Introduction

Extending Asia Pacific bioinformatics into new realms in the "-omics" era

Shoba Ranganathan*[1,2], Frank Eisenhaber[3], Joo Chuan Tong[4] and Tin Wee Tan[2]

Address: [1]Department of Chemistry and Biomolecular Sciences and ARC Centre of Excellence in Bioinformatics, Macquarie University, Sydney NSW 2109, Australia, [2]Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, 8 Medical Drive, Singapore 117597, [3]Bioinformatics Institute, 30 Biopolis Street; #07-01 Matrix, Singapore and [4]Institute for Infocomm Research, 1 Fusionopolis Way, #21-01 Connexis, Singapore

Email: Shoba Ranganathan* - shoba.ranganathan@mq.edu.au; Frank Eisenhaber - franke@bii.a-star.edu.sg; Joo Chuan Tong - jctong@i2r.a-star.edu.sg; Tin Wee Tan - tinwee@bic.nus.edu.sg

* Corresponding author

This article is available from: http://www.biomedcentral.com/1471-2164/10/S3/S1

Abstract

The 2009 annual conference of the Asia Pacific Bioinformatics Network (APBioNet), Asia's oldest bioinformatics organisation dating back to 1998, was organized as the 8th International Conference on Bioinformatics (InCoB), Sept. 7-11, 2009 at Biopolis, Singapore. Besides bringing together scientists from the field of bioinformatics in this region, InCoB has actively engaged clinicians and researchers from the area of systems biology, to facilitate greater synergy between these two groups. InCoB2009 followed on from a series of successful annual events in Bangkok (Thailand), Penang (Malaysia), Auckland (New Zealand), Busan (South Korea), New Delhi (India), Hong Kong and Taipei (Taiwan), with InCoB2010 scheduled to be held in Tokyo, Japan, Sept. 26-28, 2010. The Workshop on Education in Bioinformatics and Computational Biology (WEBCB) and symposia on Clinical Bioinformatics (CBAS), the Singapore Symposium on Computational Biology (SYMBIO) and training tutorials were scheduled prior to the scientific meeting, and provided ample opportunity for in-depth learning and special interest meetings for educators, clinicians and students. We provide a brief overview of the peer-reviewed bioinformatics manuscripts accepted for publication in this supplement, grouped into thematic areas. In order to facilitate scientific reproducibility and accountability, we have, for the first time, introduced minimum information criteria for our pubilcations, including compliance to a Minimum Information about a Bioinformatics Investigation (MIABi). As the regional research expertise in bioinformatics matures, we have delineated a minimum set of bioinformatics skills required for addressing the computational challenges of the "-omics" era.

Overview

The Asia-Pacific Bioinformatics Network (APBioNet, [1]) is the oldest regional bioinformatics society, established in 1998, to gather scientists from diverse disciplines to work together to advance the frontiers of bioinformatics. The first three annual meetings were held at the Pacific Symposium of Biocomputing (1999-2001), in Hawaii, to nucleate a bioinformatics core group in the region. With the core group reaching critical mass, the APBioNet executive committee members assisted in organizing InCoB2002 (the International Conference on Bioinformatics, 2002) in Bangkok, Thailand and adopted this meeting as their annual conference. Following this successful event, InCoB meetings have been held in Penang, Malaysia (2003); Auckland, New Zealand (2004); Busan, South Korea (2005), New Delhi, India (2006); Hong Kong/Hanoi (by videocasting) (2007) and Taipei, Taiwan (2008, [2]).

In order to sustain the growth of bioinformatics in the region, a special interest group meeting of bioinformatics educators, the Workshop on Education in Bioinformatics and Computational Biology (WEBCB) was organized at InCoB2008, attended also by life scientists participating in the symposium of the Federation of Asian and Oceanician Biochemists and Molecular Biologists (FAOBMB) at the same location. WEBCB follows on the Workshops on Education in Bioinformatics (WEB, initiated by SR) [3], to incorporate bioinformatics into mainstream life science research and its second meeting was held at InCoB2009, along with tutorials in traditional and emerging topics as part of APBioNet's training mandate.

