

### The genomes of several plant species contain endogenous geminiviral sequences

Denis Filloux, Sasha Murrell, Maneerat Koohapitagtam, Michael Golden, Charlotte Julian, Serge Galzi, Marilyne Uzest, Marguerite Rodier-Goud, Angélique d'Hont, Marie-Stéphanie Vernerey, et al.

#### ▶ To cite this version:

Denis Filloux, Sasha Murrell, Maneerat Koohapitagtam, Michael Golden, Charlotte Julian, et al.. The genomes of several plant species contain endogenous geminiviral sequences. 8. International Geminivirus Symposium, Nov 2016, New Delhi, India. 1 p., 2016. hal-02798133

#### HAL Id: hal-02798133 https://hal.inrae.fr/hal-02798133

Submitted on 5 Jun2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The genomes of several plant species contain endogenous

# geminiviral sequences





Denis Filloux,<sup>1</sup> Sasha Murrell,<sup>2,3</sup> Maneerat Koohapitagtam,<sup>1,4</sup> Michael Golden,<sup>2</sup> Charlotte Julian,<sup>1</sup> Serge Galzi,<sup>1</sup> Marilyne Uzest,<sup>1</sup> Marguerite Rodier-Goud,<sup>5</sup> Angélique D'Hont,<sup>5</sup> Marie-Stéphanie Vernerey,<sup>1</sup> Paul Wilkin,<sup>6</sup> Michel Peterschmitt,<sup>1</sup> Stephan Winter,<sup>7</sup> Ben Murrell,<sup>2,8</sup> Darren P. Martin,<sup>2</sup> and Philippe Roumagnac<sup>1</sup>

2 sequenc

<sup>1</sup> CIRAD-INRA-SupAgro, UMR BGPI, Campus International de Montferrier-Baillarguet, 34398 Montpellier Cedex-5, France. Email: filloux@cirad.fr
<sup>2</sup> Computational Biology Group, Institute of Infectious Diseases and Molecular Medicine, University of Cape Town, Cape Town 4579, South Africa.
<sup>3</sup> Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA 92037, USA.
<sup>4</sup> Department of Pest Management, Faculty of Natural Resources, Prince of Songkla University, Hat Yai campus, Thailand 90120.
<sup>5</sup> CIRAD, UMR AGAP, TA A-108/03, Avenue Agropolis, F-34398 Montpellier Cedex 5, France.
<sup>6</sup> Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AB, UK.
<sup>7</sup> DSMZ Plant Virus Department, Messeweg 11/12, 38102, Braunschweig, Germany.
<sup>8</sup> Department of Medicine, University of California, San Diego, La Jolla, CA.

Endogenous viral sequences are essentially 'fossil records' that can sometimes 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. The best studied of these events are those which yielded the geminivirus-related DNA elements (GRD) and the geminivirus-like elements (EGV) found respectively within the nuclear genomes of several *Nicotiana* species (Kenton *et al.* 1995; Bejarano *et al.* 1996; Ashby *et al.* 1997; Murad *et al.* 2004) and various *Dioscorea* spp. of the *Enantiophyllum* clade (Filloux *et al.*, 2015).

Those two new classes of endogenous plant virus sequence are apparently derived from ancient geminiviruses in the genus *Begomovirus*. GRD and EGV sequences likely became integrated millions years ago. Interestingly, we found evidence of natural selection actively favouring the maintenance of EGV-expressed replication-associated protein (Rep) amino acid sequences, which clearly indicates that functional EGV Rep proteins were probably expressed for prolonged periods following endogenization.

# The genomes of many yam (*Dioscorea* spp.) species contain transcriptionally active endogenous geminiviral sequences that may be functionally expressed





Characteristics of the EGV2 sequence: only *rep* (absence of *cp*, *ren* and *rep*)

