Series on Advances in Bioinformatics and Computational Biology – Volume 5

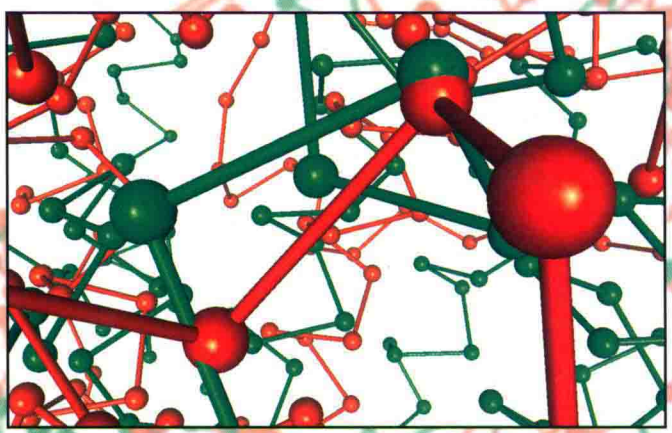
Proceedings of the 5th Asia-Pacific BIOINFORMATICS Conference

EDITORS

David Sankoff

Lusheng Wang

Francis Chin



Imperial College Press

Series on Advances in Bioinformatics and Computational Biology – Volume 5

Proceedings of the 5th Asia-Pacific

BIOINFORMATICS

Conference

Hong Kong 15 – 17 January 2007

editors

David Sankoff
University of Ottawa, Canada

Lusheng Wang
City University of Hong Kong, Hong Kong

Francis Chin
The University of Hong Kong, Hong Kong

ICP

Imperial College Press

Published by

Imperial College Press
57 Shelton Street
Covent Garden
London WC2H 9HE

Distributed 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.

PROCEEDINGS OF THE 5TH ASIA-PACIFIC BIOINFORMATICS CONFERENCE

ISBN-13 978-1-86094-783-4
ISBN-10 1-86094-783-2

Printed in Singapore by B & JO Enterprise

Proceedings of the 5th Asia-Pacific

BIOINFORMATICS

Conference

SERIES ON ADVANCES IN BIOINFORMATICS AND COMPUTATIONAL BIOLOGY

ISSN: 1751-6404

Vol. 1: Proceedings of the 3rd Asia-Pacific Bioinformatics Conference
Eds: Yi-Ping Phoebe Chen and Limsoon Wong

Vol. 2: Information Processing and Living Systems
Eds: Vladimir B. Bajic and Tan Tin Wee

Vol. 3: Proceedings of the 4th Asia-Pacific Bioinformatics Conference
Eds: Tao Jiang, Ueng-Cheng Yang, Yi-Ping Phoebe Chen and Limsoon Wong

Vol. 4: Computational Systems Bioinformatics
Eds: Peter Markstein and Ying Xu

Vol. 5: Proceedings of the 5th Asia-Pacific Bioinformatics Conference
Eds: David Sankoff, Lusheng Wang and Francis Chin

PREFACE

High-throughput sequencing and functional genomics technologies have given us the human genome sequence as well as those of other experimentally, medically and agriculturally important species, and have enabled large-scale genotyping and gene expression profiling of human populations. Databases containing large numbers of sequences, polymorphisms, structures and gene expression profiles of normal and diseased tissues are rapidly being generated for human and model organisms. Bioinformatics is thus rapidly growing in importance in the annotation of genomic sequences, in the understanding of the interplay among and between genes and proteins, in the analysis the genetic variability of species, in the identification of pharmacological targets and in the inference of evolutionary origins, mechanisms and relationships.

The Asia-Pacific Bioinformatics Conference series is an annual forum for exploring research, development, and novel applications of bioinformatics. It brings together researchers, professionals, and industrial practitioners for interaction and exchange of knowledge and ideas. The Fifth Asia-Pacific Bioinformatics Conference, APBC2007, was held in Hong Kong 15–17 January, 2007.

A total of 104 papers were submitted to APBC 2007. These submissions came from Bangladesh, China, Hong Kong, India, Japan, Korea, Malaysia, Singapore, Taiwan, Thailand, Australia, New Zealand, Denmark, France, Germany, Hungary, Italy, Israel, Portugal, UK, Canada, Mexico and USA. We assigned each paper to at least three members of the programme committee. Although not all members of the programme committee managed to review all the papers assigned to them, a total of 317 reviews were received, so that there were about three reviews per paper on average.

