The use of rolling circle amplification (RCA) for diagnosis and characterization of geminiviruses

Von der Fakultät Energie-, Verfahrens- und Biotechnik der Universität Stuttgart zur Erlangung der Würde eines Doktors der Naturwissenschaften (Dr. rer. nat.) genehmigte Abhandlung

vorgelegt von

Patricia Soares Wyant

aus Goiânia / Brasilien

Hauptberichter: Prof. Dr. Holger Jeske
Mitberichter: PD Dr. Christina Wege

Tag der mündlichen Prüfung: 10. Juni 2011

Biologisches Institut
Abt. Molekularbiologie und Virologie der Pflanzen
Universität Stuttgart
2011

To my parents

Index

Abbreviation list

aa	amino acid
APS	adenosine 5' phosphosulfate
ATP	adenosine triphosphate
bp	base pairs
c	curling
ccc	covalently closed circular
CCD	charge - coupled device
CR	common region
cs	common strain
CSN	consensus sequence number
CSR	complementary strand synthesis
CTAB	Cetyl trimethylammonium bromide
DNA	deoxyribonucleic acid
dNTPs	deoxynucleotide triphosphates
dpi	days post inoculation
dsDNA	double stranded DNA
EDTA	Ethylenediaminetetraacetic acid
GFP	Green fluorescence protein
ICTV	International Committee on Taxonomy of Viruses
IR	intergenic region
kb	kilo bases
M	molecular weight marker
MP	Movement protein
nc	nucleotide
NJ	Neighbor Joining
NSP	nuclear shuttle protein
oc	open circular
ORF	open reading frame
Ori	origin of replication
PAR	photosynthetically active radiation
PCNA	proliferating cell nuclear antigen
PCP	precoat protein
PCR	polymerase chain reaction
PPi	pyrophosphate
pRBR	retinoblastoma - like protein
RCA	Rolling cicle amplification
RCR	rolling circle replication
RDR	recombination - dependent replication
REn	Replication enhancer
Rep	Rep - protein
RFLP	Restriction Fragment Length Polymorphism
SI	Sequence identity
siRNA	small interference RNA
ssDNA	single stranded DNA
TrAP	transcriptional - activator protein
wpi	weeks post inoculation
yv	yellow vein

Virus list

AbMBoV	Abutilon mosaic Bolivia virus
AbMBV - [AL:09]	Abutilon mosaic Brazil virus - [Alagoas:2009]
AbMV	Abutilon mosaic virus
AbMV - [DE]	Abutilon mosaic virus - [Germany]
AbMV - SA	Abutilon mosaic virus - South African isolate
ACMV	African cassava mosaic virus
AGMV	Asytasia golden mosaic virus
ALCuCMV	Ageratum leaf curl Cameroon virus
AYVSLV	Ageratum yellow vein Sri Lanka virus
BCTV	Beet curly top virus
BDMV - [CO:87]	Bean dwarf mosaic virus - [Colombia:1987]
BGMV	Bean golden mosaic virus
BGMV - [BR:Cam1:78]	Bean golden mosaic Virus - [Brazil:Campinas1:1978]
BGMV - [BR:SAdG7:soybean:08]	Bean golden mosaic virus - [Brazil:SAdG 7:soybean:2008]
BGMV - AL	Bean golden mosaic virus - Alagoas
BGYMV	Bean golden yellow mosaic virus
BGYMV - [DO:87]	Bean golden yellow mosaic virus - [Dominican Republic:1987]
BGYMV - [US:Hom:05]	Bean golden yellow mosaic virus - [United States of America:Homestead:2005]
BlYSV	Blainvillea yellow spot virus
BlYSV - [BR:Coi25:07]	Blainvillea yellow spot virus - [BR:Coi25:07]
CabLCuJV - [JM:Cuc3:05]	Cabbage leaf curl Jamaica virus - [Jamaica:CUc3:2005]
CabLCuJV - [JM:Cuc32:05]	Cabbage leaf curl Jamaica virus - [Jamaica:CUc32:2005]
ChiLCuV	Chilli leaf curl virus - India
ClLCrV	Cleome leaf crumple virus
ClLCrV[BR:AL:2009]	Cleome leaf crumple virus - [Brazil:Alagoas:2009]
CoYSV	Corchorus yellow spot virus
CoYSV - [MX:Yuc:05]	Corchorus yellow spot virus - [Mexico:Yucatan:2005]
CYVMV	Croton yellow vein mosaic virus
EACMCV	East African cassava mosaic Cameroon virus
EACMKV	East African cassava mosaic Kenya virus
EACMV	East African cassava mosaic virus
EACMZV	East African cassava mosaic Zanzibar virus
EuMV - [MGS1:07]	Euphorbia mosaic virus [Mato Grosso do Sul1:2007]
EuMV - [MGS2:07]	Euphorbia mosaic virus [Mato Grosso do Sul2:2007]
EuMV - [MX:Yuc: 04]	Euphorbia mosaic virus - A [Mexico:Yucatan:2004]
EuMV - EF2H	Euphorbia mosaic virus clone EF2H
EuMV - Peru	Euphorbia mosaic virus - Peru
EuMV - Wis	Euphorbia mosaic virus - [Jamaica:Wissadula]
HYMV	Horsegram yellow mosaic virus
KuMV	Kudzu mosaic virus
MaMPRV - [PR:90]	Macroptilium mosaic Puerto Rico vírus - [Puerto Rico:1990]
MaYMV - [US:Flo:85]	Macroptilium yellow mosaic Florida virus - [United States of America:Florida:1985]
MaYMV - HP	Macroptilium yellow mosaic virus isolate Hope Pastures
MaYMV - pmrsrfb14	Macroptilium yellow mosaic virus isolate pmrsrfb14
MeMV - PR[PR:]	Merremia mosaic virus - Puerto Rico [Puerto Rico:]
MeMV - PR4	Merremia mosaic virus isolate PR4
MeMV - PR80	Merremia mosaic virus isolate PR80
MGMV - JM:W:AT	Macroptilium golden mosaic virus - [Jamaica:Wissadula:August Town]

MGMV - JM1:ST	Macroptilium golden mosaic virus - [Jamaica1:SpanishTown]
MiYLCV	Mimosa yellow leaf curl virus
MSV	Maize streak virus
MYMV	Mungbean yellow mosaic virus
OMoV - [BR:Ok:6319]	Okra mottle virus - [Brazil:okra:6319]
OMoV - [BR:Ok:6328]	Okra mottle virus - [Brazil:okra:6328]
OMoV - [BR:Sag8:Soy:08]	Okra mottle virus strain - [Brazil:Sag8:Soy:08]
OYMMV - [MX:ST:3abg]	Okra yellow mosaic Mexico virus - [Mexico:Santa Teresa 2 :3abg]
PaLCuCNV	Papaya leaf curl China virus
PepGMV - US [MX:Tam]	Pepper golden mosaic virus - United States of America [Mexico:Tamaulipas]
PepYVMV	Pepper yellow vein Mali virus
PSLDV - [BR:LNS2:Pas:01]	Passionfruit severe leaf distortion virus - [Brazil:LNS2:Pas:01]
PYMPV - [PA:Div:Tom]	Potato yellow mosaic Panama virus - [Panama:Divisa:Tomato]
PYMV - To [GP:Tom]	Potato yellow mosaic virus - Tomato [Guadeloupe:Tomato]
PYMV - To [PR:Tom:04]	Potato yellow mosaic virus - Tomato [Puerto Rico:Tomato:2004]
PYMV - TT [TT:Tom]	Potato yellow mosaic virus - Trinidad [Trinidad & Tobago:Tomato]
RhYMV	Rhynchosia yellow mosaic virus
SbBMV	Soybean blistering mosaic virus
SbBMV - [AR:NOA:05]	Soybean blistering mosaic virus - [Argentina:NOA:2005]
SiCMV - [BR:Coi4:07]	Sida common mosaic virus - [BR:Coi4:07]
SiGMHV - [HN]	Sida golden mosaic Honduras vírus - [Honduras]
SiGMV	Sida golden mosaic virus
SiMAV	Sida mosaic Alagoas virus.
SiMBV	Sida mosaic Brazil virus
SimMV	Sida micrantha mosaic virus
SimMV - [5157]	Sida micrantha mosaic virus - [Brazil:Okra] isolate 5157
SimMV - [BR:A2B2]	Sida micrantha mosaic virus - [Brazil:A2B2]
SimMV - [BR:A1B3]	Sida micrantha mosaic virus - [Brazil:A1B3]
SimMV - [SP77]	Sida micrantha mosaic virus - [SP77]
SimMV - [BR:Sag3:Soy:08]	Sida micrantha mosaic virus -[BR:Sag3:Soy:08]
SimMV - [MGS1:07]	Sida micrantha mosaic virus - [Mato Grosso do Sul1:2007]
SimMV - [MGS2:07]	Sida micrantha mosaic virus - [Mato Grosso do Sul2:2007]
SimMV - rho[Bo:CF2:07]	Sida micrantha mosaic virus - Rhombifolia [Bolivia:Cerro Fraile 2:2007]
SiMoV - mic[BR:A1B3]	Sida mottle virus - Micrantha [Brazil:A1B3]
SiMoV - rho[BR:Vic1:99]	Sida mottle virus - Rhombifolia [Brazil:Vicosa1:1999]
SiYLCV - [BR:Coi3:07]	Sida yellow leaf curl virus - [BR:Coi3:07]
SiYMV - [BR:Vic2:99]	Sida yellow mosaic virus - [Brazil:Vicosa2:1999]
SiYMYuV	Sida yellow mosaic Yucatan virus
SiYVV - [HN]	Sida yellow vein virus - [Honduras]
SoMBoV	Solanum mosaic Bolivia virus
TbCSV - [Y41]	Tobacco curly shoot virus - [Y41]
TbLCZV	Tobacco leaf curl Zimbabwe virus
TGMV	Tomato golden mosaic virus
TGMV - [BR:Com:84]	Tomato golden mosaic virus - [Brazil:Common:1984]
TGMV - [BR:YV]	Tomato golden mosaic virus - [Brazil:Yellow Vein]
TMV	Tobacco Mosaic Virus
ToCMoV - BA [BR:Sea1:96]	Tomato chlorotic mottle virus - Bahia [Brazil:Seabra1:1996]
ToCMoV - MG [BR:Iga1:96]	Tomato chlorotic mottle virus - Minas Gerais [Brazil:Igorape1:1996]
ToCMV - [BR:Coi22:07]	Tomato common mosaic virus - [BR:Coi22:07]
ToCSV - [ZA]	Tomato curly stunt virus - [South Africa]
ToLCAnjV	Tomato leaf curl Anjouan virus
ToLCAntV	Tomato leaf curl Antsiranana virus

