Abstract

Genome sequencing methods that produce sequences shorter than 50 nucleotides (nt) present significant challenges to genome assembly and sequence alignment algorithms. When attempting to assemble these short reads, most assemblers will produce highly fragmented assemblies, with breaks occurring at the location of every repetitive sequence longer than a read. We have developed a new assembly algorithm that overcomes most of the major difficulties of short-read assembly. One of the key innovations is the use of predicted genes to span gaps, which we call gene-boosted assembly [1]. This method is particularly effective for gene-dense species including bacteria and viruses. Using our new algorithm in conjunction with several other techniques, we assembled over 8.6 million reads from a bacterial genome sequenced with an Illumina Genome Analyzer. Each read was 33 nt in length. We were able to assemble the genome into fewer than 100 large contigs. The consensus sequence accuracy is >99.97%, and over 97% of the genes are contained within contigs. In the second part of my talk, I will address the problem of rapid alignment of short reads to the human genome. We have developed a new program, Bowtie, based on the Burroughs-Wheeler Transform, that aligns short reads at very high speed with very modest memory requirements. Bowtie is able to align reads to the human genome using only a standard desktop workstation, with performance benchmarks that are dozens to hundreds of times faster than competing systems. This talk describes joint work with Dan Sommer, Daniela Puiu, Ben Langmead, and Cole Trapnell.


Biography of Speaker

Steven Salzberg is the Director of the Center for Bioinformatics and Computational Biology (CBCB) and the Horvitz Professor of Computer Science at the University of Maryland, College Park. From 1997 to 2005 he was at The Institute for Genomic Research (TIGR) in Rockville, Maryland, where he was the Senior Director of Bioinformatics, in charge of TIGR's bioinformatics research as well as its production operations. During that time he was also a Research Professor of Computer Science and Biology at Johns Hopkins University in Baltimore. Dr. Salzberg received his B.A. degree in English and M.S. and M.Phil. degrees in Computer Science from Yale University, and his Ph.D. in Computer Science from Harvard University. Following his Ph.D. studies, he joined the Computer Science Department at Johns Hopkins as an Assistant Professor in 1989. Salzberg's research on gene finding in the 1990s included the development of the Glimmer system for bacterial gene-finding program, which has become one of the world's most successful and widely-used gene finders. Glimmer has been used in hundreds of bacterial, archaeanal, and viral genome projects, including the bacteria that cause tuberculosis, Lyme disease, cholera, syphilis, and anthrax. Eukaryotic gene finders developed by Salzberg's group have been used for animals (including human), plants, and eukaryotic parasites including Plasmodium falciparum (malaria), Brugia malayi, and Trypanosoma brucei. Salzberg and his group have developed and continue to improve software for large-scale genome sequence alignment and genome assembly, including the open-source genome assembler package, AMOS. Their open-source systems have been distributed to thousands of scientific laboratories.
around the globe. Dr. Salzberg has authored or co-authored two books and over 150 publications in leading scientific journals. He is a Fellow of the American Association for the Advancement of Science (AAAS) and a member of the Board of Scientific Counselors of the National Center for Biotechnology Information at NIH. He currently serves on the Editorial Boards of the journals BMC Biology, Journal of Computational Biology, PLoS ONE, BMC Genomics, BMC Bioinformatics, Biology Direct, Applied Bioinformatics, and is a member of the Faculty of 1000 Biology.
Abstract
While most data mining techniques do not directly depend on ontologies, data mining generally benefits from cleaner, crisper, broader datasets, abstracted away from unnecessary details. Using examples from the biomedical domain, I will review the potential benefits of ontologies for data mining, namely through normalization, integration and aggregation of biomedical data. Normalization contributes to the creation of more homogeneous datasets. Integration enables queries across datasets, thus broadening the scope of investigations. Finally, aggregation helps gain statistical power by abstracting away from details specific to a particular dataset. Biomedical text mining has long relied on ontologies, because they provide both a source of names for identifying biomedical concepts in text and the domain knowledge necessary for validating the relations extracted from text corpora. Data mining can take advantage of ontologies in a similar fashion. Some limitations of this approach will be discussed as well, including the limited availability and lack of integration of some ontologies. Olivier Bodenreider is a Staff Scientist in the Cognitive Science Branch of the Lister Hill National Center for Biomedical Communications at the U.S. National Library of Medicine. His research interests include terminology, knowledge representation and ontology in the biomedical domain, both from a theoretical perspective and in their application to natural language understanding, reasoning, information visualization and integration.

