Virology 493 (2016) 142–153

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

Molecular characterization and prevalence of two capulaviruses: Alfalfa leaf curl virus from France and Euphorbia caput-medusae latent virus from South Africa

Pauline Bernardo [a,1], Brejnev Muhire [b], Sarah François [a,c,d], Maëlle Deshoux [a], Penelope Hartnady [b], Kata Farkas [e], Simona Kraberger [e], Denis Filloux [a], Emmanuel Fernandez [a], Serge Galzi [a], Romain Ferdinand [a], Martine Granier [a], Armelle Marais [f,g], Pablo Monge Blasco [h], Thierry Candresse [f,g], Fernando Escriu [h,i], Arvind Varsani [e,j,k], Gordon W. Harkins [l], Darren P. Martin [b], Philippe Roumagnac [a,*]

[a] CIRAD-INRA-SupAgro, UMR BGPI, Campus International de Montferrier-Baillarguet, Montpellier Cedex-5, France
[b] Computational Biology Group, Institute of Infectious Disease and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Observatory, South Africa
[c] INRA, UMR 1333, DGIMI, Montpellier, France
[d] CNRS-IRD-UM1-UM2, UMR 5290, MIVEGEC, Avenue Agropolis, Montpellier, France
[e] School of Biological Sciences and Biomolecular Interaction Centre, University of Canterbury, Private Bag 4800, Christchurch, New Zealand
[f] INRA, UMR 1332 Biologie du Fruit et Pathologie, Villenave d'Ornon Cedex, France
[g] Université de Bordeaux, UMR 1332 Biologie du Fruit et Pathologie, Villenave d'Ornon Cedex, France
[h] Unidad de Sanidad Vegetal, Centro de Investigacion y Tecnologia Agroalimentaria de Aragon (CITA), Av. Montañana 930, 50059 Zaragoza, Spain
[i] Unidad de Sanidad Vegetal, Instituto Agroalimentario de Aragón IA2 (CITA - Universidad de Zaragoza), Av. Montañana 930, 50059 Zaragoza, Spain
[j] Department of Plant Pathology and Emerging Pathogens Institute, University of Florida, Gainesville, USA
[k] Structural Biology Research Unit, Department of Clinical Laboratory Sciences, University of Cape Town, Observatory, South Africa
[l] South African National Bioinformatics Institute, MRC Unit for Bioinformatics Capacity Development, University of the Western Cape, Cape Town, South Africa

ARTICLE INFO

Article history:
Received 20 December 2015
Returned to author for revisions
22 March 2016
Accepted 23 March 2016
Available online 31 March 2016

Keywords:
Geminiviridae
Alfalfa leaf curl virus
Euphorbia caput-medusae latent virus
Prevalence
Recombination
Secondary structure
Genome organization
France
Spain
South Africa

ABSTRACT

Little is known about the prevalence, diversity, evolutionary processes, genomic structures and population dynamics of viruses in the divergent geminivirus lineage known as the capulaviruses. We determined and analyzed full genome sequences of 13 Euphorbia caput-medusae latent virus (EcmLV) and 26 Alfalfa leaf curl virus (ALCV) isolates, and partial genome sequences of 23 EcmLV and 37 ALCV isolates. While EcmLV was asymptomatic in uncultivated southern African *Euphorbia caput-medusae*, severe alfalfa disease symptoms were associated with ALCV in southern France. The prevalence of both viruses exceeded 10% in their respective hosts. Besides using patterns of detectable negative selection to identify ORFs that are probably functionally expressed, we show that ALCV and EcmLV both display evidence of inter-species recombination and biologically functional genomic secondary structures. Finally, we show that whereas the EcmLV populations likely experience restricted geographical dispersion, ALCV is probably freely moving across the French Mediterranean region.