Species	No.	Section	Clade	Country	Origin of the species	EGV1 sequence	EGV
D. wallichii Hook.f.	14472	Enantiophyllum	Enantiophyllum	Thailand	South East Asia	+	+
D. inopinata Prain & Burkill	15674	Enantiophyllum	Enantiophyllum	Thailand	South East Asia	+	+
D. oryzetorum Prain & Burkill	15671	Enantiophyllum	Enantiophyllum	Thailand	South East Asia	+	+
D. alata L.	313	Enantiophyllum	Enantiophyllum	India	South East Asia	+	+
D. persimilis Prain & Burkill <sup>a</sup>	271	Enantiophyllum	Enantiophyllum	Vietnam	South East Asia	+	+
D. nummularia Lam.	206	Enantiophyllum	Enantiophyllum	Vanuatu	Melanesia	+	+
D. transversa R.Br.	336	Enantiophyllum	Enantiophyllum	Vanuatu	Melanesia	+	+
D. glabra Roxb.	21051	Enantiophyllum	Enantiophyllum	Thailand	South East Asia	+	+
D. calcicola Prain & Burkill	6215	Enantiophyllum	Enantiophyllum	Thailand	South East Asia	+	+
D. hamiltonii Hook.f.ª	6210	Enantiophyllum	Enantiophyllum	Thailand	South East Asia	+	+
D. brevipetiolata Prain & Burkill	14475	Enantiophyllum	Enantiophyllum	Thailand	South East Asia	+	+
D. opposita Thunb.	265	Enantiophyllum	Enantiophyllum	France	China	+	+
D. decipiens Hook.f.	6481	Enantiophyllum	Enantiophyllum	Thailand	South East Asia	+	+
D. cirrhosa Lour.	15672	Enantiophyllum	Enantiophyllum	Thailand	South East Asia	+	+
D. lanata Bail	6181	Enantiophyllum	Enantiophyllum	Socotra	Arabian Peninsula	_	_
D. abyssinica Hochst. ex Kunth	109	Enantiophyllum	Enantiophyllum	Benin	West Africa	+	+
D. cayenensis Lam.	1	Enantiophyllum	Enantiophyllum	Haïti	West Africa	+	+
D. praehensilis Benth.	255	Enantiophyllum	Enantiophyllum	Benin	West Africa	+	+
D. rotundata Poir.	389	Enantiophyllum	Enantiophyllum	Benin	West Africa	+	+
D. schimperiana Hochst. ex Kunth	21044	Enantiophyllum	Enantiophyllum	Malawi	East Africa	+	_
D. schimperiana Hochst. ex Kunth	22295	Enantiophyllum	Enantiophyllum	Malawi	East Africa	+	_
D. togoensis R.Knuth	114	Enantiophyllum	Enantiophyllum	Guinea	West Africa	_	+
D. minutiflora Engl.	031	Enantiophyllum	Enantiophyllum	Madagascar	West Africa	+	+
D. bulbifera L.	272	Opsophyton	Compound leafed	Papua New Guinea	South East Asia	_	_
D. dumetorum (Kunth) Pax	67	Lasiophyton	Compound Leafed	Burkina Faso	West Africa	_	_
D. pentaphylla L.	Dp1038	Lasiophyton	Compound-Leafed	?	?	_	_
D. maciba Jum. & H.Perrier	14348	?	Malagasy	Madagascar	Madagascar	_	_
D. sansibarensis Pax	269	Macroura	Malagasy	Benin	West Africa	_	_
D. birmanica Prain & Burkill	15677	Stenophora	Birmanica	Thailand	South East Asia	_	_
D. esculenta (Lour.) Burkill	002	Combilium	Birmanica	Madagascar	South East Asia	_	_
D. buchananii Benth.	15073	Rhacodophyllum	Africa	Zambia	Southern Africa	_	_
D. elephantipes (L'Hér.) Engl.	328	Testudinaria	Africa	France	Southern Africa	_	_
D. communis (L.) Caddick & Wilkin	310	?	Europe	France	Europe	_	_
D. amaranthoides C.Presl	16523	Strutantha	New World	Bolivia	South America	_	_
D. galeottiana Kunth	6204	?	New World	Mexico	North America	_	_
D. membranacea Pierre ex Prain & Burkill	21050	?	Stenophora	Thailand	South East Asia	_	_
D. balcanica Koanin	266	Stenophora	Stenophora	France	Europe	_	_
D. villosa L.	267	Stenophora	Stenophora	France	North America	_	_
D. trifida L.f.	78	Macrogynodium	Macrogynodium	French Guyana	South America	_	_
D. melastomatifolia Uline ex Prain	368	?	?	French Guyana	South America	_	_
D. pubescens Poir.	367	2	2	French Guyana	South America	_	_

Distribution of EGV1 and EGV2 sequences among members of the *Dioscoreacea* family





siRNA mapping on EGV1 sequence suggests that rep and



Western blot on the total protein extracts of several *Dioscorea* species using an antibody directed to Rep peptide of EGV1



Maximum likelihood tree describing the evolutionary relationships between clades of *Dioscoreacea* based on *rbcL* and *matK* nucleotide sequences





### New D. alata sequence resources available to further explore yam EGVs

ren genes are transcriptionally

Draft genome of D. alata now available on GenBank (CZHE0200000)

active

Four D. alata BAC clones containing EGVs are now sequenced (CNRGV, INRA, Toulouse, France)

### Several other plant species contains EGVs

We recently found using *in silico* searches that other EGVs are included within complete or draft genomes of various plant species, including apple tree (*Malus domestica*), black cottonwood (*Populus trichocarpa*), several *Coffea* spp, eggplant (*Solanum melongena*), lettuce (*lactuca sativa*), and Tepary bean (*Phaseolus acutifolius*), which suggests that endogenous geminiviruses may be more common in plant genomes than has previously been appreciated.

## Phylogenetic relationships between EGV1, EGV2 and the geminiviruses

substitutions

### References

Ashby, M. K., Warry, A., Bejarano, E. R., Khashoggi, A., Burrell, M., Lichtenstein, C. P. (1997). Analysis of multiple copies of geminiviral DNA in the genome of four closely related Nicotiana species suggest a unique integration event. Plant molecular biology, 35(3), 313-321. Bejarano, E. R., Khashoggi, A., Witty, M., Lichtenstein, C. P. (1996). Integration of multiple repeats of geminiviral DNA into the nuclear genome of tobacco during evolution. Proceedings of the National Academy of Sciences, 93(2), 759-764. Filloux, D., Murrell, S., Koohapitagtam, M., Golden, M., Julian, C., Galzi, S., Uzest, M., Rodier-Goud, M., D'Hont, A., Vernerey, M.S., Wilkin, P., Peterschmitt, M., Winter, S., Murrell, B., Martin, D.P., Roumagnac, P., 2015. The genomes of many yam species contain transcriptionally active endogenous geminiviral sequences that may be functionally expressed. Virus Evol. 1, 1–17. Kenton, A., Khashoggi, A., Parokonny, A., Bennett, M. D., Lichtenstein, C. P. (1995). Chromosomal location of endogenous geminivirus-related DNA sequences in Nicotiana tabacum L. Chromosome Research, 3(6), 346-350. Murad, L., Bielawski, J. P., Matyasek, R., Kovarik, A., Nichols, R. A., Leitch, A. R., Lichtenstein, C. P. (2004). The origin and evolution of geminivirus-related DNA sequences in Nicotiana. Heredity, 92(4), 352-358.







