



Vol. 1 No. 1
January 2008

Bioinformatics Organization

Big plans for '08

A message from the president

In this issue



Project spotlight

.....Page 2



Book review: Bioinformatics

.....Page 2



Franklin Award nominees

.....Page 3



Job search

.....Page 3



Upcoming events

.....Page 4

Copyright (c) 2008
Bioinformatics Org., Inc.

Happy New Year, and welcome to the first issue of the Bioinformatics Organization newsletter. In this monthly publication, we will include the best of our various online forums and detail some of our internal (and external) activities. From time-to-time we'll even include a featured article by a guest contributor.

To begin with, these newsletters will be short, since we don't have a staff of writers. But, as the saying goes, the best things in life come in small packages!

Last year was the best on record for the Organization. We started offering online courses, featuring Shailender Nagpal from Tanisha Systems, Prashanth Survajhala from Roskilde University, and Warren Lathe from OpenHelix. In all, more than 100 people attended our courses. (For more information on our educational program, please visit <http://edu.bioinformatics.org/>.)

We also received a sponsorship from the fine folks at eXludus Technologies (<http://www.exludus.com/bioinformatics.php>), helping us to continue the development of our flagship software application, Pipet, which has an amalgamation of features found in wikis, Web forms, and visual programming languages. And it will soon run the Bioinformatics.Org website. (More information on Pipet will be provided in the next issue of this newsletter.)

In 2008, we're also offering the long-planned Professional Membership (<http://wiki.bioinformatics.org/Membership>). This membership has all of the benefits of the free membership plus a 20% discount on educational services provided by the Bioinformatics Organization, 10-20% discounts on

conferences and materials sponsored by the Organization, printed (mailed via post!) issues of this monthly newsletter, an IMAP4 and POP3 email account using the bioinformatics.org domain name, and a Linux shell account with some storage space and a Web folder.

Our educational program will also expand this year. We're actively seeking additional instructors to help increase the number of courses offered. Also, we're putting together an "educational conference" for later in the year. Unlike typical conferences, all of our sessions will provide hands-on training and use open-source software for bioinformatics. (More information on this conference will come in the following issues.) If you are interested in becoming an instructor at Bioinformatics.Org, please contact us at edu@bioinformatics.org.

Our online Jobs Forum (<http://bifx.org/jobs>) reached new levels of popularity in 2007, making it to the top of the list at Google. And, in 2008 we plan to expand and improve the site, forming the Bioinformatics Career Center. Planned features include integrated *curriculum vitae* posts and search features for employers.

One more thing: 2008 marks our 10th anniversary! It will be a great year. Stay tuned!

-- Jeff



J.W. Bizzaro

Project spotlight

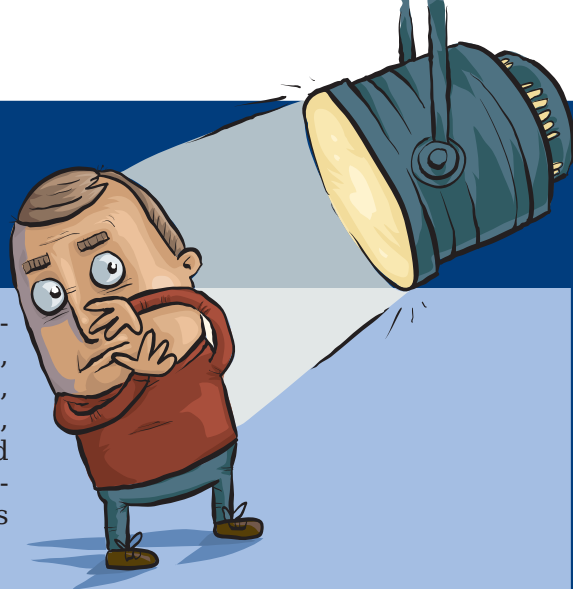
GeneZilla

GeneZilla is a state-of-the-art program for computational prediction of protein-coding genes in eukaryotic DNA, and is based on the Generalized Hidden Markov Model (GHMM) framework, similar to GENSCAN and GENIE. It is highly reconfigurable and includes software for re-training by the end-user. It is written in highly optimized C++ and runs under most UNIX/Linux platforms. The run time and memory requirements are linear in the sequence length, and are in general much better than those of competing systems, due to GeneZilla's novel decoding algorithm. Graph-theoretic representations of the high scoring open reading frames are provided, allowing for exploration of sub-optimal gene models. It utilizes Interpolated Markov Models (IMMs), Maximal Dependence Decomposition (MDD), and includes

