Algorithms and Methods for Biological Data

Department of Biotechnology and Bioinformatics, School of Life Sciences, University of Hyderabad, Hyderabad- 500046, INDIA
February 09– February 24, 2019

Overview:

Algorithms and Methods for Biological Data Analysis is a graduate level course that provides a broad overview of the field of bioinformatics, algorithmic solutions for biological data analysis, and applications in genomics and proteomics. The course is based on a course offered at the University of Cincinnati to students in computational and interdisciplinary programs, such as Biomedical Informatics, Biostatistics, Computer Science, Electrical and Computer Engineering, Biomedical Engineering, as well as Systems Biology & Physiology, Cancer Biology, and related programs. Other students interested in learning about computational methods in biomedical research are also encouraged to take this course.

This course will expose participants to current algorithms in biological data analysis and modelling, and teach how to build and modify an algorithm from a mathematical and logical standpoint to a computationally executable form. Furthermore, the course aims to expose students to the importance of statistical evaluation of biological data and development of various types of statistical modeling approaches in biological data mining, discovery and validation.

Objectives

The primary objectives of the course are as follows:

- Students will acquire basic concepts underlying sequence approximation to molecular biology.
- Students will become familiar with various aspects of processing and analyzing biological data sets, including those generated by modern DNA sequencing platforms.
- Students will become familiar with a number of algorithmic approaches that provide efficient and scalable solutions to central computational challenges in the context of DNA sequencing and related applications.
- Students will learn effectively using state-of-the-art databases, analysis tools and pipelines, as well as other bioinformatics resources.
- Through hands-on, students will learn developing effective strategies for applying computer science and computational modeling techniques to the analysis of biomedical data.

Course details

- Sequence approximation in molecular biology; Overview of next generation sequencing approaches
- Building reference genomes: physical maps and genome assembly
- Genome browsers and other resurces for biological sequence analysis.
- Bioinformatics of re-sequencing projects
- Exact string matching for sequencing with short reads
- Applications of large-scale (re-) sequencing with short reads: variants, mutations, gene expression, BOWTIE and related software packages
- Genes, regulatory sequences and other genomic features
- Motif discovery; Finding coding and regulatory sequences
- Gene prediction in prokaryotic genomes
- Dynamic Programming (DP) for biological sequence analysis
- Sequence alignment and homology search
- Alignment methods quiz and overlap (prefix-suffix) matches using DP
- Sensitive and efficient methods for biological sequence alignment: BLAST
- Models of random alignments and statistical significance of BLAST hits
- Statisical models for biological sequences quiz; Randomizing data to assess the noise level and derive empirical p-values
- Detecting remote homologs using Psi-BLAST and related approaches
- Hidden Markov Models for CpG islands identification
- Profile HMMs for multiple sequence alignment
- Transcription regulation by methylation and CpG islands; Analysis of epigenetic changes using deep sequencing
- mRNA-seq and related assays for gene abundance
- Going beyond sequence approximation: Macromolecular structure and function
- Molecular interactions in drug development
- Tools and resources for macromolecular structure analysis
- Machine and statistical learning for structural bioinformatics
- Protein structure and function prediction
- Network analysis and statistical models of co-regulated modules

Who can attend?

- Executives, engineers and researchers from manufacturing, service and government organizations including R&D laboratories.
- Student students at all levels (BTech/MSc/MTech/PhD) or Faculty from reputed academic institutions and technical institutions.

How to Apply

Interested candidates must login GIAN-MHRD website (http://www.gian.iitkgp.ac.in/) to fill the application. Please submit your detailed CV to biodata2019@gmail.com.

Registration Fees: It will include all instructional materials, lunches, computer use for tutorials, local transport and 24hrs internet facility. Accommodation in guest house will be provided on the basis of payment.

Participants from Abroad	US$ 100
U.G./ P.G. Students	INR 1500
Postdocs/ PhD Students	INR 2000
Faculty/ Scientists	INR 5000
Industry	INR 10000

The demand draft may be drawn in favour of "UOH Temp 2018/010 " or deposit SBI HUC Branch Account No. 38010884305 (IFSC Code. SBIN0005916).