ToLCCMV	Tomato leaf curl Cameroon virus
ToLCHnV	Tomato leaf curl Hainan virus
ToLCKeV	Tomato leaf curl Kerala virus
ToLCKMV	Tomato leaf curl Comoros virus
ToLCMGV	Tomato leaf curl Madagascar virus
ToLCSinV - [NI:SL]	Tomato leaf curl Sinaloa vírus - [Nicaragua:Santa Lucia]
ToLDV - [BR:Pda4:05]	Tomato leaf distortion virus - [BR:Pda4:05]
ToMLCV	Tomato mosaic leaf curl virus
ToMMV - [BR:Pda58:05]	Tomato mild mosaic virus - [BR:Pda58:05]
ToMoTV - [CU]	Tomato mottle Taino virus - [Cuba]
ToMoV - [PR:04]	Tomato mottle virus - [Puerto Rico:2004]
ToMoV - [US:Flo:89]	Tomato mottle virus - [United States of America:Florida:1989]
ToMYLCAV - [VE:10]	Tomato mild yellow leaf curl Aragua virus - [Venezuela:10]
ToSLCV - GT [GT:San:96]	Tomato severe leaf curl virus - Guatemala [Guatemala:Sansirisay:1996]
ToSLCV - GT [MX:Rio1:05]	Tomato severe leaf curl virus - Guatemala [Mexico:Rioverde1:2005]
ToSRV - [BR:Ube2:2000]	Tomato severe rugose virus - [Brazil:Uberlandia2:2000]
ToSRV - PetrolinadeGoias	Tomato severe rugose virus strain Petrolina de Goias
ToSRV - PJU - SP	Tomato severe rugose virus strain PJU - Sao Paulo
ToSRV - Sumare	Tomato severe rugose virus strain Sumare
ToYLDV	Tomato yellow leaf distortion virus clone EF5N
ToYSV - [AR]	Tomato yellow spot virus - [Argentina]
ToYSV - [BR:Bic2:99]	Tomato yellow spot virus - [Brazil:Bicas2:1999]
ToYVSV	Tomato yellow vein streak virus
ToYVSV - Ba - 3	Tomato yellow vein streak virus isolate Ba - 3
ToYVSV - G22	Tomato yellow vein streak virus isolate G - 22 from Brazil
ToYVSV - T06	Tomato yellow vein streak virus isolate T06
TPCTV	Tomato pseudo curly top virus
TYLCDiV	Tomato leaf curl Diana virus
TYLCV	Tomato yellow leaf curl virus
TYLCV - OM [IR:Ji31:06]	Tomato yellow leaf curl virus - [Iran:Jiroft]
TYMLCV - [VE:Mer:57]	Tomato yellow margin leaf curl virus - [Venezuela:Merida:57]
WGMSTV	Wissadula golden mosaic St Thomas Virus
WGMSTV - W132BFL5	Wissadula golden mosaic St Thomas Virus - W132BFL5
WmCSV - [SD]	Watermelon chlorotic stunt virus - [SD]

Zusammenfassung

Geminiviridae ist die größte Familie der phytopathogenen DNA-Viren, die eine Vielzahl an Nutzpflanzen infizieren. Hierbei verursachen sie erhebliche Einbußen bei der Produktion von wirtschaftlich bedeutsamem Ausmaß. Der Eigenbedarf an Ernteerträgen zur Sicherung der Existenz in den Tropen und Subtropen ist dadurch gefährdet. Zunehmendes Wissen über Epidemiologie, Sequenz- und Bio-Diversität ist daher überaus bedeutend, um vorbeugende Strategien zu bewerkstelligen.

In dieser Arbeit wurde die Rolling-Circle-Amplifikation (RCA) in Kombination mit dem Restriktionsfragmentlängenpolymorphismus (RFLP) zusätzlich zu anderen, in der Molekularbiologie angewandten Methoden, zur technisch entscheidenden Verbesserung der direkten Sequenzierung, der biolistischen Klonierung und der Pyrosequenzierung genutzt. In den ersten zwei Teilen dieser Studie wurde die RCA/RFLP angewandt, um symptomatische Pflanzenproben aus Südamerika auf eine geminivirale Infektion hin zu untersuchen. So wurden alle Proben, insgesamt sieben Unkraut- aus Bolivien und 21 von 22 Nutzpflanzen (u.a. Bohnen) aus Brasilien, als viruspositiv diagnostiziert. Zudem wurden die RCA-Produkte unter Verwendung von zwei verschiedenen Methoden sequenziert. Hierfür wurde für die bolivianischen Proben eine neuartige und sehr effiziente Klonierungsstrategie genutzt, bei der die zu inserierende DNA als Tandem-Wiederholung über limitierte Restriktion mit *Sau*3AI generiert wurde. Bei den brasilianischen Proben wurden die RCA-Produkte gemischt und über die kommerziell erhältliche Methode 454 sequenziert, um ein schnelles Detektions- und Charakterisierungssystem zu entwickeln. Mit Hilfe dieser Methoden wurden die vollständigen Sequenzen der Genome aller nachgewiesenen Viren ermittelt. Die Analyse der Sequenzen ergab, dass alle diese Viren Neuwelt-*Begomoviren* mit einer bipartiten Genomorganisation sind. Dies wurde weiterhin mittels phylogenetischer Analyse bestätigt. Hierdurch wurden fünf neue Virusspezies, zwei neue Stämme sowie fünf neue Varianten, die zu bereits beschriebenen Viren gehören, nachgewiesen. Im dritten Teil dieser Arbeit wurden zwei Varianten des *Tomato golden mosaic virus*, common strain (cs) und yellow vein (yv), die anfänglich aus Tomatenpflanzen isoliert wurden, aber zu keiner Zeit wieder in Tomaten nachweisbar waren, komplett sequenziert und überdies erfolgreich mittels Partikelbeschuss in Tomaten zurückgeführt. Diese induzierten nur milde Symptome mit einer extrem niedrigen Infektionsrate. Im letzten Teil dieser

Arbeit wurden die in den Achtzigerjahren in Westafrika gesammelten *Asystasia gangetica* Pflanzenproben analysiert. Diese zeigten Symptomsegregation mit Blattmosaik und Vergilbung der Blattadern, die auf zwei verschiedene geminivirale Infektionen zurückzuführen sind. Eine Analyse der RCA-Produkte mittels an die RCA/RFLP-Technik angeschlossenes *deep sequencing* bekräftigte die Annahme, dass diese Pflanzen tatsächlich mit zwei verschiedenen Geminiviren infiziert waren, die zu einer beträchtlichen Segregation in derselben Pflanze führten.

Summary

Geminiviridae is the largest family of plant DNA viruses that infect a broad range of plants causing a limitation to the production of economically important or staple food crops within tropical and subtropical countries. Increasing knowledge about its epidemiology, sequence diversity and biodiversity is highly important in order to implement preventative strategies.

With the use of rolling circle amplification (RCA) combined with restriction fragment length polymorphism (RFLP), in addition to other methods applied in molecular biology, technical improvements on direct sequencing, shot gun cloning and pyrosequencing were achieved in this work. In the first two parts of this study RCA/RFLP was used to diagnose geminiviral infection in symptomatic plant samples originally collected in South America. All samples from Bolivia, in total 7 weeds and 21 of 22 samples from Brazil, including beans samples, were diagnosed virus positive when analyzed using RCA/RFLP. RCA products were sequenced using two different methods and the complete genome sequences of all detected viruses were obtained. For the Bolivian samples, a novel and efficient cloning strategy with tandem repeat inserts obtained by limited *Sau*3AI digestion was used and for the Brazilian samples, in order to develop a faster detection and characterization system for geminiviral infection, the RCA products were pooled and sequenced by the commercially available 454 method. The sequences analysis showed that all viruses had a genomic organization of bipartite New World begomoviruses. Phylogenetic analysis revealed amongst the detected viruses, five distinct new virus species, two new strains and five variants of previously described viruses. In the third part of the study, two variants of Tomato golden mosaic virus (TGMV), common strain (cs) and yellow vein (yv), originally extracted from tomato but never detected in tomatoes again, were completely sequenced for the first time and successfully bombarded back into tomato plants. The infection rate was extremely low and symptoms induction was very mild. Finally, in the last part of this work, *Asystasia gangetica* plants, collected in West Africa in the 1980s with segregation of symptoms, including mosaic and yellow veining, typical of geminiviral infections were analyzed. Also in this case using the RCA/RFLP technique followed by deep sequencing of the RCA products, the results confirm the idea of infection by two distinct geminiviruses, showing a remarkable segregation of geminiviruses in a single plant.

Introduction

Geminiviruses constitute the largest and most economically important family of plant DNA viruses and is represented by a diverse group of viruses that infect a broad range of plants causing devastating diseases for important crops, like for example, beans, cassava, cotton, maize, pepper, sugar beet, sweet potato, and tomato (Mansoor et al., 2003; Moffat, 1999; Morales and Anderson, 2001).

Over the last four decades, agricultural intensification and the emergence and prevalence of a new and more aggressive biotypes of the insect vector (*Bemisia tabaci* biotype-B) facilitated the expansion of begomovirus populations and their movement into new plant hosts, contributing to the emergence of new and sometimes more virulent viruses, producing an increase in frequency and severity of the diseases produced by begomoviral infection (Argüello-Astorga et al., 1994). Because the diseases caused by geminiviral infection are a major limitation to the production of economically important and staple foods, which provide subsistence agriculture of tropical and subtropical countries, combating geminivirus epidemics is an important task for social and economic stability in these areas (Moffat, 1999). Therefore, increasing knowledge about the viruses, in particular their epidemiology and biodiversity is vital in order to implement preventative strategies.

Isothermal DNA amplification methods are widely used in basic and applied research (Haible et al., 2006; Inoue-Nagata et al., 2004; Jeske et al., 2010; Lizardi et al., 1998; Paprotka et al., 2010a; Paprotka et al., 2010b; Paprotka et al., 2010c). The rolling circle amplification (RCA) technology revolutionized molecular diagnostic, because of its simplicity, high sensitivity, and proof reading against misincorporation of nucleotides. It can be used to amplify sequences without the necessity of previous knowledge of the sequence, once it dispenses the usage of specific oligonucleotides, and its easy compatibility with other detection techniques (Haible et al., 2006; Inoue-Nagata et al., 2004).

Small circular DNA are preferentially amplified by RCA, what makes geminiviruses, which have circular single stranded (ss) DNA genome, ideal substrates for RCA. In addition, during viral replication, through complementary strand synthesis (CSR) of the circular ssDNA, rolling circle replication (RCR), and recombination-dependent replication (RDR), various DNA intermediates are produced (Jeske et al., 2001).

The RCA methodology in combination with restriction fragment length polymorphism (RFLP) recently has been suggested by Haible et al (2006) to be a highly reproducible tool for geminivirus diagnosis. The diagnosis of geminiviral infection by using RCA/RFLP is largely independent of source plant type and origin, viral genome organization or sample preparation method, since neither specific primers nor expensive equipment, like a thermocycler are needed for the reaction.

