30. Aligning intron-exon structures of alternatively spliced genes

Michael Sammeth [1,2], Eduardo Eyras[1] and Roderic Guigo[1]

1 Genome Research Laboratory (GRIB/IMIM/UPF/CRG)
2 Genome Informatics, Bielefeld University, Germany

With the ever growing availability of gene annotations an evolutive analysis of alternative splicing events across different species has become possible. However, accurate alignments of intron-exon structures are crucial for correct analyzes. We investigate multiple alignment algorithms applied to evolutionary related alternatively spliced genes in human, mouse and rat.

The availability of the genome sequence for multiple species has made possible comparative studies, which have helped to obtain a better understanding of the biology and evolution of the human and other genomes. Additionally, annotations for these genomes have been systematically generated and made publicly available. Consequently, we now have for the first time a lot more information about the intron-exon structures of genes across multiples species, which opens up some new approaches for comparative analysis. The intron-exon structure of a gene is an attribute as essential as the protein encodes in order to understand the regulation of the gene. However, while we know much about how proteins evolve, little is yet known about how this exon-intron structure of the gene evolves.

One major aspect regarding the evolution of intron-exon structures is the biological phenomenon of alternative splicing (AS). In several experimental studies, AS has been shown to be a crucial process for regulating gene expression. It enables genes to adapt the expression levels of variants to the corresponding cellular context, developmental stage or state of disease. Moreover, recent analyzes of the complete human genome sequence have revealed that AS is the most important diversity generator, creating the contrast between the molecular complexity of the cell and the apparently low number of genes.

In order to discover and investigate the mechanisms involved in the evolution of AS events, accurate sequence alignments of orthologous genes across various species are needed. However, several obstacles, as large intronic areas with low sequence similarity, are to be circumvented before we can generate optimal gene alignments. Another problem arises from the fact that all currently used alignment programs that are dealing with sequences of realistic sizes use heuristical shortcuts, where the need for heuristics stems from the fact that the multiple alignment problem is known to be NP-complete.

We aim to understand whether the exon-intron structures tell us anything that sequence alignments based on simple character comparison cannot. To this end, we compare large-scale sequence alignments of genes across different species. In particular, we focus on the comparison of orthologous human, mouse and rat genes where the AS pattern has changed during evolution. We call the AS observed in the corresponding genes an 'evolutionary alternative splicing event' (EAS), since its pattern evolved from one to the other species. EAS genes are aligned using character-based comparison alignment methods (Muscle, T-Coffee and Dialign) as well as methods that take into account exon-

intron structures (M-Lagan, Mavid, T-Coffee) and the achieved results are compared to each other.