

An Online Hands-on Workshop on
“Genomics, Metagenomics & Bioinformatics in Microbial Ecology”

Organized by

Indian Network for Soil Contamination Research & PhiXgen Pvt. Ltd.
under the aegis of International Society for Microbial Ecology
Organized by the International Society for Microbial Ecology (ISME)

DATE: November 7, 2022

ISME's mission is to promote Microbial Ecology research and teaching. On these lines, Dr. Jasvinder Kaur, Gargi College, University of Delhi, opened the session with a welcoming note and an introduction to ISME and the ISME ambassador program. Prof. Rup Lal, Senior Ambassador, ISME, made remarks after that (Indian Ocean Region). He discussed the duties of an ISME Ambassador and talked about the “Role of Computational Biology in Microbial Ecology”. Dr. Jillian Peterson, the Editor-in-Chief of ISME, delivered the plenary talk on " Publishing with ISME, an Editor-in-Chief's view”.

The inauguration session was followed by a series of workshop sessions as follows:

10:30-11:05 am- Role of Computational Biology in Microbial Ecology

Prof. Rup Lal (INSA Senior Scientist, Acharya Narendra Dev College, DU)

Dr. Lal launched the workshop by highlighting to the students how bioinformatic analyses enable the mining of massive metagenomic datasets and the identification of broad patterns underlying microbial ecosystems. He talked about the fundamental relevance of microbial ecology, as microorganisms account for the vast majority of the genetic and metabolic diversity on Earth and power the bulk of crucial ecological processes that recycle matter and energy. He talked about how microorganisms play a key part in a number of ecosystem processes, including oxygen production, decomposition, evolution, and symbiotic interactions. He also discussed his role as an ISME ambassador and how he and his team have been promoting ISME and microbial ecology throughout India and surrounding regions.

11:05-11:40 am- PLENARY TALK: Publishing with ISME, an Editor-in-Chief's view

**Dr. Jillian Petersen (Department of Microbiology and Ecosystem Science,
Division of Microbial Ecology, Vienna, Austria)**

Dr. Peterson started the talk with a tad introduction about her research interest, i.e., the most ancient chemosynthetic symbiosis and how environment affects it. She then talked about the ISME society, its beneficial membership and the prestigious ISME communications journal. She also informed the audience of the next ISME symposium to be held in Cape Town. Dr. Peterson emphasized that the ISME journal is ranked at number 7 worldwide amongst 173 ecology journals. The credit goes to a robust peer-review process and adherence to quality.

Dr. Peterson discussed about the rejection rates of the ISME journal and the common causes that underlie it. She stressed upon the importance of originality in research. In the later part of her lecture, she talked about the publication ethics and ways to maximize visibility of your work post publication. An overview of publication process was discussed, a good peer-reviewing process forms the basis of which. In the end, Dr. Peterson discussed her experience in the assessment of articles as per the editor-in-chief.

11:40-01:00 pm- Optimising the process of Biological Investigations - Computational Approach using Biopython

Resource Person: Satyam Sangeet (University of Auckland)

Mr. Satyam Sangeet Talked about the importance of optimising the investigation of biological sequences using BioPython. The session dealt with theoretical understanding of the importance of optimization when dealing with large datasets. The concept of optimisation was showed by taking an example of SARS-CoV-2 genomic sequence. Actual workflow was demonstrated where the genomic sequence of the virus was taken and the comparative analysis was performed between SARS- CoV and SARS-CoV-2.

01:00-01:30 pm LUNCH BREAK

01:30-02:00 pm- File Formats in Computational Studies

Resource Person: Dr. Renu Solanki (Deen Dayal Upadhyaya College, DU)

Dr. Solanki began the session with a talk on “File Formats in Computational Studies”. She apprised the participants about the requirement of having data formats in Computational studies. She mentioned that there are two types of data formats (i) sequence data format and (ii) structure data format. She explained the key features and identification of the formats including Fasta, Fastq,

Gff, Gbk, ENA-EMBL, SWISS-PROT, Clustal, Pdb, Mol2. At the end of the session she mentioned that the data format can be an input for a program, for example, Fasta format. Data format including Gbk, ENA-EMBL, SWISS-PROT are representation of information on different databases and some of the formats are output including FastQ, Clustal.