Geminiviruses

Viruses belonging to the family *Geminiviridae* are composed of circular ssDNA molecules encapsidated in a twinned icosahedral capsid (Fig. 1). Based on their genome organization, sequence similarity, host plants and insect vectors, geminiviruses are classified into four genera: *Mastrevirus*, *Curtovirus*, *Topocuvirus* and *Begomovirus*, the names of which were created from the abbreviations of the type members Maize streak virus (MSV), Beet curly top virus (BCTV), Tomato pseudo curly top virus (TPCTV) and Bean golden mosaic virus (BGMV), respectively (Jeske, 2009). Geminiviruses are transmitted by insects and they can infect monocotyledonous or dicotyledonous plants (Stanley et al., 2005).
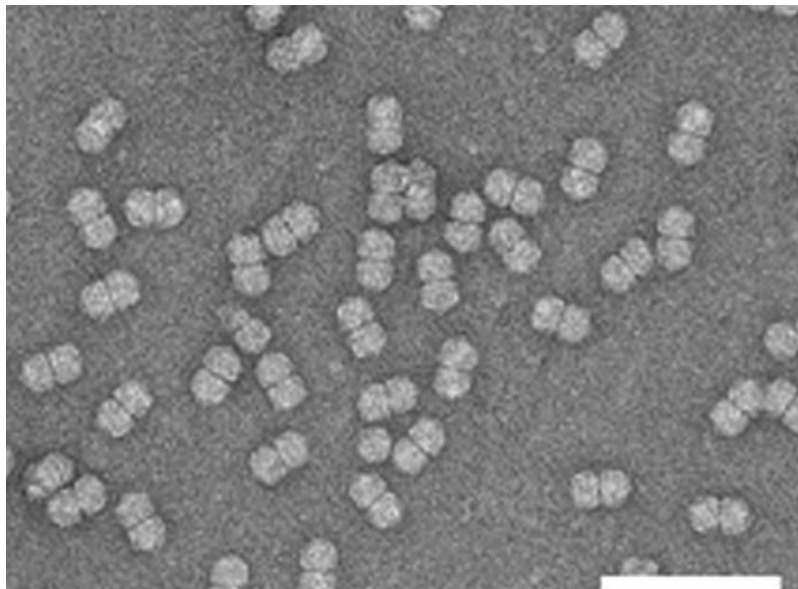

Figure 1. Structure of geminivirions. Purified, negatively stained African cassava mosaic virus (ACMV) particles. Bar = 100 nm. Figure from Jeske (2009).

Begomoviruses

The largest genus of the plant virus family *Geminiviridae* is *Begomovirus*, with members that infect only dicotyledonous plants and are transmitted exclusively by

Bemisia tabaci (whitefly). Based on their genome organization, genetic diversity and geographical distribution, begomoviruses can be divided into two groups, Old World, resembling Europe, Africa, Asia and Australia and New World resembling the American continents (Nawaz-ul-Rehman and Fauquet, 2009).

All New World and some Old World begomoviruses are bipartite, presenting two components DNA A and DNA B. Both components are essential for successful systemic infection. The DNA A harbors information for replication, transcription and encapsidation, whereas DNA B provides movement functions (Jeske, 2009; Stanley et al., 2005). Members of the genera *Mastre-*, *Curto-*, *Topocuvirus* and some members Old World *Begomovirus* carry the necessary genetic information on one DNA molecule, therefore classified as monopartite geminiviruses.

Structural and genomic organization of New World begomoviruses

The name geminivirus, from Latin *geminus*: twin, was given due to the unique twin geometry of these viruses, in which two incomplete icosahedra are attached at a waist, where both halves are twisted against each other by 20 degrees (Fig. 2). On bipartite begomoviruses, each component is encapsidated within one twin particle. Both of the twin particles, one containing the DNA A and the other one containing the DNA B component (Fig. 2) are needed to obtain a full infection (Böttcher et al., 2004).
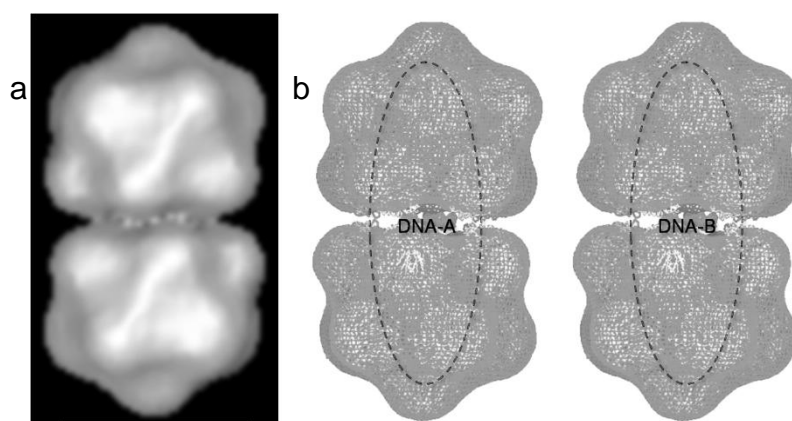
Figure 2. Geminiviral particles. (a) Representations of the three-dimensional ACMV particle. Bar, 20 nm (b) Scheme of bipartite begomovirus. Modified from Böttcher et al (2004).

Bipartite begomoviruses contain a common region (CR) within the intergenic region. The CR has about 200 nucleotides (nt) and usually shows more than 96% identity between DNA A and DNA B components of the same virus, but little similarity between different viruses. All geminiviruses have a highly conserved sequence of nine nucleotides (TAATATTAC), called nonanucleotide sequence, which is located

within the CR, constituting the origin of replication (Ori) as shown on Figure 3 (Jeske, 2009).
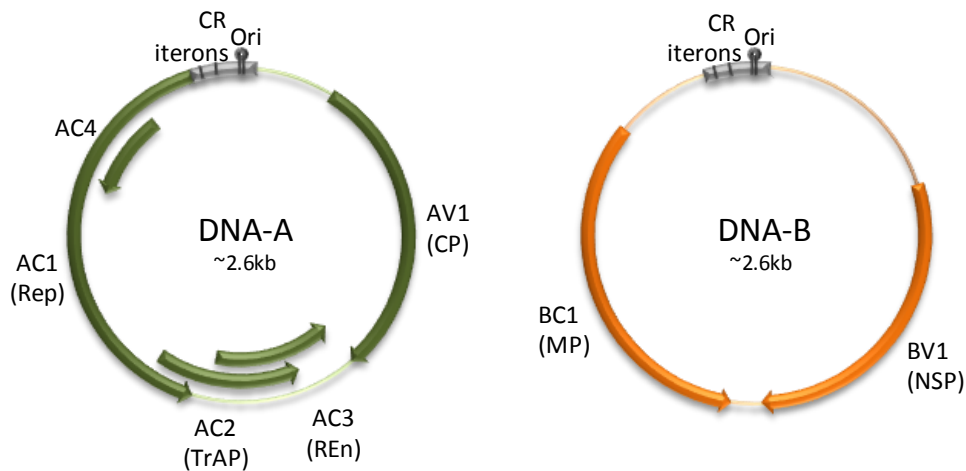
Figure 3. Genomic organization representative for New World begomoviruses: ORF positions and directions of translation are indicated by arrows in DNA A and DNA B, AV1 for the coat protein (CP), AC1 for the Replication-associated protein (Rep), AC2 for the transcriptional-activator protein (TrAP), AC3 for a replication enhancer (REn), AC4 with unknown function, BV1 for a nuclear shuttle protein (NSP) and BC1 for a movement protein (MP). CR indicates the common region, "Ori" the origin of replication and "iterons" locate repeated DNA sequences that represent Rep binding sites.

DNA A encodes five protein open reading frames (ORFs), AC1, AC2, AC3, AC4 and AV1, involved in replication, transcription and packaging, whereas DNA B encodes two protein ORFs, BC1 and BV1, involved in viral movement (Fig. 3). The replication-associated protein (Rep) encoded by the ORF AC1, also named C1 in monopartite geminiviruses, is multifunctional. Rep binds at the "iterons", Rep-binding DNA motifs species-specific in New World begomoviruses (Argüello-Astorga et al., 1994; Argüello-Astorga and Ruiz-Medrano, 2001), which are direct or indirect multiple repeats located in the CR, up and downstream of the hairpin loop of replication origin, auto-regulating its own transcription. Rep is also responsible for the initiation of the replication through helicase (Choudhury et al., 2006; Clerot and Bernardi, 2006) and nicking-closing activity at the Ori (Laufs et al., 1995). In addition it stimulates the DNA synthesis by controlling the cell cycle through interaction with the pRBR factor (plant retinoblastoma-related protein) leading G1-arrested cells to S phase in differentiated tissues (Grafi et al., 1996; Xie et al., 1996; Xie et al., 1995).

The protein encoded by AC2 ORF of bipartite begomoviruses are involved with activation of transcription of the genes AV1 and BV1 (Sunter and Bisaro, 1992), therefore designated as transcription-activating protein (TrAP). In addition, TrAP has been identified as a silencing suppressor (Trinks et al., 2005). The C2 ORF of curto- and topocuviruses, although at the same genomic position, is not homologous to

C2/AC2 of begomoviruses and regardless of having similar function on silencing suppression it does not have transcription activation activity (Hormuzdi and Bisaro, 1995).

The AC3/C3 ORF encodes a replication enhancer protein, REn (Settlage et al., 2005).

The function of AC4 has not being identified for New World begomoviruses so far and has been suggested to be a vestigial gene, which remains conserved due to sequence constraints imposed by the essential overlapping AC1 gene (Pooma and Petty, 1996). In contrast, C4 of other geminiviruses has been suggested to act as silencing suppressor and as determinant of symptoms severity (Bisaro, 2006; Rigden et al., 1994; Vanitharani et al., 2004).

The coat protein (CP; AV1 or V1) is the only protein that composes the virion and has been suggested to play role on the insect transmissibility determination (Briddon et al., 1990). In monopartite geminiviruses, the CP additionally fulfills the nuclear shuttle protein (NSP) transport function (Jeske, 2009). Also in bipartite viruses, CP has been shown to complement some of the respective defects in defective NSP mutants (Qin et al., 1998).

Old World begomoviruses have an AV2 (or V2) ORF, which encodes the pre-coat protein (PCP). Studies using green fluorescent protein (GFP) fused with AV2 suggest that the PCP may function as a redundant transport protein and is possibly a genetic relic of monopartite begomoviruses (Rothenstein et al., 2007).

The DNA B contains two ORFs, one BC1 is encoding the movement protein (MP) for cell-to-cell and long distance transport, the last one via the phloem, and a second, BV1, encoding for the NSP, responsible for transport from the nucleus into the cytoplasm and vice versa (Hehnle et al., 2004; Jeske, 2009; Rojas et al., 2005). Both MP and NSP have been shown to be determinants of pathogenicity and play role on symptom induction (Jeske, 2009; Zhou et al., 2007).

Viral replication

The insect vector introduces the virus to the host plant phloem cells and after uncoating, monomeric circular ssDNA are directed to the nucleus, where viral replication takes place. The ssDNA is then complemented to double-stranded DNA (dsDNA) by host enzymes via complementary strand replication (CSR) generating an

open circular (oc) dsDNA, which can be packaged into nucleosomes, giving rise to covalently closed circular (ccc) dsDNA (Jeske et al., 2001; Preiss and Jeske, 2003).

