BMC Genomics

INTRODUCTION

InCoB celebrates its tenth anniversary as first joint conference with ISCB-Asia

Christian Schönbach[1*], Tin Wee Tan[2], Janet Kelso[3], Burkhard Rost[4], Sheila Nathan[5,6], Shoba Ranganathan[7,2]

From Asia Pacific Bioinformatics Network (APBioNet) Tenth International Conference on Bioinformatics – First ISCB Asia Joint Conference 2011 (InCoB/ISCB-Asia 2011)
Kuala Lumpur, Malaysia. 30 November - 2 December 2011

Abstract

In 2009 the International Society for Computational Biology (ISCB) started to roll out regional bioinformatics conferences in Africa, Latin America and Asia. The open and competitive bid for the first meeting in Asia (ISCB-Asia) was awarded to Asia-Pacific Bioinformatics Network (APBioNet) which has been running the International Conference on Bioinformatics (InCoB) in the Asia-Pacific region since 2002. InCoB/ISCB-Asia 2011 is held from November 30 to December 2, 2011 in Kuala Lumpur, Malaysia. Of 104 manuscripts submitted to BMC Genomics and BMC Bioinformatics conference supplements, 49 (47.1%) were accepted. The strong showing of Asia among submissions (82.7%) and acceptances (81.6%) signals the success of this tenth InCoB anniversary meeting, and bodes well for the future of ISCB-Asia.

Introduction

The tenth InCoB (International Conference on Bioinformatics), an official conference of the Asia-Pacific Bioinformatics Network (APBioNet) [1], was held in Kuala Lumpur, Malaysia as a joint conference with the ISCB-Asia meeting of the International Society for Computational Biology (ISCB) [2]. InCoB/ISCB-Asia 2011 is the result of ISCB's mission to host meetings in locations that are not well served by the world's largest bioinformatics conference ISMB (International Conference on Intelligent Systems for Molecular Biology). Owing to the success of previous InCoB conferences [3,4], and as ISCB's first regional affiliate, APBioNet was selected to co-organize the first ISCB-Asia conference.

Manuscript submission and review

We offered authors three tracks to submit manuscripts for potential publication in the supplement issues of *BMC Bioinformatics* or *BMC Genomics* (BMC track) and *Immunome Research* (IR) [5]. Of 110 submitted manuscripts from 19 countries 104 were designated for the BMC track. Immunome Research received three submissions. Shorter manuscripts representing research in progress were also sought for Bioinformation (BI), which received three submissions. Most manuscripts received three reviews from the 76 member strong Program Committee, supported by 39 additional reviewers (Additional File 1). The first round of reviews resulted in the provisional acceptance of 32 (30.8%) manuscripts, with minor revisions. The authors of 28 manuscripts, including three that were transferred by the Program Committee Co-chairs to BI and IR tracks, had to address major concerns raised by the reviewers. After a second round of review, 26 (25%) manuscripts were provisionally accepted pending minor revisions. A statistics of submissions and final acceptances by countries is shown in Additional File 2. The manuscript revision and re-review policy of InCoB/ISCB-Asia 2011 resulted in a more than twofold greater acceptance rate compared to ISMB/ECCB 2011 [6] which considers only manuscripts requiring minor revisions.

The 25 articles selected for this issue represent a cross-section of scientific and technological innovations accelerating bioinformatics in the areas of databases, software tools, RNomics, next-generation sequencing,

* Correspondence: schoen@bio.kyutech.ac.jp
[1]Department of Bioscience and Bioinformatics, Graduate School of Computer Science and Systems Engineering, Kyushu Institute of Technology, Fukuoka 820-8502, Japan
Full list of author information is available at the end of the article

sequence analysis, evolution, proteome analysis and disease informatics. The remaining 24 accepted BMC articles are published as a BMC Bioinformatics supplement [7].

Biological databases and software tools

Papers presenting a wide variety of databases and software tools were accepted [8-12]. Particular focus areas include the development of software for the alignment of apicomplexan sequences [8] which have an unusual sequence composition, software for the identification of gene regulatory modules from microarray data [9], and software for the identification of taxonomic identity from metagenomic samples [10]. Liverome [11] and Detoxiprot [12] represent databases for collecting and sharing of network modules enriched in liver cancer-pathways as well as enzymes, substrates, inducers and inhibitors of detoxification.

RNomics and next-generation sequencing

Four papers make use of next-generation sequencing (NGS) [13-16] to address a range of challenges from the identification of the taxonomic origin of organism sequences in metagenomic studies [13] to the identification of unique and novel genes in metatranscriptomic data [14,15].

