11th Annual Bioinformatics Open Source Conference

BOSC 2010

1. Welcome

Welcome to BOSC 2010! This is the 11th annual Bioinformatics Open Source Conference held as a Special Interest Group (SIG) meeting in conjunction with the Intelligent Systems for Molecular Biology Conference.

BOSC is sponsored by the Open Bioinformatics Foundation (O|B|F), a non-profit group dedicated to promoting the practice and philosophy of Open Source software development within the biological research community. Many open source bioinformatics packages are widely used by the research community across many application areas and form a cornerstone in enabling research in the genomic and post-genomic era. Open source bioinformatics software has facilitated rapid innovation, dissemination, and wide adoption of new computational methods, reusable software components, and standards.

A major theme of BOSC this year is using cloud-based approaches to improving software and data accessibility. The emergence of cloud computing has made highly scalable cluster computing available to computational biologists. Services such as the Amazon Elastic Compute Cloud combined with publicly available datasets promise to lower the overhead to participate in large scale data analyses. We have a number of talks about how the community can build up resources and datasets for cloud infrastructure, share insights, and contribute implemented workflows. The pre-conference Codefest session had a focus on cloud computing as well. We also feature talks that discuss approaches to analyzing high-throughput 'omics data, including projects that use the MapReduce framework for parallelized analysis of the possibly terabyte size data sets from next-generation sequencing and projects that parallelize bioinformatics algorithms in general. We also have a series of talks involving the semantic web and data visualization tools. Finally, we will hear updates from the Bio* projects hosted by the O|B|F.

One of the hallmarks of BOSC is the coming together of the open source developer community in one location to meet face-to-face. This creates synergy where participants can work together to create use cases, prototype working code, or run bootcamps for developers from other projects as short, informal, and hands-on tutorials in new software packages and emerging technologies. In short, BOSC is not just a conference for presentations of completed work, but is a dynamic meeting where collaborative work gets done and attendees can learn about new or on-going developments that they can directly apply to their own work. We have two formats to facilitate these collaborations: lightning talks (short, 5 minute talks for focused ideas and demos) and Birds of a Feather sessions where small groups gather around a common interest.

BOSC is made possible by a community effort. We thank the abstract reviewers and session chairs. We invite you to join the O|B|F (application form enclosed) and the BOSC 2011 planning committee (e-mail bosc@open-bio.org). Finally, the keynote speakers and student travel awards to BOSC 2010 were sponsored by Eagle Genomics, Ltd. and an anonymous donor. Congratulations to the student winners: Nobuaki Kono, Eric Talevich, and Jens Lichtenberg!

Organizing Committee
Kam Dahlquist (Chair), Nomi Harris, Brad Chapman, Michael Heuer, Darin London, Steffen Moeller, Jim Procter, and Ron Taylor, Chris Dagdigian, Hilmar Lapp, and Jason Stajich

