

Meeting report

Open Access

Highlights from the Fourth International Society for Computational Biology Student Council Symposium at the Sixteenth Annual International Conference on Intelligent Systems for Molecular Biology

Lucia Peixoto*¹, Nils Gehlenborg^{2,3} and Sarath Chandra Janga*⁴

Address: ¹Biology Department and Penn Genomics Frontier Institute, University of Pennsylvania, Philadelphia, PA, USA, ²European Bioinformatics Institute, Hinxton, UK, ³Graduate School of Life Sciences, University of Cambridge, Cambridge, CB2 1RX, UK and ⁴MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK

Email: Lucia Peixoto* - luciap@sas.upenn.edu; Sarath Chandra Janga* - sarath@mrc-lmb.cam.ac.uk

* Corresponding authors

from Fourth International Society for Computational Biology (ISCB) Student Council Symposium
Toronto, Canada. 18 July 2008

Published: 30 October 2008

BMC Bioinformatics 2008, **9**(Suppl 10):11 doi:10.1186/1471-2105-9-S10-11

This article is available from: <http://www.biomedcentral.com/1471-2105/9/S10/11>

© 2008 Peixoto et al; licensee BioMed Central Ltd

Abstract

In this meeting report we give an overview of the talks and presentations from the Fourth International Society for Computational Biology (ISCB) Student Council Symposium held as part of the annual Intelligent Systems for Molecular Biology (ISMB) conference in Toronto, Canada. Furthermore, we detail the role of the Student Council (SC) as an international student body in organizing this symposium series in the context of large, international conferences.

Introduction

The Student Council (SC) of the International Society for Computational Biology (ISCB) is a student-led, worldwide network of young researchers in computational biology and bioinformatics. The ISCB Board of Directors officially approved the SC in July 2004 at the joint European Conference on Computational Biology (ECCB) and Intelligent Systems for Molecular Biology (ISMB) conference in Glasgow, Scotland. The major aims of the SC are to organize events and facilitate networking opportunities for SC members while nurturing soft skills to complement the normal academic program, such as organizational, teamwork and networking skills. Since its inception, the SC has organized an annual student symposium for the benefit of the student community in the field of computational biology and bioinformatics. This year the 4th ISCB

Student Council Symposium was held in Toronto on the 18th of July, as a satellite meeting of the Annual International conference ISMB 2008, with the participation of over 100 delegates.

The symposium opened with a keynote by ISCB president Burkhard Rost (Columbia University, USA) on "Evolution teaches protein prediction". Later in the day Mark Gerstein (Yale University, USA) talked on computational proteomics emphasizing the study of protein motions from a database perspective during his lecture. The symposium ended with a closing keynote by Timothy Hughes (University of Toronto, Canada), who gave an overview of the complexity and diversity of the protein-DNA Interactome to the audience, concluding with an impressive array of talks by internationally well-acclaimed scientists. A

research and industry partners session provided a chance for students to become familiar with the opportunities outside the academic world. The session included presentation by Louisa Wright (EBI-EMBL), Jong Bhak, (Korean BioInformation Center, KOBIC) and Richard Wintle (The Centre for Applied Genomics, Canada). A panel discussion on "Career Paths in Bioinformatics and Computational Biology" led by Manuel Corpas with the participation of Philip E. Bourne (University of California San Diego, USA), Alfonso Valencia (Spanish National Research Council, Spain), Jong Bhak, (Korean BioInformation Center, KOBIC) and Richard Wintle (The Centre for Applied Genomics, Canada) gave the attendees the opportunity to ask questions about their opinion on different choices for a future in the field, to both people in academic and non-academic sectors.

Proceedings

Students were given the chance to present their work during two student presentation sessions, which included a total of nine talks, and a poster session at the end of the symposium. The goal of the proceedings of the symposium is to present a selection of the best abstracts selected from a total of 75 submissions. The abstracts that have been selected for this supplement comprise eight abstracts chosen from oral presentations and a set of six best-ranked abstracts from the posters presented at the symposium during the poster session. Rigorous peer review was carried out by members of the program committee, which comprised of an international group of students and young researchers.

All the abstracts selected for the proceedings can be broadly grouped into the following major themes:

Protein-protein interactions

Last century has seen a revolution in our understanding of biological sciences due to the availability of several high-throughput technologies generating numerous large-scale data sets, which have provided us with an unprecedented knowledge of biological systems. Protein-protein interaction networks, which are an outcome of these technological advancements, have proven to be an invaluable tool to mine functional genomics data and to understand protein function. One of the crucial features of these networks is the presence of functional modules (groups of proteins involved in common biological processes). In line with efforts to identify these groups of functionally related proteins, Cho *et al.* [1] presented a functional flow-based approach to efficiently identify overlapping modules that outperforms the graph-theoretic or data-mining techniques currently available. Another overlooked aspect of protein-protein interactions is that proteins tend to aggregate seemingly as frequently as they form complexes with other proteins. Pechmann *et al.* [2] showed that

interface regions of protein complexes are on average more aggregation prone than other surface regions and that there is indeed a competition between aggregation and protein complex formation. Although understanding protein interactions is extremely useful not only to characterize the functional role of a protein but also to decipher its cellular and biochemical context, large-scale interaction networks are only available for a few model organisms. With that in mind, Michaut *et al.* [3] have developed an automated tool (InterPORC) to predict and transfer the annotations of interacting proteins using the currently available interaction network data for model systems, thereby facilitating the identification of conserved interaction networks for any species with a completed genome sequence.