Biography of Speaker
Dr. Bodenreider is a Fellow of the American College of Medical Informatics. He received a M.D. degree from the University of Strasbourg, France in 1990 and a Ph.D. in Medical Informatics from the University of Nancy, France in 1993. Before joining NLM in 1996, he was an assistant professor for Biostatistics and Medical Informatics at the University of Nancy, France, Medical School.
Functions of Intrinsically Disordered Proteins and Relationship with Human Disease Network

Zoran Obradovic

Abstract

About 10 years ago we developed the first prediction based method to show that thousands of proteins lack fixed structure or are disordered (or unfolded) under physiological conditions. In this talk we will first briefly present the initial and our more recent predictor rated as the best model in disorder category at the seventh critical assessment of structure prediction experiments (CASP7). Next, we will describe how we used this predictor to provide a leap jump in answering a challenging question of understanding relationship between protein disorder and protein function by discovering that 238 Swiss-Prot functional categories are strongly positively correlated with predicted long intrinsically disordered regions. We will also describe a text mining method that can help further characterizations by identifying PubMed publications with relevant protein-disorder related experimental evidence. Motivated by observed strong correlation between intrinsic disorder and molecular functions known to be involved in cancer, we performed a large scale analysis of intrinsic disorder in genes implicated in Human Disease Network. We found significant differences in disorder content distributions for genes related to 20 human disease classes. Prediction of Molecular Recognition Features (MoRFs) suggests that intrinsic disorder in disease genes is mainly involved in protein-protein interactions. Genes related to several classes of diseases were found to have significantly higher occurrence of alternative splicing (AS), and given that we found AS regions to be highly disordered, we conclude that intrinsic disorder, together with AS, likely plays an important role in these classes of diseases.

Biography of Speaker

Zoran Obradovic’s research interests focus on developing data mining and statistical learning methods for knowledge discovery at large databases. He has authored about 200 articles addressing data analysis challenges in bioinformatics, medical informatics and other domains. For example, Obradovic co-authored a number of “first of” informatics studies on protein disorder providing strong support for a hypothesis that intrinsic protein disorder lies at the basis of signaling, regulation, and control. He also served as the team leader for the best predictor in protein disorder category at the fifth, the sixth and the seventh Critical Assessments of Structure Prediction experiments (CASP). Obradovic is currently journal editorial board member at seven journals. He was track chair at seven and program committee member at about 40 informatics conferences. Last year Obradovic co-chaired 39th Symposium on the Interface of Statistics, Computing Science and Applications, ACM First International Workshop on Text Mining in Bioinformatics, and IEEE International Conference on Bioinformatics and Biomedicine. Currently he serves as the steering committee member of 2008 IEEE International Conference on Bioinformatics and Biomedicine and is the program chair of 2009 SIAM International Conference on Data Mining.
Telemedicine: Wearable Intelligent Sensors and Systems (WISS)
for Mobile Healthcare
Yuan-Ting Zhang

Abstract
The steady-growing global market of medical devices and biosensors over the past decade has brought a great deal of attention to the academia and industry everywhere in the world. To meet the needs of those with the chronic diseases of aging, wearable medical devices are becoming an important part of solution. In order to monitor human health constantly without disturbing users’ normal daily activities, the ideal wearable medical devices and biosensors should be designed to be so smart and autonomous that they will be operable by any individuals from a child to an aged person for their own health management, and to be so small that they will not affect the appearance and function of the wearable carrier in which they are embedded. One of major challenges is to develop new principle of physiologic measurements in incorporation with effective bio-signal processing and medical information fusion techniques enabling the design of small and smart wearable devices. We are currently developing a series of wearable intelligent sensors and systems (WISS) with a body sensor network (BSN) forming the communication infrastructure for telemedicine and mobile healthcare. This talk will review some of the more recent developments in these areas, highlight the evolving MINDS concepts of wearable design, and discuss how the emerging wearable technologies can meet the growing demand for future homecare and mobile health services. With the wearable cuffless blood pressure meter as an example, the talk will also attempt a living proof on how the consumer personal electronic devices can be used in creating innovative and high-value-added healthcare products in a relatively short duration, which could rapidly open up a new product sector for industry and offer great benefits for the society.