2. Schedule for Friday, July 9 (Day 1)*

9:00-9:15	Kam Dahlquist	[Welcome] Introduction to the O|B|F and Welcome to BOSC 2010
9:15-10:15	Guy Coates	[Keynote] Clouds: all fluff and no substance?
10:15-10:45	**Coffee Break**	
10:45-11:05	Ronald Taylor	[Cloud Computing/HT 'omics Data] An overview of the Hadoop/MapReduce/HBase framework and its current applications in bioinformatics
11:05-11:25	M. Hanna	[Cloud Computing/HT 'omics Data] The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data
11:25-11:45	Brian D. O'Connor	[Cloud Computing/HT 'omics Data] SeqWare Query Engine: Storing and Searching Sequence Data in the Cloud
11:45-12:00	Judy Qiu	[Cloud Computing/HT 'omics Data] Hybrid Cloud and Cluster Computing Paradigms for Life Science Applications
12:00-12:15	Ben Langmead	[Cloud Computing/HT 'omics Data] Cloud-scale genomics: examples and lessons
12:15-12:30	Enis Afgan	[Cloud Computing/HT 'omics Data] Deploying Galaxy on the Cloud
12:30-2:00	**Lunch**	
2:00-2:20	Steffen Möller	[Cloud Computing] Community-driven computational biology with Debian and Taverna
2:20-2:40	Darin London	[Cloud Computing/HT 'omics Data] Dealing with the Data Deluge: What can the Robotics Community Teach us?
2:40-3:00	Nyasha Chambwe	[HT 'omics Data] The Goby framework: towards efficient next-generation sequencing data analysis
3:00-3:15	Dana Robinson	[HT 'omics Data] BioHDF: Open binary file formats for NGS data - current status and future directions
3:15-3:30	Jens Lichtenberg	[HT 'omics Data] Concurrent Bioinformatics Software for Discovering Genome-wide Patterns and Word-based Genomic Signatures
3:30-3:38	Chris Hemmerich	[HT 'omics Data] Automated Annotation of NGS Transcriptome Data using ISGA and Ergatis
3:38-4:00	**Coffee Break**	
4:00-4:20	Mark Wilkinson	[Semantic Web] From Moby to SADI - Modeling Semantic Web Services with the Semantic Automated Discovery and Integration Framework
4:20-4:35	Aravind Venkatesan	[Semantic Web] ONTO-ToolKit: enabling bio-ontology engineering via Galaxy
4:35-4:45	Kazuharu Arakawa	[Semantic Web] G-language Bookmarklet: a gateway for Semantic Web, Linked Data, and Web Services
4:45-4:55	Christian M Zmasek	[Semantic Web] Connecting TOPSAN to Computational Analysis
4:55-5:10	Jianjiong Gao	[OS Software] Musite: Global Prediction of General and Kinase-Specific Phosphorylation Sites
5:10-6:00	Lightning Talks and Birds of a Feather	
6:30	O|B|F Board Meeting and No-host Dinner (location TBA)	

* Schedule updates are posted at http://www.open-bio.org/wiki/BOSC_2010_Schedule

3. Schedule for Saturday, July 10 (Day 2)*

9:00-9:15	Kam Dahlquist	[Announcements]
9:15-10:15	Ross Gardler	[Keynote] Community Development at the Apache Software Foundation
10:15-10:45	**Coffee Break**	
10:45-11:05	Christopher Fields	[Bio* Update] Towards a Modern BioPerl: BioPerl Update 2010
11:05-11:20	Raoul J.P. Bonnal	[Bio* Update] BioRuby 2010 updates: moving to agile bioinformatics
11:20-11:35	Peter Rice	[Bio* Update] EMBOSS: The European Molecular Biology Open Software Suite
11:35-11:50	Brad Chapman	[Bio* Update] Biopython Project Update
11:50-12:05	Tiago Antao	[Bio* Update] interPopula: Database and tool integration for population genetics using Python
12:05-12:20	Laurent Gautier	[Bio* Update] Bioconductor with Python, What else ?
12:20-12:30	Eric W Talevich	[Bio* Update] Bio.Phylo: A unified phylogenetics toolkit for Biopython
12:30-2:00	**Lunch**	
2:00-2:20	Simon Mercer	[OS Software Framework] The Microsoft Biology Foundation
2:20-2:40	William Crawford	[OS Software/Code Generation] Building Bioinformatics Web Application with Clickframes
2:40-3:00	Morris A. Swertz	[OS Software/Code Generation] MOLGENIS: rapid prototyping of biosoftware at the push of a button
3:00-3:10	Alexandros Kanterakis	[OS Software/Code Generation] Generating a data platform for microarray gene expression experiments using MOLGENIS and MAGE-TAB
3:10-3:25	Sebastian J. Schultheiss	[Web Services] Long-term availability of bioinformatics web services
3:25-4:00	**Coffee Break**	
4:00-4:15	Lincoln Stein	[Visualization] GBrowse2
4:15-4:35	Gary D. Bader	[Visualization] Cytoscape Web: An interactive, customizable web-based network browser
4:35-4:50	Nobuaki Kono	[Visualization] Pathway Projector: Web-Based Zoomable Pathway Browser Using KEGG Atlas and Google Maps API
4:50-5:05	James A. Morris	[Visualization] Evoker: a visualization tool for genotype intensity data
5:05-5:20	Pavel Tomancak	[Visualization] Fiji Is Just ImageJ - an Open Source platform for biological image analysis
5:20-5:30	Iddo Friedberg	[Visualization] IPRStats: visualization and analysis of InterProScan Results
5:30-6:00	Birds of a Feather	