Fast-growing trees from plantations are used in pulp and fiber production. Lignin, an essential component of the tree structure, is undesired in this process. Ong and Wickneswari [16] applied NGS to profile the expression of *Acacia mangium* small RNAs in secondary xylem samples of varying lignin content. The obtained small RNA profiles, which include 82 novel miRNAs may open venues to silence monolignol biosynthesis-associated genes without compromising the fitness of the tree.

Secretory proteins are important players in host-pathogen interactions. Existing prediction tools do not support the identification of secretory protein candidates from NGS data, thus hampering the identification of non-classical secreted protein candidates. Garg and Ranganathan [17] developed an assembly protocol for NGS-derived sequence data to cluster, translate and homology search an in-house dataset of experimentally determined parasitic helminth excretory/secretory proteins. Another aspect of the cellular secretome is exosomes or vesicles used in intracellular communication and transport of cellular components such as RNA. Since exosomes could be used as vectors for therapeutic drug delivery, it is of interest to identify motifs involved in targeting RNA for secretion. Bagatov and colleagues [18] determined *ab initio* various short linear motifs that are specific for certain exosome-enriched RNAs.

Evolution and sequence analysis

The growing number of complete genome drafts and protein sequence data allows to re-visit questions on the origin of vertebrate promoters, evolutionary conservation of enzymes, as well as the generation of functional diversity through gene duplication and alternative splicing. Profiling of ascidian promoters revealed a primordial vertebrate promoter type that appears to be partially methylated with a high but limited extent of CpG scores and G+C content [19]. Phylogenetic profiles of rate-limiting enzymes with inhibiting relations showed higher conservation across human, rat, mouse, budding yeast and *E. coli* than common enzymes [20]. Chen *et al.* [21] analysed the relationship between the two mechanisms using protein sequence and paralog information of seven organisms ranging from human to *C. elegans*. The results indicate a duplication age-dependent relationship of alternative splicing and an evolutionary constraint among duplicates. Alternative splicing was more frequently observed among ancient than recent duplicates.

The increasing scale and speed of whole-genome sequencing requires faster and high-accuracy genome mapping methods. A new method that utilizes perfect Hamming Code with a hash table significantly reduces the number of hash keys required for searching genome positions [22]. The Universal Automatic SNP Identification System (UASIS) [23], cross-references SNP identifiers of various databases to bridge potential SNP nomenclature ambiguities and to promote compliance with Human Genome Variation Society guidelines.

Proteome analysis

Knowledge of protein-protein interaction (PPI) data quality and their limitations is critical for interpreting results and essential for using them as a reference set. Zhou and Wong [24] found that *M. tuberculosis* H37Rv bacterial two-hybrid (B2H) data seemed to contain numerous false positives and false negatives, while the H37Rv STRING predicted PPIs included many none-direct interactions. Yet, stringently predicted H37Rv STRING PPI data appeared to be suitable as a reference for analysing functional associations rather than physical interactions. To understand the behaviour of proteins in PPI networks, PPIs are often correlated with gene expression parameters. The correlation of high-confidence human PPIs derived from the HitPredict database with gene expression stability improved the classification of transient obligatory hubs and revealed functionally distinct, previously unknown type of hubs [25].

All protein domains common to mouse-infecting viruses and mice were analyzed and published as MusVirus database and tool [26]. Biochemical and functional interpretations of the shared domains indicated that the

analysed viruses prefer to acquire genes of the innate immune response pathways. Granzymes are serine proteases that are implicated in the pathogenesis of several chronic inflammatory and cardiovascular disorders. Wee *et al.* [27] implemented a webserver to predict the potential granzyme B degradome using support vector machine classifiers derived from experimentally verified substrate cleavage sites.

Beta-D-mannosidase mutations cause lysosomal storage disease. A comparative structural bioinformatics analysis of inherited mutations in β-D-mannosidase across multiple species revealed a geno-phenotpye correlation [28]. The identification of five mutational hotspots indicated that the closeness of mutations to the active site correlates with the severity of phenotypes and thus allowing phenotype severity predictions.

Disease informatics

Papers on the discovery and analysis of biomarkers and disease gene candidates focussed on copy number variations (CNVs) [29], cut-circularize-linearize-and-paste (CCLP) models [30] of chromosomal rearrangements and an extreme class-discrimination (ECD) [31] of lung adenocarcinoma gene expression data. Shia *et al.* [29] examined CNVs in Taiwanese hyperlipidemia and myocardial infarction patient cohorts. A multistage analysis revealed seven CNV regions that were associated with both disorders and deserve further exploration as potential biomarkers for early-stage diagnosis. To study the human immune response to tumours and phylogenetic tree reconstructions, Huang *et al* [30] developed an algorithm that minimizes the number of CCLP operations to sort permutations of a number of genes in a given chromosome.

