TATA-variant identification, characterization and functional classification in plant genomes

Virginie BERNARD, Véronique BRUNAUD, Alain LECHARNY

URGV, UMR INRA 1165 -CNRS 8114 -UEVE, 2 rue Gaston Crémieux,
91057 Evry Cedex, France
Université Paris-Sud, IBP, UMR CNRS 8618 -UPS, bâtiment 630,
91405 Orsay Cedex, France

The TATA-box is a regulatory element involved in the Transcription Initiation Complex (TIC) formation and is conserved through the evolution at a strict location in promoters. Some TATA-variants are also TIC binding sequences. Nevertheless, few studies have been dedicated to their sequences and functions. We assume that the more TATA-variants share features with the TATA-box, the more they may be recently diverged motifs. We propose to identify, characterize and compare TATAvariants, in order to investigate their putative evolutionary links with the canonical TATA-box.

In plants, a description at the genomic level of the promoter architecture is possible for both Arabidopsis thaliana and Oryza sativa due to the large availability of transcripts. We developed an ab-initio approach using the preferential location of motifs to identify biologically relevant regulatory elements (ftp://urgv.evry.inra.fr/Publications/BernardV_et_al_JOBIM_5to7juli2006_Bordeaux _2006_17-28.pdf).

We identified TATA-variants conserved between Arabidopsis thaliana and Oryza sativa and exhibiting the TATA-box topological constraints. Only some of the possible TATA-box substitutions are observed in the variants and thus have been conserved. A third of the plant genes contain a TATA-box or a TATA-variant. Feature analysis of gene sub-sets containing one of these motifs led to a functional TATA-variant classification. While some containing-variant gene sets are highly divergent in expression, function, gene and promoter structures from the TATA-box containing genes, other are close.
Our results might be indicative of the TATA-box expected region adaptation. With few mutations the region could promote rapid changes in expression of functionally diversified duplicate genes. We plan to combine this analysis to a global proximal promoter analysis in order to investigate the question of a functional link between regulatory elements specifically organized in promoters.