02:00-02:30 pm- Computational Databases

Resource Person: Dr. Shailly Anand (Deen Dayal Upadhyaya College, DU)

Dr. Shailly Anand continued the discussion with a brief introduction to databases, their importance in the present context of big data, and the general architecture of any database. She explained with suitable examples how a biological database is a collection of data that is structured, searchable, updated periodically, and cross-referenced. A database helps to store, maintain, annotate and curate the data. She went ahead to elaborate on the importance of databases for storing and communicating large datasets, distributing resources, and making biological data available to scientists in computer-readable form along with ease of access. She then discussed the different basis of classification of biological databases as proposed by various workers. Thereafter, she gave an insight into the available primary and secondary databases and how each is involved in keeping nucleotide and protein sequence and structural data in an organized form. Towards the end, she not just gave an overview of organism-specific databases such as FlyBase, ZFIN, TAIR, etc but also conducted a hands-on simple exercise on retrieving a protein sequence from a primary database and fetching information in different formats.

02:30-03:30 pm- Protein Modelling

Resource Person: Dr. Nirjara Singhvi (School of Allied Sciences, DBUU)

Dr. Singhvi continued the session by talking about role of the cell molecules - the proteins that perform various cellular functions. She highlighted the significance of understanding the relationship between sequence, structure and function of proteins in order to understand the functioning of cells under various conditions. She elaborated on the procedure of homology modelling for predicting the structure of proteins from their sequences based on the homology of these sequences with the sequences of proteins whose structures are known and deposited in protein databank. The participants learnt about the SWISS-MODEL interface to predict the structure of an example protein sequence. The information provided as the output was analysed.

Further, structure assessment and validation methods, for e.g. by reading the Ramachandran plot of the obtained structure were also discussed. The obtained structure was learnt to be visualized in UCSF Chimera software.

03:30-04:30 pm- Sequence alignment

Resource Person: Dr. Mansi Verma (Hansraj College, DU)

Dr. Verma continued the session with “Sequence alignment” where she started with the basics of similarity and identity. Participants were taught about the alignment matrices PAM and BLOSUM and how these matrices are used for generating scores of alignment. Further the session dealt with Global and local alignment, and types of BLAST. The pairwise alignment was extended further to Multiple sequence alignment. Students performed hands on session on BLAST where they learnt about BLASTp and BLASTN alignment for proteins and nucleotide respectively. Clustal Omega was used for performing Multiple sequence alignment on protein sequences.

04:30-05:15 pm- Genome Annotation by RAST

Resource Person: Dr. Princy Hira (Maitreyi College, University of Delhi)

Dr. Hira talked about “Genome Annotations”. She dealt with the basics of annotation like structural and functional annotations, levels of annotation and different tools, hands on exercise on RAST (Rapid Annotation using Subsystem Technology) Server and how it provides a high-quality genome annotation for prokaryotes across the whole phylogenetic tree were explained. Once annotation is completed, genomes can be downloaded in a variety of formats or viewed online. The genome annotation provided does include a mapping of genes to subsystems and a metabolic reconstruction. In addition, students were also taught how to predict closest neighbors for the bacterial genome using RAST.

05:15-06:15 pm- Culturomics and Metagenomics: Two sides of the same coin!

Resource Person: Dr. Roshan Kumar (PG Department of Zoology, Magadh University)

The session came to an end with a talk by Dr. Kumar on Culturomics and metagenomics. Dr. Kumar started briefing the advantages of metagenomics. He mentioned the significance of next generation sequencing approach to deliver the diversity composition of any ecological niche whether it is soil, water, fecal sample etc. He then emphasized the genome binning process from the metagenomics read. Later, he focused on the high throughput culturomics approach and use