Annotation and data mining

With increasing number of completely sequenced genomes and large-scale data sets, the task of annotating these genomes and mining the data becomes progressively more complex. This is particularly true for large-scale gene-expression and genotyping experiments done with microarrays on a variety of platforms derived from new technologies, as well as for proteomics data. For most biologists, the low correlation between mRNA and protein abundance estimated by high-throughput methods reported previously is surprisingly low. With a clear need for better technologies to address these questions, Bitton *et al.* [4] showed that the advent of exon arrays, as well as the application of enhanced bioinformatics filtering and peptide mapping techniques supports a tighter integration of quantitative proteomics and microarray data. Another challenge posed by these ever increasing numbers of large-scale studies is our ability to integrate them and annotate related experiments. Ruau *et al.* [5] developed a new way of integrating microarray data and doing semantic annotation with ontologies by employing text mining and expression profile correlation.

Comparative genomics and evolution

One of the most common ways to explore genome-wide data is by studying the evolutionary presence of different components involved in the system, whether that is comparing different species, different variants within a population of the same organism or variation within the same genome. Jain *et al.* [6] illustrated the use of comparative genomics to identify non-classically secreted proteins in fungal genomes and experimentally validated their predictions in *Botrytis cinerea* and symbiotic fungus *Laccaria bicolor*. Bohnert *et al.* [7], on the other hand explored sequence variation within the same species, *Oryza sativa* (rice), using machine learning methods and defined an inventory of sequence variation that constitutes an unprecedented resource for further functional studies and modern breeding of rice. While variation of homologous

sequences between and within species are common ways to enhance adaptation to the environment, duplications play a major role in the evolution of genomes by creating and modifying molecular functions. Abraham *et al.* [8] developed a program to detect repeats in DNA, protein sequences, and in 3D structures and found many symmetrical repeats in 3D structures that may drive protein evolution.

Novel algorithms

Novel technologies for generating high-throughput data require novel algorithms. Most recent sequencing technologies produce huge amounts of reads that are short and error prone, however they come with considerable advantages. Both De Bona *et al.* [9] and Rausch *et al.* [10] developed algorithms to circumvent some of the problems faced with short sequence reads. The first approach employed the additional information, in terms of position-wise quality scores, to improve the alignments using machine learning methods, while the latter developed a new consensus method that is robust to data with high coverage, short reads and genomic variation. Novel sequencing technologies however are limited in their ability to remediate the challenge imposed by repeat elements that are so pervasive in eukaryotic genomes. While a number of algorithms are available for discovering novel dispersed repeats, a significant amount of analysis is required to characterize the new elements. Saha *et al.* [11] presented a novel algorithm for mining spatial relationships among repeat families to yield clusters of repeat regions on the genome.

Bioinformatics of Health and Disease

The utility of bioinformatics tools in understanding complex diseases and public health issues has proven invaluable in the genomic era. However, integrating large-scale data sets to gain meaningful understanding of disease dynamics and disease-causing organisms is challenging. Identifying Single Nucleotide Polymorphisms (SNPs) that are responsible for common and complex diseases such as cancer is of major interest in current molecular epidemiology. Hyoun Lee *et al.* [12] developed a novel framework for prioritizing SNPs based on their possible deleterious effects in a probabilistic framework and thus help prioritize SNPs possibly linked to disease. Understanding the causative agents of human diseases, implies an improvement in our understanding of their genomes, as illustrated by the work of Chukualim *et al.* [13] in establishing a metabolic pathway database that will facilitate research on *Trypanosoma brucei*, the causative agent of african sleeping sickness. Drug target development remains an important challenge in battling disease causing organisms as well as cancer and chronic diseases. Hence, it is not surprising that drug-target prediction using bioinformatics approaches is an extremely fast-growing field, but most of

the available tools are very intuitive and qualitative. Xiao *et al.* [14] presented a work in which they characterized quantitative systems-level determinants of human drug targets and found that genes associated with successful FDA-approved drugs have a number of properties at the network, sequence, and tissue-expression levels that significantly distinguish them from other human genes.

Conclusion

A total of 75 abstracts were received for the symposium. All abstracts were rigorously reviewed by at least two referees to select the best abstracts for oral presentation. Eventually, nine authors were invited to give oral presentations of their abstracts and the remaining were invited to present their work as posters. The research presented covered a wide variety of scientific themes including (1) Protein-protein interactions, (2) Annotation and data mining, (3) Comparative genomics and evolution, (4) Novel algorithms, as well as (5) Bioinformatics of Health and Disease. A selection of 14 outstanding abstracts has been compiled for this supplement.