states for signal peptides, branch points, TATA boxes, CAP sites, and will soon model CpG islands as well.

GeneZilla is an open-source project hosted at Bioinformatics.Org and currently consists of ~20,000 lines of code. GeneZilla evolved out of the *ab initio* eukaryotic gene finder TIGRscan, which was developed at The Institute for Genomic Research over a 3-year period under NIH grants R01-LM06845 and R01-LM007938, and which served as the basis for the comparative gene finder TWAIN.

<http://www.genezilla.org/>



Book review

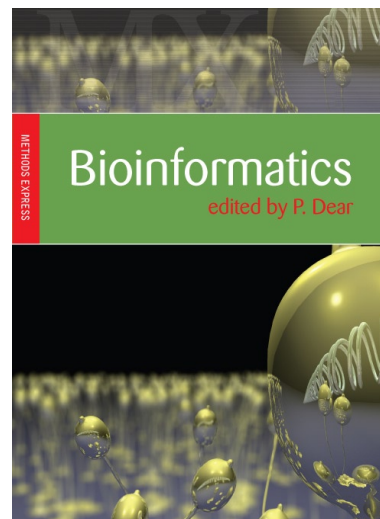
Bioinformatics: Methods Express

Reviewed by Peter St. Onge

The progression of bioinformatics over the last twenty years, from initial programs and simulations to the rich set of tools now available, is mirrored in the rapid growth in specialized fields (genomics, proteomics, metabolomics, with seemingly more appearing daily) and the drive to sequence more and more species every year.

With most of the development in bioinformatics revolving around software, a majority of texts have thus far focused on software programming and development in various languages. While large, genome-level projects require this approach, few outside such large projects would invest such time. This book stands out from many other bioinformatics texts as it emphasizes the opposite - using existing software tools without requiring programming. The book is not aimed at the programmer, but rather at the non-programmer - the researcher wanting to focus on particular biological questions relevant to their research.

The book's chapters deal with sequenced genomes, sequence similarity searches, gene prediction, noncoding transcripts prediction, finding regulatory elements in DNA sequence, ESTs, protein structure, gene ontology, protein function prediction, multiple sequence alignment, and inferring phylogenetic relationships from sequence data. Each of the chapters stands alone well, with a section providing abbreviations and other useful context. Along with lists of additional online resources, the copious and detailed references at the end of each chapter provide the reader ample opportunity to explore the subject area further. The book also has a companion web site with data files and other useful information to supplement the book.



Bioinformatics: Methods Express
Edited by Paul H. Dear
ISBN: 978-1-904842-16-3

Continued on page 3

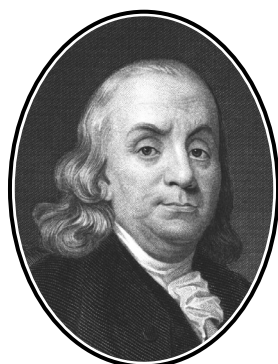
This book provides an excellent overview and synthesis of the best and most accessible bioinformatics resources currently available for the biologist, and provides an excellent introduction to the field in general.

Peter D. St. Onge
BIOTA Group
Depts of Cell and Systems Biology &
Ecology and Evolutionary Biology
University of Toronto



Peter St. Onge

<http://www.scionpublishing.com/bioinformaticsoffer>



Benjamin Franklin Award

Here are the candidates for this year's Award...

