Analysis of Alternative Splicing and Alternative Transcriptional Initiation Pattern: Diverse Mechanism of Gene Expression

Hideki Nagasaki
AIST Research Staff
Algorithm Team

The genome sequencing of various organisms have been finished in recent years, and the total number of genes of each organism has turned up. As to human, the number of genes has calculated to be approx. 25,000, which was much smaller than earlier estimates. Therefore scientists' interests get on to the structure of transcriptional products and diverse manner of gene expression that are influenced by alternative splicing (AS) and alternative transcriptional initiation (ATI). The consensus sequences existing around splice sites of AS, and in promoter/enhancer region of ATI, have been the target to search in a mathematical statistical analysis for the last decade. Some base sequences common to particular species and genes have been elucidated by these efforts; however, it is still short to draw a complete picture. Accumulating sequence data, higher performance computer system, and the appearance of revolutionary algorithms may contribute to solve these problems. It will be, however, still challenging to elucidate the complexity of AS and ATI (hereinafter ASTI) even with such progress.

We elucidated the exon-intron structure corresponding to each transcripts by comparing genome and cDNA sequences for six species (humans, mammals, insects, nematoda and plants). We have further detected ASTI by comparing these exon-intron structures derived from respective gene. The exon-intron structure is partially modified by ASTI, generally showing complex patterns. We have developed an algorithm that converts the exon-intron structures to binary (0/1) description, and detects ASTI patterns, classifies distinct types (e.g. cassette type) (Fig.1).[*1] The ratio of basic ASTI types detected by this method showed characteristic features specific to each species: an especially significant difference was detected between animals and plants.[*2] This is considered that it reflects the difference of the ASTI production mechanism and its role in animals and plants. Future efforts will be made focusing more on detailed analysis of individual species and the comparison between closely related species. We hope that our approach - classifying diverse ASTI patterns - will help accelerate the analysis of common sequences, responsive for the complex mechanism of ASTI events.

These results from ASTI pattern analysis are open to public through ASTRA (Alternative Splicing and TRanscription Archives, (Fig.2), a database equipped JAVA applet viewer capable of displaying the most complex patterns (http://alterna.cbrc.jp).[*1] ASTI manifests the complexity of lives, where the transcription and expression of eukaryote genes diversifies the structure of transcription product. Following genome analysis, the focus will be shifted to transcriptome, proteome, or individual ASTI gene analysis. We view ASTRA as a useful tool that provides a solid foundation for these studies.

References

*1 Nagasaki, H., Arita, M., Nishizawa, T., Suwa, M., Gotoh, O.: "Automated classification of alternative splicing and transcriptional initiation and construction of visual database of classified patterns", *BIOINFORMATICS*. (accepted)

*2 Nagasaki, H., Arita, M., Nishizawa, T., Suwa, M., Gotoh, O.: "Species-specific variation of alternative splicing andtranscriptional initiation in six eukaryotes", *GENE*, **364**, pp.53-62 (2005).

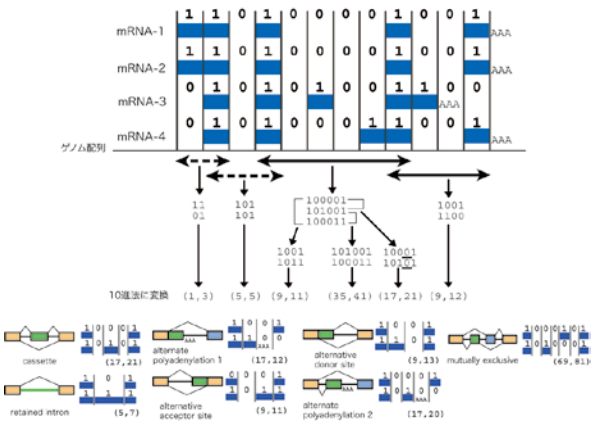

Fig.1 A sketch of the algorithm to detect the patterns of Alternative Splicing and Alternative Transcriptional Initiation (ASTI)

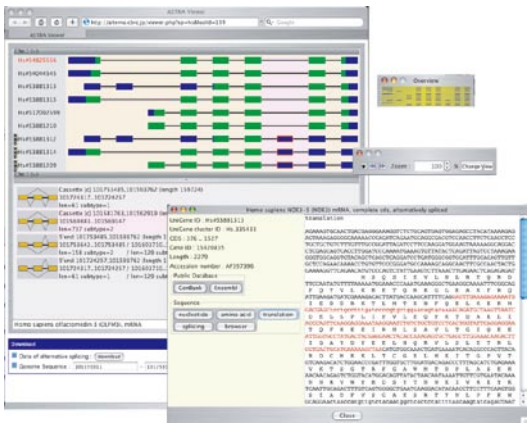

Fig.2 The viewer of "ASTRA", a database of ASTI patterns (http://www.alterna.cbrc.jp)