhan

Florida International University, Florida, USA

Sergei Petoukhov

Russian Academy of Sciences, Moscow, Russia



World Scientific

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

Series in Mathematical Biology and Medicine — Vol. 8

ADVANCES IN BIOINFORMATICS AND ITS APPLICATIONS

Proceedings of the International Conference

Copyright © 2005 by World Scientific Publishing Co. Pte. Ltd.

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

ISBN 981-256-148-X

Proceedings of the International Conference
ADVANCES IN BIOINFORMATICS
AND ITS APPLICATIONS

SERIES IN MATHEMATICAL BIOLOGY AND MEDICINE

Series Editors: **P. M. Auger and R. V. Jean**

Published

- Vol. 1: Stochastic Models of Tumor Latency and their Biostatistical Applications
A. Yu. Yakovlev and A. D. Tsodikov
- Vol. 2: Volterra-Hamilton Models in the Ecology and Evolution of Colonial Organisms
P. L. Antonelli and R. H. Bradbury
- Vol. 3: The Hierarchical Genome and Differentiation Waves:
Novel Unification of Development, Genetics and Evolution
R. Gordon
- Vol. 4: Symmetry in Plants
Ed. by R. V. Jean and D. Barabe
- Vol. 5: Computational Medicine, Public Health and Biotechnology: Building a Man in
the Machine — Proceedings of the First World Congress
Vols. 1 & 3 — ed. by M. Witten
Vol. 2 — ed. by D. Joan and M. Witten
- Vol. 6: Advances in Mathematical Population Dynamics — Molecules, Cells and Man
Proceedings of the 4th International Conference on Mathematical Population
Dynamics
Ed. by O. Arino, D. Axelrod and M. Kimmel
- Vol. 7: Neuronal Information Processing
From Biological Data to Modelling and Applications
Ed. by G. Burdet, P. Combe and O. Parodi

FOREWORD

Recent progress in sequencing of genomes has resulted in a spectacular increase in the size of genomic databases. But what is it that these sequences tell us and what are the generalizations and rules that they are governed by? It seems that we understand too little about the genetic contexts to be able to “read” them correctly. Analytical, statistical, mathematical, and computational tools help in the process of decoding the secrets of biology.

The advent of bioinformatics and biotechnology has changed the way biomedical research is conducted and now plays a critical role in the discovery process. Biological data comes in many forms and often in large volumes. Analyzing this data poses many interesting challenges, whether it is for the theoretical computer scientist to design efficient algorithms, or the machine learning expert to design learning systems that use existing knowledge, or the statistician to uncover interesting correlations, or the mathematical modeler to investigate novel models of these natural systems, or the mathematician to investigate symmetries in the natural system, or the computer scientist to design smart databases and knowledge-bases. The results of these investigations help the practicing biologist in many ways: in identifying unknown connections, in narrowing down possibilities for a search, in suggesting new hypotheses, designing new experiments, proposing new models. Bioinformatics has already had far-reaching impact on drug discovery, on the fight against disease, in the design of new nanomaterials, and more generally in understanding the natural world around us.

The articles in these proceedings “Advances in Bioinformatics and its Applications (ABA)” were selected and peer-reviewed from contributions submitted to the “The International Conference on Bioinformatics and its Application” (ICBA), held from December 16th to 19th, 2004 at Fort Lauderdale, Florida, USA. Also included are articles based on invited talks given at the conference by keynote speakers. The articles published in these proceedings present major developments and trends in bioinformatics. These papers cover sequences, structures and functions; pathways, biological

networks, system biology, microarray data analysis, medical informatics, and applications. It contains a collection of approximately 50 high quality scientific research papers and survey articles. In addition, this book also contains 6 extended abstracts of ICBA invited talks by Jimenez-Montaña; Yang; Štambuk, Konjevoda and Gotovac; Dam, Su, Olman, and Xu; and Srinark and Kambhamettu. The full-length of these 6 papers are available at the Journal of Biological Systems, Volume 12, Special Issue, 2004. The ICBA conference was organized by the Nova Southeastern University Bioinformatics Group, the Florida International University Bioinformatics Group (BioRG), the International Society of Symmetry in Bioinformatics (ISSB), and was supported by Nova Southeastern University, Florida International University, and the International Symmetry Association (ISA).

