Dr.
Ikuo Uchiyama

National Institute for Basic Biology,
Nishigonaka 38, Myodaiji,
Okazaki 444-8585, Japan
Tel: +81-564-55-7629
Fax: +81-564-55-7625
E-Mail: uchiyama [at] nibb.ac.jp

Toward a systematic approach to microbial genome comparison: identification of genomic core among moderately related genomes

Recent accumulation of genomic sequences enables us to investigate diversity and evolution of microorganisms living in various habitats. To facilitate systematic comparison of hundreds of microbial genomes, we have been developing a microbial genome database named MBGD, in which genes in published microbial genomes are exhaustively classified into orthologous groups (OGs). However, in prokaryotic genome comparison, effectiveness of orthology analysis as well as general phylogenetic analysis seems to be limited due to the existence of horizontal transfers (HGT). In fact, the genes constituting a prokaryotic genome appear to be divided into two classes: a "core gene pool" that comprises intrinsic genes encoding the proteins of basic cellular functions, and a "flexible gene pool" that comprises HGT-acquired genes encoding proteins which function under particular conditions, such as genomic islands. Therefore, the identification of the set of intrinsically conserved genes, or the genomic core, among a taxonomic group is crucial not only for establishing the identity of each taxonomic group, but also for understanding prokaryotic diversity and evolution.

Here, we consider the core structure of related genomes as a set of sufficiently long segments in which gene orders are conserved among multiple genomes so that they are likely to have been inherited mainly through vertical transfer. To identify such structure, we developed a method for aligning conserved regions of multiple genomes, which finds the order of pre-identified OGs that retains to the greatest possible extent the conserved gene orders. The method was applied to genome comparisons of two well-characterized families, *Bacillaceae* and *Enterobacteriaceae*, and identified their core structures comprising 1438 and 2125 OGs, respectively, which correspond to a third of the number of the *B. subtilis* genes and half of the *E. coli* genes, respectively. The core sets contained most of the essential genes (>90%) and their related genes, which were primarily included in the intersection of the two core sets comprising around 700 OGs. We also investigated the core structures in terms of G+C content homogeneity and phylogenetic congruence, and found that the core genes primarily exhibited the expected characteristic, i.e., being indigenous and sharing the same history, more than the non-core genes.

The program, named CoreAligner, is available as a part of the RECOG system, a general workbench for comparative genomics. On this system, users can connect to the MBGD database, choose an appropriate set of genomes to compare, identify OGs by the DomClust program, and then construct a core genome alignment by the CoreAligner program.

CV

1990	B. Sc in biology, Kyoto University
1992	M. Sc in biophysics, Kyoto University
2007	Ph. D. in bioinformatics, University of Tokyo
1992-1995	Research associate in Kyoto University
1995-1999	Research associate in University of Tokyo
1999-	Research associate/Assistant professor in National Institute for Basic Biology