

# **9th International Conference on Bioinformatics and Computational Biology (BICOB 2017)**

Honolulu, Hawaii, USA  
20 – 22 March 2017

## **Editors:**

**Oliver Eulenstein  
Qin Ding  
Hisham Al-Mubaid**

ISBN: 978-1-5108-3667-9

**Printed from e-media with permission by:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571



**Some format issues inherent in the e-media version may also appear in this print version.**

Copyright© (2017) by the International Society for Computers and Their Applications  
All rights reserved. Reproduction in any form without the written consent of ISCA is prohibited.

Original ISBN: 978-1-943436-07-1 (Out of Print)  
Reprint ISBN: 978-1-5108-3667-9

Printed by Curran Associates, Inc. (2017)

For permission requests, please contact the International Society for Computers and Their Applications  
at the address below.

International Society for Computers and Their Applications  
64 White Oak Court  
Winona, MN 55987  
USA

Phone: (507) 458-4517

isca@ipass.net

**Additional copies of this publication are available from:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571 USA  
Phone: 845-758-0400  
Fax: 845-758-2633  
Email: curran@proceedings.com  
Web: www.proceedings.com

# Table of Contents

<b>Keynote</b>	<b>1</b>
Phylogenetic Analyses: Past, Present, and Future <i>Bernard M.E. Moret</i> .....	1
<b>Genomics I</b>	<b>3</b>
A Bayesian Framework for Marker Gene Taxonomic Classification Based on Sequence Alignment <i>Xiang Gao, Huaiying Lin, Kashi Revanna, Qunfeng Dong</i> .....	3
GraphPE: Refining Metagenome Binning by Use of Paired-end Graph of Contigs <i>Xianghui Liu, Rohan B. H. Williams</i> .....	9
Clustering Metagenome Sequences Using Canopies <i>Mohammad Arifur Rahman, Nathan LaPierre, Huzefa Rangwala, Daniel Barbara</i> .....	17
Improving Microbiome Classification with Dirichlet-Multinomial Models <i>Xiang Gao, Huaiying Lin, Qunfeng Dong</i> .....	25
<b>Structure and Function</b>	<b>31</b>
Evaluation of Protein Structural Models Using Random Forests <i>Renzhi Cao, Jie Hou, Taeho Jo, Jianlin Cheng</i> .....	31
Visual Experiments of Geometric Combinatorics for Neural Stem Cells and Their Derivatives <i>T. J. Peters, J. C. Conover, D. McManus, K. Pratt, K. D. Williams</i> .....	39
Classification of DNA Structure through Structure Network Analysis <i>Tanmay Kulshrestha, Ankita Gupta, Shibasish Chowdhury</i> .....	45
A Fast Algorithm for Finding Largest Common Substructures in Multiple RNAs <i>Jithendar Anandan, Eric Fry, Keith Monschke, Abdullah N. Arslan</i> .....	51
<b>Systems Biology I</b>	<b>59</b>
Approximate Frequent Subpath Mining Applied to Pangenomics <i>Alan Cleary, Brendan Mumeey, Thiruvarangan Ramaraj, Joann Mudge</i> .....	59
Determining Multifunctional Genes and Diseases in Human Using Gene Ontology <i>Hisham Al-Mubaid, Sasikanth Potu, M. Shenify</i> .....	65
A Local Search Algorithm for Clustering Large Biological Networks <i>Guy Coccimiglio, Salimur Choudhury</i> .....	73
Reconstructing Protein and Gene Phylogenies by Extending the Framework of Reconciliation <i>Esaie Kuitche, Manuel Lafond, Aïda Ouangraoua</i> .....	79
ChIP-GMM: A Gaussian Mixture Model for Inferring Binding Regions in ChIP-seq Profiles <i>Sharmi Banerjee, Xi Chen, Xiaowei Wu, Hehuang Xie, Jianhua Xuan, Wu-chun Feng</i> .....	87
<b>Healthcare and Bioinformatics Applications</b>	<b>93</b>
An In-Memory Database Platform for Systems Medicine <i>Milena Kraus, Matthieu-P. Schapranow</i> .....	93
A Pilot Study of Breast Cancer Patients: Can Machine Learning Predict Healthcare Professionals' Responses to Patient Emotions? <i>Luke Barracliffe, Ognjen Arandjelović, Gerry Humphrisy</i> .....	101

