Virus Research 244 (2018) 235–251

Contents lists available at ScienceDirect

Virus Research

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

Next generation sequencing elucidates cacao badnavirus diversity and reveals the existence of more than ten viral species

E. Muller[a,b,*], S. Ravel[a,b], C. Agret[c,d], F. Abrokwah[e], H. Dzahini-Obiatey[f], I. Galyuon[g], K. Kouakou[h], E.C. Jeyaseelan[i], J. Allainguillaume[j], A. Wetten[k,1]

[a] CIRAD, UMR BGPI, 34398, Montpellier, France
[b] BGPI, Univ Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France
[c] CIRAD, UMR AGAP, 34398, Montpellier, France
[d] AGAP, Univ Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France
[e] Department of Biochemistry, University of Cape Coast, Ghana
[f] Cocoa Research Institute of Ghana, P. O. Box 8, Akim Tafo, Ghana
[g] Department of Molecular Biology and Biotechnology, University of Cape Coast, Ghana
[h] World Cocoa Foundation, Abidjan, Cote d'Ivoire
[i] University of Jaffna, Faculty of Science, Department of Botany, Jaffna, Sri Lanka
[j] University of the West of England, Frenchay Campus, Coldharbour Lane, Bristol, BS16 1QY, United Kingdom
[k] School of Agriculture, Policy and Development, University of Reading, Whiteknights, Reading, RG6 7BE, United Kingdom

ARTICLE INFO

Keywords:
Cacao swollen shoot virus
Complete genomes
Cacao
Phylogeny
Illumina sequencing

ABSTRACT

Cacao swollen shoot virus is a member of the family *Caulimoviridae*, genus *Badnavirus* and is naturally transmitted to *Theobroma cacao* (L.) by several mealybug species. CSSV populations in West African countries are highly variable and genetically structured into several different groups based on the diversity in the first part of ORF3 which encodes the movement protein. To unravel the extent of isolate diversity and address the problems of low titer and mixed viral sequences in samples, we used Illumina MiSeq and HiSeq technology. We were able to reconstruct *de novo* 20 new complete genomes from cacao samples collected in the Cocoa Research Institute of Ghana (CRIG) Museum and from the field samples collected in Côte d'Ivoire or Ghana. Based on the 20% threshold of nucleotide divergence in the reverse transcriptase/ribonuclease H (RT/RNase H) region which denotes species demarcation, we conclude there exist seven new species associated with the cacao swollen shoot disease. These new species along with the three already described leads to ten, the total number of the complex of viral species associated with the disease. A sample from Sri Lanka exhibiting similar leaf symptomology to West African CSSD-affected plants was also included in the study and the corresponding sequence represents the genome of a new virus named cacao bacilliform SriLanka virus (CBSLV).

1. Introduction

Cacao swollen shoot disease (CSSD) which results from cacao swollen shoot virus (CSSV) infection is now regarded as the major viral disease affecting cacao and has been recognized as one of the most important diseases in West Africa limiting cacao production. CSSD was first described in Ghana at Effiduase in the New Juabeng district of the Eastern region in 1936 (Steven) although the disease was probably present in the nearby Nankese township of Ghana from 1922 (Paine, 1945). The disease subsequently appeared in all major cacao growing areas in West Africa with CSSD reported in Côte d'Ivoire in 1943 (Burle, 1961; Mangenot et al., 1946), in Nigeria in 1944 (Thresh, 1959), in Togo in 1949 (Partiot et al., 1978) and in Sierra Leone in 1958 (Attafuah et al., 1963). In addition, West African Amelonado cacao, planted uniformly throughout West Africa, appeared to be highly susceptible and sensitive to CSSV and has favored the rapid spread of the disease. CSSD has always been described as a disease endemic to West Africa, as it has never been reported in South America, the cacao tree's

Abbreviations: CSSV, cacao swollen shoot virus; CSSCDV, cacao swollen shoot CD virus; CSSTAV, cacao swollen shoot Togo A virus; CSSD, cacao swollen shoot disease; PCR, polymerase chain reaction; RT, reverse transcriptase; RNase H, ribonuclease H; ORF, open reading frame; $^{t}RNA^{Met}$, methionyl transfer RNA; CRIG, Cocoa Research Institute of Ghana; ICTV, International Committee on Taxonomy of Viruses; CBSLV, cacao bacilliform SriLanka virus; NGS, next generation sequencing