The first ever Clinical Bioinformatics Symposium (CBAS) was held prior to the InCoB2009 scientific conference, while the Singapore chapter of the Regional Student Group of the International Society for Computational Biology organized the Singapore Symposium on Computational Biology (SYMBIO) to provide students and young investigators an opportunity to present their ongoing research work.

In order to provide opportunities for international peer-reviewed high-impact factor journal publications, APBioNet embarked on further raising the standards for the region by publishing a dedicated *BMC Bioinformatics* supplement, since 2006 [2,4,5]. In 2009, the manuscripts from APBioNet members have diversified and increased in quality and sophistication with computational biology articles addressing analysis pertaining to "-omics" data published in this supplement and a *BMC Bioinformatics* supplement [6], focussing on protein sequence analysis, genetic and population analysis, structural bioinformatics, text mining and ontology, chemoinformatics and bio-

diversity informatics as well as a case study on the impact of e-learning tools in bioinformatics education.

Reviewing standards

Papers submitted to these proceedings were peer-reviewed by at least three reviewers, from the APBioNet/InCoB program committee and invited external experts as required (listed in Additional File 1). InCoB2009 also provided multi-track submissions, with the inclusion of research highlights from recent publications and for showcasing technology developments along with posters. With tutorials, a specialist workshop on bioinformatics education and two special interest research symposia, InCoB2009 provided a comprehensive five-day international bioinformatics meeting in the Asia Pacific.

The editors carefully screened the 90 full paper submissions and relegated two of these to the poster session, in order to select only the best papers from more than a dozen Asia Pacific countries, as well as UK and USA. From the 88 full paper submissions reviewed, 49 were short listed for oral presentation. This supplement features 34 papers addressing "-omics" data analysis, while 10 papers were accepted for publication in the BMC Bioinformatics supplement [6], reflecting an overall acceptance rate of 50%, with five more appearing in the online journal, *Bioinformation* [7]. Extensive collaboration in bioinformatics research in the region is evident from co-authorship of papers involving Australia, China, Hong Kong, India, Japan, Korea, Singapore, Taiwan, Thailand, UK and USA. A brief review of the different themes is provided below.

Next-generation sequence analysis

With the advent of next-generation sequencing technologies, nucleotide sequencing has become inexpensive, albeit resulting in short sequence reads, requiring new and efficient bioinformatics tools. Zhao et al. [8] present an alignment tool, BOAT, capable of mapping large volumes of short reads to reference sequences with better sensitivity and lower memory requirement than other currently existing algorithms. Chung and Park [9] propose an empirical method for choosing efficient discriminative seeds for oligonucleotide design, while Piriyapongsa et al. [10] have developed a new integrated primer design tool. Venkatachalam et al. [11] have predicted the occurrence of peroxisome proliferator response elements, which are promising targets for cancer treatment, followed by *in vitro* experimental validation. In order to accurately predict the caspase degradome on a systems-wide basis, Wee et al. [12] have developed a multifactor model incorporating cleavage site prediction and structural factors.

Genome analysis

To track genome-wide transposon-based insertional mutagenesis experiments, Yang et al. [13] have developed

MP-PBmice, a web-based application for large-scale insertional mutation mapping onto the mouse genome, while Lim et al. [14] present BioBarcode, a database resource for Asian biodiversity resources. Jenjaroenpun and Kuznetsov [15] tackle the problem of identifying triplex DNA forming region, towards discovering biologically meaningful genome modules and to optimize experimental design of anti-gene treatments, while Saeed and Halgamuge [16] propose a one-dimensional signature derived from oligonucleotide frequency, for efficient grouping or "binning" of metagenome fragments, crucial to the success of microbial genome consortia. Chacko and Ranganathan [17] present the genome-wide analysis of alternative splicing in the bovine genome, with implications for using the cow as a model for specific human diseases.

Genetic and population analysis
For genotype-phenotype correlations, understanding mitochondrial variations of each individual, haplogroup or geographical location is of primary importance. Mito-Variome (Lee at al. [18]) provides a human mitochondrial variation resource for researchers in this area. To identify splice variants and SNPs in short sequence reads from next-generation sequencing approaches, Bao et al. [19] have developed a software tool, MapNext.

