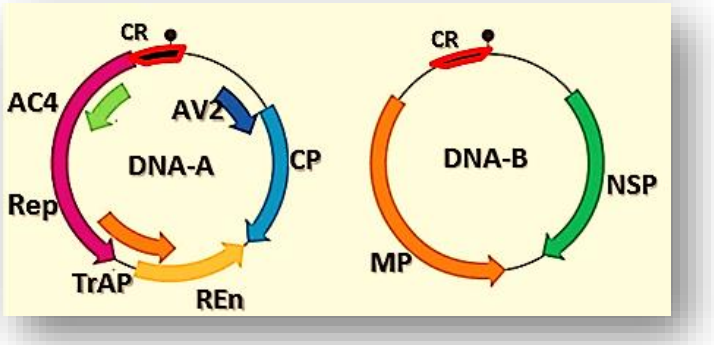
Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Department of Plant Biology

Molecular Detection and Characterization of Begomoviruses from Burkina Faso and Nicaragua

Suneel Sekhar Rachapudi



Independent project in Biology, 30 hp, EX0564
Master's programme in Biotechnology
Examensarbete / Institutionen för växtbiologi, SLU
Nr. 147
Uppsala, 2019

Molecular Detection and Characterization of Begomoviruses from Burkina Faso and Nicaragua

Suneel Sekhar Rachapudi

Supervisor: Prof. Anders Kvarnheden, Department of Plant Biology, Uppsala BioCenter, Swedish University of Agricultural Sciences and Linnean Center for Plant Biology, P.O. Box 7080, SE-750 07 Uppsala, Sweden

Assistant supervisor: Dr. Muhammad Naeem Sattar, Central laboratories, King Faisal University, Box 400, Al-Hasa 3192, Saudi Arabia.

Examiner: Prof. Johan Meijer, Swedish University of Agricultural Sciences, Department of Plant Biology

Credits: 30 hp
Level: A1E
Course title: Independent project/ Degree project in Biology
Course code: EX0564
Programme/education: Biotechnology- Master's Programme

Place of publication: Uppsala
Year of publication: 2019
Cover picture: Begomovirus Genome Organization
Title of series: Examensarbete / Institutionen för växtbiologi, SLU
Part number: Nr. 147
Course coordinating department: Department of Plant Biology, SLU
Online publication: https://stud.epsilon.slu.se

Keywords: Geminivirus, begomovirus, *Calopogonium mosaic virus, Bean golden mosaic virus*, *Phaseolus vulgaris*, Polymerase chain reaction, Rolling circle amplification.

Swedish University of Agricultural Sciences
Faculty of Natural Resources and Agricultural Sciences
Department of Plant Biology

Popular Science

Agriculture plays a major role in the economy of countries in Africa and America. Viruses have emerged as major plant pathogens during the last few decades. Tomato plants from Burkina Faso have been severely infected with some kind of viruses, which caused huge economic loss. A similar pattern of viral infection was noted in bean plants in Nicaragua decreasing the production of beans. The occurrence of these viral infections in Africa and Central America motivated researchers to identify and characterize the viruses that are causing the infections. Analysis of the available information suggested that the infection was caused by viruses of the genus *Begomovirus.* This genus belongs to the family *Geminiviridae,* which is the largest family of plant infecting viruses. Symptomatic plant samples obtained from Burkina Faso and Nicaragua were analyzed in the lab at SLU to confirm the presence of begomoviruses. The results showed that the common bean plants from Nicaragua were infected with two kinds of begomoviruses, *Bean golden yellow mosaic virus* (BGYMV) and *Calopogonium golden mosaic virus* (CalGMV). Unfortunately, the tomato samples from Burkina Faso seemed to have been contaminated in the laboratory with DNA from another sample and it was therefore not possible to identify the virus in these samples in a correct way.

"The ultimate goal of farming is not the growing of crops, but the cultivation and perfection of human beings." **— Masanobu Fukuoka**

Abstract

Agriculture plays a major role in the economy of countries in Africa and America. Viruses have emerged as major plant pathogens during the last few decades on agriculturally important field crop plants. The genus *Begomovirus* belongs to the family *Geminiviridae,* which is the largest family of plant-infecting viruses. Geminiviruses can infect both monocot and dicot plants. Begomoviruses have circular, single-stranded DNA (css-DNA) genomes with one or two equally sized genome components (~2.7 kb each). In previous studies, the begomoviruses, tomato leaf curl Mali virus (ToLCMLV), tomato yellow leaf curl Mali virus (ToYLCMV) and pepper yellow vein Mali virus (PeYVMLV) were found in tomato (*Solanum lycopersicum)* plants from Burkina Faso, while bean golden yellow mosaic virus (BGYMV) and calopogonium golden mosaic virus (CalGMV) had been found in common bean (*Phaseolus vulgaris*) plants from Nicaragua.

The aim of this study was to confirm the reported infections by these viruses in tomato plants as well as common beans plants. Another aim of this study was to amplify the full-length genomes of CalGMV and BGYMV since only partial DNA sequences were determined in the previous study on bean-infecting begomoviruses from Nicaragua. Sequencing results of cloned amplification products derived from rolling circle amplification (RCA) of tomato plants from Burkina Faso showed the highest nucleotide identity to okra yellow crinkle virus (OkYCrV) and okra yellow mosaic Mexican virus (OYMMV), and contamination was probably the most likely reason for this result. Sequencing results of PCR products of the coat protein gene from two samples of common beans from Nicaragua showed highest nucleotide (nt) identity at 98% with BGYMV (sample 2a) or 86-88% with CalGMV (sample 15c). Moreover, the corresponding full-length sequences of DNA-A components showed 99% identity with BGYMV or 94% with the partial sequence of CalGMV available in GenBank (AF439402). Previously no complete sequence of CalGMV existed so the full-length sequences of 15c can be considered as the first full-length sequences of the DNA-A component of CalGMV.

Contents

1. Introduction

Agriculture is economically very important for many developing countries. It is one of the main pillars for development and plays a major role in most of the tropical and subtropical regions in Africa, Asia and America. Approximately 80% of the farming is being carried out by farmers of small scale holdings and employ about 175 million people directly (Alliance for a Green Revolution in Africa, 2014). Agriculture stands for up to 30-40% of the total gross of Africa's economy. However many plant diseases and insect pests are responsible for crop yield losses and deteriorate the quality.

Viruses have emerged as major plant pathogens to the global crop production and about 47% of all emerging infectious diseases of plants have been assumed to be caused by viruses (Anderson et al., 2004). As an example, viruses have been causing significant losses of bean (*Phaseolus vulgaris*) production in Nicaragua and other Central American countries (Singh & Schwartz, 2010).

