

Endogenous Dionyvirus sequences are widespread in plant genomes

Andrew D.W. Geering¹, Nathalie Choisne², Simone Scalabrin³, Matthias Zytnicki², Silvia Vezzulli⁴, Riccardo Velasco⁴, Hadi Quesneville² & Pierre-Yves Teycheney⁵

¹ Queensland Alliance for Agriculture and Food Innovation (QAAFI), Ecosciences Precinct, 41 Boggo Road, Dutton Park Qld 4102 Australia

² URGI, INRA Versailles, Route de Saint Cyr, 78026 Versailles Cedex, France

³ Istituto di Genomica Applicata, Parco Scientifico e Tecnologico di Udine Luigi Danieli, Via J Linussio 51, 33100 Udine, Italy.

⁴ IASMA Research and Innovation Centre, Fondazione Edmund Mach, Via E. Mach 1, 38010 San Michele a/Adige (TN), Italy

⁵ CIRAD-Bios, UPR75, Station de Neufchâteau, Sainte-Marie, F-97130 Capesterre Belle-Eau, Guadeloupe, France

andrew.geering@deedi.qld.gov.au; teycheney@cirad.fr

Endogenous sequences from members of the *Caulimoviridae* and *Geminiviridae* families have been identified in the genome of several plant species. They are thought to result from illegitimate recombination events and are generally replication-defective. However, some caulimovirid sequences of tobacco, petunia and banana are capable of causing infection.

We performed *in silico* analyses on nucleotide sequences from plant genome databases and reconstituted 12 full length and potentially infectious viral genomes from endogenous viral sequences embedded in the genomes of monocotyledonous and dicotyledonous plant species. Sequence comparisons show that the corresponding viruses belong to a new genus in the family *Caulimoviridae*, tentatively named Dionyvirus. Mapping of endogenous Dionyvirus sequences was achieved in the fully sequenced genomes of grape, poplar, peach and rice, allowing for the first time the study of the distribution pattern of endogenous viral sequences at the host plant genome scale. Using the TEannot pipeline from the REPET package, we sensitively detected virus fragments and were able to join them to recover fragmented virus sequences. Full length and partial virus sequences were found in all four genomes, evenly distributed along chromosomes; they appear to be repeated throughout the whole genome.

Dionyvirus-specific primers were designed and used for a PCR-based large scale screening of plant germplasm. It showed that endogenous Dionyvirus sequences belonging to distinct viral species are widespread among plants of temperate, tropical and arctic origins, and that endogenization of viral sequences is therefore a common phenomenon in plants. siRNAs homologous to endogenous Dionyvirus species were identified in host plants and mapped along reconstituted viral genomes, providing evidence that the expression of such endogenous sequences is tightly regulated and/or might trigger RNAi-based antiviral defence.

Keywords: endogenous pararetrovirus; dionyvirus; mapping; screening