Transcriptome analysis
DNA microarrays have led to a multitude of gene expression analysis studies. Nguyen and Lió [20] have developed a new metric, BayesGen, to measure the similarity between gene expression profiles, for constructing genome-wide co-expression networks, and for clustering cancer human tissues into subtypes. Kuem et al. [21] present a pathway-based gene expression similarity measure, which outperforms other commonly used similarity measures, while Ho et al. [22] have used quantile regression models to characterize gene-expression patterns underpinning the molecular regulatory mechanisms leading to mammalian ageing.

Interactome analysis
Interaction data from biology present enormous challenges to bioinformatics in the "-omics" era and contribute to our understanding the functions of biological molecules and systems. Kadupitige et al. [23] present a tool for exploratory analysis of gene interaction networks, while Park et al. [24] have developed a prototype database and server for analyzing protein interaction networks. Maulik et al. [25] have applied protein interaction data to understand plant defense mechanisms, while Reja et al. [26] have developed MitoInteractome, a mitochondrial protein interaction resource for characterizing the ageing process.

Structural genomics
Towards understanding biological function, large-scale biological structure determination projects are currently underway, leading to structure analysis into protein domains by Yoo et al. [27]. Where experimental structures are not available, structure prediction is still an important area of research, with Liu et al. [28] presenting a new sequence-based hybrid predictor to identify conformationally ambivalent regions in proteins. Huang et al. [29] have used sequence and structural information to predict DNA-binding residues in transcription factors, while Jongkon et al. [30] have predicted the binding preference of various strains of avian influenza A to cognate human receptors. Prasad et al. [31] have employed RNA secondary structure and sequence motifs for the phylogenetic analysis of flukes.

Networks, pathways and systems biology
With the growing interest in systems biology, six studies in computations approaches to networks, pathways and systems were presented at InCoB2009. Kim and Gelenbe [32] propose the G-network for the steady-state behaviour of gene regulatory networks for identifying disease-causing genes while Le et al. [33] have applied rule induction learning to characterise nucleosome dynamics from genomic and epigenetic data. Chaturvedi and Rajapakse [34] have developed a skip-chain model to study time-delayed regulation in gene regulatory networks. Rhee et al. [35] identifed cell cycle-related regulatory motifs using a kernel canonical correlation analysis. Mapping virus-host protein interactions to host signalling pathways has been carried out [36] to characterize alternative pathways that can be targeted by drugs, while Min and Hong [37] has analysed how rate-limiting enzymes influence metabolic flux.

Disease informatics
The human genome was primarily sequenced to understand the genetic basis of diseases. In this area, Yang et al. [38] have compiled a database of Parkinson's Disease-related genes and genetic variation. Khan and Ranganathan [39] have used a multi-species comparative structural approach of inherited mutations to correlate genotype to phenotype in an inherited disease.

Clinical informatics
Ultimately, bioinformatics will have to process clinical data to address the quest for personalized medicine. As medical and biological data are commonly presented usually as small datasets without balanced class distribution, Yang et al. [40] have developed a particle swarm approach to address this imbalance for the discovery of new pathological conditions and disease subtypes. Jho et al. [41] have developed a data deposition system integrated with

automated bioinformatics tools for mutation detection and analysis starting from raw patient data.

Conclusion

Judging from the breadth and depth of topics covered in this issue, Asia Pacific bioinformatics research has not only maintained its level of research achievement, but extended computational approaches to systems biology and medical and clinical informatics. Computational biology is now acknowledged to be a core research discipline in our educational and research institutions. We also note that several papers bearing graduate students as first authors, testifying to the quality of research training imparted. In order to effectively train new entrants to undertake bioinformatics research in the "-omics" era, WEBCB has, over the past two years, run discussion fora to define the key elements constituting a minimum set of bioinformatics skills. Based on the outcome of these discussions, Tan et al. [42] present the minimum skill set in bioinformatics and computational biology that an ideal graduating student in life sciences would possess.

