

Sourav Sarker

[✉ sourav39.dev@gmail.com](mailto:sourav39.dev@gmail.com) | [🌐 iamsourav.ml](http://iamsourav.ml) | [🔗 po95](https://po95.com) | [🌐 sourav-sarker-6a9029b8](https://sourav-sarker-6a9029b8.netlify.app)

Having a colorful footprint in software engineering spanning over four years, I have cultivated a robust skill set that includes clean coding, system design, settling R&Ds and debugging abilities; also, my commitment to problem-solving is demonstrated through successful project completions. Now, I am fueled by an unwavering passion for delving into challenging research topic which makes me an ideal candidate for contributing in.

Education

Bachelor of Science in Biochemistry and Molecular Biology

SHAHJALAL UNIVERSITY OF SCIENCE AND TECHNOLOGY(SUST)

2016-2019

Sylhet, Bangladesh

- CGPA: 3.36 out of 4.00
- Project (2019): Prediction of Missing DNA Methylation from Whole Genome Bisulfite Data Using KNN
 - Advisor: Rashedul Islam - PhD in Bioinformatics - University of British Columbia, Canada

Research Experiences

Child Health Research Foundation

RESEARCH OFFICER, BIOINFORMATICIAN - GENOMICS TEAM

Dhaka, Bangladesh

01 Aug 2022 - Present

- CZI A Global Pediatric Cell Atlas of Nasal and Oral Mucosa.
 - Single cell isolation, Single cell capture, Whole transcriptomic amplification, Library preparation using HoneyComb, 10x Genomics, Parse Biosciences and loading on NextSeq2000.
 - Computational analysis using Scanpy and Seurat.
- Training: From the team members of Ordovas-Montañes Lab, Assistant Professor Harvard Medical School.
- Workshop: Attended 2024 HCA Single-Cell Omics in-person Computational and Experimental Design Workshop, Bangkok, Thailand
- Collaboration: Ordovas-Montañes Lab - Boston Childrens Hospital, MIT
- A computational framework to automatically interpret pediatric chest X-rays and diagnose pneumonia.
 - Model development using CNN and Pytorch.
 - Funded by: NIHR Global Health Research Unit on Respiratory Health
- Parv4 detection in children with suspected meningitis is associated with high mortality.
 - Metagenomics analysis—consensus sequence generated using CZ ID, mapping, MAFFT, RAxML, and annotation.
- Global Pneumococcal Sequencing Project.
 - Development of Nextflow workflow.
 - Advanced Bioinformatics Workshop developed for GPS and JUNO projects by Wellcome Sanger Institute in Bangalore, India.
- RSV Vaccine Impact Monte Carlo Simulation.
 - RSV vaccination mortality impact analysis via Monte Carlo simulation.

National Gene Bank, National Institute of Biotechnology(NIB)

RESEARCH FELLOW

Dhaka, Bangladesh

5 Sept 2021 - 30 June 2022

- Revealing the Disease complexity and Population Structure of Human Gut Microbiota from Diverse Cohort of Bangladesh Population.
 - Metagenomic analysis of 16S rRNA sequences using qiime2.

Teaching Experience

Child Health Research Foundation (CHRF)

TEACHING ASSISTANT, BUILDING SCIENTISTS FOR BANGLADESH

Dhaka, Bangladesh

2023-2024

- Advanced Next Generation Sequencing
 - Linux OS, NGS Data Analysis: Quality checking, Quality filter, Genome Assembly, Genome Annotation
 - Building Consensus using CZID, Sample Sheet Preparation for NextSeq2000
- Bacterial Genomics Antimicrobial Resistance
 - Linux OS, Bash Scripting Genome Assembly, Genome Mapping, Genome Annotation, Genome Submission, MLST Prediction
- AI in Public Health
 - Data cleaning and preprocessing, Data Visualization, Architecture of ML and DL Libraries- Numpy,Pandas, Keras,TensorFlow
 - Dengue prediction in Puerto Rico using LSTM and ARIMA
 - Predicting Pneumonia in X-rays using CNN
 - Advanced Modeling Techniques- PyCaret, GRU

Bioinformatics Division, National Institute of Biotechnology(NIB)

INSTRUCTOR, BIOINFORMATICS DIVISION

Dhaka, Bangladesh

2021

- Delivered training on Bioinformatics for Biotechnology Research
 - Linux OS, Bash Scripting and Biopython for NGS Data Analysis

Technical Skills

- **Languages:** Bash, Python, C, R, Nextflow, Java
- **Database Management Systems:** SQL, Redis, SQLite3
- **Machine Learning and Deep Learning:** Proficient in various frameworks and techniques: Numpy, Pandas, Matplotlib, Seaborn, Keras, SciKit-Learn, TensorFlow, PyTorch
- **Bioinformatics:**
 - Proficient in Docker for pipeline buildup.
 - Microarray and Bulk RNA sequencing Analysis using DESeq2, stats, ggplot, corrplot, pheatmap, EDASeq, gProfileR
 - Genome quality checking and assembly: **Illumina**: BCL2fastq, fastQC, MultiQC, Quast, Trimmomatic, fastp, Unicycler, Spades, Megahit; **Nanopore**: fast52fasta, Pilon, Flye
 - Genome annotation: Kraken2 , Prokka, Seroba, AMRFinderPlus, Abricat, SRST, MLST, Snippy, Mafft, fasta2phylip, Raxml-ng, PopPunk, PlasmidFinder, ResFinder
 - Molecular Dynamics Simulation using GROMACS.
 - Metagenomics analysis using qiime2 and CZID.

