DOE GRANT NO. SC0012396

FINAL REPORT 2014

ISMB Conference Funding for Students and Young Scientists

Historical Description

The Intelligent Systems for Molecular Biology (ISMB) conference has provided a general forum for disseminating the latest developments in bioinformatics on an annual basis for the past 22 years. ISMB is a multidisciplinary conference that brings together scientists from computer science, molecular biology, mathematics and statistics. The goal of the ISMB meeting is to bring together biologists and computational scientists in a focus on actual biological problems, i.e., not simply theoretical calculations. The combined focus on “intelligent systems” and actual biological data makes ISMB a unique and highly important meeting. 21 years of experience in holding the conference has resulted in a consistently well-organized, well attended, and highly respected annual conference.

"Intelligent systems" include any software which goes beyond straightforward, closed-form algorithms or standard database technologies, and encompasses those that view data in a symbolic fashion, learn from examples, consolidate multiple levels of abstraction, or synthesize results to be cognitively tractable to a human, including the development and application of advanced computational methods for biological problems. Relevant computational techniques include, but are not limited to: machine learning, pattern recognition, knowledge representation, databases, combinatorics, stochastic modeling, string and graph algorithms, linguistic methods, robotics, constraint satisfaction, and parallel computation. Biological areas of interest include molecular structure, genomics, molecular sequence analysis, evolution and phylogenetics, molecular interactions, metabolic pathways, regulatory networks, developmental control, and molecular biology generally. Emphasis is placed on the validation of methods using real data sets, on practical applications in the biological sciences, and on development of novel computational techniques.

The ISMB conferences are distinguished from many other conferences in computational biology or artificial intelligence by an insistence that the researchers work with real molecular biology data, not theoretical or toy examples; and from many other biological conferences by providing a forum for technical advances as they occur, which otherwise may be shunned until a firm experimental result is published. The resulting intellectual richness and cross-disciplinary diversity provides an important opportunity for both students and senior researchers. ISMB has become the premier conference series in this field with refereed, published proceedings, establishing an infrastructure to promote the growing body of research.

General Program

The ISMB 2014 meeting was July 11-15, 2014 in Boston, Massachusetts, USA. The science presented was exceptional, and in the course of the three main meeting days, 36 original scientific papers, 48 highlights from recently published papers, 18 Late Breaking Research papers, 21 Oral Poster Presentations, 527 posters, 38 technology track presentations, and 6 special sessions were presented. Additionally, 2 satellite meetings, 11 special interest group meetings, a Student Council symposium and 2 tutorials filled the two days preceding the main meeting.

ISMB 2014 attracted 1617 total participants from 43 countries, including 1307 attendees for the main meeting. There were 22 organizations exhibiting and 3 Silver sponsors and 6 general sponsors. Four (4) media participants attended and reported on the conference.

In addition to the main conference, special interest group pre-meetings attracted 963 participants with 555, (57%) attending the main meeting, 2 satellite pre-conference meetings attracted 147 participants, 94 (64%) attending the main meeting. The one-day Student Council symposium attracted 64 participants (39 or 78% of whom stayed for the main conference). The meeting began with an ice-breaker orienteering event with 60 registrants.

The ISMB conference is especially important because the cultures of computer science and biology are so disparate. ISMB, as a full-scale technical conference with refereed proceedings that have been indexed by both MEDLINE and Current Contents since 1996, bridges this cultural gap. Since 2001 Oxford University Press has published the *ISMB Proceedings* in *Bioinformatics*. The 2014 proceedings are now available online as a fully open access issue, freely available worldwide.

At the main meeting each session was introduced by a keynote address (see below). The keynote speakers are especially selected to emphasize new areas related to the conference, biological problems ripe for investigation by computational means, and exciting research areas that are less well known to the ISMB audience. The keynote addresses, therefore, always play a key role in the continuing education of all attendees, and most especially the younger scientists who are often participating and gaining exposure to the international bioinformatics community and its leaders for the first time.

