HACKATHON PROBLEMS

1. **Interactive Visualization of Taxonomic Diversity of all species with DNA Barcode Data from Nepal**

DNA barcoding refers to the use of DNA sequences from a signature region of the genome to make species-level identifications. The process involves the use of reference sequence libraries from individual specimens identified by experts. There are many databases openly available where you can access the DNA barcode and use those information for various purposes. One of the databases is Bold Systems (https://www.boldsystems.org/index.php). BOLD is a cloud-based data storage and analysis platform developed at the Center for Biodiversity Genomics in Canada. It consists of four main modules, a data portal, an educational portal, a registry of BINs (putative species), and a data collection and analysis workbench.

You are chosen to utilize the data from the Barcode of Life Data Systems (BOLD) where you must extract records of all species from Nepal, then create a web-based interactive map of Nepal which can be used to locate (i.e. pin) the geo-spatial location of the sample collected and link the species to its respective record in the BOLD database.

HACKATHON PROBLEMS

2. A computational/bioinformatics pipeline to identify toxic bacteria.

A hypothetical scenario is presented here. Kathmandu University recently conducted a county-wide survey of the state and quality of drinking water reservoirs. Microbiological and microscopic analysis of water samples collected from Sundarijal, Kathmandu, detected the presence of gram-negative bacteria. However, the scientists could not identify the species of the bacteria. For identification, the National Public Health Laboratory (NPHL) has recently performed whole genome sequencing of the bacterial isolates using both the Illumina Next-seq and Oxford Nanopore sequencing platforms.

Imagine you are hired as the chief bioinformatician of the newly established Center for Disease Control (CDC), Nepal. You have been assigned a task to identify if the bacteria are toxic to the human population. Construct a computational pipeline starting from the raw sequencing data to identify toxic bacteria. How would you complete the task?

HACKATHON PROBLEMS

3. GWAS : Genome Wide Association Study

In genomics, a genome-wide association study (GWA study, or GWAS), also known as whole genome association study (WGA study, or WGAS), is an observational study of a genome-wide set of genetic variants in different individuals to see if any variant is associated with a trait. GWAS studies how genotype is related to phenotype, and typically focus on associations between single-nucleotide polymorphisms (SNPs) and traits like major human diseases, but can equally be applied to any other genetic variants and any other organisms.

Given the map and ped files, find the genotypes which are associated significantly with the phenotype. You are free to use any models or software, but justify your reasons.