A total of 35 papers (33%) were accepted for presentation and publication in the proceedings of APBC 2007. Based on the affiliation of the authors, 1.25 of the accepted papers were from China, 1.46 were from Hong Kong, 3 were from Japan, 0.83 were from Korea, 1 were from Singapore, 2.30 were from Australia, 2 were from Denmark, 0.15 were from France, 4.5 were from Germany, 0.5 were from Italy, 1 were from Portugal, 1.66 were from UK, 4.57 were from Canada, 10.78 were from USA.

In addition to the accepted papers, the scientific programme of APBC 2006 also included three keynote talks, by Jennifer A. Marshall Graves, Joseph H. Nadeau and Pavel A. Pevzner, as well as tutorial and poster sessions. The presentations were of very high quality. Almost a third focused on evolution and phylogeny, largely at the genome level, a similar number dealt with protein structure and proteomics more generally, and a good proportion studied various aspects of pathways,

networks, transcriptomics and microarray technology. A range of other topics in bioinformatics and computational biology were also covered, ranging from motif and gene recognition, through haplotypes and population genetics, to databases and text mining. Much of this work featured techniques of sequence analysis, while many of the papers included applications to biology and medicine.

We had a great time in Hong Kong, enhancing the interactions between many researchers and practitioners, and reuniting the Asia-Pacific bioinformatics community in the context of an international conference with worldwide participation.

Finally, we wish to express our gratitude to the authors of the submitted papers, the members of the programme commitee and their subreferees, the members of the organizing committee, Phoebe Chen and Limsoon Wong (our liaisons in the APBC steering committee), the keynote speakers, our generous sponsors, and supporting organizations for making APBC 2007 a great success.

David Sankoff
Lusheng Wang
Francis Chin

17 January 2007

APBC2007 ORGANIZATION

Conference Chair

Francis YL Chin, The University of Hong Kong, Hong Kong

Organizing Committee

Francis YL Chin (Chair), The University of Hong Kong
David Smith (Poster Session), The University of Hong Kong
H.F. Ting (Finance and Registration), The University of Hong Kong
Lusheng Wang (Publication), The City University of Hong Kong
Xiaowen Liu (Publication), The City University of Hong Kong
S.M. Yiu (Local Arrangement and Tutorial), The University of Hong Kong
Daniel Hung (Webmaster), The University of Hong Kong
Samson Sin (Webmaster), The University of Hong Kong

Steering Committee

Phoebe Chen (Chair), Deakin University, Australia
Sang Yup Lee, KAIST, Korea
Satoru Miyano, University of Tokyo, Japan
Mark Ragan, University of Queensland, Australia
Limsoon Wong, National University of Singapore

Program Committee

David Sankoff (Chair), The University of Ottawa
Lusheng Wang (Chair), The City University of Hong Kong
Tatsuya Akutsu, Kyoto University
Miguel Andrade, Ottawa Health Research Institute
Stephane Aris-Brosou, University of Ottawa
Joel Bader, Johns Hopkins University
Serafim Batzoglou, Stanford University
David Bryant, University of Auckland
Jeremy Buhler, Washington University in St. Louis
Peter Donnelly, University of Oxford
Dannie Durand, Carnegie Mellon University
Nadia El-Mabrouk, University of Montreal
Robert Giegerich, Bielefeld University
Carole Goble, University of Manchester
Concettina Guerra, Universita' di Padova
Dan Gusfield, University of California, Davis
Michael Hallett, McGill University
Sridhar Hannenhalli, University of Pennsylvania
Daniel Huson, Tbingen University
Gavin Huttley, Australian National Univeristy
Jenn-Kang Hwang, National Chiao Tung University
Tao Jiang, University of California - Riverside
Uri Keich, Cornell University
Anand Kumar, University of Leipzig
Tak Wah Lam, The University of Hong Kong
Doheon Lee, KAIST, Korea
Jinyan Li, Institute for Infocomm Research
Wentian Li, Feinstein Institute for Medical Research
Guohui Lin, University of Alberta
Michal Linial, The Hebrew University of Jerusalem
Zhijie Liu, University of Georgia
Bin Ma, University of Western Ontario
Satoru Miyano, The University of Tokyo
Laxmi Parida, IBM Thomas J. Watson Research Center
Mark Ragan, University of Queensland
Marie-France Sagot, University Claude Bernard, Lyon I
Akinori Sarai, Kyushu Institute of Technology
Vincent Schachter, Genoscope
Steven Skiena, State University of New York at Stony Brook
Edward Susko, Dalhousie University
Yun Song, University of California, Davis
Robert Stevens, University of Manchester
Alfonso Valencia, Centro Nacional de Biotechnologia
Michael Waterman, University of Southern California