1.1. Classification of viruses

The viruses have been taxonomically classified by the International Committee on Taxonomy of Viruses (ICTV)(Zerbini et al., 2017) on the basis of their genome organization and genetic material (DNA or RNA). The classified viruses are distributed into 87 families, 349 genera, and ~20,000 species (https://talk.ictvonline.org/taxonomy/). Plant viruses are distributed into 20 families, 90 genera and ~800 species (Fauquet et al., 2008).

1.2. Family *Geminiviridae*

Geminiviridae is the family of plant-infecting viruses with the largest number of species (Briddon et al., 2010). Geminiviruses infect both monocot and dicot plants, causing major crop losses leading to a serious threat to food security in tropical and subtropical regions of the world (Moffat, 1999). Geminivirus alone have been estimated to cause economic losses up to 2300 million dollars in Africa (Thresh & Cooter, 2005) and 300 million dollars in India (Shivaprasad et al., 2005). Geminiviruses are circular, single stranded DNA (cssDNA) viruses with monopartite or bipartite genomes. Monopartite geminiviruses have a single cssDNA molecule and bipartite geminiviruses have two cssDNA molecules (Levy & Tzfira, 2010). Geminiviruses have incomplete double capsid icosahedron virions, which contain capsid proteins and each capsid contains only a single cssDNA molecule (Zerbini et al., 2017).

Viruses in the family *Geminiviridae* are grouped into 9 genera based on their genome organization, sequence identity, plant host and mode of transmission: *Becurtovirus, Begomovirus, Capulavirus, Curtovirus, Eragrovirus, Grablovirus, Mastrevirus, Topocuvirus* and *Turncurtovirus* (Zerbini et al., 2017).

All geminiviruses have similarities in their genome organization, but differs from each other in various aspects. Geminivirus genomes have up to eight protein-coding genes, which are organized in the mode of bidirectional transcription (Varsani et al., 2014). Geminiviruses can replicate either through a rolling circle replication (RCR) or recombination-dependent replication (RDR) mechanism within the nuclei of infected host plant cells. Geminiviruses encode a replication-associated protein (Rep) to initiate the replication process but still the replication process is dependent on the host DNA replication machinery (Rojas et al., 2005).

1.3. Begomoviruses and their genome organization

The genus *Begomovirus* is the largest among the nine genera in the family *Geminiviridae*, consisting of ~ 300 species (Zerbini et al., 2017). Begomoviruses are causative agents of diseases like leaf curl in cotton, pepper, tomatoes, pulses and beans, and during the past three decades begomoviruses have emerged as important plant pathogens (Brown et al., 2012). The DNA-A and DNA-B components of bipartite begomoviruses are similar in size (2.5-2.6 kb), but differ in their organization. Most of the monopartite begomoviruses have been reported from the Old World (OW) and, have a single DNA component of about 2.8 kb. The two DNA components of bipartite begomoviruses have different nucleotide sequences except for 200 nt in the region termed “common region” (CR) (Fig. 1). The CR contains the essential sequences for replication and transcription (Rojas et al., 2005).

The DNA-A component contains five to six open reading frames (ORFs) (Fig. 1) (Zerbini et al., 2017). The *AV1* or coat protein (*cp*) gene encodes the coat protein (CP) for insect transmission and encapsidation. The *AV2* gene encodes the pre-coat protein for movement in plants and also for RNA silencing suppressing activity. The *AC1* gene encodes the replication-associated protein (Rep), which initiates replication by identifying the iteron sequences (Hanley-Bowdoin et al., 2013). The *AC2* gene encodes the transcription activator protein (TrAP) which is required for upregulating the late genes and also regulate viral pathogenicity. The *AC3* gene encodes the replication enhancer protein (REn), which is required for enhancing and stimulating DNA replication and ORF *AC4* sometimes acts as a pathogenicity determinant.

The DNA-B component contains *BC1*, which encodes the nuclear shuttle protein (NSP) involved in transportation of viral DNA and *BV1* that encodes the movement protein (MP) involved in viral movement and symptom development (Hanley-Bowdoin et al., 2013).

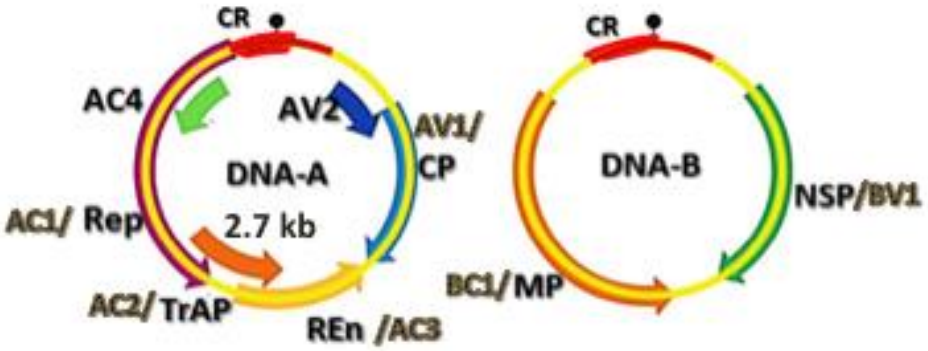


Figure 1. Bipartite begomovirus genome composition; DNA-A: AC1/Rep – replication-associated protein; TrAP/AC2 - transcription activator; REn/AC3 - replication enhancer; CP/AV1 - coat protein; AV2 - pre-coat protein; AC4 - pathogenicity determinant. DNA-B: MP/BC1 - movement protein; NSP/BV1 - nuclear shuttle protein.

1.4. Vector transmission

Begomoviruses are efficiently transmitted by whiteflies of the *Bemisia tabaci* species complex in a persistent circulative manner (Hanssen et al., 2010). The whitefly vector is widely distributed in subtropical and tropical regions and viruses transmitted by whiteflies have emerged as a major disease threat to agriculture in many parts of the world causing major economic losses (Tzanetakis et al., 2013). For virus transmission to a healthy plant, the whitefly acquires virus from infected plants when it feeds on sap from the phloem tissue. The viral particles accumulate in the salivary glands and mix with the saliva. The virus is then transmitted to healthy plant phloem cells when the whitefly injects saliva.