1. Introduction

Next Generation Sequencing and metagenomics-based study designs have impacted our appreciation of the prevalence, pervasiveness and diversity of environmental single stranded DNA (ssDNA) viruses (Candresse et al., 2014; Filloux et al., 2015b; Kraberger et al., 2015; Ng et al., 2014, 2009, 2011; Roossinck et al., 2015; Rosario and Breitbart, 2011). Among the best studied of these ssDNA viruses have been the plant-infecting viruses in the family *Geminiviridae*. The past four decades have seen the worldwide emergence of several major plant diseases caused by geminiviruses and also the discovery of hundreds of previously unknown, and sometimes highly divergent, geminiviral species (Bernardo et al., 2013; Briddon et al., 2010; Liang et al., 2015; Loconsole et al.,

* Corresponding author.
E-mail address: philippe.roumagnac@cirad.fr (P. Roumagnac).
[1] P.B. and B.M. contributed equally to this work.

http://dx.doi.org/10.1016/j.virol.2016.03.016
0042-6822/

2012; Ma et al., 2015; Roumagnac et al., 2015; Varsani et al., 2009; Yazdi et al., 2008). The rate at which new geminiviruses are being discovered has recently been accelerated through the development and application of sequence-non-specific virus discovery approaches (Roossinck et al., 2015; Rosario et al., 2012) such as rolling circle amplification (RCA) based virus cloning and sequencing (Haible et al., 2006; Inoue-Nagata et al., 2004; Shepherd et al., 2008) and VANA (virion-associated nucleic acids) or siRNA based metagenomics (Candresse et al., 2014; Filloux et al., 2015a; Kreuze et al., 2009). In studies exploring the diversity of these viruses, such approaches have facilitated the expansion of sampling efforts to include both insects (both virus vectors and their predators) and uncultivated host species (Bernardo et al., 2013; Ng et al., 2011; Rosario et al., 2011, 2015). The application of these improved sampling and virus discovery strategies have revealed that geminivirus diversity exceeds that which is currently known (Haible et al., 2006; Ng et al., 2011; Schubert et al., 2007). Such studies have also led to a reevaluation of the known geographical ranges of the different geminivirus genera with, for example, the overturning of the long-held view that members of the genus *Mastrevirus* do not occur in the Americas (Agindotan et al., 2015; Candresse et al., 2014; Kreuze et al., 2009).

Since some of the geminiviruses discovered over the past few years are highly divergent, and, in some cases, have unique genome architectures (Briddon et al., 2010; Loconsole et al., 2012; Varsani et al., 2009; Yazdi et al., 2008), three new geminivirus genera were created and approved in 2014 (*Becurtovirus*, *Turncurtovirus*, and *Eragrovirus*; Varsani et al., 2014b) and at least three others are likely to follow (Bernardo et al., 2013; Krenz et al., 2012; Loconsole et al., 2012); among which are the capulaviruses (Bernardo et al., 2013). Three distinct species of this new lineage were discovered between 2010 and 2011 infecting, respectively, a wild spurge, *Euphorbia caput-medusae* in South Africa (Euphorbia caput-medusae latent virus, EcmLV; Bernardo et al., 2013), alfalfa (*Medicago sativa*) in France (Alfalfa leaf curl virus, ALCV; Roumagnac et al., 2015), and French bean (*Phaseolus vulgaris*) in India (French bean severe leaf curl virus, FbSLCV; Accession number NC_018453). In addition to their high degree of sequence divergence, these viruses also exhibit a genome organization that is unique among the geminiviruses (Bernardo et al., 2013; Roumagnac et al., 2015). ALCV and FbSLCV cause severe symptoms in alfalfa and French bean, respectively, and it has recently been shown that ALCV is transmitted by *Aphis craccivora* (Roumagnac et al., 2015); an invasive aphid species with an almost global distribution (CIE, 1983).

We here collected hundreds of alfalfa and *E. caput-medusae* plants from France and South Africa, respectively, and comparatively analyze the genome sequences of 13 isolates of EcmLV and 26 ALCV along with that of FbSLCV. We show that both EcmLV and ALCV have high prevalence (12–13%) in their respective host species in the Western Cape region of South Africa and in three Southern regions of France. We also present a new codon-model based natural selection detection approach to reveal open reading frames that are probably functionally expressed in capulaviruses genomes. We further demonstrate that, as with other geminiviruses, capulavirus genomes display evidence of both inter-species recombination and biologically functional secondary structures.

2. Materials and methods

2.1. Plant sampling

In 2014, 238 *M. sativa* plants were randomly collected (i.e. irrespective of the presence of potential symptoms) from three regions of Southern France, including the Rhône delta (seven sampling locations), the Montpellier region (four sampling locations) and the Toulouse region (two sampling locations; Supplementary Fig. 1). The symptom status of the 238 plants was assessed and plants were then stored at −80 °C prior to virus detection and characterization. A further 43 symptomatic and 15 asymptomatic *M. sativa* plants were collected later in 2014 from the same three areas (Supplementary Fig. 1). An additional four symptomatic alfalfa plants collected in 2012 and 2013 from the Ebro valley region of the Zaragoza province of Spain were also included for further analysis. Collectively, 300 alfalfa plants were obtained from France and Spain between 2012 and 2014.