It has been our honor to organize this conference in such an exciting and promising field. The conference and the proceedings would not have been possible without the contributions of many excellent scientists and researchers. We would like to thank Dr. Roger V. Jean, Chief Editor of the Journal of Biological Systems, for his critical role in making the conference proceedings reality. We are also thankful to Ms. Chelsea Chin, responsible for the proceedings at World Scientific Publishing in Singapore, for her expert handling of this publication on a tight schedule. Also our gratitude goes to our panel of reviewers for lending their valuable time and energy to the review process.

Our special thanks to the ICBA International Program Committee members: Dr. Creig Beam, Dr. Krzysztof Cios, Dr. Alexander P. Dubrov, Dr. Steve Eschrich, Dr. Dmitry Goldgof, Dr. Laurence I. Gould, Dr. Matthew He (Co-chair), Dr. Chun-Hsi Huang, Dr. Miguel A. Jimenez-Montano, Dr. Chandra Kambhamettu, Dr. Vladimir A. Karasev, Dr. Jie Liang, Dr. Giri Narasimhan (Co-Chair), Dr. Jun Ni, Dr. Victor Olman, Dr. Yi Pan, Dr. Sergei Petoukhov (Chair), Dr. Paolo Ricci, Dr. Alexey Stakhov, Dr. Junping Sun, Dr. Chi-Ming Yang, Dr. Bo Yuan, and Dr. Ya-xiang Yuan, for their professional support and guidance to the ICBA'04 conference.

We would also like to thank the ICBA Local Organizing Committee members: Dr. Naomi D'Alessio, Dr. Edye Groseclose, Dr. Matthew He (Chair), Dr. Ed. Keith (Co-Chair), Dr. Sawsan Khuri, Dr. Emil Kozarov, Dr. Leonard Levy, Dr. Jennie Lou, Dr. Giri Narasimhan (Co-Chair), Dr. Don Rosenblum, Dr. Emily Schmitt, Dr. Mahmood Shivji, Dr. Junping Sun (Co-Chair), Dr. Raisa Szabo, Dr. Claire Thuning-Roberson, for their tremendous contribution of time and effort in organizing the ICBA'04 conference. We also thank Dr. Naomi

D'Alessio, Dr. Leonard Levy, and Dr. Don Rosenblum for their constant support to the conference planning.

We are extremely grateful to the following outside reviewers who helped the program committee with the scientific review of the papers: Gulsah Altun, E. Olusegun George, Juntao Guo, Chindo Hicks, Tao Li, Kalai Mathee, Tom Milledge, Stephen Pellicer, Meera Sitharam, Hongwei Wu, Chengyong Yang, Erliang Zeng, Gaolin Zheng, and Wei Zhong. Finally, our thanks to Fernando L. Niño for help with the last minute editing.

We would specially like to thank all ICBA organizers and collaborators. Finally we thank the sponsors: Farquhar College of Arts and Sciences of NSU, Oceanographic Research Center of NSU, Florida International University Bioinformatics Group (BioRG), NSU Bioinformatics Group, Graduate School of Computer and Information Sciences of NSU, Health Professions Division of NSU, NSU President's Faculty Research and Development Grant, International Symmetry Association, International Journal of Bioinformatics Research and Applications (IJBRA), International Society of Symmetry in Bioinformatics, Journal of Bioinformatics and Computational Biology (JBCB), Journal of Biological Systems (JBS), National Center for Biotechnology Information (NCBI), and World Scientific Publishing Company.

Dr. Matthew He (Nova Southeastern University, Florida, USA)

Dr. Giri Narasimhan (Florida International University, Florida, USA)

Dr. Sergei Petoukhov (Russian Academy of Sciences, Moscow, Russia)