Furthermore, computational biology is now considered an essential area of research for supporting to "-omics" and health research, as reported by scientists from Malaysia [43] and Singapore [44] and elsewhere. However, the situation is plagued by author ambiguity especially for Asian names, broken links for web tools, disappearing databases and inadequate disclosure, not enough for reproducibility.

New initiatives from the Asia-Pacific Bioinformatics network include compliance standards such as Minimum Information about a Bioinformatics investigation (MIABi), currently under development. This MIABi compliance will require, firstly, for authors to be issued with unique author identifiers (e.g. see prototype at http://aid.apbionet.org/) for identity disambiguation and accountability purposes. Authors with multiple identifiers issued by various publishers (e.g. Schopus author ID, researcherID) can now be resolved to a unique individual through indentifiers cross-referencing.

Secondly, it will eventually require deposition of scientific datasets through a central portal (e.g. http://docid.apbionet.org/) for persistence, provenanc, accessibility and reproducibility. All databses, datasets and codes cited in papers published through our processes may be mandated to be archived in this way, supported by distributed repository nodes, such as that of the Asian Bioinformation Centers [45] initiative. Moreover, a database on a pre-configured operating system (OS) such as BioSlax (http://www.bioslax.com) can also be archived as an image, stored at such repositories. Even though the original database server may no longer be available, the database-OS image can be dynamically re-instantiated on demand via a cloud computing virtualized platform.

Competing interests

The authors declare that they have no competing interests.

Additional material

> ### Additional file 1
> *APBioNet InCoB2009 Program Committee members and reviewers.*
> Click here for file
> [http://www.biomedcentral.com/content/supplementary/1471-2164-10-S3-S1-S1.pdf]

References

1. Ranganathan S, Subbiah S, Tan TW: **APBioNet: the Asia-Pacific regional consortium for bioinformatics.** *Appl Bioinformatics* 2002, **1:**101-105.
2. Ranganathan S, Hsu WL, Yang UC, Tan TW: **Emerging strengths in Asia Pacific bioinformatics.** *BMC Bioinformatics* 2008, **9(Suppl 11):**S2.
3. Ranganathan S: **Bioinformatics Education--Perspectives and Challenges.** *PLoS Comput Biol* 2005, **1:**e52.
4. Ranganathan S, Tammi M, Gribskov M, Tan TW: **Establishing bioinformatics research in the Asia Pacific.** *BMC Bioinformatics* 2007, **7(Suppl 5):**S1.
5. Ranganathan S, Gribskov M, Tan TW: **Bioinformatics Research in the Asia Pacific - an update.** *BMC Bioinformatics* 2007, **9(Suppl 1):**S1.
6. Ranganathan S: **Towards a career in bioinformatics.** *BMC Bioinformatics* 2009, **10(Suppl 15):**S1.
7. **Bioinformation** [http://www.bioinformation.net/]
8. Zhao SQ, Wang J, Zhang L, Li JT, Gu X, Gao G, Wei L: **BOAT: Basic Oligonucleotide Alignment Tool.** *BMC Genomics* 2009, **10(Suppl 3):**S2.
9. Chung WH, Park SB: **An empirical study of choosing efficient discriminative seeds for oligonucleotide design.** *BMC Genomics* 2009, **10(Suppl 3):**S3.
10. Piriyapongsa J, Ngamphiw C, Assawamakin A, Wangkumhang P, Suwannasri P, Ruangrit U, Agavatpanitch G, Tongsima S: **REx-Primer: an integrated primer designing tool increases PCR effectiveness by avoiding 3' SNP-in-primer and mis-priming from structural variation.** *BMC Genomics* 2009, **10(Suppl 3):**S4.
11. Venkatachalam G, Kumar AP, Yue LS, Pervaiz S, Clement MV, Sakharkar MK: **Computational identification and experimental validation of PPRE motifs in NHE1 and MnSOD genes of Human.** *BMC Genomics* 2009, **10(Suppl 3):**S5.
12. Wee LJK, Tong JC, Tan TW, Ranganathan S: **A multi-factor model for caspase degradome prediction.** *BMC Genomics* 2009, **10(Suppl 3):**S6.
13. Yang W, Jin K, Xie X, Li D, Yang J, Wang L, Gu N, Zhong Y, Sun LV: **Development of a database system for mapping insertional**

mutations onto the mouse genome with large-scale experimental data. *BMC Genomics* 2009, **10(Suppl 3):**S7.

