**\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*ARGhmm library \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***

ARGhmm library contained 254 different types of antibiotic resistance genes (ARGs) by profile hidden markov model (HMMs) of protein and nucleotide called as p/nARGhmm.

**\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* Description and Usage \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***

The described antibiotic resistance analysis tool is based on HMM profiles of ARGs and is named as BacARscan. The proposed tool can be used to monitor and annotate bacterial antibiotic resistance genes in both genomics and metagenomic dataset. One of the most notable improvements of BacARscan over other ARG annotation methods is its ability to work on complete genomic and as well as small reads sequence library with equal efficiency. It is anticipated that this ability of BacARscan would be helpful in rapid monitoring, characterization and surveillance of ARG repertoire in bacterial communities at an early stage of infection/outbreak. We also expect that BacARscan would be helpful to the scientific community for quick monitoring of ARGs in a microbial population.

**\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* Methods Overview \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***

The data to build ARGhmms is taken from (i) existing antibiotics resistance gene databases (e.g. CARD, ARDB, LacED, Resfams etc.), (ii) UniProtKB and (iii) by comprehensive literature survey. All genes were first binned on the basis of their antibiotic inactivation profile. Genes in each bin were then divided into distinct cluster on the basis of their sequence homology with the help of BLASTclust program with identity threshold 90% and query coverage 95%. After clustering, all clusters were manually analyzed, inspected and curated. Sequences of each selected cluster, was multiply aligned using Muscle 3.8 program with default parameters. Subsequently each MSA was converted into profile hidden markov models (profile HMMs). HMM models were building using hmmbuild function of the HMMER tool (hmmer.wustl.edu/) (version 3.1) using default parameters. For nucleotide profile HMMs first all protein sequences were back-translated into nucleotide sequences using backtranseq module of EMBOSS. Afterward for building of ARGhmm library the similar procedure follows again which I used with protein sequences. Each HMM model is manually annotated and curated with its different properties namely, information and descriptions about the gene/protein name and their corresponding families; where particular gene/protein belongs, associated enzyme commission (EC) numbers, organism name, interactions and pathway information related with specific gene/protein, 3D structural information of each gene/protein, and their subcellular location, functions, resistance mechanism and gene ontology (cellular component, molecular function & biological process) as well as their UniProt accession number. Scanning of AR genes using ARGhmm library (254 hmm profiles) nucleotide or protein profile HMMs against gene or protein sequence using nhmmscan or hmmscan module of HMMER respectively. Upon search it gives the alignment score and E-value of each query sequence, additionally user can retrieved complete annotation of each ARGhmms.