* Corresponding author at: CIRAD, UMR BGPI, TA A – 54/K, Campus international de Baillarguet, 34398, Montpellier Cedex 5, France.
E-mail address: emmanuelle.muller@cirad.fr (E. Muller).
[1] Present address: University of the West of England, Frnechay Campus, Clodharbour lane, Bristol, BS16 1QY, United Kingdom.

https://doi.org/10.1016/j.virusres.2017.11.019
Received 7 September 2017; Received in revised form 15 November 2017; Accepted 18 November 2017
Available online 21 November 2017
0168-1702/ © 2017 Elsevier B.V. All rights reserved.

centre of origin. Additionally, CSSV has not been reported in Sao Tome, nor in Fernando Po (Tinsley, 1971), islands which were the main stepping stones of cacao introduction from the American continent towards West Africa. A viral disease causing similar leaf symptoms was reported in Trinidad (Kirkpatrick, 1953; Swarbrick, 1961), but it is not associated with swellings. Following a government mandated-eradication program, this disease reappeared 14 years ago in the International Cocoa Genebank, Trinidad (ICGT) and two new badnaviruses have been characterized (Chingandu et al., 2017). The existence of CSSD in Malaysia, Indonesia and Sri Lanka (Kenten and Woods, 1976; Peiris, 1953; Crop Protection Compendium, 2002) has been mentioned but only as an attenuated form of CSSD. Additionally, in Malaysia, the disease is likely due to the importation of infected clones (Liu and Liew, 1979). To date, beyond West African cacao, swellings were only reported in Sri Lanka (Orellana and Peiris 1957).

Symptom variability between many different viral isolates has been noted from the first description of the disease in parallel with distinct designations in the different West African countries. However, isolate description by symptomology alone is inadequate for an understanding of the biology, origin and relationships between these different viruses.

Different types of serological diagnosis have been developed but to date these have insufficient polyvalence and sensitivity to address the high variability of the virus (review in Muller, 2008). PCR-based diagnosis holds more promise for the detection of latent infections (Muller et al., 2001) and distinct molecular groups corresponding to different viral species (Kouakou et al., 2012; Abrokwah et al., 2016) though even the use of degenerate primers has not consistently achieved viral fragment amplification from symptomatic leaves (Kouakou et al., 2012; Abrokwah et al., 2016). Furthermore, PCR diagnosis with degenerate primers cannot resolve the existence of mixed viral infections.

PCR primers have been designed in different parts of the genome, particularly, the first and third part of ORF3 corresponding respectively to the putative movement protein (ORF3A area) and to the reverse transcriptase/ribonuclease H (RT/RNase H) region. The first part of ORF3 (primers ORF3A- putative movement protein) was found to be highly conserved between the first six complete CSSV genomes and was therefore used both as a source of diagnostic primers (Muller and Sackey, 2005) and in variability studies to describe different CSSV molecular groups (Kouakou et al., 2012; Abrokwah et al., 2016). To date, using the 80% nucleotide identity threshold in the RT/RNase H region (primers Badna 1/4 CSSV) and according to the recommendations of the International Committee on Taxonomy of viruses (ICTV) (https://talk.ictvonline.org/ictv-reports/ictv_online_report/), we have described five different species responsible for CSSD: A (CSSTAV), B (CSSV), D (CSSCDV), G and M. However, for some samples, discrepancies between the two reconstructed phylogenies from ORF3A and RT/RNase H regions were observed indicating either recombination events between the two regions or the presence of mixed infection. There is a need therefore to study the complete genomic sequences corresponding to these isolates to characterise this diversity. Next generation sequencing technologies now offer an opportunity to resolve this dilemma and to complete the detection of cacao viruses without a priori sequence knowledge.

In 1944, twenty years after the discovery of CSSD in Ghana, the West African Cocoa Research Institute (WACRI) was established at Tafo, Ghana (later becoming the Cocoa Research Institute of Ghana-CRIG). Surveys across Ghana for CSSD began at the same time and a collection of viral isolates based on symptom description was established and named the CRIG Museum. This museum currently comprises more than 70 isolates in the form of potted, symptomatic cacao plants var. Amelonado (up to ten plants per isolate), which are maintained by regularly regrafting. This collection represents a valuable inventory of CSSV isolates from a range of past infected sites (albeit a collection that will have experienced sequence evolution in the intervening period).