Attendee demographics

The participants at ISMB 2014 represented a broad spectrum of backgrounds, with 43 countries represented at the conference, thus emphasizing its important international aspect. Participation based on geographical origin of attendees (regardless of citizenship) included 60% from the United States, 9% from Asia, and 8% from Europe with the remainder well spread from other parts of the world. Among all attendees, 56% were from academia/non-profit/government sectors (combination of researchers and faculty), 31% were student registrants and 12% were from the commercial sector. In all, 72% of the conference participants were male, 25% were female, with 3% delegates declining to state their gender. The percentage of female participants seems to be consistent with reported averages of women working and studying within the interdisciplinary sciences that make up the fields of bioinformatics and computational biology.

Previous and Future Meetings

The ISMB conference is the annual meeting of the International Society for Computational Biology (ISCB). The core mission of the ISCB is dedicated to advancing the scientific understanding of living systems through computation. The Society was formally incorporated in 1997 and has been granted nonprofit tax-exempt status in the United States as a 501(c) (3) corporation. The Society now has an annual budget of approximately two million dollars, with the majority of that allocated to producing the ISMB conference. At the time of ISMB 2014 there were nearly 3,100 ISCB members from 74 countries including more than 1,500 student and postdoctoral trainee members.

The ISMB conference series was established in 1993, and after its fifth year gave rise to the formation of the professional society to support and coordinate the conference. The conference has been held annually and grown in numbers of participants from 201 in 1993 to over 2,100 in 2004, when, for the first time, it was held as a joint meeting with the European Conference on Computational Biology. The conference purposefully alternates its venue between North American and European sites, occasionally adding non-North American/European sites to foster international exchange and cooperation. Attendance in years outside the North American/European rotation has been significantly lower due to the increased travel time and costs for many participants. The locations of previous and future ISMB conferences were/are as follows:

Past and Future Performance Sites and Attendance Figures:

1995: Cambridge, UK – 270 attendees	2006: Fortaleza, Braila – 868 attendees
1996: St. Louis, USA – 279 attendees	2007: Vienna, Austria – 1809 attendees
1997: Halkidiki, Greece – 254 attendees	2008: Toronto, Canada – 1395 attendees
1998: Montreal, Canada – 413 attendees	2009: Stockholm, Sweden – 1359 attendees
1999: Heidelberg, Germany – 580 attendees	2010: Boston, USA – 1643 attendees
2000: San Diego, USA – 1245 attendees	2011: Vienna, Austria – 1906 attendees
2001: Copenhagen, Denmark – 1251 attendees	2012: Long Beach, USA – 1467 attendees
2002: Edmonton, Canada – 1585 attendees	2013: Berlin, Germany – 1528 attendees
2003: Brisbane, Australia – 927 attendees	2014: Boston, USA - 1617 attendees
2004: Glasgow, Scotland – 2136 attendees	2015: Dublin, Ireland- 1403 attendees
2005: Detroit, USA – 1728 attendees	2016: Orlando, USA - 1525 anticipated

*indicates anticipated attendance for future years based on regional trends of past conferences.

Specific Scientific Program

Keynote Addresses 2014

A total of six keynote addresses were delivered during the three main conference days of ISMB 2014. One lecture started the morning program, and a second lecture was delivered each afternoon of the conference. The keynote speakers and the titles of their talks were as follows:

KEYNOTE SPEAKER	TITLE
Michal Linial, PhD Hebrew University of Jerusalem Israel	Good Things Come in Small Packages – Replicators and Innovators

Eugene Myers Max Planck Institute of Molecular Cell Biology and Genetics Germany	DNA Assembly: Past, Present, and Future
Isaac Kohane MD, PhD Harvard Medical School and Children's Hospital Boston United States	Biomedical Quants of the World Unite! We only have our disease burden to lose
Dana Pe'er Columbia University United States	A multidimensional single cell approach TO understand cellular behavior
Dr. Robert Langer Massachusetts Institute of Technology United Kingdom	Biomaterials and biotechnology: From the discovery of the first angiogenesis inhibitors to the development of controlled drug delivery systems and the foundation of tissue engineering
Russ B. Altman MD, PhD Stanford University United States	Informatics for understanding drug response at all scales.