Open circular dsDNAs as well as ccc dsDNA can be used as template for rolling circle replication (RCR). In order to begin replication of ccc dsDNA, the Rep protein produces a single strand DNA break (nick) in the hairpin loop sequence (TAATATT|AC), binds covalently at the free 5' end and after one complete round of replication it ligates the 5' and 3' end (nicking-closing activity) of the novel DNA molecule (Fig. 4), resulting in circular ssDNA (Preiss and Jeske, 2003).

ssDNAs which have been replicated only partially or have been digested by host nucleases can be repaired via recombination-dependent replication (RDR). In this process, ssDNA is transferred into cccDNA through homologous recombination guaranteeing a full-size genomic template, the ssDNA is then elongated. Concomitantly with RDR, the ssDNA newly synthesized serve as template for replication via CSR, producing dsDNA products (Jeske et al., 2001; Preiss and Jeske, 2003).
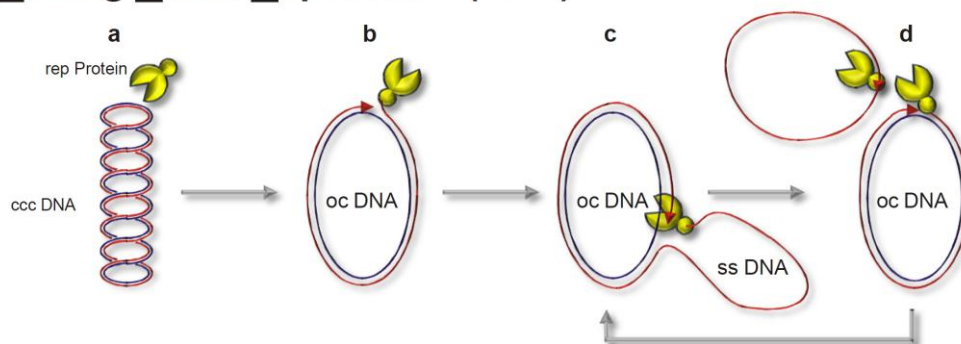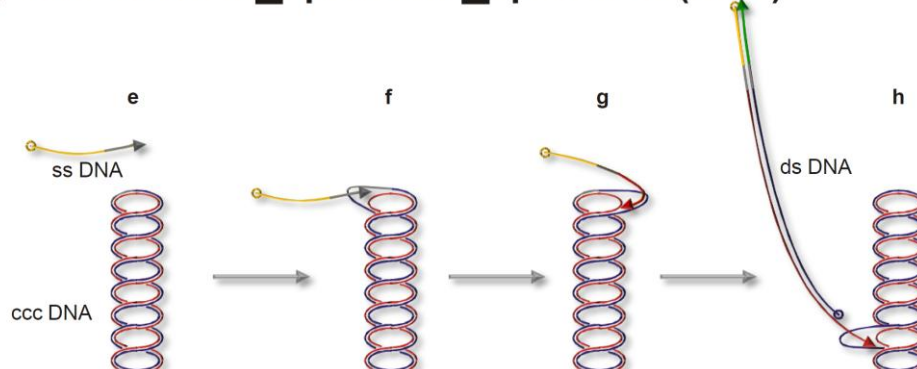
Figure 4. Models of rolling circle replication (RCR) and recombination-dependent replication (RDR). Step a: binding of a replication-associated protein (Rep, encoded by the AC1 ORF) to the origin of replication (ori). Step b: nicking of DNA and covalently binding of Rep to the 5'-end of DNA. Step c: ssDNA displacement and replication. Step d: new nicking, ssDNA closing and Rep release. Step e: incomplete ssDNA interacts with cccDNA at homologous sites. Step f: homologous recombination. Step g: loop migration and ssDNA elongation. Step h: ssDNA elongation and complementary strand synthesis resulting in dsDNA (from: Jeske et al, 2001).

Rolling circle amplification (RCA)

RCA is an isothermal method which uses the DNA polymerase of the *Bacillus subtilis* bacteriophage phi29 (Φ29), a monomeric enzyme with a molecular mass of about 66 kDa, that possesses a polymerase activity located on the C-terminal domain and two degradative activities: pyrophosphorolysis (depolymerization), a process for which its physiological significance is still unclear, and a 3'–5'-exonuclease activity within its N-terminal domain, involved in proofreading function, resulting in an error rate of only 1 in 10^6–10^7, approximately 100 times more accurate than Taq DNA polymerase, which gives an error rate of 1 error in 10^4–10^5 amplified base pairs (Eckert and Kunkel, 1990).

The enzyme has two other intrinsic properties: high processivity (70 kb) and strand displacement ability making unnecessary the participation of accessory proteins and DNA helicases, thus allowing templates as small circular ssDNA to be replicated to nearly unlimited extent using the rolling circle replication mechanism (Blanco et al., 1989; Blanco and Salas, 1996; Dean et al., 2001; Garmendia et al., 1992).

During RCA, random hexamer oligonucleotides bind to the DNA, the polymerase binds to the 3' end of the oligonucleotides and starts the complementary strand synthesis. When the polymerase reaches the 5' end of another hexamer oligonucleotide, it displaces the strand and continues the amplification process. The nascent ssDNA is now target for more hexamer oligonucleotides, where more polymerases can start other amplification processes (Fig. 5).
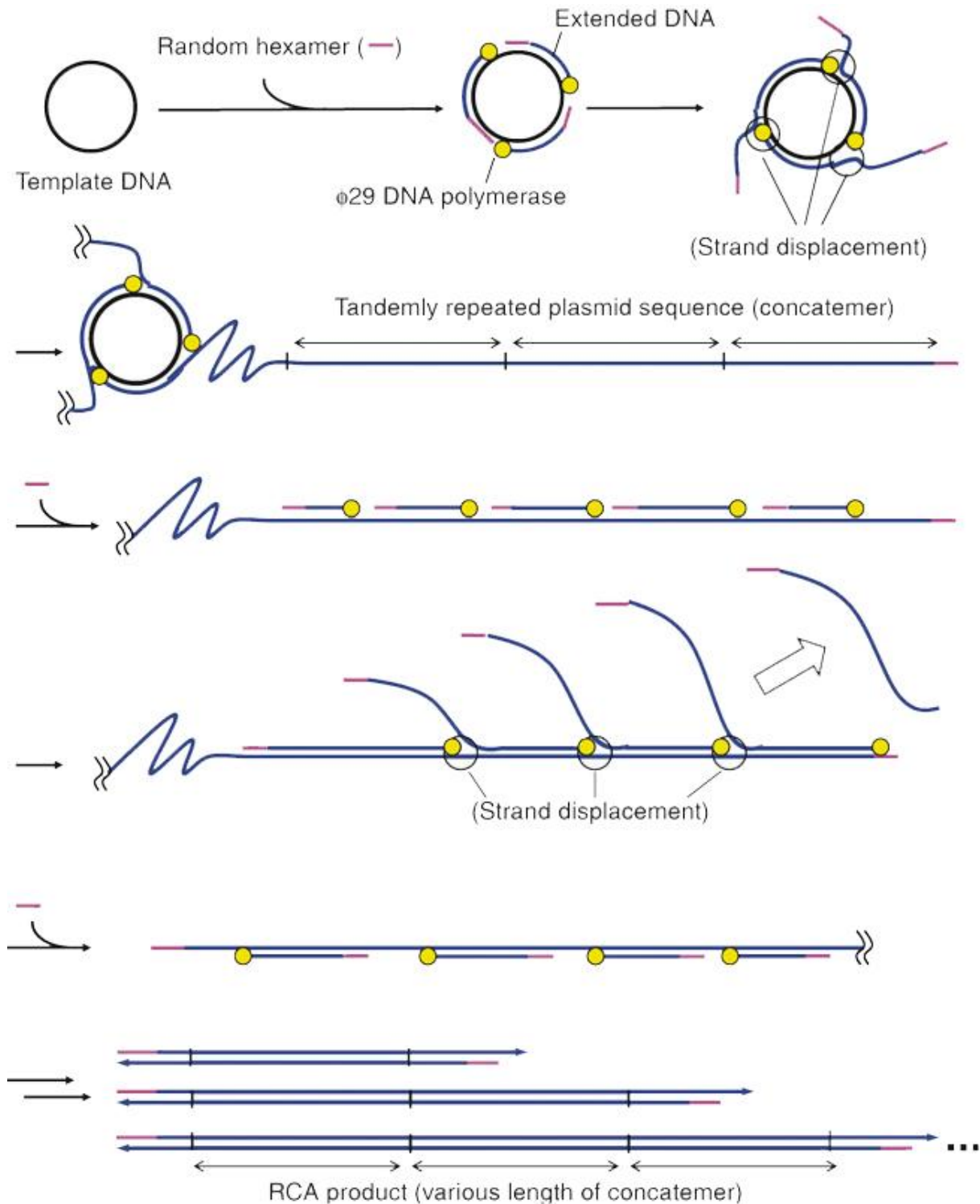
Figure 5. Mechanism of rolling circle amplification (RCA). Random hexamers (NNNNNN) are hybridized with circular DNA, and the resulting double-strand segments function as oligonucleotides in the polymerization reaction carried out by Φ29 DNA polymerase, a unique enzyme with very high strand-displacement activity. As the 'front' of the extending complementary strand of the plasmid encounters double-stranded portions of DNA, the advancing new strand displaces the old one from the template. This extension process covers the entire length of the circular DNA multiple times, resulting in the formation of repeated sequences of the template, called concatemers. The hexamers also hybridize with these concatemers, which become templates in their own right. This extension, however, proceeds only until the terminus of the linear concatemer is reached. The result is the formation of various lengths of double-stranded DNA consisting of repeats of the template sequence. Figure from Fujii et al. (2006).

Phylogenetics

"Phylogenies, or evolutionary trees, are the basic structures necessary to think clearly about differences between species, and to analyze those differences statistically." (Felsenstein, 2004).

Phylogenetics can be defined as the systematic study of the relationships between organisms that leads to a taxonomical classification based on how closely they are related in terms of evolutionary differences. The use of DNA and amino acid sequences to estimate evolutionary history is denominated molecular phylogenetics (Archer, 2008). Thus, molecular characteristics are used to classify organisms placing them on a map of evolutionary relationships known as the phylogenetic tree (Fig. 6), that shows the probable evolution of various organisms. This classification is made through the analysis of rates and patterns of changes observed in DNA and protein sequences (Anonymus, 2004).
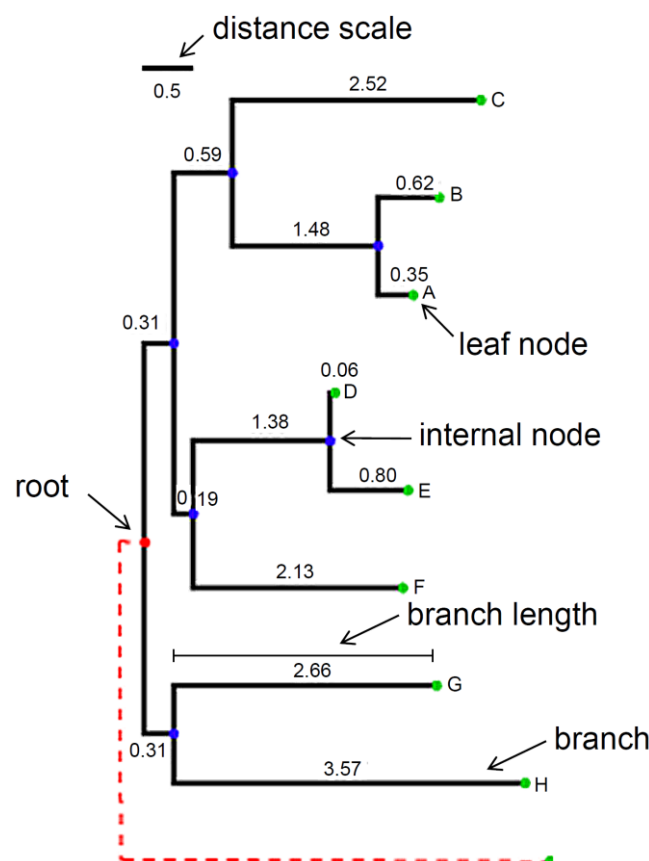
Figure 6. Example of a phylogenetic tree. Randomly generated tree consisting of 8 hypothetical strains labeled A – H. Internal nodes are marked in blue and represent the ancestral relationships between the outer leaf nodes (green). The branch lengths represent the amount of evolutionary change between each of the nodes. The red dotted line represents a hypothetical outgroup that could be used to find the root of the tree (red dot). Modified from Archer (2008).