December 16, 2004

Fort Lauderdale, Florida, USA

CONTENTS

1. Microarray Analysis and Regulatory Networks

Microarray Normalization using Proportional Abundance <i>S. A. Eschrich and C. A. Beam</i>	3
Emerging Pattern Based Subspace Clustering of Microarray Gene Expression Data using Mixture Models <i>Y. B. Kim, J. H. Oh and J. Gao</i>	13
Thermodynamic Feasibility of Metabolic Networks <i>R. Nigam and S. Liang</i>	25
Computational Reconstruction of the Carbon Fixation Pathway in <i>Synechococcus</i> sp. WH8102 <i>P. Dam, Z. Su, V. Olman and Y. Xu</i>	37
Modular Neural Networks and their Application in Exon Prediction <i>Y. Pan and C. W. Sensen</i>	47
Gene Definition in Different Databases and its Implication in Microarray Data Interpretation <i>P. Wang, M. Dai, R. Thompson, S. J. Watson and F. Meng</i>	62
Exploring Important Issues in the Implementation of Gene Set Enrichment Analysis <i>Y. Wang, S. J. Watson and F. Meng</i>	71
Querying A Database of Regulatory Elements <i>C. Yang, E. Zeng, K. Mathee and G. Narasimhan</i>	81

2. Machine Learning and Statistical Analyses

Computational Epidemiology: Bayesian Disease Surveillance <i>K. Abbas, A. R. Mikler, A. Ramezani and S. Menezes</i>	95
Non-Equilibrium Fluctuations in a Stochastic Planktonic Model: Periodicity of Phytoplankton Bloom <i>R. Bhattacharyya and B. Mukhopadhyay</i>	107
Fuzzy Decoding in Profile Hidden Markov Models for Protein Family Identification <i>N. P. Bidargaddi, M. Chetty and J. Kamruzzaman</i>	119
Screening Potential Allele-Frequency Disequilibrium in Data from Affymetrix Gene Expression Arrays <i>M. Dai, Y. Wang, E. Sliwerska, M. Burmeister, S. J. Watson, H. Akil and, F. Meng</i>	132
The Application of Self Organizing Maps for the Analyses of DNA Sequences <i>H. Dozono, S. Hara, H. Tokushima and Y. Noguchi</i>	140
Gaussian Kernel Width Generation for Support Vector Clustering <i>S.-H. Lee and K. Daniels</i>	151
Serum Proteomic Pattern Analysis for Early Cancer Detection <i>Y. Liu</i>	163
A Linear Solution for Haplotype Perfect Phylogeny Problem (Extended Abstract) <i>Y. Liu and C.-Q. Zhang</i>	173
<i>In Silico</i> Biology: Haplotype Inference, Linkage Disequilibrium Analysis, and Microarray-Based Gene Selection Techniques <i>T. Niu</i>	185

The Role of Feature Redundancy in Tumor Classification <i>C. H. Ooi, M. Chetty and I. Gondal</i>	197
Clustering of Eukaryotic Orthologs based on Sequence and Domain Similarities using the Markov Graph-Flow Algorithm <i>H. Ozer, J. Chen, F. Zhang and B. Yuan</i>	209
Rapid Drug Target Identification in Microbial Genomes – An <i>in silico</i> Approach with Special Reference to <i>Salmonella</i> Species <i>K. R. Sakharkar, C. Verma, M. K. Sakharkar and V. T. K. Chow</i>	221
A Collimator Neural Network Model for the Classification of Genetic Data <i>M. S. B. Sehgal, I. Gondal and L. Dooley</i>	229
Training Set Design for Pattern Discovery with Applications to Protein Motif Detection <i>Y. Sun, Z. Deng, G. Narasimhan and K. Mathee</i>	240
Gene Selection by Weighted Fisher Criterion for Multiclass Prediction <i>J. Xuan, Y. Dong, J. Khan, R. Clarke, E. Hoffman and Y. Wang</i>	252
Sequential Bifurcation Approach to Learning Protein Functional Classes <i>M. Q. Yang, O. K. Ersoy and J. Y. Yang</i>	264
Identification of Contaminants in MALDI-TOF Proteomics Spectra via Data Mining <i>C.-Y. Yen, K. J. Cios and M. W. Duncan</i>	276
Analysis of the Effects of Recombinant Human Erythropoietin (EPO) on Murine Bone Marrow Populations using Mixture Modeling Techniques <i>G. Zheng, R. Capocasale, B. Amegadzie, P. Bugelski and J. Lu</i>	285