Ken Wolfe, University of Dublin
Stacia Wyman, Williams College
Hong Yan, City University of Hong Kong
Qiang Yang, Hong Kong University of Science and Technology
Kaizhong Zhang, University of Western Ontario
Liqing Zhang, Virginia Tech
Louxin Zhang, National University of Singapore

Additional Reviewers

Shandar Ahmad
Iris Bahir
Shihyen Chen
Tobias Dezulian
James Eales
Morihiro Hayashida
Helen Hulme
Gunnar W. Klau
Jingping Liu
Julia Mixtacki
Hiroshi Nakashima
Magnus Rattray
Baozhen Shan
Ashish V. Tendulkar
Balaji Venkatachalam
Katy Wolstencroft
Zheng Yuan
Marcos Jesus Arauzo Bravo
Richard Bean
Matteo Comin
Zhihong Ding
Logan Everett
Cornelia Hedeler
Shane Jensen
Kiyoung Lee
Yaniv Loewenstein
Jose Carlos Nacher
Kay Nieselt
Jonathan Schug
Yun S. Song
Victor Tong
Li-San Wang
Yufeng Wu
Guanglan Zhang
Alexander Auch
Pierre-Yves Bourguignon
Mike Cornell
Maxime Durot
Paul Fisher
Duncan Hull
Noam Kaplan
Yaoyong Li
Suryani Lukman
Niranjan Nagarajan
Elon Portugaly
Charles Semple
Kristian Stevens
Roy Varshavsky
Michael Wilson
Lei Xin

CONTENTS

EXPLORING GENOMES OF DISTANTLY RELATED MAMMALS

JENNIFER A. MARSHALL GRAVES

ARC Centre for Kangaroo Genomics, Research School of Biological Sciences
Australian National University, Canberra, ACT 2601, Australia

There are three groups of extant mammals, two of which abound in Australia. Marsupials (kangaroos and their relatives) and monotremes (echidna and the fabulous platypus) have been evolving independently for most of mammalian history. The genomes of marsupial and monotreme mammals are particularly valuable because these alternative mammals fill a phylogenetic gap in vertebrate species lined up for exhaustive genomic study. Human and mice (~70MY) are too close to distinguish signal, whereas mammal/bird comparisons (~310MY) are too distant to allow alignment. Kangaroos (180 MY) and platypus (210 MY) are just right. Sequence has diverged sufficiently for stringent detection of homologies that can reveal coding regions and regulatory signals. Importantly, marsupials and monotremes share with humans many mammal-specific developmental pathways and regulatory systems such as sex determination, lactation and X chromosome inactivation.

The ARC Centre for Kangaroo Genomics is characterizing the genome of the model Australian kangaroo Macropus eugenii (the tammar wallaby), which is being sequenced by AGRF in Australia, and Baylor (funded by NIH) in the US. We are developing detailed physical and linkage maps of the genome to complement sequencing, and will prepare and array cDNAs for functional studies, especially of reproduction and development. Complete sequencing of the distantly related Brazilian short-tailed opossum Monodelphis domestica by the NIH allows us to compare distantly related marsupials. Sequencing of the genome of the platypus, Ornithorhynchus anatinus by Washington University (funded by the NIH) is complete, and our lab is anchoring contigs to the physical map. We have isolated and completely characterized many BACs and cDNAs containing kangaroo and platypus genes of interest, and demonstrate the value of comparisons to reveal conserved genome organization and function, and new insights in the evolution of the mammalian genome, particularly sex chromosomes.