The crucial components of a molecular phylogenetic tree are nodes, branches, leaves, and the root. Each leaf represents a sequence. Each node represents an ancestral relationship between the sequences. The root represents the common ancestor's sequence of all sequences represented on the tree. Branch length represents the number of changes that have occurred in determined sequence. Branch lengths connecting the nodes represent the amount of change that occurs between each node. The lengths are calculated directly from the alignment (Archer, 2008). Distance scale represents the number of differences between sequences in that scale. The distance between two sequences is the sum of all branch lengths back to the closest common node, for example the distance between "A" and "C" in Figure 6 is $0.35 + 1.48 + 2.52 = 4.35$ divergences between sequences "A" and "C" for instance. Outgroup is a sequence that is less closely related to the sequences represented on the tree than these sequences are to each other. Outgroups enable the rooting of the tree and the identification of correct evolutionary pathways (Anonymus, 2004).

Neighbor joining method

The Neighbor joining (NJ) is one of various methods to reconstruct phylogenetic trees. Trees are constructed by clustering neighboring sequences in a stepwise manner. It begins with an unresolved star-like tree. Each pair of neighbors is joined and the sum of all branches length is calculated of the resultant tree. The pair that yields the smallest sum is considered as the closest neighbors and is thus joined. A new branch is inserted between them and the rest of the tree and the branch length is recalculated. Each step of sequence clustering minimizes the sum of branch lengths and thus examines multiple topologies producing in the end only one possible tree (Saitou and Nei, 1987; Tamura et al., 2004).

To check the reliability of branch order, one of the most commonly used tests is the bootstrap test. Bootstrapping is performed by randomly rearranging columns of the input multiple alignments (Fig. 7). These now constitute a new set of sequences. A tree is then reconstructed with these new sequences. The branch order on the new tree is then compared to the original tree, and each interior branch of the original tree that is different from the bootstrap tree is given a score of 0; all other interior branches are given the value 1. This procedure of resampling the sites and the

subsequent tree reconstruction is repeated several hundred times, and the percentage of times each interior branch is given a value of 1 is noted. This is known as the bootstrap value. In general, the more random samples that support a given branch order on the tree the more reliable that branch order is (Archer, 2008).
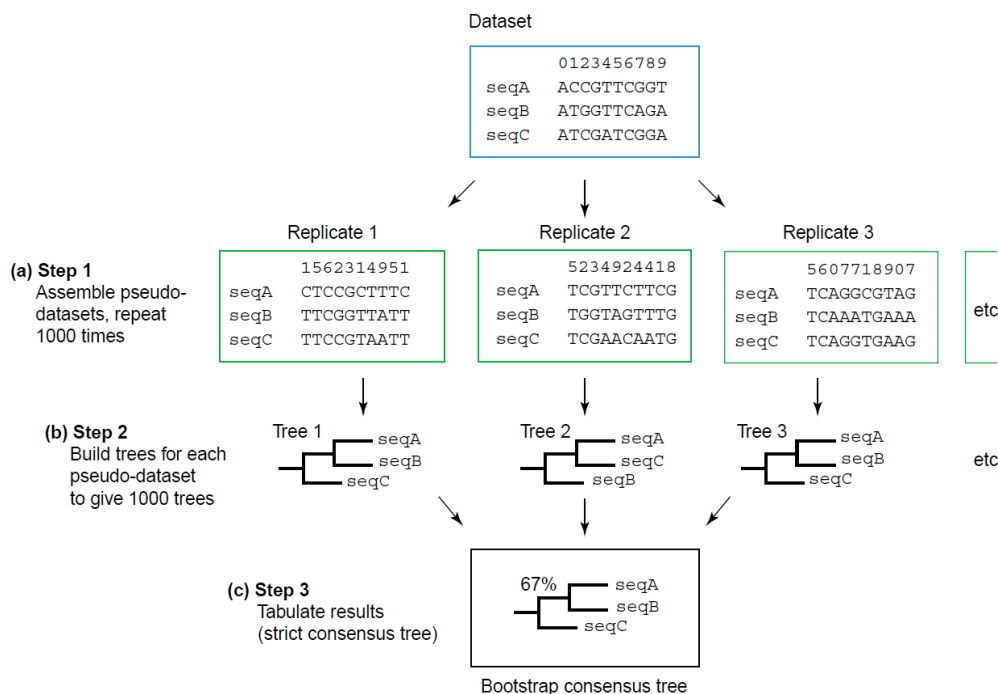
Figure 7. Bootstrap analysis proceeds in three steps. The dataset is randomly sampled with replacement to create multiple pseudo-datasets of the same size as the original (a), three are shown in this example. (b) Individual trees are constructed from each of the pseudo-datasets. (c) Each of the pseudo-dataset trees are scored for which nodes (groupings) appear and how often. In this case, a node uniting seqA plus seqB is found in two of the three replicate trees. This gives a bootstrap support for this grouping of 2/3 or 67%. (Baldauf, 2003).

Pyrosequencing

Pyrosequencing is based on the detection of released pyrophosphate (PPi) during enzymatic DNA synthesis. This process is initiated by the hybridization of a sequencing oligonucleotide to a single-stranded PCR amplicon (the DNA template), which is incubated with several enzymes: DNA polymerase, ATP sulfurylase, luciferase, and apyrase (a nucleotide-degrading enzyme) as well as the substrates, adenosine 5' phosphosulfate (APS) and luciferin. One by one, the four deoxynucleotide triphosphates (dNTPs) are then separately added to the reaction mixture in a known order. When the dNTP is the correct complement to the target strand, the DNA polymerase incorporates that nucleotide onto the end of the nascent strand, resulting in release of PPi at a concentration equimolar to the nucleotide incorporation, whereas non-complementary dNTPs result in no PPi release. ATP

sulfurylase converts PPi to ATP in the presence of (APS). This ATP drives the luciferase-mediated conversion of luciferin to oxyluciferin that generates detectable light signal proportional to the amount of ATP. The amount of light produced in the luciferase-catalyzed reaction can readily be estimated by a suitable light-sensitive device such as a luminometer or a charge-coupled device (CCD) camera and seen as a peak in the raw data output. The height of each peak (light signal) is proportional to the number of nucleotides incorporated. As the process continues, and repeated cycles of deoxynucleotide addition are performed, the complementary DNA strand is formed and the nucleotide sequence is determined from the signals peaks in the data output, as shown in Figure 8 (Marziali and Akeson, 2001; Ronaghi, 2001; Ronaghi et al., 1998).
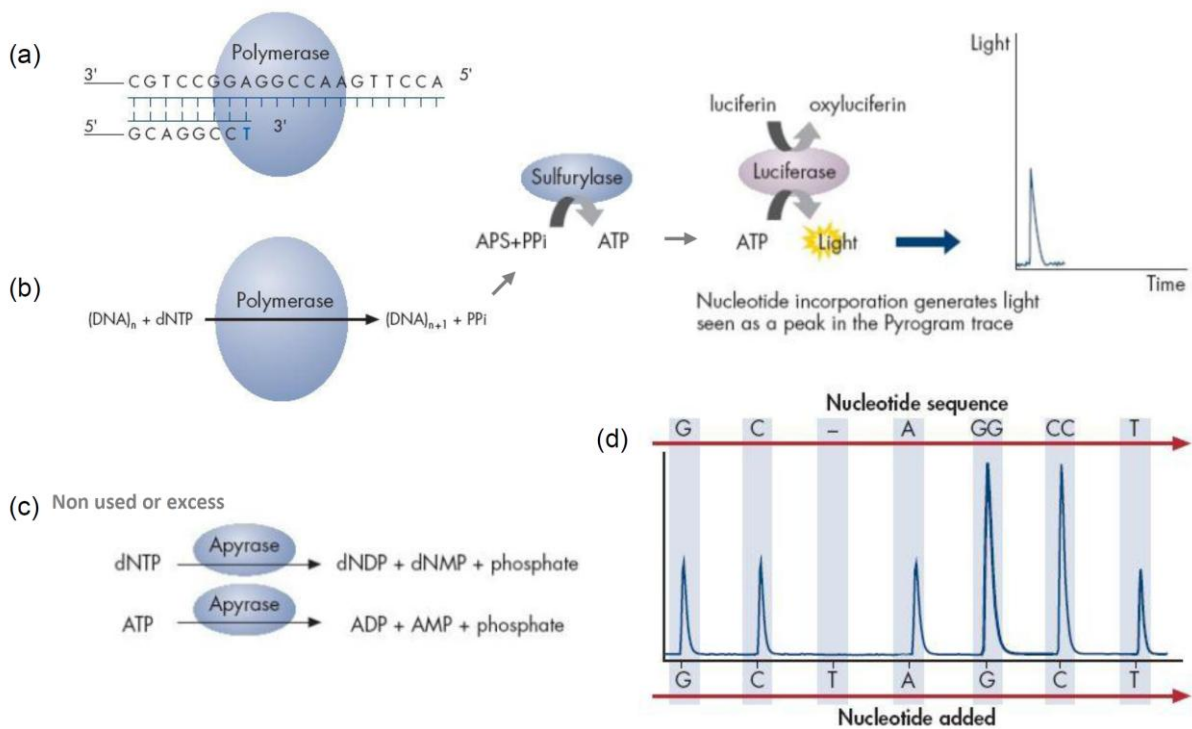
Figure 8. The enzyme cascade of pyrosequencing technology. (a) Incorporation of a nucleotide by polymerase. (b) The conversion of the released PPi to light by enzymatic cascade followed by detection of the generated light by CCD sensors and represented as a peak in the raw data output. (c) Continuous degradation of unincorporated nucleotides and ATP by apyrase. (d) Representation of sequence quantitative decoding according to the sequential addition of nucleotides and the intensity of light, represented by the height of each peak, which is proportional to the number of nucleotides incorporated. (Modified from: Qiagen).

