

# Syllabus of

## Next Generation Sequencing Tools and Algorithms (SNGS)

### (Lecture-1, Tutorial-2, Practical-1)

Component	Unit	Topics
C1	Unit 1	DNA sequencing, strings, and matching: DNA sequencers and working principle, DNA as a string. Parsing and manipulating real genome sequences and real DNA sequencing data. Naive exact matching, homology detection; optimal pair-wise sequence alignment, alignment score statistics, efficient database searches (BLAST), Data science of metabolomics, pathway models
	Unit 2	Preprocessing, indexing and approximate matching: Improving on naive exact matching with Boyer-Moore. Preprocessing and indexing. Indexing through grouping and ordering, k-mers and k-mer indexes. Approximate matching and the pigeonhole principle. Edit distance, assembly, overlaps: Hamming and edit distance. Algorithms for computing edit distance. Dynamic programming. Global and local alignment. De novo assembly. Overlaps and overlap graphs.
C2	Unit 3	Algorithms for assembly: Shortest common superstring and the greedy version. How repetitive DNA makes assembly difficult. De Bruijn graphs and Eulerian walks. How real assemblers work. The future of assembly.
	Unit 4	Data variability and replication, Data transforms, Clustering, Dimension reduction, Pre-processing and normalization, Linear models with categorical covariates, Logistic regression, Null and alternative hypotheses analysis, false discovery rate, permutation and bootstrapping, Gene expression repository (GEO).

#### Books and references

1. DNA Sequencing From Experimental Methods To Bioinformatics by *Alphey, Luke*
2. Analytical Techniques In DNA Sequencing by *Veena Kumari*
3. Next-Generation Sequencing Data Analysis by *Xinkun Wang*
4. Primer to Analysis of Genomic Data Using R (Use R!) by *Cedric Gondro*