Scientific Papers 2014

Of the 2014 conference proceedings that were submitted; the following 36 were selected for presentation and for publication in *Bioinformatics* volume 30 issue 12, as a fully open-access, online issue of the journal (http://bioinformatics.oxfordjournals.org/content/30/12.toc).

PRESENTING AUTHOR	**TITLE**
Cristina G. Ghiurcuta	Evaluating synteny for improved comparative studies
Ya'ara Arkin	EPIQ—efficient detection of SNP–SNP epistatic interactions for quantitative traits
Nelle Varoquaux	A statistical approach for inferring the 3D structure of the genome
Michael Kramer	Inferring gene ontologies from pairwise similarity data
Terumasa Tokunaga	Automated detection and tracking of many cells by using 4D live-cell imaging data
Yu-Keng Shih	A single source k-shortest paths algorithm to infer regulatory pathways in a gene network
Anika Oellrich	Using association rule mining to determine promising secondary phenotyping hypotheses
Nagarajan Natarajan	Inductive matrix completion for predicting gene–disease associations
Kourosh Zarringhalam	Robust clinical outcome prediction based on Bayesian analysis of transcriptional profiles and prior causal networks
Iman Hajirasouliha	A combinatorial approach for analyzing intra-tumor heterogeneity from high-throughput sequencing data
Ran Libeskind-Hadas	Pareto-optimal phylogenetic tree reconciliation
Andrei Todor	Large scale analysis of signal reachability

Christoph Bernau	R Cross-study validation for the assessment of prediction algorithms
Tarmo Äijö	Methods for time series analysis of RNA-seq data with application to human Th17 cell differentiation
Michael K. K. Leung	Deep learning of the tissue-regulated splicing code
Matthew E. Studham	Functional association networks as priors for gene regulatory network inference
Wei Cheng	Graph-regularized dual Lasso for robust eQTL mapping
Anke Penzlin	Pipasic: similarity and expression correction for strain-level identification and quantification in metaproteomics
Huibin Shen	Metabolite identification through multiple kernel learning on fragmentation trees
Masaaki Kotera	Metabolome-scale prediction of intermediate compounds in multistep metabolic pathways with a recursive supervised approach
A. Ercument Cicek	MIRA: mutual information-based reporter algorithm for metabolic networks
Richard Leslie	GRASP: analysis of genotype–phenotype results from 1390 genome-wide association studies and corresponding open access database
Hsin-Ta Wu	Detecting independent and recurrent copy number aberrations using interval graphs
Farhad Hormozdiari	Privacy preserving protocol for detecting genetic relatives using rare variants
Ladislav Rampášek	Probabilistic method for detecting copy number variation in a fetal genome using maternal plasma sequencing
Mengfei Cao	New directions for diffusion-based network prediction of protein function: incorporating pathways with confidence
Lei Huang	DrugComboRanker: drug combination discovery based on target network analysis
C Marc Hulsman	Scale-space measures for graph topology link protein network architecture to function
Marinka Žitnik	Gene network inference by probabilistic scoring of relationships from a factorized model of interactions
Yichao Zhou	An efficient parallel algorithm for accelerating computational protein design
Jing Ren	Tertiary structure-based prediction of conformational B-cell epitopes through B factors
Pavankumar Videm	BlockClust: efficient clustering and classification of non-coding RNAs from short read RNA-seq profiles
Zhaojun Zhang	RNA-Skim: a rapid method for RNA-Seq quantification at transcript level
Andrey D. Prjibelski	ExSPAnder: a universal repeat resolver for DNA fragment assembly
Mikhail Kolmogorov	Ragout—a reference-assisted assembly tool for bacterial genomes
Alastair M. Kilpatrick	Stochastic EM-based TFBS motif discovery with MITSU