One requirement of this technology is that excess of dNTPs and ATP is removed after each cycle before the next dNTP addition step for incorporation. Therefore and in order to avoid any washing step, a nucleotide-degrading enzyme (apyrase) is used. This enzyme possesses the critical properties of hydrolyzation of all deoxynucleotide triphosphates at approximately the same rate and also hydrolyzation

of ATP. The time for nucleotide degradation is slower than nucleotide incorporation by the polymerase and the rate of ATP synthesis by the sulfurylase is faster than the rate of ATP hydrolysis. This is important to obtain ATP concentrations and light production proportional to the amount of the released PPi (Ronaghi et al., 1998).

A relatively new sequencing method called 454 sequencing, which became commercially available in 2004, uses the principle of pyrosequencing to sequence about 400-600 million base pairs with a single sequencing run, which takes approximately 10 hours to complete. This method is illustrated and explained on Figure 9.
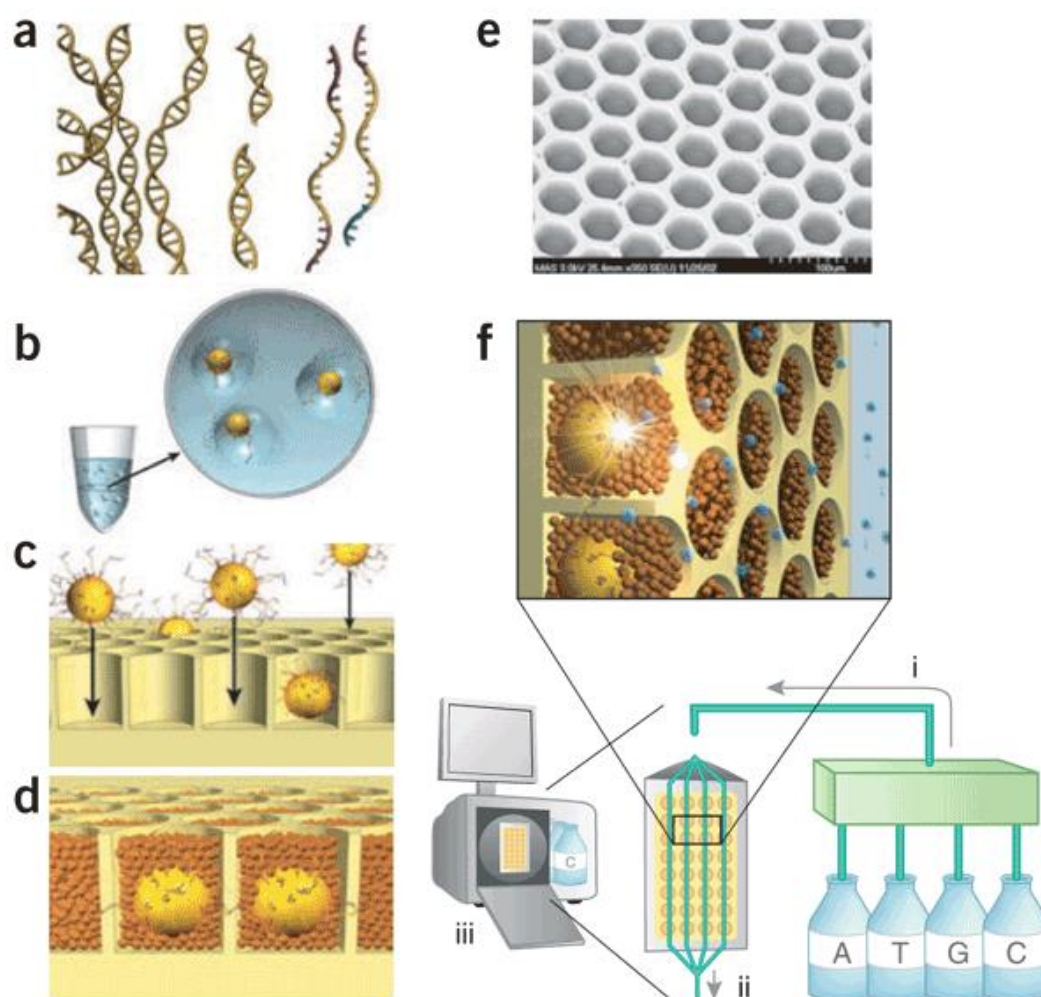
Figure 9. Overview of the 454 sequencing technology (a) Genomic DNA is isolated, fragmented, ligated to adapters and separated into single strands. (b) Fragments are bound to beads under conditions that favor one fragment per bead, the beads are isolated and compartmentalized in the droplets of a PCR-reaction-mixture-in-oil emulsion and PCR amplification occurs within each droplet, resulting in beads each carrying ten million copies of a unique DNA template. (c) The emulsion is broken, the DNA strands are denatured, and beads carrying single-stranded DNA templates are enriched (not shown) and deposited into wells of a fiber-optic slide. (d) Smaller beads carrying immobilized enzymes required for a solid phase pyrophosphate sequencing reaction are deposited into each well. (e) Scanning electron micrograph of a portion of a fiber-optic slide, showing fiber-optic cladding and wells before bead deposition. (f) The 454 sequencing instrument consists of the following major subsystems: a fluidic assembly (i), a flow cell that includes the well-containing fiber-optic slide (ii), a CCD camera-based imaging assembly with its own fiber-optic bundle used to image the fiber-optic slide (iii), and a computer that provides the necessary user interface and instrument control (iii). From: (Rothberg and Leamon, 2008).

Objectives of the study

The objective of this study was to develop, optimize and improve methods to accelerate geminiviruses diagnosis, identification and characterization with the use of RCA/RFLP technique. As a proof of concept of the newly developed and/or improved methods the works described below were successfully carried out. As result information about sequence diversity and biodiversity was generated, improving knowledge about this important group of viruses. The specific objectives of the projects executed were:

- To detect geminiviral infection in symptomatic plant samples originating from South America and Africa; to sequence, using different methods, including 454 sequencing; analyze and characterize the detected viruses by comparison with other previously described begomoviruses.

- To describe the complete sequences of two variants, "yellow vein" and "common strain" of Tomato golden mosaic virus (TGMV) isolated from agroinfected tomato plants (*Solanum lycopersicum*); to investigate the divergence between the two variants and compare the obtained sequences with previously described sequences; to re-investigate TGMV transmission to tomato plants.

- To investigate geminiviruses infecting *Asystasia gangetica* plants causing symptom segregation, in different leaves on a single plant or within a single leaf, and to investigate the infectivity and symptom induction of the detected viruses in alternative host plants *Nicotiana benthamiana* and *Datura stramonium*.

26

Results and discussion

In the first part of this study, described in **Wyant et al.(2011)**, seven symptomatic weed samples collected at three different sites in Bolivia, were analyzed for begomoviral infection by RCA combined with RFLP. All samples were diagnosed positive for begomoviral infection and sequenced by the use of random fragmentation of RCA products followed by cloning and sequencing with universal primers. The complete genomes of eight New World begomoviruses were sequenced.

The sequence analysis showed the presence of four new begomovirus species with the following suggested names: Solanum mosaic Bolivia virus (SoMBoV), Sida mosaic Bolivia virus 1 (SiMBoV1), Sida mosaic Bolivia virus 2 (SiMBoV2) and Abutilon mosaic Bolivia virus (AbMBoV), isolated from a putative *Solanum* plant, *Sida micrantha* (mixed infection), and *Abutilon sp.*, respectively. In addition, one new strain of Sida micrantha mosaic virus (SimMV) isolated from three *S. rhombifolia* samples, for which the name Sida micrantha mosaic virus – Rhombifolia is suggested, and one variant of SimMV were detected. These results were confirmed by phylogenetic analysis.

In this work the use of the partial digestion by restriction enzymes of RCA products combined with a cloning strategy showed to be a very efficient method to obtain full genome sequences. Along with RCA/RFLP analysis, this technique allowed the easy detection of a mixed infection and the identification of the respective begomoviruses on the same sample. Another advantage of this technique in comparison to the classical cloning and sequencing method (Inoue-Nagata et al., 2004; Paprotka et al., 2010a; Paprotka et al., 2010c), in which it is necessary to find specific single cutters for each component to be sequenced and to design new sets of primers after each sequencing step for primer walking, is that only one enzyme and the same procedure are applied for every sample, allowing several samples to be analyzed simultaneously as a routine laboratory procedure. In addition, as the result of partial digestion, random fragments overlapping DNA segments of different sizes are generated, which cover altogether the viral genome. Therefore it is possible after insertion into a vector to obtain very fast the whole genome sequence by simultaneous sequencing of clones using only a single pair of primers located close to the multiple cloning site.

In the second part of the study (**Manuscript 2**), a new approach was used to sequence directly the RCA products from 21 out of 22 symptomatic samples, including beans (*Phaseolus lunatus* and *Phaseolus vulgaris*) and weeds (*Blainvillea rhomboidea, Sida rhombifolia, Sida spp.*, *Cleome affinis, Macropitilium lathyroides*), collected in Northeast of Brazil, which were previously detected to be begomovirus-infected by RCA/RFLP analysis. In this case, the RCA products of positive samples were pooled and sequenced by pyrosequencing. The generated sequences were assigned to the original sample by comparison of RCA/RFLP fragment patterns either predicted by *in silico* digestion or obtained by biochemical digestion. In order to confirm this assignment, fragments of RCA products were cloned and sequenced in a conventional way, as described in **Wyant et al. (2011)** and compared to the pool of sequences obtained from pyrosequencing. Using this new combination of techniques, the complete genomes of New World begomoviruses infecting 20 of the 21 positive samples were determined.

The analysis of the sequences, confirmed by phylogenetic analysis, showed the presence of a total of 5 distinct begomoviruses species infecting the 21 samples, including a new virus species, isolated from a *Sida spp.* sample, for which the name Sida mosaic Alagoas virus (SiMAIV) is suggested. A new strain of Bean golden mosaic virus (BGMV) infecting beans and weed samples was detected. For the new strain, the name Bean golden mosaic virus - Brazil:Alagoas (BGMV-BR:AL) is suggested. In addition, Cleome leaf crumple virus (ClLCrV), Abutilon mosaic Brazil virus (AbMBV) and Blainvillea yellow spot virus (BlYSV) infecting *Cleome affinis*, *S. rhombifolia* and *Blainvillea rhomboidea* samples, respectively, were detected. BGMV was detected in most of the samples, i.e. 13 of 21. This virus is one of the major problems in bean cropping in several countries, including Brazil (Castillo-Urquiza et al., 2008; Jovel et al., 2004), and was not only found in infected bean samples, but also in infected weeds. Like in the first part, also in this part, weeds were investigated for begomoviral infection. Weeds can retain the virus that can be transmitted by the insect vector back to crop plants (Assunção et al., 2006) causing devastation of the crops. Additionally, because they act like virus reservoirs, recombination and generation of new viral genomes is facilitated (Frischmuth et al., 1997; Jovel et al., 2007; Morales and Anderson, 2001). Therefore, the results presented here emphasize the importance of investigating viral infection in weeds.

The pyrosequencing of a pool of RCA products positive for RCA/RFLP analysis has shown to be convenient, quick, efficient and a very sensitive method to diagnose and characterize geminiviruses in several samples simultaneously, including the detection of mixed infection in a single plant sample and the representation of quasispecies population on the pool of generated sequences.