Lack of clinically applied biomarkers for the early diagnosis and aggressiveness of lung adenocarcinoma inspired the development ECD to select features in paired gene expression samples, and to identify genes that are essential for reprogramming lung tissue cells [31]. Compared to commonly used methods, ECD produced highly discriminative variables when the number of samples was small. Nevertheless, prioritizing and experimentally verifying new candidates can be time-consuming and expensive. Hsu *et al.* [32] proposed a parameter-free interconnectedness method to rank disease candidates by evaluating their distance to known disease genes in a network.

Conclusion

In recent years numerous Asian bioscience conferences have added bioinformatics sessions to their program. Yet, higher-profile annual conferences dedicated to bioinformatics remained largely in North America and Europe. The high number of manuscript and poster submissions to InCoB/ISCB-Asia 2011 suggests that the joint conference model ISCB-Asia may indeed invigorate regional bioinformatics-themed meetings in Asia, which are necessary to advance bioinformatics research, particularly in South and Southeast Asia. The strengthening of these regional conferences should lead to a gradual and healthy consolidation of smaller meetings that will benefit participants. Towards this end, we expect to see you again at the 11th InCoB in Bangkok or at the 2nd ISCB-Asia conference.

Additional material

> **Additional File 1: List of Program Committee Members and Additional Reviewers in Alphabetical Order**
>
> **Additional File 2: A. Submission and Acceptance Statistics by Regions and Countries B. Acceptance Rate for Submissions to BMC track**

Author details

[1]Department of Bioscience and Bioinformatics, Graduate School of Computer Science and Systems Engineering, Kyushu Institute of Technology, Fukuoka 820-8502, Japan. [2]Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, , Singapore 117597. [3]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany. [4]Department for Bioinformatics and Computational Biology, Faculty for Informatics, Technical University Munich, 85748 Garching, Germany. [5]School of Biosciences and Biotechnology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi Selangor D.E., Malaysia. [6]Malaysia Genome Institute, Jalan Bangi, 43000 Kajang Selangor D. E., Malaysia. [7]Department of Chemistry and Biomolecular Sciences and ARC Centre of Excellence in Bioinformatics, Macquarie University, Sydney NSW 2109, Australia.

Authors' contributions

CS and SR (Program Committee Co-chairs) wrote the introduction and managed the review and editorial processes. CS, SR, JK (Chair, ISCB Conferences Committee), BR, TWT and SN (Conference Chair) jointly contributed to the scientific program development and its implementation. TWT supported the post-acceptance manuscript processing.

Competing interests

The authors declare that they have no competing interests.