- Nextstrain Built and maintained from Bangladesh for SARS-CoV-2.
- **Single Cell RNA sequencing Analysis using Scanpy and Seurat: Data pre-processing, Clustering and Cell annotation, Integration and Batch correction, Cell-cell communication, BCR background and 10x analysis, Trajectory inference, Differential abundance, Multiomic scATAC**
- **Wet Lab:**
 - Single Cell RNA sequencing Analysis using HoneyComb, 10x Genomics, Parse Biosciences, Library preparation, Sequence on Illumina and Nanopore
 - Molecular techniques: Conventional and Real-time PCR, DNA & RNA extraction, Agarose & Polyacrylamide Gel Electrophoresis and Imaging, Blotting techniques, Molecular cloning, ELISA & ICT
 - Microbiological techniques: Culture, Biochemical test

Publications

Dev,P.,C., Rehman,A., Hossain,M.,M., Kazi,E., Goswami,S., Rehman,H., Hooda,Y., Tanmoy,A.,M., Saha,S.,K., Saha, S. Genomic epidemiology of Salmonella Paratyphi B isolates from Bangladesh.(In-process)

Debnath, J., P., Hossen,K., Khandaker,M., S., Sayed,S., B., **Dev, P., C.,*** Sarker, S.,* Hossain,T.* Influenza and Other Respiratory Viruses. **Decoding the Unique Gene Signatures Upon 2022 MPXV Infection: Predictive Biomarkers and Drug Targets via Transcriptomic Network Analysis and Machine Learning Models.(Submitted)**

Islam,S., Kabiraj,R., Sarkar,H., **Dev,P.,C.,** Akter,A.T., Purna,D.K., Rajib,A.M., Mohammad,A.T., Kumar,S.S., Hooda,Y., Saha,S. Microbiology Resource Announcements: **Genome sequences of bacteriophages that infect Salmonella Typhi from Bangladesh.(Submitted)**

Raju,R.,S., Nahid,A.,A., **Dev,P.,C.,** Islam,R. **VirusTaxo: Taxonomic classification of viruses from the genome sequence using k-mer enrichment.** Genomics. 2022 Jul;114(4):110414. doi: 10.1016/j.ygeno.2022.110414. Epub 2022 Jun 17. PMID: 35718090.

Chowdhury,Z.,M., Bhattacharjee,A., Ahammad,I., Hossain,M.,U., Jaber,A.A., Rahman,A., **Dev,P.,C.,** Salimullah,M., Keya,C.,A. **Exploration of Streptococcus core genome to reveal druggable targets and novel therapeutics against S. pneumoniae.** PLoS One. 2022 Aug 18;17(8):e0272945. doi: 10.1371/journal.pone.0272945. PMID: 35980906; PMCID: PMC9387852.

Conference Presentations

Dev,P.,C.,* Malaker, A.,R., Keya,D.,P., Hooda,Y., Saha,S., Building a single-cell atlas of the nasopharyngeal mucosa to investigate SARS-CoV-2 infection. 2024 HCA Single-Cell Omics in-person Computational and Experimental Design Workshop, Bangkok, Thailand.

Dev,P.,C.,* Tanmoy,A.,M., Rahman,A., Saha,S.,K., Hooda,Y., Saha,S., Paratype v1.1: recent updates to the genotyping tool for paratyphoid fever surveillance. 13th International Conference on Typhoid and Other Invasive Salmonelloses, Kigali, Rwanda.

Accomplishment

6th Position from Bangladesh: Rosalind, SUST
2018-2024 Bioinformatics problem solving using Python.

100+ Problem solved: Leetcode, SUST
2022-2023 Problem solving in competitive programming using Python.

220 out of 13558: DrivenData, CHRFBD Datascience Team - Dhaka, Bangladesh
2023-2024 Predicting Disease Spread of Dengue using LSTM.

Thesis Supervision: Child Health Research Foundation (CHRF), Dhaka, Bangladesh
2022-2024 Assist **4 students and 1 intern** with executing the bioinformatics pipeline for their projects.

Academic Honor: Dept. of Biochemistry and Molecular Biology(SUST), Sylhet, Bangladesh
2016-2019 Merit-based scholarship for academic excellence in the 1st and 4th year.

References

Tanvir Hossain

ASSISTANT PROFESSOR, DEPT. OF BIOCHEMISTRY AND MOLECULAR BIOLOGY
PhD Candidate, Australian Institute for Bioengineering and Nanotechnology
The University of Queensland, Brisbane, Australia

Sylhet, Bangladesh

SUST

Mohammad Uzzal Hossain
SENIOR SCIENTIFIC OFFICER, BIOINFORMATICS DIVISION
MSc, Oxford University

Dhaka, Bangladesh

National Institute of Biotechnology(NIB)

Dr.Yogesh Hooda
SCIENTIST
Post-doc, MRC lab-Cambridge University
Email: yhooda@chrfbd.org

Dhaka, Bangladesh

CHRF