14. Lim J, Kim SY, Kim S, Eo HS, Kim CB, Paek WK, Kim W, Bhak J: **BioBarcode: a general DNA barcoding database and server platform for Asian biodiversity resources.** *BMC Genomics* 2009, **10(Suppl 3):**S8.

15. Jenjaroenpun P, Kuznetsov VA: **TTS Mapping: integrative WEB tool for analysis of triplex formation target DNA Sequences, G-quadruplets and non-protein coding regulatory DNA elements in the human genome.** *BMC Genomics* 2009, **10(Suppl 3):**S9.

16. Saeed I, Halgamuge SK: **The oligonucleotide frequency derived error gradient and its application to the binning of metagenome fragments.** *BMC Genomics* 2009, **10(Suppl 3):**S10.

17. Chacko E, Ranganathan S: **Genome-wide analysis of alternative splicing in cow: implications in bovine as a model for human diseases.** *BMC Genomics* 2009, **10(Suppl 3):**S11.

18. Lee YS, Kim WY, Ji M, Kim JH, Bhak J: **MitoVariome: a variome database of human mitochondrial DNA.** *BMC Genomics* 2009, **10(Suppl 3):**S12.

19. Bao H, Xiong Y, Guo H, Zhou R, Lu X, Yang Z, Zhong Y, Shi S: **MapNext: a software tool for spliced and unspliced alignments and SNP detection of short sequence reads.** *BMC Genomics* 2009, **10(Suppl 3):**S13.

20. Nguyen VA, Lió P: **Measuring similarity between gene expression profiles: a Bayesian approach.** *BMC Genomics* 2009, **10(Suppl 3):**S14.

21. Keum C, Woo JH, Oh WS, Park SN, No KT: **Improving gene expression similarity measurement using pathway-based analytic dimension.** *BMC Genomics* 2009, **10(Suppl 3):**S15.

22. Ho JWK, Stefani M, Remedios CGd, Charleston MA: **A model selection approach to discover age-dependent gene expression patterns using quantile regression models.** *BMC Genomics* 2009, **10(Suppl 3):**S16.

23. Kadupitige SR, Leung KC, Sellmeier J, Sivieng J, Catchpoole DR, Bain ME, Gaëta BA: **MINER: exploratory analysis of gene interaction networks by machine learning from expression data.** *BMC Genomics* 2009, **10(Suppl 3):**S17.

24. Park SJ, Choi JS, Kim BC, Jho SW, Ryu JW, Park D, Lee KA, Bhak J, Kim SI: **PutidaNET: Interactome database service and network analysis of Pseudomonas putida KT2440.** *BMC Genomics* 2009, **10(Suppl 3):**S18.

25. Maulik A, Ghosh H, Basu S: **Comparative study of protein-protein interaction observed in PolyGalacturonase-Inhibiting Proteins from Phaseolus vulgaris and Glycine max and PolyGalacturonase from Fusarium moniliforme.** *BMC Genomics* 2009, **10(Suppl 3):**S19.

26. Reja R, Venkatakrishnan AJ, Lee J, Kim BC, Ryu JW, Gong S, Bhak J, Park D: **MitoInteractome: Mitochondrial protein interactome database, and its application in 'aging network' analysis.** *BMC Genomics* 2009, **10(Suppl 3):**S20.

27. Yoo PD, Zhou BB, Zomaya AY: **A modular kernel approach for integrative analysis of protein domain boundaries.** *BMC Genomics* 2009, **10(Suppl 3):**S21.

28. Liu YC, Yang MH, Lin WL, Huang CK, Oyang YJ: **A sequence-based hybrid predictor for identifying conformationally ambivalent regions in proteins.** *BMC Genomics* 2009, **10(Suppl 3):**S22.