RCA products of the samples from Bolivia were also included in the pool of sequences subjected to pyrosequencing, and hence, the fidelity of the sequences generated by pyrosequencing could be validated by comparison with the sequences obtained in the first part of the study (**Wyant et al., 2011**). Furthermore, this comparison showed a limitation of the method in detecting all sample sequences included in the pool (for details refer to **Manuscript 2**), possibly due to the large number of samples subjected to deep sequencing at the same time.

In the third part of this study, primer walking strategy was used to sequence directly RCA products of two variants of Tomato golden mosaic viruses (TGMV): "common strain" (cs) and "yellow vein" (yv) with the first inducing local up to extensive chlorosis in leaves, and the second, veinal yellowing and chlorosis in close association with the vascular bundles. The two variants were regained from agroinfiltrated tomato plants. This approach allowed the analysis of the complete genome of the original virus sequences without further purification or cloning steps.

Here, the DNA A sequence of csTGMV was described for the first time. All generated sequences were compared with reference sequences from the GenBank to evaluate the differences in genomic sequences of these two variants that induce different phenotypes (see **Manuscript 3** for details). The reference sequence of yvTGMV DNA B shows a truncated ORF of the movement protein BC1, whereas the sequence generated in this study shows an undisrupted ORF of BC1, which was not described before.

The RCA products were also used to biolistically inoculate tomato with TGMV, and despite of the low infectivity ratio (see **Manuscript 3**) of the virus in tomato plants in conformity with the observed by agroinoculation, the bombardment was successful. The generated sequences were directly derived from TGMV extracted from tomato plants, and proved to be different from the cloned sequences, which were derived from TGMV propagated in *Nicotiana benthamiana* plants (Hamilton et al., 1984; von Arnim and Stanley, 1992). These results show that RCA products can be

conveniently and successfully used for direct sequencing with the primer walking strategy, avoiding selection of individual clones, thus allowing to access directly the virus population of the plant, and therefore, to obtain a sequence that reflects the sequence of the majority of virus genomes present in the plant. Moreover plant infection can also be performed without any further purification steps with the viruses represented in the original extraction.

Finally, in the last part of the work, described in **Manuscript 4**, *Asystasia gangetica* plants, originally collected in West Africa, showing different symptoms typical of geminiviral infection, such as mosaic and yellow veining, segregated in different branches of a plant or even on the same leaf, were analyzed for the presence of geminiviruses. The approach used in this part was the same as described in the second part of this study (details in **Manuscript 2**) namely, by the use of RCA/RFLP followed by pyrosequencing of pooled samples.

The sequence analysis revealed the presence of four begomoviral components. Two cognate DNAs A and B with genomic organization, that resembled bipartite Old World begomoviruses and an additional DNA A that due to the genomic organization and the absence of the DNA B cognate is suggested to be a monopartite begomovirus, and its defective DNA. According to the sequence analysis, which was confirmed by phylogenetic analysis, the two begomoviruses belong to different species and were classified as new viruses, for which the names Asystasia begomovirus 1 (ABgV1) for DNA A1 and DNA B1 and Asystasia begomovirus 2 (ABgV2) for DNA A2 and its defective DNA are suggested.

Investigation of the relation between induced symptoms and the presence of the different components showed that plants with mosaic symptoms harbored all components, whereas in samples showing only yellow vein symptoms, DNA A2 was clearly predominant, either alone or together with DNA dA2.

The influence of different light regimes in symptom induction over six months on new cuttings showing predominance of distinct symptoms was also evaluated (**Manuscript 4**). In all cases, irrespective of the symptom predominance observed initially, after six months the plantlets developed both mosaic and yellow vein symptoms. The presence of both viruses was confirmed again by RCA/RFLP. The symptoms and components were again segregated as described for the original plant from which the cuttings were made.

The RCA/RFLP combined with pyrosequencing was shown again to be a useful and sensitive tool to characterize and generate information about geminiviral infection.

As for TGMV, the RCA products from samples with mixed symptoms were successfully used to infect other plant species biolistically in order to evaluate the viral infectivity in other hosts, more specifically *Datura stramonium* and *Nicotiana benthamiana*. RCA/RFLP analysis confirmed the systemic infection and the presence of the components DNA A1, B1 and A2 in four out of ten inoculated *D. stramonium* and one of ten *N. benthamiana* plants. The infected plants showed stunting symptoms and *D. stramonium* plants showed additionally curling and neither of them showed segregation of symptoms, suggesting that this segregation of the virus within the plant is related to the host physiology. As presented in the third part of this study and **Manuscript 3***,* the direct bombardment of plants with RCA products has shown to be a convenient and quicker tool for biolistic inoculation, in agreement with previous studies that successfully used RCA products for biolistic inoculation of geminiviruses (Haible et al., 2006; Jeske et al., 2010).

Specific antibodies (Givord et al., 1994; Harrison et al., 2002; Perveen et al., 2010) and polymerase chain reaction (PCR) using either specific or degenerated primers (Briddon and Markham, 1995; Li et al., 2004; Rouhibakhsh et al., 2008) are alternative methods used for detection of geminiviral infection in plant samples. RCA revolutionized the diagnosis of geminiviral infection and moreover combined with RFLP has been shown to be a suitable and sensitive tool for geminiviral infection screening (Haible et al., 2006; Inoue-Nagata et al., 2004; Kumar et al., 2008; Paprotka et al., 2010c) due to the several advantages over other methods. The most striking advantage, when screening for viral infection, is that it is not necessary to know the viral sequence to be amplified since this technique enables amplification of all small circular DNAs present in the sample.

It is known that RCA/RFLP can give false positive results because RCA amplifies every small circular DNA present in the sample, and therefore, it amplifies plant material like mitochondrial mini plasmids (Homs et al., 2008), which is true for some plants. However, in this study, such false positive results were not observed. All samples with a positive RCA/RFLP pattern were confirmed to be positive for viral infection by sequencing.

Taken together, the results obtained in this work demonstrate the usefulness and convenience of the methods developed and/or improved on direct sequencing, shot gun cloning and pyrosequencing.

Presentation of the results as manuscripts for publication

Manuscript 1:

The genomes of four novel begomovirus species and a new *Sida micrantha* mosaic virus strain from Bolivian weeds (published)

Patrícia Soares Wyant, Diether Gotthardt, Benjamin Schäfer, Björn Krenz and Holger Jeske

Manuscript 2:

Circomics of geminiviruses as exemplified for begomoviruses infecting bean crops and weeds in the Northeast of Brazil

Patricia Soares Wyant, Stephan Strohmeier, Benjamin Schäfer, Björn Krenz, Iraildes Pereira Assunção, Gaus Silvestre de Andrade Lima and Holger Jeske

Manuscript 3:

Tomato golden mosaic virus back in tomatoes

Patricia Soares Wyant, Sigrid Kober, Holger Jeske, Christina Wege

Manuscript 4:

Segregating geminiviruses in Asystasia gangetica induce distinct symptoms

Patricia Soares Wyant, Stephan Strohmeier, Benjamin Schäfer, Annika Allinger, Björn Krenz, Rob W Briddon and Holger Jeske

ANNOTATED SEQUENCE RECORD

The genomes of four novel begomoviruses and a new Sida micrantha mosaic virus strain from Bolivian weeds

**Patrícia Soares Wyant · Diether Gotthardt ·
Benjamin Schäfer · Björn Krenz · Holger Jeske**

Abstract *Begomovirus* is the largest genus within the family *Geminiviridae* and includes economically important plant DNA viruses infecting a broad range of plant species and causing devastating crop diseases, mainly in subtropical and tropical countries. Besides cultivated plants, many weeds act as virus reservoirs. Eight begomovirus isolates from Bolivian weeds were examined using rolling-circle amplification (RCA) and restriction fragment length polymorphism (RFLP). An efficient, novel cloning strategy using limited *Sau*3A digestion to obtain tandem-repeat inserts allowed the sequencing of the complete genomes. The viruses were classified by phylogenetic analysis as typical bipartite New World begomoviruses. Four of them represented distinct new virus species, for which the names Solanum mosaic Bolivia virus, Sida mosaic Bolivia virus 1, Sida mosaic Bolivia virus 2, and Abutilon mosaic Bolivia virus are proposed. Three were variants of a new strain of Sida micrantha mosaic virus (SimMV), SimMV-rho [BoVi07], SimMV-rho[Bo:CF1:07] and SimMV-rho [Bo:CF2:07], and one was a new variant of a previously described SimMV, SimMV-MGS2:07-Bo.

Geminiviruses constitute a large family of plant DNA viruses infecting a broad range of plants. They cause

P. S. Wyant · D. Gotthardt · B. Schäfer · B. Krenz ·
H. Jeske (✉)
Department of Molecular Biology and Plant Virology,
Institute of Biology, University of Stuttgart,
Pfaffenwaldring 57, 70550 Stuttgart, Germany
e-mail: holger.jeske@bio.uni-stuttgart.de

devastating diseases in important crops that provide staple food in subsistence agriculture or are otherwise economically important, such as beans, cassava, cotton, maize, pepper, sugar beet, sweet potato, and tomato [19, 21, 22]. Within this family, the genus *Begomovirus* is composed of viruses that infect dicotyledonous plants, are transmitted by one whitefly species (*Bemisia tabaci*), and possess one or two genome components, named DNA-A and DNA-B [24]. Over the last four decades, agricultural intensification and the emergence and prevalence of a new and more aggressive biotype of the insect vector (*Bemisia tabaci* biotype-B) have facilitated an increase in begomovirus populations and their expansion to new plant hosts in Latin America. This has contributed to the emergence of new and more virulent viruses, producing an increase in frequency and severity of disease [3]. Besides cultivated plants, many weed species are hosts for begomoviruses [6], from which they can be transmitted to crops. Thus, weeds may act as virus reservoirs, facilitating recombination and generation of new viral genomes [11, 16, 22]. The characterization of weed-infecting begomoviruses is, therefore, important for elucidating their ecological and evolutionary behavior [6]. Within Latin America, Bolivia has climate zones that are suitable for begomovirus dissemination by its insect vector. However, even though beans, soybeans, cotton, and potatoes [2, 22], all of which are known to be geminivirus hosts, are amongst the main agricultural products cultivated in Bolivia, reports on begomoviral incidence are limited to a bean-infecting virus (bean golden mosaic virus - BGMV) [22]. A survey of Bolivian weeds using rolling-circle amplification (RCA)-based detection technology has now increased the number of known begomovirus isolates to eight.