* Schedule updates are posted at http://www.open-bio.org/wiki/BOSC_2010_Schedule

4. Keynote Speakers

Guy Coates

Guy Coates, Group leader of the Informatics Systems Group at the Wellcome Trust Sanger Institute, is an expert in large Linux/Unix systems architecture, management and application optimization, next-gen DNA sequencing pipelines, very large database systems (>50TB), and high performance compute clusters (1500+ CPUs). He will lead off the conference, with a talk entitled: *Clouds: all fluff and no substance?* Cloud architectures have made plenty of impact in the press over the past 18 months, but have they made any impact in solving real bioinformatics challenges? This talk will detail the Sanger Institute's experiences with the cloud, both good and bad, and ask what needs to be done to prepare the cloud for the era of commodity sequencing and $500 genomes.

Ross Gardler

Ross Gardler is Vice President for Community Development at the Apache Software Foundation and Manager for OSS Watch (University of Oxford), which helps institutions and projects in the UK that are using or developing free and open source software. He has also performed Computer Science and Management research and teaching in both the UK and West Indies. In his talk, he will share with the O|B|F community the insights he has gained in community building for Apache and other open source projects.

5. O|B|F Membership Application

The O|B|F is actively inviting professionals, scientists, students, and others active in the Open Source Software arena in the life sciences to join the O|B|F as a member. The membership body was formally established at the 2005 Board of Directors meeting. As laid out in the bylaws, officers in the Board of Directors will be elected by the membership among nominees, and candidates for future Directors will be nominated from the membership when seats are added or a term expires.

The eligibility criteria are met by anyone who is "interested in the objectives of the OBF", and there are no dues at present. You can join the O|B|F at BOSC by filling out the application form on the following page, signing it, and giving it to a Board member. You may also e-mail the scanned form to the current Parliamentarian, Hilmar Lapp, at hlapp@gmx.net.

The Board established the membership body as the platform from which major future changes to the O|B|F's mission, agenda, and scope will originate, through active participation, discussion, and possibly most importantly, through providing the future leadership of the O|B|F. If you have ideas and the energy to advocate them then we encourage you to join as a member if you haven't already, and to step up and nominate yourself for future elections to the Board. The bylaws also define terms for Directors as well as Officers, so seats will come up for election every year.

If you wish to find out more, please come to the O|B|F Board Meeting and No-host Dinner (location TBA) in the evening of Day 1 of the conference. All are welcome to attend.

6. Abstracts

Abstracts appear in the following pages in the order that they are presented at the conference.

O|B|F – Open Bioinformatics Foundation

Membership Application

I wish to apply for membership in the Open Bioinformatics Foundation (O|B|F).

First and Last Name: _____

Street Address: _____

City, State, Zip Code: _____

Country of Residence: _____

Email Address: _____

All fields are mandatory. The O|B|F will treat all personal information strictly confidential and will not share personal information with anyone except members of the O|B|F Board of Directors, or entities or persons appointed by the Board to administer membership communication. This may be subject to change, please see below.

I am an attendee of BOSC 200__: ☐ Yes ☐ No

If you answered No, please state why you meet the membership eligibility requirement of being interested in the objectives of the O|B|F:

_____

_____
(use back of page if you need more space)

I understand that membership rights and duties are laid down in the O|B|F Bylaws which may be downloaded from the O|B|F homepage at http://www.open-bio.org/. I understand that if the O|B|F's privacy statement changes I will be notified at my email address (as known to O|B|F), and if I do not express disagreement with the proposed change(s) by terminating my membership within 10 days of receipt of the notification, I consent to the change(s).

_____
(Signature)