S5-3

Comparative Genomics of Vertebrate Hox Gene

KIM Chang-Bae
National Genome Information Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon, Korea

Recent development of powerful genome technologies capable of producing sequence data in large quantities from diverse organisms is empowered mainly by Human Genome Project. In the postgenome era, genome structure and function analyses can be strengthened through examining evolutionary history. The natural partnership between the co-evolving disciplines of genomics and evolution is now emerging in the form of comparative genomics and phylogenomics. In this study, I show how two separate principles make possible new ways of comparative analyses of vertebrate Hox clusters as an example for the new trend. This analysis aims to examine relations of different Hox genomes in vertebrates and identify regulatory regions of Hox genes. This approach will improve the understanding how the architecture of development contributes to evolvability with accurate estimates of changes in evolutionary patterns and selective constraints. In addition this will contribute to understand the relationsbetween Hox gene duplication and body plan complexity. Finally, a new phylogenetic way to detect enhancer elements of the Hoxc8 gene that were characterized wellby previous biochemical and genetic studies is presented.

S5-4

ECgene: Genome-based EST Clustering and Gene Modeling for Alternative Splicing

KIM Namshin[1,2], SHIN Seokmin[2], and LEE Sanghyuk[1]
[1] Division of Molecular Life Sciences, Ewha Womans University, Seoul, Korea
[2] School of Chemistry, Seoul National University, Seoul, Korea

EST clustering has played a central role in finding and characterizing unknown genes. It gives ample insight into gene structure, expression, and gene function. Most EST clustering algorithms are based on pairwise comparison of EST sequences. We developed a new method of EST clustering based on the recently completed human genome map.

Brief outline of the algorithm is as follows. mRNA and ESTs in the UniGene are aligned against the genome map using BLAT and SIM4 programs. Sequences that share exon-intron splice sites are clustered initially, which produces the 'primary EST clusters'. The genomic exon alignments are analyzed in graph-theoretic fashion to produce possible gene models of alternative splicing events. Singleton ESTs are added at later stage.

ECgene (gene by EST Clustering, available at http://genome.ewha.ac.kr/ECgene) has significant advantages over conventional pairwise EST clustering. First of all, since the genomic al ignment is precisely known, ample information from public genome annotation - for example, promoter, SNP, expression data, repeats, conserved regions between species - is readily available (via the genome browser). Secondly, identification of alternative splicing is naturally incorporated in the algorithm. ECgene models are currently being analyzed for gene expression and function with emphasis on alternative splicing. We also plan to expand ECgene for mutiple species in the near future.