Transposable element products, functions, and regulatory networks in Arabidopsis thaliana

Carles Borredá*[1], Basile Leduque[1], Vincent Colot[2], and Leandro Quadrana[1]

[1]Institute of Plant Sciences Paris-Saclay (IPS2) – Université d'Évry-Val-d'Essonne, Université Paris-Saclay, Centre National de la Recherche Scientifique, Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement – Institute of Plant Sciences - Paris-Saclay Bâtiment 630, rue de Noetzlin 91190 - Gif-sur-Yvette, France

[2]Institut de biologie de l'ENS Paris – Département de Biologie - ENS Paris, Institut National de la Santé et de la Recherche Médicale, Centre National de la Recherche Scientifique – 46, Rue d'Ulm75005 Paris, France

Abstract

Transposable element (TE) mobilization is primarily catalyzed by self-encoded factors, yet these factors have been poorly investigated due to difficulties in defining TE genes in genomes. Here we leveraged extensive long- and short-read transcriptome data, together with structural predictions, to build a comprehensive atlas of TE transcripts and TE-encoded products in the model organism Arabidopsis thaliana. We uncovered hundreds of transcriptionally competent TEs, each of which potentially encoding multiple proteins either through distinct genes, alternatively splicing, or post-translational processing of polyproteins. Structural-based protein analyses revealed dozens of hitherto anonymous domains of unknown functions, and allowed us to build a comprehensive guide of TE-encoded products in Arabidopsis thaliana. Some of these domains displayed significant structural similarity with domains from cellular proteins, suggesting protein co-option or convergent evolution. Furthermore, we identified potential DNA binding and multimerization domains involved in the formation of macromolecular complexes such as transpososomes. In addition, using transcription factor binding site identification and large-scale transcriptome data we demonstrate that TE expression is highly intertwined with the transcriptional network of cellular genes, and identify transcription factors and cis-regulatory elements associated with their coordinated expression during development or in response to environmental cues. This comprehensive atlas of TE-genes and TE-proteins provides a valuable resource for studying the mechanisms involved in the TE-driven evolution of genome function.

Keywords: long read, transcriptomics, functional annotation, protein structure

*Speaker