After 4 years of successful events we are confident that the Student Council Symposium series and related proceedings are a necessary and well-received addition to the program of major conferences in our field. The Student Council leadership is determined to continue and expand its efforts to promote the career-development of the next generation of computational biologists.

Outlook

The 5th ISCB Student Council Symposium will be held in Stockholm, Sweden in conjunction with the joint European Conference on Computational Biology and Intelligent Systems for Molecular Biology (ECCB/ISMB) conference in July 2009. Further information about the Student Council Symposium will be available on our web site at the address <http://www.iscb.org>.

Acknowledgements

We would like to thank ISCB Executive Officer BJ Morrison McKay and ISMB conference organizer Steven Leard for logistic support and invaluable advice in organizing the symposium. We are indebted to all speakers for their time and effort in particular Burkhard Rost, Mark Gerstein and Timothy Hughes. We would also like to thank the ISCB board of directors for their enthusiastic commitment to our mission. We are grateful to the referees for their dedication and effort in reviewing the submitted abstracts. We also thank the *BMC Bioinformatics* Editorial Office for their help in preparing this supplement.

The 4th ISCB Student Council Symposium was made possible through generous financial support from the European Bioinformatics Institute, Korean Bioinformation Center, Ontario Genomics Institute, Centre for Applied Genomics, Institute for Systems Biology, Iowa State University and *BMC Bioinformatics*. The organizers are also grateful to Oxford University Press for sponsoring best poster and best presentation awards.

References

1. Cho YR, Shi L, Zhang A: **Functional module detection by functional flow pattern mining in protein interaction networks.** *BMC Bioinformatics* 2008, **9(Suppl 10)**:O1.
2. Pechman S, Levy E, Gaetano Tartaglia G, Vendruscolo M: **Competition between protein aggregation and protein complex formation.** *BMC Bioinformatics* 2008, **9(Suppl 10)**:O2.
3. Michaut M, Kerrien S, Montecchi-Palazzi L, Cassier-Chauvat C, Chauvat F, Aude JC, Legrain P, Hermjako H: **InteroPORC: an automated tool to predict highly conserved protein interaction networks.** *BMC Bioinformatics* 2008, **9(Suppl 10)**:P1.
4. Bitton DA, Okoniewski MJ, Connolly Y, Miller CJ: **Exon level integration of proteomics and microarray data.** *BMC Bioinformatics* 2008, **9(Suppl 10)**:P2.
5. Ruau D, Kolarik C, Mevissen HT, Müller E, Assent I, Krieger R, Seidl T, Hofman-Apitius M, Zenke M: **Public microarray repository semantic annotation with ontologies employing text mining and expression profile correlation.** *BMC Bioinformatics* 2008, **9(Suppl 10)**:O5.
6. Jain P, Podila G, Davis M: **Comparative analysis of non-classically secreted proteins in *Botrytis cinerea* and symbiotic fungus *Laccaria bicolor*.** *BMC Bioinformatics* 2008, **9(Suppl 10)**:O3.
7. Bohnert R, Zeller G, Clark RM, Childs KL, Ulat V, Stokowski R, Ballinger D, Frazer K, Cox D, Bruskiwich R, Buell CR, Leach J, Leung H, McNally KL, Weigel D, Ratsch G: **Revealing sequence variation patterns in rice with machine learning methods.** *BMC Bioinformatics* 2008, **9(Suppl 10)**:O8.
8. Abraham AL, Pothier J, Rocha E: **Protein evolution driven by symmetric structural repeats.** *BMC Bioinformatics* 2008, **9(Suppl 10)**:P3.
9. De Bona F, Ossowski S, Schneeberger K, Ratsch G: **Optimal spliced alignments of short sequence reads.** *BMC Bioinformatics* 2008, **9(Suppl 10)**:O7.
10. Rausch T, Emde AK, Reinert K: **Robust consensus computation.** *BMC Bioinformatics* 2008, **9(Suppl 10)**:P4.
11. Saha S, Bridges S, Magbanua Z, Peterson P: **Discovering relationships among dispersed repeats using spatial association rule mining.** *BMC Bioinformatics* 2008, **9(Suppl 10)**:O4.
12. Hyoun Lee P, Shatkay H: **Ranking single nucleotide polymorphisms by potential deleterious effects.** *BMC Bioinformatics* 2008, **9(Suppl 10)**:O6.
13. Chukualim B, Peters N, Hertz Fowler C, Berriman M: **TrypanoCyc – a metabolic pathway database for *Trypanosoma brucei*.** *BMC Bioinformatics* 2008, **9(Suppl 10)**:P5.
14. Yao L, Rzhetsky A: **Quantitative systems-level determinants of drug targets.** *BMC Bioinformatics* 2008, **9(Suppl 10)**:P6.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