In 2012, 302 asymptomatic *Euphorbia caput-medusae* were randomly collected from seven separate locations within the Western Cape region of South Africa (Supplementary Fig. 1). These samples were stored at −80 °C. In 2015, 14 additional samples were collected from an eighth location at the University of the Western Cape Nature Reserve (Supplementary Fig. 1).

2.2. DNA extraction, amplification, cloning and sequencing

Total DNA from alfalfa and *E. caput-medusae* plant samples was extracted as previously described by Bernardo et al. (2013).

PCR-mediated detection of ALCV from the 296 alfalfa plants collected in France, and four plants from Spain was performed using two pairs of PCR primers (ALCV-187F forward primer 5'-TGG AAT ATT GTG CTG CTT GG-3' and ALCV-971R reverse primer 5'-ATT TTG GGA CTT GTG CTC CA-3'; and ALCV-986F forward primer 5'-ATG ATG GAT AAT TCA AAC CC-3' and ALCV-1202R reverse primer 5'-TTC TTC TGG GTA TTT GCA TA-3'). Amplification conditions consisted of 94 °C for 2 min; 30 cycles at 94 °C for 1 min, 58 °C for 1 min (primer pair 1)/55 °C for 30 s (primer pair 2), 72 °C for 50 s; and 72 °C for 5 min. Amplicons were directly sequenced using automated Sanger sequencing (Beckman Coulter Genomics). Circular DNA molecules from samples that tested positive by at least one of the two PCR assays were enriched using RCA (using TempliPhi™, GE Healthcare, USA) as previously described (Shepherd et al., 2008). The RCA products were used as a template for PCR using an abutting pair of primers designed from the 44-1E ALCV complete genome (Accession number KP732474; Roumagnac et al., 2015); Cap-ncoIF: 5'-CCA TGG CCT TCA AAG GTA GCC CAA TTC AAY ATG G-3' and Cap-ncoIR: 5'-CCA TGG GGC CTT ATY CCT CKG YGA TCG-3' using KAPA HiFi Hotstart DNA polymerase (Kapa Biosystems, USA). Amplification conditions consisted of: 96 °C for 3 min, 25 cycles at 98 °C for 20 s, 60 °C for 30 s, 72 °C for 165 s, and 72 °C for 3 min. The amplicons were gel purified, cloned into pJET2.1 (Thermo Fisher, USA) and Sanger sequenced by primer walking at Macrogen Inc. (Korea). In addition, RCA products were digested with *Eco*RI, *Bam*HI, *Dra*I, *Nco*I or *Nde*I for 3 h at 37 °C in order to screen for the presence of a potential DNA-B geminiviral component or satellite sequences.

PCR-mediated detection of EcmLV from the 316 *E. caput-medusae* plants was performed using two pairs of PCR primers: (i) Dar-136F forward primer 5'-CGA AGA GGT CAT TGG GAC AT-3' and Dar-730R reverse primer 5'-CGG GTC TGG CTA AGA GAG TG-3' as previously described by Bernardo et al. (2013) and (ii) Dar-1775F forward primer 5'-TTG AAT TGC ATG GGC ACT TA-3' and Dar-2433R reverse primer 5'-GCC CTT TTG GTC ATT TTG AA-3'. Amplification conditions consisted of: 95 °C for 5 min; 30 cycles at 94 °C for 1 min, 56 °C for 1 min, 72 °C for 50 s; and 72 °C for 5 min. Circular DNA molecules from samples that tested positive by at least one of the two PCR assays was enriched using RCA as described above for alfalfa. RCA products were all digested with *Eco*RI for 3 h at 37 °C. Subsequently, samples that could not be cleaved using *Eco*RI were digested with *Bam*HI for 3 h at 37 °C. Geminivirus-like genomes from 13 *E. caput-medusae* samples were cloned in pGEM-T Easy (Promega Biotech) using methods described by Bernardo et al. (2013).

Prevalence was defined as the proportion of alfalfa or *E. caput-medusae* plants being infected by ALCV or EcmLV from the alfalfa or *E. caput-medusae* populations that were randomly collected in France or South Africa, respectively.