Discovery of Local Protein Sequence Motifs using Improved K-Means Clustering Technique	297
<i>W. Zhong, G. Altun, R. Harrison, P. C. Tai and Y. Pan</i>	
3. Biomolecular Sequence and Structure Analysis	
Predicting and Characterizing Metal-Binding Sites using Support Vector Machines	307
<i>L. Bobadilla, F. Niño and G. Narasimhan</i>	
MTMM — A Matlab Toolbox for Macromolecular Modeling	319
<i>F. Cui, T. Bibi and Z. Wu</i>	
Stabbing Balls and Simplifying Proteins	329
<i>O. Daescu and J. Luo</i>	
Step-Wise Mutations of mRNA Sequences Lead to Progressive Changes in Calculated Folding Free Energies	341
<i>D. Digby and W. Seffens</i>	
Aligning Sequences with Non-Affine Gap Penalty: Plains Algorithm, A Practical Implementation, and Its Biological Applications in Comparative Genomics	351
<i>O. Gill, Y. Zhou and B. Mishra</i>	
Using the Length Constraints of Helix to Evaluate Protein Secondary Structure Prediction for Helix	363
<i>J. He and Y. Lu</i>	
An Algorithm for Finding Similarities in Protein Active Sites	373
<i>K. Kupas, A. Ultsch and G. Klebe</i>	
Parallel Tempering in Rosetta Practice	380
<i>Y. Li, C. E. M. Strauss and A. Gorin</i>	

An Application of Association Rule Mining to HLA-A*0201 Epitope Prediction	390
<i>T. Milledge, G. Zheng and G. Narasimhan</i>	
A Two-Way Searching Algorithm for <i>De Novo</i> Peptide Sequencing via Tandem Mass Spectrometry	402
<i>J. H. Oh, Y. B. Kim and J. Gao</i>	
An Algorithm for Determining Glycan Structures from MS/MS Spectra	414
<i>B. Shan, K. Zhang, B. Ma, C. Zhang and G. Lajoie</i>	
Counting and Enumeration of Self-Assembly Pathways for Symmetric Macromolecular Structures	426
<i>M. Sitharam and M. Bóna</i>	
Improving Multiple Whole Genomes Alignment by Considering Rearrangement	437
<i>J. Yang and J. S. Deogun</i>	
A Filtering Approach to Splice Site Predictions in Human Genes	447
<i>K. Yoon and S. Kwek</i>	
The Construction of Structural Templates for the Modeling of Conserved Protein Domains	459
<i>F. Zhang, J. Chen, Z. Liu and B. Yuan</i>	
 4. Symmetry in Sequences	
Applications of Hyper Genetic Code to Bioinformatics	473
<i>M. A. Jimenez-Montaño</i>	
A Dodecahedron-Based Model of Spatial Representation of the Canonical Set of Amino Acids	482
<i>V. A. Karasev, V. V. Luchinin and V. E. Stefanov</i>	

Molecular Mappings: Group Theory, Coding Theory, and the Emergence of Replication <i>D. A. MacDónaill</i>	494
Symmetry and Information in the Genetic Code <i>T. Negadi</i>	502
The Rules of Degeneracy and Segregations in Genetic Codes – The Chronocyclic Conception and Parallels with Mendel's Laws <i>S. V. Petoukhov</i>	512
Symbolic Coding of Amino Acid and Nucleotide Properties <i>N. Štambuk, P. Konjevoda and N. Gotovac</i>	533
Analyzing Patterns of Tripeptides using Statistical Approach and Neural Network Paradigms <i>R. Szabo, M. He, E. Burnham and J. Jurani</i>	544
Molecular Versus Atomic Information Logic behind the Genetic Coding Contents Constrained by Two Evolutionary Axes and the Fibonacci-Lucas Sequence <i>C. M. Yang</i>	554
5. Signal Processing, Image Processing and Visualization	
Use of EEG Recordings as Valid Measures of Kinesthetic and Visual Imagery Ability <i>J. G., Cremades, M. Tito, M. Adjouadi and A. Barreto</i>	565
The Merits of the Third Dimension for Visual Analysis of Multiple Anatomy Ontologies <i>A.-S. Dadzie and A. Burger</i>	576

Localization of Brain Activation through Integration of Optical Topography and Magnetic Resonance Imaging	588
<i>M. Mourad, M. Adjouadi and I. Y aylali</i>	
A Distributed Multimedia Database Visualization Within an Immersive Environment for Bioinformatics	600
<i>P.-E. Gros, N. Férey, J. Hérisson and R. Gherbi</i>	
Author Index	613

SECTION 1

Microarray Analysis and Regulatory Networks

