Program and Abstracts

November 9 – 13, 2019

48

Hidden diversity of endogenous geminiviral sequences across plant genomes and transcriptomes

Vikas Sharma[1], Philippe Roumagnac[2], Denis Filloux[2], Pierre Lefeuvre[3], Pierre Yves Teycheney[4], Darren P. Martin[5], Florian Maumus[1]

[1]URGI, INRA, Versailles, France. [2]UMR BGPI, Cirad, Montpellier, France. [3]UMR PVBMT, Cirad, Saint Pierre de la Réunion, France. [4]UMR AGAP, Cirad, Capesterre Belle-Eau, France. [5]University of Cape Town, Cape Town, South Africa

Abstract

Endogenous viral elements (EVE) can be used as 'fossil records' to reveal the genomic features of long extinct virus species. Although numerous known instances exist of single-stranded DNA (ssDNA) genomes becoming stably integrated within the genomes of bacteria and animals, there remain very few examples of such integration events in plants. Most of the EVEs that have been characterized so far belong to family *Caulimoviridae*. However the first plant EVEs to be discovered were geminivirus derived sequences in the nuclear genomes of various *Nicotiana* species. Since then, endogenous geminivirus-like elements (EGV) have also been identified in the genomes of several plants, including yam (several *Dioscorea* species), apple (*Malus domestica*), lettuce (*Lactuca sativa*), cottonwood (*Populus trichocarpa*) and coffee (*Coffea canephora*). We therefore search for evidence of EGVs within 134 plant genome sequences and 797 plant transcriptome sequences. We detected homologues of geminivirus replication-associated protein (*rep*) genes from 17 genomes and 39 transcriptomes from angiosperms. Copy numbers of EGVs within these genomes varied widely with the highest copy numbers, approximately 1000, being found in two varieties of tea (*Camellia sinensis*). Phylogenetic and similarity-based analyses revealed multiple taxonomically novel geminivirus lineages, including two in *Camellia* species which might represent novel genera. We found that some of the *Camellia* and *Dioscorea* EGVs are transcriptionally active, and display evidence of purifying selection, suggesting that expressed geminivirus proteins were, and may still be, functionally active in certain host plants. Collectively our analysis expands the known breadth of past geminivirus diversity, provides a first large-scale view of EGV prevalence, and strengthens support for the hypothesis that EGVs impact the biology of their hosts.