1.5. Background of the project

Tomato is one of the main food crops in Burkina Faso and it is affected severely by infection with begomoviruses (Konate et al., 1995). Symptoms of leaf curl virus infection along with invasion of the whitefly *B. tabaci* were observed in tomato crops in different parts of Burkina Faso. Sequence analyses of amplification produces from symptomatic tomato samples from Burkina Faso revealed infection with tomato leaf curl Mali virus (ToLCMLV), tomato yellow leaf curl Mali virus (TYLCMLV) and pepper yellow vein Mali virus (PeYVMLV) (Sattar et al., 2015). In this study, attempts were made to complement and extend these analyses.

Another part of this study was to identify and characterize begomoviruses in bean plants in Nicaragua. In a previous B.Sc. thesis project, samples of common bean plants from Nicaragua were analyzed for the presence of begomoviruses. The preliminary sequence characterization of the cloned *cp* amplification products revealed highest nucleotide (nt) identity to calopogonium golden mosaic virus (CalGMV) and bean golden yellow mosaic virus (BGYMV) (Karlsson, 2012).

2. Aim

1. To analyze the DNA sequences of begomoviruses infecting tomato plants in Burkina Faso as well as to confirm their infectivity.

2. To identify begomoviruses infecting common beans in Nicaragua and to characterize their DNA sequences.

3. Material and Methods

3.1. Sample collection

Leaves of symptomatic tomato plants from Burkina Faso (samples 10 and 20) and bean plants from Nicaragua (samples 2a and 15c) had previously been collected (Karlsson, 2012; Sattar et al., 2015) and were stored as tissue prints on FTA cards (Whatman).

3.2. DNA extraction from FTA cards

Viral DNA was extracted from FTA cards by the following procedure: a sterile puncher was used to remove 2-6 mm discs from the FTA cards and the discs were transferred to 1.5 ml Eppendorf tubes. To the tubes, 300 µl of ethanol (70%) was added and the tubes were centrifuged for 5 minutes. The discs were subsequently washed after centrifugation using 300 µl TE-buffer (EDTA and Tris HCl pH 8.0). Discs were subsequently washed with 300 µl of FTA purification buffer and then left to dry at room temperature for one hour. For elution of DNA, 20 µl of TE-buffer was added and the discs were incubated at room temperature for 20 minutes. Purified DNA was stored at -20°C.

3.3. Viral DNA amplification by rolling circle amplification

Rolling circle amplification (RCA) was performed to increase the amount of circular viral DNA for further analysis. Circular viral DNA from the samples was amplified by RCA in a single reaction using "IllustraTempliPhi 100 Amplification Kit" (GE Healthcare) as described by Just et al. (2014). RCA products were analyzed by gel electrophoresis at 80 V in 1% agarose gel with ethidium bromide as fluorescent DNA label.

3.4. Amplification of coat protein gene by polymerase chain reaction

Polymerase chain reaction (PCR) was used to amplify the core region of the viral *cp* gene. The same methods were used for samples of Burkina Faso and Nicaragua for identifying the *cp* gene. RCA product was used as a template for PCR. The universal begomovirus degenerate primers AV494/AC1048 were used with an expected product size of ~575 bp including primers (Wyatt & Brown, 1996). PCR was performed with 1 µl (100 ng) of RCA product, 0.5 µl of dNTP mix (10 mM), 0.5 µl of forward and reverse primer (10 mM), 2.5 µl of 10x DreamTaq buffer (Fermentas), 0.25 µl of DreamTaq DNA Polymerase (5U/µl) (Fermentas) and 19.75 µl of sterile water. As negative control, sterile water was used instead of RCA product. The PCR tubes were placed in a thermal cycler (BioRad) and the conditions below were followed.

1. Initial denaturation, 94°C for 2 minutes.
2. Denaturation, 94°C for 30 seconds.
3. Annealing, 50°C for 1 minute.
4. Elongation, 72°C for 2 minutes.

Steps 2, 3 and 4 were repeated for 35 cycles followed by an extension at 72°C for 10 minutes. The PCR products were analyzed by 1 % agarose gel electrophoresis.

3.5. Gel extraction

Positive PCR products were visualized in agarose gels under UV light and were collected in 1.5 ml Eppendorf tubes. The DNA was extracted using GeneJET (ThermoScientific) gel extraction kit following the manufacturer's instructions. Purified DNA product was quantified by a Nanodrop ND-1000 spectrophotometer (NanoDrop Technologies).

3.6. Cloning

The purified amplicons were cloned into the cloning vector pJET 1.2 available in the CloneJet PCR cloning kit (ThermoScientific) using T4 DNA ligase (ThermoScientific). Total reaction volumes of 10 µl were prepared according to manufacturer's instructions. Insert and vector (pJET1.2) at a molar ration of 3:1 was used in the ligation followed by overnight incubation at 16°C.

3.7. Bacterial transformation

Five µl of ligation mixture was added to sub-cloning efficiency *Escherichia coli* DH5-α competent cells (100 µl) followed by incubation on ice for 30 minutes and heat shock at 42°C for one minute. The cells were then incubated on ice for 2 minutes. Liquid broth (900 µl) was added to the mixture and the cells were incubated at 37°C with shaking (225 rpm) for one hour. One hundred µl of transformation mixture was spread on LB agar plates containing ampicillin and X-gal. The rest of the transformation mixture was centrifuged for 3 min at 5,000 rpm. The pellet was dissolved in 100 µl of LB and spread on LB plates containing ampicillin and X-gal. The plates were incubated overnight at 37°C.

3.8. Preparation of cultures and plasmid isolation

The positive bacterial colonies containing the recombinant plasmid were identified through blue-white screen selection. Colonies were picked and transferred to culture tubes containing 4 ml of sterile LB and ampicillin. Culture tubes were incubated at 37°C overnight with shaking. Two ml of each overnight culture was transferred into a 2 ml Eppendorf tube followed by centrifugation for 4 minutes at 8000 rpm. The obtained pellets were used for plasmid isolation using GeneJet plasmid miniprep kit (ThermoScientific) according to the manufacturer's instructions. To check the presence of insert, restriction analysis was performed. Two enzymes, *Xba*I and *Xho*I, were selected based on their unique restriction sites in the vector pJET 1.2. Both enzymes were used for restriction digestion following the manufacturer's instructions (ThermoScientific). Restriction digestion was analyzed by gel electrophoresis on 1% agarose gel.

3.9. Restriction fragment length polymorphism

Restriction fragment length polymorphism (RFLP) analysis was used to characterize RCA products to identify viral full-length DNA components. For RFLP, 2 µl of RCA product of sample 2a was digested with enzyme *Eco*RI at 37°C for one hour, followed by 20 minutes at 65°C to terminate the reaction. Digested product was analyzed by gel electrophoresis. Putative full-length fragment (2.7 kb) of sample 2a was ligated to alkaline phosphatase - treated pBluescript-II KS^+ vector using T4 DNA ligase (ThermoScientific) following the manufacturer's instructions. Transformation as well as plasmid purification and analysis were carried out as described above and the presence of insert in plasmid was confirmed by restriction analysis with *Eco*RI.