Ergude Bao	AlignGraph: algorithm for secondary de novo genome assembly guided by closely related references
Serghei Mangul	Accurate viral population assembly from ultra-deep sequencing data

Tutorials 2014

Tutorials are offered as additional learning opportunities the day before the main conference begins to enable participants to come quickly up to speed on specific core and emerging areas of the science. Two tutorials were presented in 2014:

Computational Metagenomics Techniques and Challenges

Curtis Huttenhower, Harvard School of Public Health, Boston, United States
Titus Brown, Michigan State University, East Lansing, United States

Microbial ecology is one of many fields that have benefitted greatly from technical advances in DNA sequencing. In particular, low-cost culture-independent sequencing has made metagenomic and metatranscriptomic surveys of microbial communities practical, including bacteria, archaea, viruses, and fungi associated with the human body, other hosts, and the environment. The resulting data have stimulated the development of many new computational approaches to meta'omic sequence analysis, including metagenomic assembly, microbial identification, and gene, transcript, and pathway metabolic profiling. This tutorial will present a high-level introduction to computational metagenomics, highlighting the state-of-the-art in the field as well as outstanding challenges. The tutorial will include standardized protocols for microbial profiling, functional profiling, and metagenome/metatranscriptome assembly with benchmarks and examples.

Level: General to intermediate

Wikipedia: WikiProject Computational Biology/ISMB 2014

Daniel Mietchen, Museum für Naturkunde Berlin, Germany
Ben Moore, University of Edinburgh, Scotland

In this tutorial, we will provide a practical introduction to editing Wikipedia for scientists. By the end of the course, participants will understand the basics of editing and have made actual contributions to Wikipedia articles. Each participant will be required to bring some pre-prepared written material, images or multimedia files to add to an article of their choice in any language version of Wikipedia. In addition, we will teach the participants about the etiquette of being a Wikipedia editor as well as helping them to avoid some of the pitfalls that new editors sometimes come across. We will also introduce the PLOS Computational Biology Topic Pages destined for the English Wikipedia, inform about the ISCB Wikipedia Competition and provide an overview of some of the sister platforms of Wikipedia, notably Wikidata and Wikimedia Commons, and their coverage of Computational Biology.

Level: General

Special Interest Group Meetings

Over the years other related, self-organized events have become a key component of the ISMB meeting series: In 2014 there were eleven special interest group (SIG) meetings and two Satellite Meetings organized by researchers, who used the presence of a considerable fraction of the scientific community to spawn off more specialized meetings. More information about the SIG meetings and Satellite Meetings can be found at http://www.iscb.org/ismb2014-program/ismb2014-sigs-satellite-meetings.

Satellite Meetings

3Dsig: Structural Bioinformatics & Computational Biophysics
CAMDA 2014 Critical Assessment of Massive Data Analysis

SIGS

Automated Function Prediction (AFP-SIG)
Bio-Ontologies
BioVis SIG
BOSC: The 15th Annual Bioinformatics Open Source Conference
HitSEQ: High Throughput Sequencing Algorithms & Applications
Integrative RNA Biology SIG
Mass Spectrometry SIG (MS-SIG)
NetBio SIG
New Challenges in Computational Single Cell Biology
Regulatory Genomics Special Interest Group (RegGenSIG)
Vari SIG (formerly SNP-SIG)

Student Travel Fellowship Support

Student travel support funds were secured for this conference from DOE, NIH, and NSF. ISCB allocated funding support based on applications received by the respective individuals. Preference and amount of reimbursement was determined based on the following factors:

- Status as co-author of a paper submitted or accepted. (A poster or paper presentation is required of funding student recipients);
- Quality of abstract received;
- Status as woman, minority, or disabled scientist;
- Degree of financial hardship;
- Status as junior scientist (student or post-graduate researcher within 5 years of degree);
- Amount of reimbursement required;
- Geographic distribution of awards (emphasis on greater distance/cost of travel).