Since 1999, attempts to describe the viral diversity present in the collection have been made using PCR amplification of specific regions of the CSSV genome followed by Sanger sequencing. This strategy has been hindered by a number of issues including the absence of symptoms for many plants, the lack of young leaves and the existence of mixed infections within the collection. However, plant screening with next generation sequencing technologies (Hiseq Illumina) can potentially address these problems of low titer and mixed viral sequences.

In the present study, 31 samples from the CRIG Museum and 14 samples recently collected from field sites in Côte d'Ivoire and Ghana and corresponding to new species and/or corresponding to samples with sequence inconsistency between the first and third part of ORF3 were analysed *via* Illumina sequencing. A sample from Sri Lanka exhibiting similar leaf symptomology to West African CSSD-affected plants was also included in the study.

We were able to reconstruct 21 new complete genome sequences corresponding to species A (CSSTAV), B (CSSV), D (CSSCDV), E, M, N, R, Q along with the viral species responsible for the cacao disease in Sri Lanka. We therefore confirm that cacao swollen shoot disease is caused by a complex of species, all of which should be taken into account for effective control of the disease in the different West African countries.

2. Material and methods

2.1. Sample description, DNA extraction and PCR-sequencing analysis

Multiple samples from the CRIG Museum were collected in 1999, 2000, 2012, 2015, and 2016 (Table 1). Total DNA was extracted from symptomatic dried leaves with the Plant DNeasy kit (Qiagen) according to manufacturer's recommendations. Twenty milligrams of dried leaves was ground in a microcentrifuge tube in the presence of ceramic beads with a MP disrupter. To confirm the presence of CSSV in the samples, CSSV sequences were obtained by direct Sanger sequencing (Eurofins, MWG Operon) of PCR products amplified from the two regions ORF3A and Badna1/4 CSSV according to Abrokwah et al. (2016). When direct sequencing failed, PCR products were cloned and several clones sequenced. The genome sequences have been deposited at DDBJ/EMBL/GenBank under the accession MF783897-MF784080.

In 2016, the number of plants harbouring the same isolate was recorded in the CRIG collection. To assess potential variability between replicate plants, two to ten leaves from separate plants were collected from 9 isolates/accessions (Gha26, Gha28, Gha30, Gha36, Gha39, Gha40, Gha53, Gha72 and Gha73).

Thirty one samples from the CRIG Museum collected in 1999, 2012, 2015 or 2016 were selected to be sequenced *via* Illumina technology. In addition, fourteen field samples from Ghana analysed in Abrokwah et al. (2016) and from Côte d'Ivoire analysed in Kouakou et al. (2012), corresponding to groups B, D, E, F, J, K or L have been included in this analysis (Table 2). A sample from Sri Lanka, supplied by the University of Jaffna's Department of Botany, exhibiting similar leaf symptomology to West African CSSD-affected plants was sourced from the Matale district (7°27′25.0″N 80°38′15.7″E) in 2015 and also included in the analysis.

2.2. Illumina DNA sequencing and de novo assembly

Extracted DNA underwent rolling circle amplification (RCA, TempliPhi kit, GE Healthcare Life Science) to concentrate/enrich the sample with circular forms and was sent to Fasteris S.A. (Geneva, Switzerland) for library preparation and sequencing using Illumina MiSeq along with Illumina HiSeq rapid run technology which resulted in paired-end reads of 250-bp mean length (Tables 3 and 4). Paired-end reads were trimmed using the cutadapt script (Martin, 2011) to remove adaptors and filter for quality. The resulting reads were mapped with BWA (Li and Durbin, 2010) against the *Theobroma cacao* reference genome (Argout et al., 2011; http://cocoa-genome-hub.southgreen.fr/). Unmapped reads were assembled using SPAdes v3.6.2 (Bankevich et al.,

Download English Version:

https://daneshyari.com/en/article/8752026

Download Persian Version:

https://daneshyari.com/article/8752026

Daneshyari.com