3.10. Amplifying full-length DNA components by PCR

From the partial sequencing results for sample 15c, two pairs of back-to-back primers were designed to amplify complete DNA-A molecules for sample 15c: Cal-F (5'-TGACTAACCAGAAGATGACG -3') and Cal-R (5'- GAGACAGGAGACCCTATGG -3'), TGM-F (5'-ACAGGAGACCCTATGGCAC-3') and TGM-R (5'-CTCTGACTAACCAGAAGATG-3'). PCR with Dream Taq DNA polymerase (ThermoScientific) was performed with RCA product of sample 15c as template. Initial denaturation at 95°C for 3 min was followed by 35 cycles of denaturation at 95°C for 30 s, annealing at 52°C for 30 s, elongation at 72°C for 3 min, and a final elongation at 72°C for 10 min. As negative control, PCR was run with sterile water instead of RCA product. PCR products of the expected size were cloned into pJET1.2 vector using T4 DNA ligase enzyme as described above.

3.11. Sequence analysis

The recombinant plasmids with inserts of the expected size of the *cp* gene (575 bp) or complete sequence of genomic components (2.7 kb) were completely sequenced in both forward and reverse directions at Macrogen Inc (the Netherlands). The complete nucleotide sequences of DNA-A were determined by consecutive primer walking. The obtained full-length DNA-A sequences were analyzed initially using the BLASTn tool available in NCBI GenBank database ((https://www.ncbi.nlm.nih.gov/) to compare with the available sequences in the database.

Evolutionary analyses were conducted using MEGA7 software (Kumar et al., 2015) with 21 selected sequences of DNA-A, retrieved from NCBI GenBank database (https://www.ncbi.nlm.nih.gov/). The evolutionary relationships were inferred by constructing phylogenetic dendrograms with 1000 bootstrap replications using Neighbour-Joining algorithm in MEGA7 software (Kumar, et al., 2015). Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated.

4. Results of coat protein gene amplification

4.1. Rolling circle amplification and polymerase chain reaction

Total genomic DNA was extracted from the tomato samples 10 and 20 from Burkina Faso and the bean samples 15c and 2a from Nicaragua using FTA-card extraction procedure. The extracted DNA was employed to perform RCA to increase the amount of viral DNA. Viral DNA was successfully amplified by RCA for all samples (Fig. 2). To check the identity of the begomoviruses, PCR was performed with RCA template for both tomato and bean samples using the begomovirus universal degenerate primers AV494 (F) and AC1048 (R). Universal degenerate primers was used since the plants has shown begomovirus symptoms such as vein clearing, vein thickening, upward curling of leaves, reduction in the leaf size, stunted growth with reduced was observed. The *cp* gene of 10, 20, 15c and 2a was successfully amplified giving a product of around 580 bp (Fig. 3).

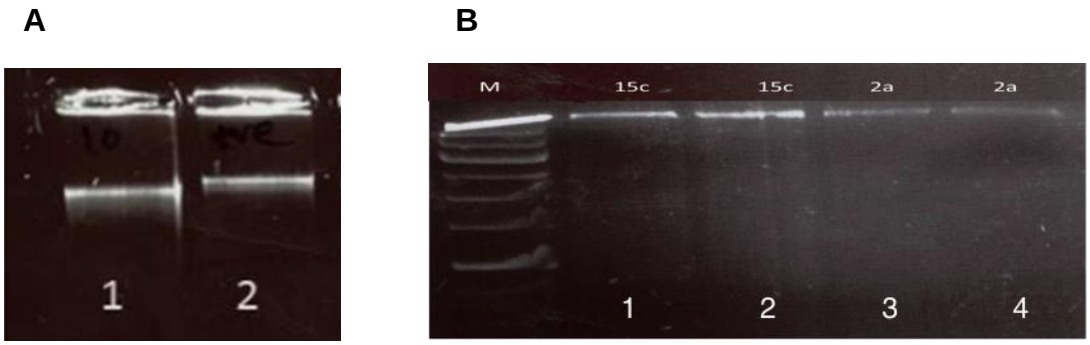


Figure 2: A) Gel image of RCA product of tomato plants from Burkina Faso. Lane 1 shows the amplification for sample 10; lane 2 is amplification for sample 20. B) Gel image of RCA products for bean samples 15c and 2a from Nicaragua. Lanes 1 and 2 show the amplification for sample 15c, and lanes 3 and 4 show the amplification product for sample 2a. M represents a 1kb DNA marker.

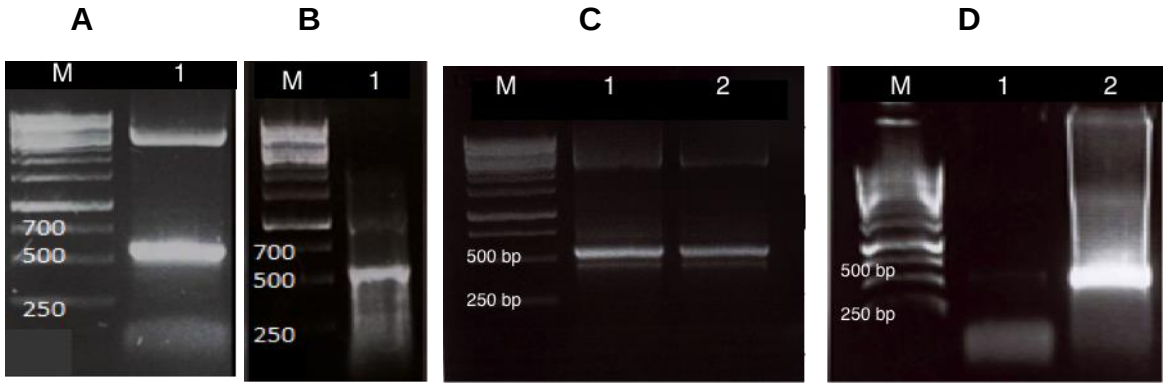


Figure 3: PCR detection of begomovirus infection in samples 10, 20, 15c and 2a. **A)** Lane M represents the 1kb DNA ladder and lane 1 represents the amplification for sample 10. **B)** Lane M represents the 1kb DNA ladder and lane 1 represents the amplification for sample 20. **C)** Lanes 1 and 2 represent the amplification product for 15c, M represents 1kb DNA ladder. **D)** Lanes 1 and 2 represent the amplification for sample 2a with only a weak positive result in lane 1. M represents 1kb DNA ladder.