Published: 30 November 2011

References

1. The Asia-Pacific Bioinformatics Network. [http://www.apbionet.org].
2. The International Society for Computational Biology. [http://www.iscb.org].
3. Ranganathan S, Schönbach C, Nakai K, Tan TW: Asia Pacific bioinformatics: a 2010 update. *BMC Genomics* 2010, **11**(Suppl 4):S1.
4. Schönbach C, Nakai K, Tan TW, Ranganathan S: InCoB2010 – 9th International Conference on Bioinformatics at Tokyo, Japan, September 26-28, 2010. *BMC Bioinformatics* 2010, **11**(Suppl 7):S1.
5. Immunome Research. [http://immunome-research.net/].
6. Gaasterland T, Vingron M: Editorial. *Bioinformatics* 2011, **27**(13):i1.
7. Ranganathan S, Schönbach C, Kelso J, Rost B, Nathan S, Tan TW: Towards big data science in the decade ahead from ten years of InCoB and the 1st ISCB-Asia. *BMC Bioinformatics* 2011, **12**(Suppl 13):S1.
8. Ali J, Paila U, Ranjan A: ApicoAlign: An alignment and sequence search tool for apicomplexan proteins. *BMC Genomics* 2011, **12**(Suppl 3):S6.
9. Kim M, Shin H, Chung TS, Joung JG, Kim JH: Extracting regulatory modules from gene expression data by sequential pattern mining. *BMC Genomics* 2011, **12**(Suppl 3):S5.
10. Mohammed MH, Ghosh TS, Reddy RM, Reddy CVSK, Singh NK, Mande SS: INDUS - A composition based approach for rapid and accurate taxonomic classification of metagenomic sequences. *BMC Genomics* 2011, **12**(Suppl 3):S4.
11. Lee L, Wang K, Li G, Xie Z, Wang Y, Xu J, Sun S, Pocalyko D, Bhak J, Kim C, Lee KH, Jang YJ, Yeom YI, Yoo HS, Hwang S: Liverome: a curated database of liver cancer-related gene signatures with self-contained context information. *BMC Genomics* 2011, **12**(Suppl 3):S3.
12. Yang Z, Yu Y, Yao L, Li G, Wang L, Hu Y, Wei H, Wang L, Hammami R, Razavi R, Zhong Y, Liang X: DetoxiProt: an integrated database for detoxification proteins. *BMC Genomics* 2011, **12**(Suppl 3):S2.
13. Mohammed MH, Ghosh TS, Chadaram S, Mande SS: i-rDNA: Alignment-free algorithm for rapid *in silico* detection of ribosomal gene fragments from metagenomic sequence data sets. *BMC Genomics* 2011, **12**(Suppl 3):S12.
14. Mitra S, Stärk M, Huson DH: Analysis of 16S rRNA environmental sequences using MEGAN. *BMC Genomics* 2011, **12**(Suppl 3):S17.
15. Ogura A, Lin M, Shigenobu Y, Fujiwara A, Ikeo K, Nagai S: Effective gene collection from the metatranscriptome of marine microorganisms. *BMC Genomics* 2011, **12**(Suppl 3):S15.
16. Ong SS, Wickneswari R: Expression profile of small RNAs in *Acacia mangium* secondary xylem tissue with contrasting lignin content - a potential regulatory sequence in monolignol biosynthetic pathway. *BMC Genomics* 2011, **12**(Suppl 3):S13.
17. Garg G, Ranganathan S: *In silico* secretome analysis approach for next generation sequencing transcriptomic data. *BMC Genomics* 2011, **12**(Suppl 3):S14.
18. Batagov AO, Kuznetsov VA, Kurochkin IV: Identification of nucleotide patterns enriched in secreted RNAs as putative *cis*-acting elements targeting them to exosome nano-vesicles. *BMC Genomics* 2011, **12**(Suppl 3):S18.
19. Okamura K, Yamashita R, Takimoto N, Nishitsuji K, Suzuki Y, Kusakabe TG, Nakai K: Profiling ascidian promoters as the primordial type of vertebrate promoter. *BMC Genomics* 2011, **12**(Suppl 3):S7.
20. Zhao M, Qu H: High similarity of phylogenetic profiles of rate-limiting enzymes with inhibitory relation in human, mouse, rat, budding yeast and *E. coli*. *BMC Genomics* 2011, **12**(Suppl 3):S10.
21. Chen TW, Wu TH, Wailap V, Ng WV, Lin WC: Interrogation of alternative splicing events in duplicated genes during evolution. *BMC Genomics* 2011, **12**(Suppl 3):S16.
22. Takenaka T, Seno S, Matsuda H: Perfect Hamming code with a hash table for faster genome mapping. *BMC Genomics* 2011, **12**(Suppl 3):S8.
23. Poo DCC, Cai S, Mah JTL: UASIS: universal automatic SNP identification system. *BMC Genomics* 2011, **12**(Suppl 3):S9.
24. Zhou H, Wong L: Comparative analysis and assessment of *M. tuberculosis* H37Rv protein-protein interaction datasets. *BMC Genomics* 2011, **12**(Suppl 3):S20.
25. Patil A, Nakai K, Kinoshita K: Assessing the utility of gene co-expression stability in combination with correlation in the analysis of protein-protein interaction networks. *BMC Genomics* 2011, **12**(Suppl 3):S19.
26. Kinjo AR, Kumaga Y, Dinh H, Takeuchi O, Standley DM: Functional characterization of protein domains common to animal viruses and mouse. *BMC Genomics* 2011, **12**(Suppl 3):S21.
27. Wee LJK, Er EPS, Ng LFP, Tong JC: *In silico* prediction of the granzyme B degradome. *BMC Genomics* 2011, **12**(Suppl 3):S11.
28. Hyunh T, Khan JM, Ranganathan S: A comparative structural bioinformatics analysis of inherited mutations in β-D-mannosidase across multiple species reveals a genotype-phenotype correlation. *BMC Genomics* 2011, **12**(Suppl 3):S22.
29. Shia WC, Ku TH, Tsao YM, Hsia CH, Chang YM, Huang CH, Chung YC, Hsu SL, Liang KW, Hsu FR: Genetic copy number variants in myocardial infarction patients with hyperlipidemia. *BMC Genomics* 2011, **12**(Suppl 3):S23.
30. Huang KH, Chen KT, Lu CL: Sorting permutations by cut-circularize-linearize-and-paste operations. *BMC Genomics* 2011, **12**(Suppl 3):S26.
31. Toh SH, Prathipati P, Motakis E, Yanamandra PS, Kwoh CK, Kuznetsov V: A robust tool for discriminative analysis and feature selection in paired samples impacts the identification of the genes essential for reprogramming lung tissue to adenocarcinoma. *BMC Genomics* 2011, **12**(Suppl 3):S24.
32. Hsu CL, Huang YH, Hsu CT, Yang UC: Prioritizing disease candidate genes by a gene interconnectedness-based approach. *BMC Genomics* 2011, **12**(Suppl 3):S25.