

# Postgraduate Seminar Series

*Venue: Graduate Seminar Room*

*Date & Time: March 07, 2026 at 2:45 PM*

## **Speaker Information**

Md. Shariful Islam Bhuyan (Std No. 0418054001) is a part time Ph.D. student in the department of CSE, BUET. He completed his B.Sc. Engg. in CSE in 2006 and M.Sc. Engg. in CSE in 2009, both from Bangladesh University of Engineering and Technology (BUET). His research interest lies in the field(s) of Bioinformatics, Machine Learning, and Deep Learning. He is currently doing his postgraduate thesis under the supervision of Dr. Mohammad Saifur Rahman. He will be speaking about his ongoing research in this talk.

## **The Landscape of Multi-Allelic Indels: From Quality Control to Genomic Annotation**

While bi-allelic SNPs are responsible to much of the recent progress in human genetics, efforts to sequence sizeable cohorts increasingly discover additional classes of variants. Such variants exhibit multiple alleles that arose not only by single nucleotide substitution but also involve insertions or deletions. These multi-allelic indels (MAI) present opportunities and challenges from data quality to analysis of population history and genetic association. In this study, we set out to survey MAI from large, publicly available sequencing datasets. We established a cross-platform quality control pipeline for different classes of MAI, from simple homopolymers to variants without even a clear phylogeny describing the alleles. We explored their accumulation as sample size increases and investigated their functional and genomic annotation. We found that different filtering criteria affect the composition of MAI. As expected, the number of novel MAI per sample decreases with cohort size, but not as sharply as the number of bi-allelic indels. This phenomenon is robust to different quality control filtering criteria. Multiple repeat types are enriched for MAI, while functional regions are depleted. To the best of our knowledge, this is the first comprehensive study of multi-allelic indels collected from high-coverage DNB (DNA Nanoball) whole genome sequencing data. Comparison with bi-allelic indels revealed several salient characteristics of MAI, which can be used as foundation for future exploration.