4.2. Restriction digestion analysis

All the positive PCR products were successfully cloned. Restriction digestion of plasmid purifications was carried out for samples 10, 20, 15c, and 2a to confirm the successful ligation of the inserted DNA in the vector. All 6 clones of sample 10 showed the expected positive bands (Fig. 4A). For sample 20, positive bands were obtained for clones 2, 3, 4, 5 and 6 (Fig. 4B), while clones 1 and 7 showed only empty vector. Three clones (1, 3 and 5) among six clones of sample 2a showed positive bands (Fig. 4C). For sample 15c, among six analyzed clones the five clones 1, 2, 3, 5 and 6 showed clear positive bands (Fig. 4D). For all positive clones, the insert was around 580 bp in size and the vector was around 3 kb.

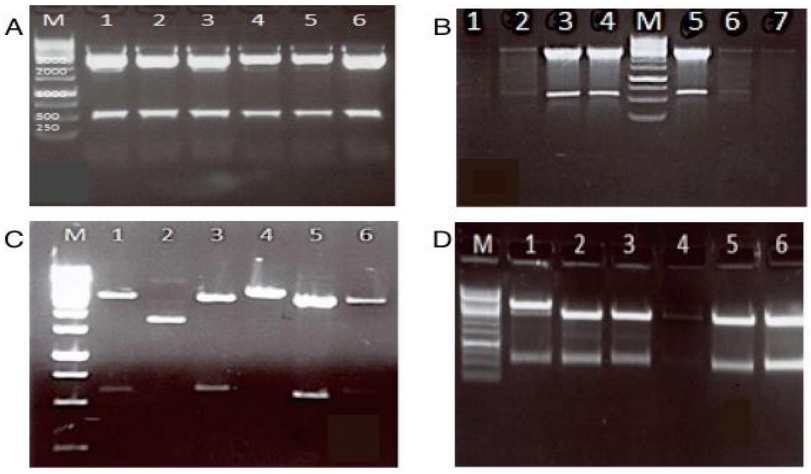


Figure 4: Gel image of plasmid restriction digests for samples 10, 20, 15c and 2a with the enzymes *Xba*I and *Xho*I. **A)** Lane marked with M represents the 1kb DNA ladder. Lanes 1, 2, 3, 4, 5 and 6 represent the positive

bands of insert and vector for sample 10. **B)** Lanes 2, 3, 4, 5 and 6 represent the positive bands of insert and vector for sample 20 and M represents the 1kb DNA ladder. **C)** Lane marked with M represents the 1kb DNA ladder and lanes 1, 3, and 5 represent positive bands of insert and vector of sample 2a. **D**) Lanes 1, 2, 3, 5, and 6 show the positive bands of insert and vector of sample 15c whereas lane marked with M represents the 1kb DNA ladder.

4.3. Sequencing results of coat protein gene

All the positive clones, which had been confirmed by restriction digestion analysis, were successfully sequenced and the obtained sequence data was used for BLASTn searches of the NCBI GenBank database to check virus nucleotide sequence identity. Sequencing results of all 6 positive clones of sample 10 showed highest identity to the *cp* gene of okra yellow crinkle virus (OYCV) from Cameroon with 94% identity and all three positive clones of sample 20 showed highest identity (98%) with the *cp* gene of okra yellow mosaic Mexico virus (OYMMV) from Mexico (Table 1).

Table 1: Sequencing results of *cp* clones for tomato samples 10 and 20 from Burkina Faso.

Clone	Virus name	Acc.no	Nucleotide identity (%)
10-1	OkYCV	FM210077	94
10-2	OkYCV	FM210077	94
10-3	OkYCV	FM210077	94
10-4	OkYCV	FM210077	94
10-5	OkYCV	FM210077	94
10-6	OkYCV	FM210077	94
20-3	OkYMMV	GU990614.1	98
20-4	OkYMMV	GU990614.1	98
20-5	OkYMMV	GU990614.1	98

All three clones of 2a-1, 2a-3, and 2a-5 showed highest identity at 98% with the *cp* gene of *Bean golden yellow mosaic virus*-[Mexico] DNA-A (Table 2). Clones 15c-1, 15c-2, 15c-3 15c-5 and 15c-6 shared highest identity at 86-88% with the *cp* gene of calopogonium golden mosaic virus (CalGMV).

Table 2: Sequencing results of *cp* clones for samples 2a and 15c from Nicaragua.

Clone	Virus name	Acc.no	Nucleotide identity (%)
2a-1	BGYMV-[Mexico] DNA-A	AF173555.1	98
2a-3	BGYMV -[Mexico] DNA-A	AF173555.1	98
2a-5	BGYMV -[Mexico] DNA-A	AF173555.1	98
15c-1	CalGMV- (CP)	EF377329.1	88
15c-2	CalGMV - (CP)	EF377329.1	87
15c-3	CalGMV - (CP)	EF377329.1	88
15c-5	CalGMV- (CP)	EF377329.1	86
15c-6	CalGMV- (CP)	EF377329.1	88

5. Results for full-length DNA components

5.1. RFLP analysis

The RCA product of sample 2a, which was positive for begomoviruses by PCR, was used for RFLP. The RCA product was digested with the restriction enzyme *Eco*RI, selected by considering previous results of partial sequencing (Karlsson, 2012). The restriction of the RCA product of 2a resulted in the expected band of 2.7 kb (full-length fragment) as well as a band of 1.7 kb (Fig. 5).

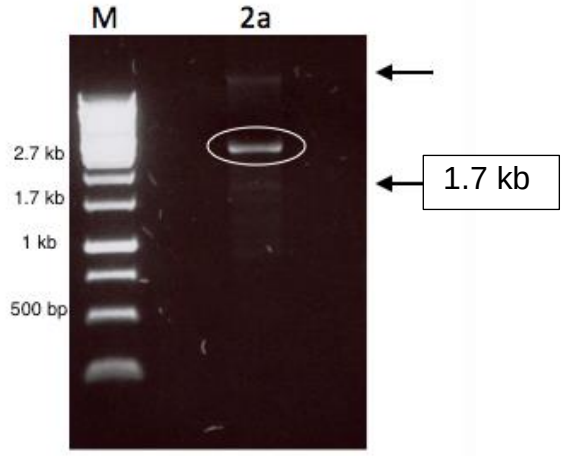


Figure 5: Digestion of RCA product from sample 2a with *Eco*RI. Undigested RCA product is visible for sample 2a indicated by arrow. A full-length fragment is visible at approximately 2.7 kb (highlighted by circle) as well as a band of 1.7 kb indicated by an arrow. Lane marked with M represents the 1kb DNA ladder.

