[SVII-3]

Identification and Application of Conserved Genes in Bacterial Genomes

Cheolmin Kim

Department of Biochemistry, College of Medicine, Pusan National University

The reemergence of bacterial infections and antibiotics resistance make it clinically important to identify quickly the pathogenic bacteria and drug resistance. Also development of new antibiotics for resistant strains of pathogens is another major concern in clinical microbiology. At first we developed diagnostic PCR methods and a oligonucleotide chip (oligo-chip) for the genotyping of *Mycobacteria* using DNA sequences common in genus *Mycobacteria.* And then we developed supportive bioinformatics system for high throughput development of universal genotyping system for human pathogenic bacteria. Recently we retrieved and analyzed conserved genes in bacteria for development of new targets for the diagnostic genotyping system and for new candidates of antibiotics targets.

The oligo-chip technology has potential to be a very useful tool in rapid and accurate genotyping of bacterial species, and in the detection of mutations related with drug resistance. In development of the oligo-chip for the genotyping of *Mycobacteria*, the probes were designed from internal transcribed spacer (ITS) sequences of *Mycobacteria* species(Fig 1). The probes consists of genus-specific and species-specific ITS sequences and sequences related with drug resistance. For developing the oligo-chip, we used basic bioinformatics tools such as sequence retrieval, sequence edition, multiple sequence analysis, homology search, primer/probe design, etc.

Though some bioinformatics tools are available in public domain services or as freely distributed softwares, the pipeline construction of analysis tools is the essential step for implementation of high-throughput development. We developed a bioinformatics pipeline to support the high-throughput development of oligo-chip. The system consists of sequence database, probe search module, and oligo-chip management module. All public bacterial whole genome data were downloaded from NCBI. Their features and sequences are parsed from GenBank flat-file to the RDBMS MySQL running on Linux. The SQL query is used to find target sequences in the database and, then, the sequences are manually put into the probe search module by using command line interface. The probe search module written in BioPerl finds highly similar and polymorphic regions from the multiple sequence analysis results. Genus-specific probes were retrieved from conserved regions and species-specific probes from the highly polymorphic regions. The BLAST, then, verifies the possibility coexistence of the similar sequences in another gene or another organism. The final sequences are used as probes of oligo-chip. The oligo-chip management module translates the chip image data to medical report for doctors. The module has connectivity to ODBC for interfacing medical information retrieval system.

At third step, we tried to find new target for development of universal diagnostic genotyping system for

pathogenic human bacteria. Conserved proteins in bacteria were retrieved from the orthologous gene list of COGs (Clusters of Orthologous Groups of proteins) in NCBI COG DB. And then amino acid sequences of the conserved orthologs from each bacteria were analyzed using a multiple alignment tool, CLUSTALW. The bacterial species were listed according to the conservativeness calculated by distance values. The GTPase-translation elongation factors (COG0050, tufB gene) was the most conserved protein according to the analysis using evolutional distance calculation. Nucleotide sequences of the tufB genes were analyzed to find conserved sequences and polymorphic sequences for development of new primers and probes for the diagnostic oligo-chip

We retrived the conserved proteins in gram positive and / or gram negative human pathogenic bacteria to find new candidates of targets for antibiotics. And we are developing a systemic analysis tool for comparative genomic analysis of microorganism. The system using supportive bioinformatics and comparative genomics shows the potential of supporting high throughput development of diagnostic oligo-chips and new candidates of antibiotics targets.