Philip E. Bourne, Protein Data Bank Co-Director, University of California San Diego

James L. Edwards, Encyclopedia of Life, Smithsonian Institution

Robert Gentleman, Fred Hutchinson Cancer Research Center

Michael Hucka, California Institute of Technology

Francis Ouellette, Ontario Institute for Cancer Research

Steven Salzberg, University of Maryland, College Park

Go here to learn more about them:
http://bifx.org/forums/forum.php?forum_id=6008

To vote, go to: <http://bifx.org/franklin/vote/2008/>

Job search

Looking for a job in the bioinformatics field?



The Broad Institute of MIT and Harvard seek an ASSOCIATE COMPUTATIONAL BIOLOGIST, under the Broad Institute-Genome Sequencing and Analysis Program, to support program staff in performing computational analysis of high-throughput epigenomic data. Will develop and support tools, maintain data sets and databases,

integrate tools and databases into existing high-throughput pipelines, and facilitate the display of analysis results; and run bioinformatics jobs to support analysis and contribute to a dynamic scientific community.

The research will concentrate on--but is not limited to--applying sequence analysis and utilizing statistical methodologies to questions in stem cell biology, epigenetic components of complex disease and human genome annotation, and in support of next-generation sequencing technology development. Duties will include performing timely and accurate bioinformatic analysis to aid researchers in performing computational analysis of epigenomic data; working with researchers to define and implement analysis tools and algorithms; maintaining, supporting, and documenting shared tools, code base, and data sets; preparing analytical results for presentation in various contexts including group meetings, funding agency reports, and journal publications; interacting with and facilitating data transfer to external collaborators and data repositories; and other tasks as requested.

Interested applicants may apply online at <http://www.broad.mit.edu/info/careers/> and reference job number MIT-00004984. Please indicate that you located this position on Bioinformatics.Org.

Full announcement:

http://bifx.org/forums/forum.php?forum_id=6029

Upcoming events



R for Biologists, Level 1; Monday, 28 January » Friday, 1 February

- Taught in the context of biological research, this course helps biologists learn how to use the statistical scripting language R for data analysis. Unlike Bioinformatics Organization course MA101A (Biostatistics with R, Part 1), this course focuses on the syntax and functionality of the language R and may be used in preparation for MA101A.

http://wiki.bioinformatics.org/CS101B_R_for_Biologists,_Level_1

Perl for Biologists, Level 2; Monday, 4 February » Friday, 8 February

- Taught in the context of biological research, this course shows biologists how to use the scripting language Perl to automate certain tasks. It is a continuation of CS101A Perl for Biologists, Level 1 and covers advanced topics and projects.

http://wiki.bioinformatics.org/CS102A_Pperl_for_Biologists,_Level_2

R for Biologists, Level 2; Monday, 11 February » Friday, 15 February

- Taught in the context of biological research, this course helps biologists understand the data analysis and visualization language R. It is a continuation of CS101B R for Biologists, Level 1 and covers advanced topics and projects.

http://wiki.bioinformatics.org/CS102B_R_for_Biologists,_Level_2

Bio-IT World Conference and Expo; Boston, MA; Monday, 28 April » Wednesday, 30 April

- Location of the Bioinformatics.Org Annual Meeting and presentation of the Benjamin Franklin Award. This premier event demonstrates how technology is transforming the life sciences throughout all discovery and development processes, including over 100 speakers. In addition to affording you access to these high quality talks this event will include an exhibit hall comprised of over 50 booths providing another great way to maximize your participation.

<http://www.bio-itworldexpo.com/>

Systems Biology in present day Bioinformatics; Orlando, FL; Sunday, June 29 » Wednesday, July 2

- CITSA 2008 is an International Multi-Conference being organized with the purpose of providing researchers, practitioners, developers, consultants, and end-users of computerized, communications and/or control systems and technologies, as well as their industrial and social applications in the private and the public sectors, an opportunity to join in a common place sharing experience and knowledge. It is intended to be a forum to expose and share current and future research work and innovations in these areas, as well as in the relationships among them.

<http://www.infocybereng.org/imeti2008/website/>



**Bioinformatics
Organization**

17 Pope Street
Hudson, MA 01749
USA