5.2. Restriction analysis

The positive band of around 2.7 kb (Fig. 5) was cloned and restriction analysis was performed to check the presence of insert in the vector. The positive clones Pb2a-2 and Pb2a-4 (Fig. 6) showed an insert at around 2.7 kb and vector around 3 kb.

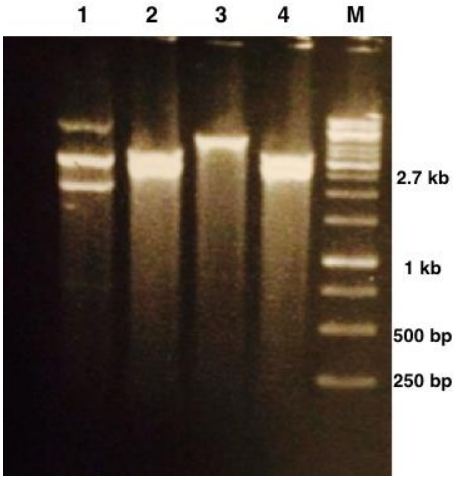


Figure 6: Gel image of plasmid restriction analysis for sample 2a with *Eco*RI enzyme. Lane marked with M represents the 1kb DNA ladder and lanes 2 and 4 show positive bands of insert and vector.

5.3. PCR amplification of full-length DNA-A components

The full-length DNA components of sample 15c were not identified by RFLP analysis, thus PCR using primers based on the determined *cp* gene sequence was used to amplify the full-

length DNA-A component for this sample (Fig. 7). Amplified products were cloned in pJET1.2 and positive clones containing inserts of the expected size (Pj15cCal1, Pj15cCal2, Pj15cCal7, Pj15cTGM1, Pj15cTGM2) were sequenced.

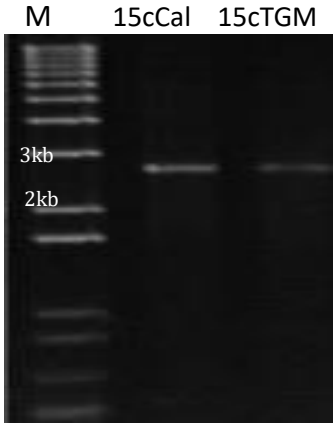


Figure 7: Gel electrophoresis showing results of the amplification of the full-length DNA-A components of begomoviral DNA with back-to-back primers (Cal F/R and TGM F/R) for sample 15c. Lane M represents the 1kb DNA ladder.

5.4. Sequencing results for full-length DNA-A components

Both the clones pb2a1 and pb2a2 of sample 2a shared highest nucleotide identity at 98.9% with BGYMV (Mexico). The length of clone pb2a1 was 2,301 nucleotides whereas pb2a2 provided the complete sequence (2.7 kb) of BGYMV DNA-A. Full-length sequences of Pj15cCal1, Pj15cCal2, Pj15cCal3, Pj15cTGM1 and Pj15cTGM2 shared highest nucleotide sequence identity at 80-82% with tomato mild yellow leaf curl Aragua virus (ToMYLCV) from Venezuela (Table 3).

A phylogenetic tree was constructed using DNA-A sequences determined in this study and of selected begomoviruses (Fig 8.). The phylogenetic analysis showed that all the full-length sequences of 15c grouped into a well-supported clade and were most closely related to Rhynchosia golden mosaic Yucatan virus (RhGMYV), cabbage leaf curl virus (CLCV), euphorbia mosaic virus (EuMV), squash mild leaf curl virus (SqMLCV) and ToMYLCV. The sequence from PB2a2 was most closely related to bean golden yellow mosaic virus (BGYMV)

Table 3: Sequencing results of full-length clones for samples 2a and 15c from Nicaragua.

Clone	Virus name	Length (nt)	Acc.no	Nucleotide identity (%)
Pb2a1	BGYMV- Mexico	2301	AF173555	98.9
Pb2a2	BGYMV- Mexico	2644	AF173555	98.9
Pj15cCal1	ToMYLCV- Venezuela	2633	AY927277	82.0
Pj15cCal2	ToMYLCV- Venezuela	2633	AY927277	80.0
Pj15cCal3	ToMYLCV- Venezuela	2633	AY927277	80.8
Pj15cTGM1	ToMYLCV- Venezuela	2633	AY927277	81.0
Pj15cTGM2	ToMYLCV- Venezuela	2633	AY927277	81.0