An ISMB 2014 Travel Fellowships Committee was formed to review applications and determine awardees and award amounts, ranking applicants based on the factors noted above. The committee was chaired by Guilherme Oliveira (FIOCRUZ, Brazil), and made up of the following members worldwide: Nicola Mulder (University of Cape Town, South Africa), Gustavo Parsi (Universidad Nacional de Quilmes, Buenos Aires, Argentina), Todd Taylor (RIKEN Quantitative Biology Center, Japan), Karyn Megy (EBI, UK), Ricardo Vencio (University of Sao Paulo, Brazil), Susana Vinga (INESC-ID LISBOA, Portugal), Ricardo Vencio (University of Sao Paulo, Brazil), Shibu Yooseph (J. Craig Venter Institute, USA).

A total of 200 eligible travel fellowship applications from students and post docs presenting a paper or poster were received. The combined funds from grants from DOE, NIH and NSF enabled a total of 84 (39%) student/post doc travel awards to be made, including 28 (28%) awards to women. These statistics show a higher percentage of female fellowship applicants and fellowship awards to women as compared to the general female population of conference attendees (25%). The DOE funds specifically provided the funding of 16 students and post-doctoral researchers, of which 25% were women.

DOE Award Funds Spent by Grantee

Student Participant Support Travel Costs for ISMB 2014, Boston, MA, USA

A total of 16 student travel support awards were made at an average of $935.00 per award.

First Name	Last Name	Funding	Title	Presentation Type
Ferhat	Ay	$500	Fit-Hi-C: Statistical confidence estimation for Hi-C data	highlights
Sara	Ballouz	$900	Guidance for RNA-seq co-expression network construction and analysis: safety in numbers	poster
Mengfei	Cao	$740	Directions for Diffusion-Based Network Prediction of Protein Function: Incorporating Pathways with Confidence	proceedings
Wei	Cheng	$1,100	Graph Regularized Dual Lasso for Robust eQTL Mapping	proceedings
A. Ercument	Cicek	$900	MIRA: Mutual Information-Based Reporter Algorithm for Metabolic Networks	proceedings
Hatice Billur	Engin	$1,925	Integrating Structure to Protein-Protein Interaction Networks that Drive Metastasis to Brain and Lung in Breast Cancer	highlights
Adenike	Fujah-Yusuf	$600	In Silico Study of Salt Tolerance in Wheat and Related Species	Highlights

Farhad	Hormozdiari	$2,200	Privacy Preserving Protocol for Detecting Genetic Relatives Using Rare Variants	Proceedings
Lei	Huang	$850	DrugComboRanker: Drug Combination Discovery Based on Target Network Analysis	Proceedings
Andre	Kahles	$600	SplAdder: Integrated Quantification, Visualization and Differential Analysis of Alternative Splicing	Poster
Noam	Kaplan	$800	From 1D to 3D and back: Genome scaffolding from DNA interaction frequency	Highlights
Guray	Kuzu	$750	Modeling protein assemblies in the proteome	Highlights
Nagarajan	Natarajan	$1,000	Inductive Matrix Completion for Predicting Gene-Disease Associations	Proceedings
Matthew	Studham	$1,200	Functional Association Networks as Priors for Gene Regulatory Network Inference	Proceedings
Haimin	Tang	$1,200	N/A	highlights
Rathi	Thiagarajan	$1,100	Transcriptional regulatory networks of single cells during in vitro hepatic differentiation of human pluripotent stem cells	Poster

TOTAL spent on Student and Keynote Travel Support using DOE funds: $15,000.00

No DOE funds were used for travel of conference organizers. No DOE funds were used for researchers of US government agencies. No DOE funds were used for the presentation of the conference itself. The grant did not fund the PI or the PI's research in any way.