Biography of Speaker
Yuan-Ting Zhang received his Ph.D. from the University of New Brunswick, Canada in 1990. He is currently Director of the Joint Research Center for Biomedical Engineering and the Founding Head of the Division of Biomedical Engineering at the Chinese University of Hong Kong. He also serves as the Director of the Key Lab for Biomedical Informatics and Health Engineering of the Chinese Academy of Sciences, the Director of the SIAT Institute of Biomedical and Health Engineering of Chinese Academy of Science and the Chairman (Adjunct) of the Department of Biomedical Engineering at Sun Yat-Sen Medical School, Guangzhou, China. Dr. Zhang was a Research Associate and Adjunct Assistant Professor at the University of Calgary, Canada from 1989 to 1994. He chaired the Biomedical Division of Hong Kong Institution of Engineers in 1996/97 and 2000/01. Dr. Zhang was the Technical Program Chair of the 20th IEEE-EMBS Annual Int'l Conference in 1998 and the General Conference Chair of the 27th IEEE-EMBS Annual Int'l Conference in 2005. He was elected as an AdCom member in 1999 and served as Vice-President (Conferences) of IEEE-EMBS in 2000. He was a Founding Co-Chair of IEEE-EMBS Summer School on Medical Devices and Biosensors. He has served as Associate Editor for IEEE Trans. on Biomedical Engineering and IEEE Trans. on Mobile Computing. He was also the Guest Editor of IEEE Communication Magazine and IEEE Trans. on Information Technology in Biomedicine. He is currently on the Editorial Board of the Book Series of Biomedical Engineering published by the IEEE press, the IEEE-EMBS Technical Committee of Wearable Systems and Sensors, and an Editorial Board Member of the Journal of China Medical Device Information. Dr. Zhang has been selected to serve as the Editor-in-Chief for IEEE Trans. on Information Technology in Biomedicine since 2008. Dr. Zhang’s research interests include neural engineering, wearable medical devices, and body sensor networks particularly for mobile health and telemedicine. He has published more than 300 scientific articles in the area of biomedical engineering and filed over 15 patent applications. Dr. Zhang and his team received more than 30 awards and recognitions which include the Fellow of International Academy of
Medical and Biological Engineering, IEEE Fellow, AIMBE Fellow, and the recipient of 2006 Distinguished Service Award of IEEE-EMBS.
Abstract

Advances in experimental techniques and instruments have resulted in increased generation of quantitative data for biological systems. These higher quality data demand better data analysis to extract the most amount of useful information from them. This talk will highlight examples that demonstrate how mathematical algorithms and computer-assisted analyses optimize the usable information that can be gleaned from an experimental data set. Examples will be taken from cancer, influenza and cardiac disease. In each example, a mathematical model of a biological experiment is paired with data from the experiment; the changes in the biological process are matched by finding the changes in model parameters that explain the experimental perturbations. The examples of this talk feature the use of a recently-published global optimization algorithm that markedly enhances the reliability of automated parameter estimation.

Biography of Speaker

Dr. Lett is a mathematician, scientist, engineer and entrepreneur, with nearly 30 years of experience developing software analysis tools for the aerospace, petroleum, environmental and biomedical industries. Dr. Lett is co-founder and CEO of The BioAnalytics Group, a New Jersey-based company that offers software products and services to the biomedical research community. He works in the areas of assay data analysis, data management, biological modeling, technology evaluation, intellectual property evaluation and custom software development. In addition to analytical tools for research, the company has developed systems for managing and sharing research data. The BioAnalytics Group’s flagship product is BioPathwise DM®, a researcher-friendly data management system, developed under contract to the National Institutes of Health. Dr. Lett’s service to the biomedical community includes membership in the Modeling and Analysis of Biological Systems (MABS) Study Section of the National Institutes of Health. He is an Adjunct Professor of Biomedical Engineering at The College of New Jersey.