29. Huang YF, Huang CC, Liu YC, Oyang YJ, Huang CK: **DNA-binding residues and binding mode prediction with binding-mechanism concerned models.** *BMC Genomics* 2009, **10(Suppl 3):**S23.

30. Jongkon N, Mokmak W, Chuakheaw D, Shaw PJ, Tongsima S, Sangma C: **Prediction of avian influenza A binding preference to human receptor using conformational analysis of receptor bound to hemagglutinin.** *BMC Genomics* 2009, **10(Suppl 3):**S24.

31. Prasad PK, Tandon V, Biswal DK, Goswami LM, Chatterjee A: **Phylogenetic reconstruction using secondary structures and sequence motifs of ITS2 rDNA of *Paragonimus westermani* (Kerbert, 1878) Braun, 1899 (Digenea: Paragonimidae) and related species.** *BMC Genomics* 2009, **10(Suppl 3):**S25.

32. Kim H, Gelenbe E: **Anomaly detection in gene expression via stochastic models of gene regulatory networks.** *BMC Genomics* 2009, **10(Suppl 3):**S26.

33. Le NT, Ho TB, Tran DH: **Characterizing nucleosome dynamics from genomic and epigenetic information using rule induction learning.** *BMC Genomics* 2009, **10(Suppl 3):**S27.

34. Chaturvedi I, Rajapakse JC: **Detecting robust time-delayed regulation in *Mycobacterium tuberculosis*.** *BMC Genomics* 2009, **10(Suppl 3):**S28.

35. Rhee JK, Joung JG, Chang JH, Fei Z, Zhangn BT: **Identification of cell cycle-related regulatory motifs using a kernel canonical correlation analysis.** *BMC Genomics* 2009, **10(Suppl 3):**S29.

36. Balakrishnan S, Tastan O, Carbonell J, Seetharaman JK: **Alternative paths in HIV-1 targeted human signal transduction pathways.** *BMC Genomics* 2009, **10(Suppl 3):**S30.

37. Zhao M, Qu H: **Human liver rate-limiting enzymes influence metabolic flux via branch points and inhibitors.** *BMC Genomics* 2009, **10(Suppl 3):**S31.

38. Yang JO, Kim WY, Jeong SY, Oh JH, Jho S, Bhak J, Kim NS: **PDbase: a database of Parkinson's Disease-related genes and genetic variation using substantia nigra ESTs.** *BMC Genomics* 2009, **10(Suppl 3):**S32.

39. Khan JM, Ranganathan S: **A multi-species comparative structural bioinformatics analysis of inherited mutations in α-D-Mannosidase reveals strong genotype-phenotype correlation.** *BMC Genomics* 2009, **10(Suppl 3):**S33.

40. Yang P, Xu L, Zhou BB, Zhang Z, Zomaya AY: **A particle swarm based hybrid system for imbalanced medical data sampling.** *BMC Genomics* 2009, **10(Suppl 3):**S34.

41. Jho S, Kim BC, Ghang H, Kim JH, Park D, Kim HM, Jung SY, Yoo KY, Kim HJ, Lee S, Bhak J: **COMUS: Clinician-Oriented locus-specific Mutation detection and deposition System.** *BMC Genomics* 2009, **10(Suppl 3):**S35.

42. Tan TW, Lim SJ, Khan AM, Ranganathan S: **A proposed minimum skill set for university graduates to meet the informatics needs and challenges of the "-omics" era.** *BMC Genomics* 2009, **10(Suppl 3):**S36.

43. Zeti AMH, Shamsir MS, Tajul-Arifin K, Merican AF, Mohamed R, Nathan S, Nor Muhammad M, Napis S, Tan TW: **Bioinformatics in Malaysia: Hope, Initiative, Effort, Reality, and Challenges.** *PLoS Comput Biol* 2009, **5(8):**e1000457.

44. Eisenhaber F, Kwoh C-K, Ng S-K, Sung W-K, Wong L: **Brief Overview of Bioinformatics Activities in Singapore.** *PLoS Comput Biol* 2009, **5(9):**e1000508.

45. **Asian Bioinformation Center**   [http://abcenter.org/]