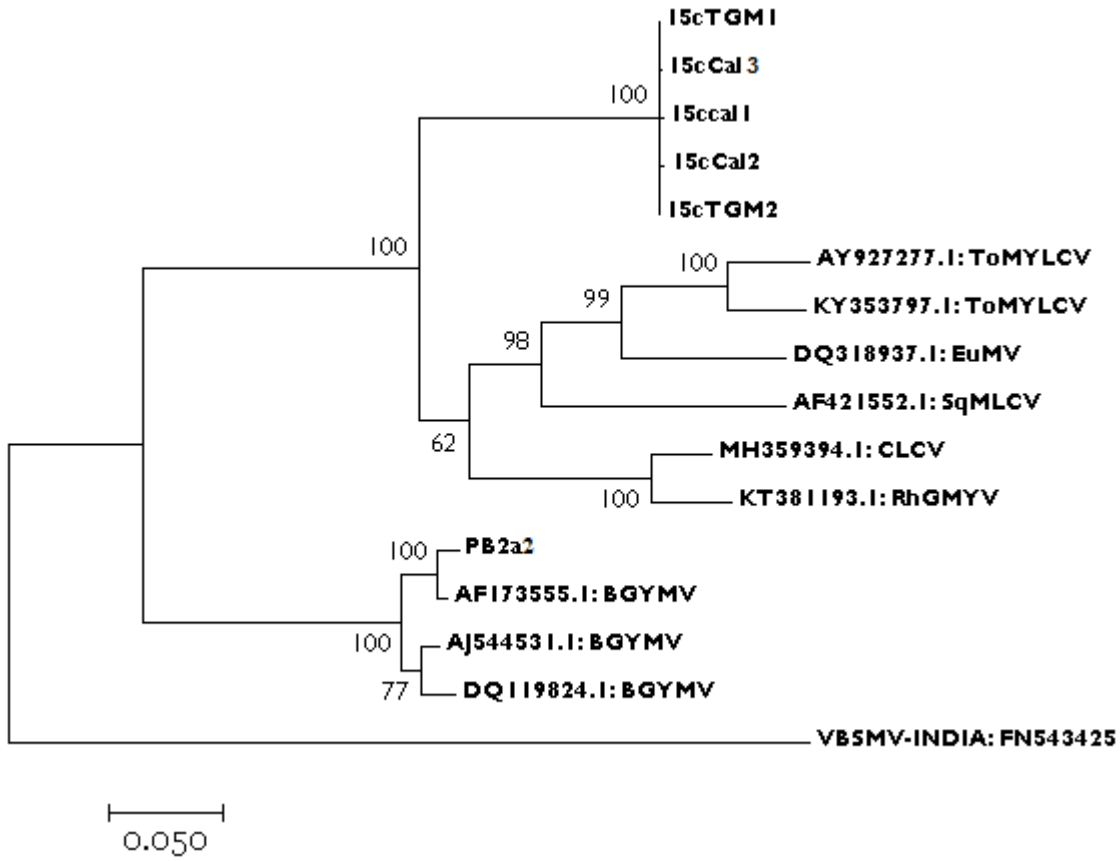


Figure 8: Phylogenetic dendrogram of full-length DNA-A sequences from this study (15cCal1, 15cCal2, 15cCal3, 15cTGM1, 15cTGM2 and PB2a2). Sequence of one begomovirus from Asia was used as an out-group. All the

retrieved sequences are represented by their respective GenBank accession numbers. The numbers on each branch are representing the bootstrap values. ToMYLCV-tomato mild yellow leaf curl Aragua virus, EuMV-euphorbia mosaic virus, SqMLCV-squash mild leaf curl virus, CLCV-cabbage leaf curl virus, RhGMYV-Rhynchosia golden mosaic Yucatan virus. BGYMV-bean golden yellow mosaic virus, VBSMV-velvet bean severe mosaic virus. The scale bar corresponds to 0.050 estimated nucleotide substitutions per site.

6. Discussion

Previous studies of tomato samples from Burkina Faso showed infections with ToLCMV, ToYLCMV and PeYVMLV (Sattar et al., 2015). In this study, complementary analyses were done with original plant sources stored on FTA cards. PCR amplification products for the *cp* gene identified OYCrV for sample 10 and OYMMV for sample 20. However, these two viruses were not identified in the previous studies and are also unlikely to occur in these samples. Most probably, the results are due to mishandling or contamination. This might be because plant samples from Cameroon (infected with OYCrV) and Nicaragua (infected with a begomovirus closely related to OYMMV) had been handled in other projects in the same lab at the same time. This kind of contamination can be avoided by following several guidelines like wearing sterile gloves, working in a laminar flow hood, sterilize the equipment with UVradiation, desinfect work areas and appropriate storage of samples. Lab premises should be designed and equipped in order to maintain maximum protection against the entry of insects or other microorganisms.

Sequence analyses of the cloned *cp* PCR products of sample 2a showed highest nucleotide identity at 98% to DNA-A of BGYMV, which confirms the previous partial sequence data (Karlsson, 2012). BGYMV infection has recently been identified in bean plants in Nicaragua (Karkashian et al., 2011). For the cloned PCR products of sample 15c, the isolates shared highest nucleotide sequence identities at 86-88% with CalGMV, which has previously been identified in bean plants in Costa Rica as well as Nicaragua (Diaz et al., 2002; Karkashian et al., 2011).

The application of RCA in combination with RFLP has been successfully used to detect and characterize infections with begomoviruses (Haible et al., 2006) and has been shown to be efficient for obtaining full-length DNA sequences (Paprotka et al., 2010). Considering the previous study by Karlsson (2012), the six cutter restriction enzyme *EcoR*I was selected for RFLP. RFLP analysis of RCA product for sample 2a with *EcoR*I showed bands at 1.7 kb and 2.7 kb as well as uncut RCA product. The reason for different bands and undigested RCA

product could be due to mixed infection and additional DNA components (Paprotka et al., 2010). Sequencing results of 2a showed highest nucleotide sequence identity at 98.9 % with BGYMV and that strengthens the partial sequence results of 2a as well as the previous study (Karlsson, 2012). Two sets of back-to-back primers, Cal-F/R and TGM- F/R, were designed from partial sequence results of 15c, to complete the DNA-A sequence of CalGMV that was identified by the sequence of the *cp* gene. Back-to-back primers successfully amplified the full-length sequences of DNA-A component for 15c by PCR. The clones Pj15cCal1, Pj15cCal2, Pj15cCal3, Pj15cTGM1 and Pj15cTGM2 grouped closely to each other in a phylogenetic analysis and shared their highest nucleotide sequence identity at 80.8% with ToMYLCV. All the clones of 15c showed highest nucleotide sequence identity (94%) with the partial sequence of CalGMV available in GenBank (AF43402). So the determined sequences are considered as belonging to members of CalGMV. Since there is no complete sequence of CalGMV, the full-length sequences of 15c can be considered as the first full-length sequences of the DNA-A component of CalGMV. The phylogenetic analysis showed that the obtained sequences of 15c share close relationship with ToYMYLCV, SqMLCV and EuYMV from Latin American countries but that they differ from those of Asian countries, like India. According to the criteria for begomovirus species demarcation by the ICTV, set at <91% when this study was made but more recently updated to 94% (https://talk.ictvonline.org/ictv-reports/ictv_online_report/ssdna-viruses/w/geminiviridae/392/genus-begomovirus), the 15c isolate in this study from a bean sample in Nicaragua is considered as being a member of CalGMV.

7. Conclusion

The studies of tomato samples from Burkina Faso were suggesting that the samples had been contaminated. So new plant samples from the fields will be required for appropriate virus identification. The study of the infected bean plants from Nicaragua showed infections with BGYMV and CalGMV. Complete genome sequences for BGYMV are available, whereas, the CalGMV genome had only been partially sequenced. In this study, the complete sequence of the DNA-A component of CalGMV was determined. The DNA sequences of this study were 94% identical to the CalGMV partial sequence, which was available in GenBank. More extensive work is needed to obtain the complete genome of CalGMV by sequencing also DNA-B and further investigation is needed to identify the virus transmission.