Teaching faculty

Prof. Jarek (Jaroslaw) Meller

Prof. Jaroslaw (Jarek) Meller is Professor of Environmental Health, Biomedical Informatics and Computer Science at University of Cincinnati, Ohio, USA. Building upon broad interdisciplinary training, he has been pursuing research at the intersection of data science, informatics and biomedicine. His primary expertise lies in the fields of computational genomics, molecular modeling, and structural bioinformatics. On the methodological side, with a team of collaborators, he has developed a number of successful prediction methods for structural and functional studies of proteins and their complexes (Adamczak et al., 2004, 2005, 2011; Porollo and Meller, 2007, 2010; Phatak et al., 2011). He has also been active in the development and applications of methods for knowledge extraction from high dimensional genomic data (Sinha and Meller, 2008; Shinde et al., 2010; Huang et al., 2012). As of January of 2016, these methods and web servers developed by his group, including Sppider, Cinteny and Polyview, have been used by over 1 million times by about 50,000 users from over 100 countries. Prof. Meller has also been involved in a number of high impact collaborative projects in the areas of basic and translational biomedical research. Examples of such interdisciplinary efforts include: sequencing and annotation of human pathogens (Cushion et al, 2007; Liu et al., 2012, Walther et al., 2015), identification of markers associated with disease subtypes and clinical outcomes (Zhang et al., 2011; Biesiada et al., 2014, 2015), modeling of signal transduction pathways in cancer and autoimmunity (Mikhaylova et al., 2012; Stratton et al., 2014), and developing inhibitors of critical protein-protein interactions (Bosco et al., 2012; Evelyn et al., 2014, 2015).

•To date, he has published over 100 peer-reviewed contributions, with the ISI Web of Science h-index of >35. Prof. Meller has served as PI, co-PI or co-Investigator on a number of NIH-funded and other grants, and has a considerable experience in coordinating large-scale multidisciplinary projects. In addition to research related activities, he has been broadly involved in quantitative and computational training efforts within the University of Cincinnati Colleges of Medicine and Engineering and Applied Sciences. He currently serves as the Director of the PhD Program in Biomedical Informatics, and co-director of the Biomedical Informatics Graduate Certificate Program. He has developed and taught a number of graduate level courses, including Introduction to Bioinformatics, Data Science for Biomedical Research and Introduction to Functional Genomics. In addition, he serves as co-director and mentor on T32 Advanced Multidisciplinary Training Program for Systems Biology, co-director of the Bioinformatics Core for the Center of Environmental Genetics at the University of Cincinnati, and director of the Protein Informatics Core at Cincinnati Children's Hospital Medical Center.

Prof. Anand K. Kondapi,

Molecular therapeutics, functional characterization of DNA topoisomerases in oncogenesis, HIV infection, neuro immune activity and brain aging.

Prof. H.A Nagarajaram,

Computational systems biology; assessment of functional impact disease causing mutations at molecular and systems level; discovery of basic structural principles governing protein functions.

Dr. D. Pankaj Singh,

Machine learning for Data Science, Mathematical/computational modelling of gene networks, Knowledge discovery in Neuronal aging/senescence and neurodegenerative diseases.

Speakers

- **Prof. Jarek (Jaroslaw) Meller, Biomedical Informatics and Computer Science** University of Cincinnati, Ohio, USA
- **Prof. Anand K. Kondapi**, Dept Biotech Bioinf, Sch Life Sciences, Univ Hyd., Hyderabad
- **Prof. H.A Nagarajaram, Dept Comput and Systems Biol.,** Univ Hyd., Hyderabad
- **Dr. D. Pankaj Singh**, Dept Biotech Bioinf, Sch Life Sciences, Univ Hyd., Hyderabad

Course coordinators

•**Prof. Anand K. Kondapi**, Dept Biotech Bioinf, Sch Life Sciences, Univ Hyd., Hyderabad

•**Dr. D. Pankaj Singh**, Dept Biotech Bioinf, Sch Life Sciences, Univ Hyd., Hyderabad

Please visit (www.sls.uohyd.ac.in) for further information or contact through e-mail (**biodata2019@gmail.com**).