Fundamental Bounds for Sequence Reconstruction from Nanopore Sequencers <i>Abram Magner, Jarosław Duda, Wojciech Szpankowski, Ananth Grama</i> .....	107
Applications of Novel Graph Theoretic Methods to Clustering Autism Spectrum Disorders Phenotypes <i>John Matta, Thy Nguyen, Gunes Ercal, Tayo Obafemi-Ajayi</i> .....	113
<b>Genomics II</b>	<b>121</b>
ParSECH: Parallel Sequencing Error Correction with Hadoop for Large-Scale Genome <i>Arghya Kusum Das, Shayan Shams, Sayan Goswami, Richard Platania, Kisung Lee, Seung-Jong Park</i> .....	121
Finding Surrogate Controls in Pathway Enrichment Analysis using RNA-Seq Expression <i>Tham H. Hoang, Yue Zhao, Pujan Joshi, Seung-Huyn Hong, Dong-Guk Shin</i> .....	129
Towards Reliable Microarray Analysis and Design <i>Max Garzon, Sambriddhi Mainali</i> .....	135
Reconstruction of Ring Chromosomes Using Paired-End Sequencing Data <i>Matthew Hayes, Jeremy Pearson</i> .....	141
<b>Systems Biology II</b>	<b>147</b>
Parameter Identification of Nonlinear Biochemical Reaction Systems Using the Statistics-based Method <i>Xiaoying Zheng, Xiufen Zou, Zishu Pan</i> .....	147
Modularity in PPI Networks: Characteristics of Existing Networks and Models of Evolution <i>Min Ye, Xiuwei Zhang, Bernard M.E. Moret</i> .....	155
Network-based Differential Analysis of Hi-C Data <i>Lu Liu, Jianhua Ruan</i> .....	165
A Parallel Algorithm for Mining Maximal Cohesive Subgraphs <i>Aditya Goparaju, Saeed Salem</i> .....	173
<b>Bioinformatics Applications I</b>	<b>179</b>
Comparative study of Differential Evolutionary-Binary Particle Swarm Optimization (DE-BPSO) algorithm as a feature selection technique with different linear regression models for analysis of HIV-1 Integrase Inhibition features of Aryl $\beta$ -Diketo Acids <i>Falguni Thakor, Ahmad Reza Hadaegh, Xiaoyu Zhang</i> .....	179
A Sequential Monte Carlo Sampler for Bayesian Phylogenetics <i>Liangliang Wang</i> .....	185
PF-Words: Biomedical Literature Based Protein Function Search <i>Wei Wei, Mike Wong, Anagha Kulkarni</i> .....	191
Detecting Haplotype Amplification in Cancer Sequencing Data <i>Weiwei Liu, Jingyang Gao, Zhongmeng Zhao, Rongrong Yang, Yu Geng, Tian Zheng, Xuanping Zhang, Xiao Xiao, Jiayin Wang</i> .....	197
<b>Bioinformatics Applications II</b>	<b>203</b>
Improving Peptide Identification by Considering Ordered Amino Acid Usage <i>Ahmed Alqurri, John Rose</i> .....	203
Non-parametric Quality Assessment of High-Content Screening Assays <i>Robert P. Trevino, Philippe C. Faucon, Thomas J. Lamkin, Steven A. Kawamoto, Ross Smith, Huan Liu</i> .....	209
Prediction of Biological Functions by Histone Modification Patterns Profiling <i>Yiou Xiao, Kishan G. Mehrotra, Chilukuri K. Mohan, Pratibha Choudhary, Ramesh Raina</i> .....	217
Genetic Information, Mutation Rates, and the Lore of the Error Threshold <i>Lee Altenberg</i> .....	223
Top-Down Correlation-Based Pattern Analysis and Feature Selection for Multi-Component FTIR Spectra <i>Ghada Badr, Najla Al-Ariefy</i> .....	231
<b>Author Index</b>	