8. Acknowledgements

Firstly, I would like to thank **Prof. Dr. Anders Kvarnheden** for giving me a chance to work in such an outstanding research environment. Special thanks to my supervisor **Dr. Muhammad Naeem Sattar** for his reliable support during my thesis. I am always thankful to **God** and **my family** for keeping my spirit ups in my difficult situations. My thanks to **my lab mates**, **my friends** in Sweden and India for their support and encourage throughout my journey. Last but not least my sincere thanks to my university **Swedish University of Agricultural Sciences** for giving me good study and good environment.

9. References

Alliance for a Green Revolution in Africa (AGRA). (2014). Africa Agriculture Status Report 2014: Climate Change and Smallholder Agriculture in Sub Saharan Africa, http://hdl.handle.net/10568/42343.

Anderson, P. K., Cunningham, A. A., Patel, N. G., Morales, F. J., Epstein, P. R., & Daszak, P. (2004). Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends in Ecology & Evolution*, 19(10), 535-544.

Briddon, R. W., Patil, B. L., Bagewadi, B., Nawaz-ul Rehman, M. S., & Fauquet, C. M. (2010). Distinct evolutionary histories of the DNA-A and DNA-B components of bipartite begomoviruses. *BMC Evolutionary Biology*, 10, 97.

Brown, J. K., Fauquet, C. M., Briddon, R. W., Zerbini, M., Moriones, E., & Navas-Castillo, J. (2012). *Geminiviridae*. In: Virus Taxonomy - Ninth Report of the International Committee on Taxonomy of Viruses, pp. 351-373. Edited by King, A. M. Q., Adams, M. J., Carstens, E. B., Lefkowitz, E. J. London., Associated Press, Elsevier.

Diaz, M., Maxwell, D. P., Karkashian, J. P., & Ramírez, P. (2002). *Calopogonium golden mosaic virus* identified in *Phaseolus vulgaris* from western and northern regions of Costa Rica. *Plant Disease*, 86(2), 188-188.

Fauquet, C. M., Briddon, R. W., Brown, J. K., Moriones, E., Stanley, J., Zerbini, M., & Zhou, X. (2008). Geminivirus strain demarcation and nomenclature. *Archives of Virology*, 153(4), 783-821.

Haible, D., Kober, S., & Jeske, H. (2006). Rolling circle amplification revolutionizes diagnosis and genomics of geminiviruses. *Journal of Virological Methods*, 135(1), 9-16.

Hanley-Bowdoin, L., Bejarano, E. R., Robertson, D., & Mansoor, S., (2013). Geminiviruses: masters at redirecting and reprogramming plant processes. *Nature Reviews Microbiology*, 11(11), 777-788.

Hanssen, I. M., Lapidot, M., & Thomma, B. P. (2010). Emerging viral diseases of tomato crops. *Molecular Plant-Microbe Interactions*, 23(5), 539-548.

Just, K., Leke, W. N., Sattar, M. N., Luik, A., & Kvarnheden, A. (2014). Detection of *Tomato yellow leaf curl virus* in imported tomato fruit in northern Europe. *Plant Pathology*, 63(6), 1454-1460.

Karkashian, J., Ramos-Reynoso, E. D., Maxwell, D. P., & Ramírez, P. (2011). Begomoviruses associated with bean golden mosaic disease in Nicaragua. *Plant Disease*, 95(8), 901-906.

Karlsson, M. (2012). Molecular identification and characterization of begomoviruses in Nicaraguan cultivars of common beans. BSc thesis, Swedish University of Agricultural Sciences.

Konate, G., Barro, N., Fargette, D., Swanson, M.M., & Harrison, B.D. (1995) Occurrence of whitefly transmitted geminiviruses in crops in Burkina Faso, and their serological detection and differentiation. *Annals of Applied Biology*, 126(1), 121-129.

Kumar, S., Stecher, G. & Tamura, K. (2015). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, 33(7): 1870-1874.

Levy, A., & Tzfira, T. (2010). Bean dwarf mosaic virus: a model system for the study of viral movement. *Molecular Plant Pathology*, 11(4): 451-461.

Moffat, A. S. (1999). Geminiviruses emerge as serious crop threat. *Science*, 286(5446): 1835-1835.

Paprotka, T., Metzler, V., & Jeske, H. (2010). The first DNA 1-like α satellites in association with New World begomoviruses in natural infections. *Virology*, 404(2): 148-157.

Rojas, M. R., Hagen, C., Lucas, W. J., & Gilbertson, R. L. (2005). Exploiting chinks in the plant's armor: evolution and emergence of geminiviruses. *Annual Review of Phytopathology*, 43: 361-394.

Sattar, M. N., Koutou, M., Hosseini, S., Leke, W. N., Brown, J. K., & Kvarnheden, A. (2015). First identification of begomoviruses infecting tomato with leaf curl disease in Burkina Faso. *Plant Disease*, 99(5): 732.

Singh, S. P., & Schwartz, H. F. (2010). Breeding common bean for resistance to diseases: A review. *Crop Science*, 50(6): 2199-2223.

Shivaprasad, P.V., Akbergenov, R., Trinks, D., Rajeswaran, R., Veluthambi, K., Hohn, T. & Pooggin, M.M. (2005) Promoters, transcripts, and regulatory proteins of mungbean yellow mosaic geminivirus. *Journal of Virology*, 79(13): 8149-8163.

Thresh, J. M., & Cooter, R. J. (2005). Strategies for controlling cassava mosaic virus disease in Africa. *Plant Pathology*, 54(5): 587-614.

Tzanetakis, I. E., Martin, R. R., & Wintermantel, W. M. (2013). Epidemiology of criniviruses: an emerging problem in world agriculture. *Frontiers in Microbiology*, 4: 119.

Varsani, A., Navas-Castillo, J., Moriones, E., Hernández-Zepeda, C., Idris, A., Brown, J. K., & Martin, D. P. (2014). Establishment of three new genera in the family *Geminiviridae: Becurtovirus, Eragrovirus* and *Turncurtovirus*. *Archives of Virology*, 159(8): 2193-2203.

Wyatt, S. D., & Brown, J. K. (1996). Detection of subgroup III geminivirus isolates in leaf extracts by degenerate primers and polymerase chain reaction. *Phytopathology*, 86(12): 1288-1293.

Zerbini, F. M., Briddon, R.W., Idris, A., Martin, D., Moriones, E., Navas-Castillo, J., Rivera-Bustamante, R., Roumagnac, P., & Varsani, A. ICTV Report Consortium. (2017). ICTV Virus Taxonomy Profile: *Geminiviridae*. *Journal of General Virology*, 98(2): 131-133.

Nr. 147
Uppsala 2019

Department of Plant Biology
SLU
BOX 7080
75007 Uppsala, Sweden