RCA utilizes the DNA polymerase of the *Bacillus subtilis* bacteriophage Φ29 [7, 8], which possesses both

🖄 Springer
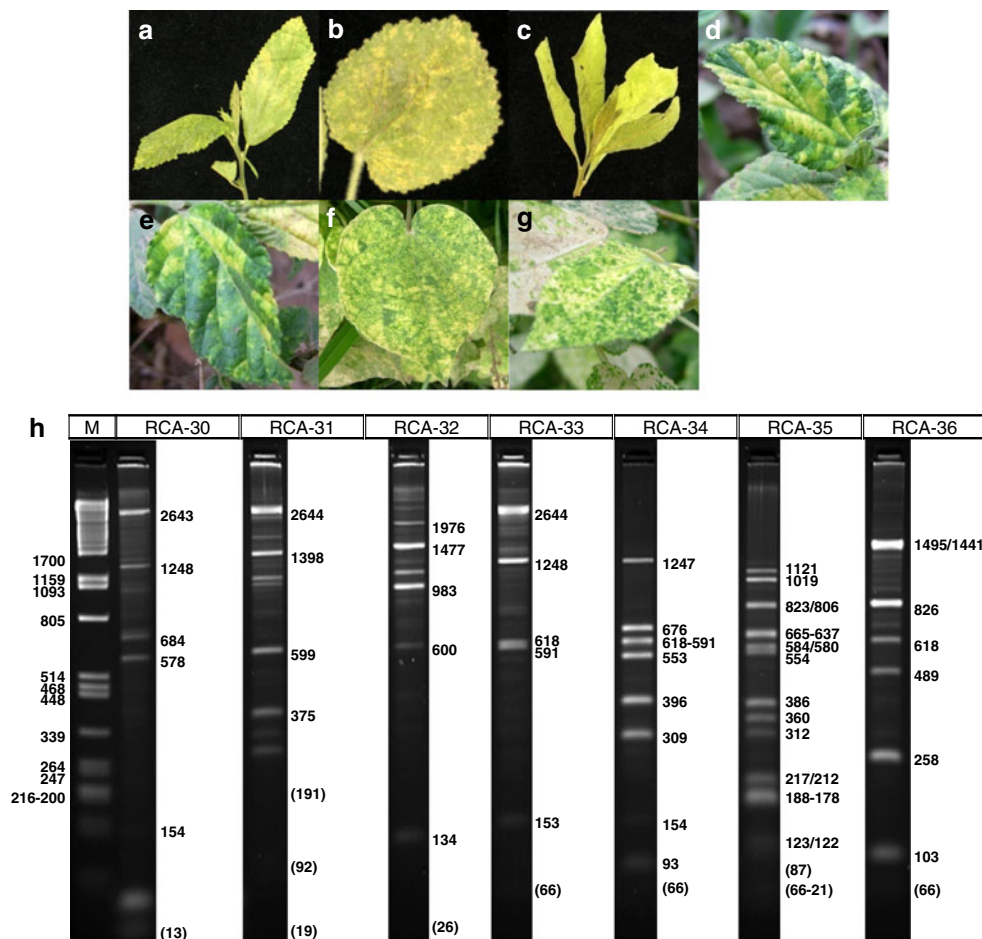
Fig. 1 Mosaic symptoms in plants from which the virus samples were isolated. *Sida rhombifolia* (**a**, **d**, **e**; source of RCAs-30, -33 and -34, respectively), *Sida micrantha* (**b**; source of RCA-31 and **f**; source of RCA-35.1 as well as RCA-35.2), a putative *Solanum* plant that was difficult to identify because it had been reduced in size by grazing animals (**c**; source of RCA-32); *Abutilon sp* (**g**; source of RCA-36). Plant samples (**a-c**) were collected in Boyuibe-Villamontes, (**d** and **e**) in Cerro Fraile and (**f** and **g**) in Santa Cruz. The respective RCA samples were diagnosed using RFLP patterns of *Hpa*II restriction (**h**). The fragments sizes predicted from the sequences are indicated where they fit to the electrophoretic mobility observed. Values in brackets represent bands too small to be resolved in this gel system; unassigned fragments may have resulted from polymorphism within the viral quasispecies or from unidentified begomovirus components. The genomic organization was determined by sequencing sample RCA-30 (i), which is representative for all the newly described viruses. ORF positions and directions of translation are indicated by arrows in DNA-A and DNA-B, AV1 for the coat protein (CP), AC1 for the Rep protein, AC2 for a transcriptional activator protein (TrAP), AC3 for a replication enhancer (REn), AC4 for a protein with unknown function in New World begomoviruses, BV1 for a nuclear shuttle protein (NSP), and BC1 for a movement protein (MP). CR indicates the common region; Ori, the origin of replication; and "iterons", repeated DNA sequences that are putative Rep binding sites

polymerase and strand-displacement activity, allowing circular templates to be amplified preferentially [18, 28, 29]. The simplicity, high sensitivity, and proofreading activity of the procedure were exploited to amplify small circular DNA sequences without initial knowledge of putative viral sequences present in the samples [12, 18] and combined with RFLP, yielding clear data on the viral agents found in different Bolivian samples [12, 23].

Seven symptomatic plant samples, all showing mosaic symptoms, were collected from weeds from different regions in Bolivia in November 2007 and analyzed by RCA/RFLP: from Boyuibe and Villamontes (S21°01'28.20''; W63°22'44.00''; 680m) RCAs-30, -31 and -32, isolated from *Sida rhombifolia*, *Sida micrantha* and a putative *Solanum* plant, respectively; from Cerro Fraile (S18°17'00.00''; W63°40'33.70''; 1338m) RCAs-33 and -34, both isolated from *Sida rhombifolia*, and from Santa Cruz (S 17° 51' 201''; W 63° 14' 465''; 1622m) RCA-35 and RCA-36, isolated from *Sida micrantha* and *Abutilon sp*, respectively (Fig. 1a-g). Total DNA was extracted using a CTAB-based method [17] and the presence of small circular DNA was confirmed by RCA followed by digestion with the restriction enzyme *Hpa*II (New England Biolabs, Frankfurt, Germany) and RFLP pattern analysis [12]. In order to obtain partial tandem repeats of viral full-length DNAs, a novel and efficient strategy was

Table 1 Virus names, abbreviations and accession numbers of the sequences identified in this study

ID	Virus name	Abbreviation	Acession numbers	
			DNA-A	DNA-B
RCA-30	Sida micrantha mosaic virus - Rhombifolia [Bolivia:Boyuibe-Villamontes:2007]	SimMV-rho[BoVi07]	HM585431	HM585432
RCA-31	Sida micrantha mosaic virus [Brazil:Mato grosso do sul:2:2007]- [Bolivia]	SimMV-MGS2:07-Bo	HM585433	HM585434
RCA-32	Solanum mosaic Bolivia virus	SoMBoV	HM585435	HM585436
RCA-33	Sida micrantha mosaic virus - Rhombifolia [Bolivia:Cerro Fraile 1:2007]	SimMV-rho[Bo:CF1:07]	HM585437	HM585438
RCA-34	Sida micrantha mosaic virus - Rhombifolia [Bolivia:Cerro Fraile 2:2007]	SimMV-rho[Bo:CF2:07]	HM585439	HM585440
RCA-35.1	Sida mosaic Bolivia virus 1	SiMBoV1	HM585441	HM585442
RCA-35.2	Sida mosaic Bolivia virus 2	SiMBoV2	HM585443	HM585444
RCA-36	Abutilon mosaic Bolivia virus	AbMBoV	HM585445	HM585446

developed, which is applicable for all geminiviruses without the necessity to search for appropriate restriction sites. The RCA products were partially digested with *Sau*3AI (or its isoschizomer *Bfu*CI, New England Biolabs): 1-1.5 μg of RCA product was treated with 0.4 units of the enzyme for 4 min at 37°C, and the reaction was stopped by adding 1 mM EDTA and 0.1% SDS (final concentration) and heating the solution for 20 min to 65°C. Of the resulting fragments, those of approximately 2.8 to 5.0 kb were eluted from a 0.8% agarose electrophoresis gel using a GFX DNA Gel Band Purification Kit according to the manufacturer's instructions (GE Healthcare, Munich, Germany) and eluted in 30 μl of sterile water. The DNA from 3 μl was inserted into *Bam*HI-cut pGreen0029 [14] plasmid, and *E. coli* DH5α cells were transformed with the recombinant DNA. Plasmids containing inserts were sequenced using universal primers (M13F/R), either in our laboratory (CEQ 8000 Genetic Analysis System, Beckman-Coulter) or commercially by Macrogen (Seoul, South Korea).

The complete sequences obtained for DNAs A and B were assembled (Contig Assembly Program, Bioedit software) [13] and analyzed with BLAST [1]. For phylogenetic analysis, the sequences were compared using codaln software [25] in order to optimize the alignment of the most closely related viral sequences from the international databases retrieved with BLAST (version 2.2.22+, January 15, 2010) [30], and neighbor joining trees were calculated using the algorithms included in MEGA 4.0 [26], with 1000 bootstrap replications each. Virus names and accession numbers are listed in Supplementary Table 1.

All seven plant samples contained small circular DNA as inferred from RCA/RFLP analysis (Fig. 1h). After partially digesting the RCA products, inserting their fragments into pGreen0029, and sequencing positive clones, all of the

viruses showed the typical genome organization of bipartite New World begomoviruses, with five open reading frames (ORFs) on their DNA-A molecules, encoding replication-associated protein (Rep), transcriptional activator protein (TrAP), replication enhancer protein (REn), AC4, and coat protein (CP), and with two ORFs on their DNA-B molecules, encoding movement protein (MP) and nuclear shuttle protein (NSP).

The sample from RCA-35, isolated from *Sida micrantha*, revealed a mixed infection with two different begomoviruses. Using the threshold level of 89% sequence identity (SI) for differentiating begomovirus species according to ICTV [9], and comparing the sequences to other geminiviral sequences in the database, one isolate from RCA-32, isolated from a putative *Solanum* plant, the two isolates from sample RCA-35 and one isolate from RCA-36, extracted from *Abutilon sp*, were identified as members of four new begomoviral species (Table 1 and Supplementary Table 2). Their component A/B pairs shared >96% SI within the common region (CR) and had identical iterons, the Rep-binding DNA motifs, which are species-specific in New World begomoviruses [4, 5]. These results support the conclusion that each isolate belongs to a new begomoviral species with two cognate DNA components. The other four isolates, one from *Sida micrantha* (RCA-31) and three from *Sida rhombifolia* (RCAs-30, -33 and -34) showed >90% DNA-A SI to previously reported sequences of Sida micrantha mosaic virus (SimMV) that had been detected in Brazil and therefore belong to this species. The isolate from sample RCA-31(*S. micrantha*) is a variant of Sida micrantha mosaic virus [Mato Grosso do Sul2:2007] (SimMV-MGS2:07; FN436005) with 95% SI. Isolates from samples RCA-30, -33 and -34 (*S. rhombifolia*) shared 94% and 93% SI with SimMV isolate 5157 (accession number EU908733.1) and 96% SI to each other

and are variants of a new strain. We did not observe any statistically significant recombination events within the viral sequences described in this study using TOPALI software [20].

Comparison of the translated amino acid sequences for the ORFs of the new isolates (Supplementary Table 3) revealed that, for the DNA-A molecules, AV1 is the most conserved ORF (77-99% SI), and AC4 the most variable one (40-94% SI), and that for the DNA-B molecules, BC1 (83-98% SI) is more conserved than BV1 (72-93% SI).

In a phylogenetic analysis of the DNA-A components, RCA-35 (*S. micrantha*) and RCA-36 (*Abutilon sp*) group with cleome leaf crumple virus, but the bootstrap value was not significant. RCA-32 (*Solanum*) was found to be closely related to soybean blistering mosaic virus (SbBMV), which has been detected in soybean crops in Argentina [10], forming a distinct cluster with a bootstrap value of 99% (Fig. 2). The three variants of the new SimMV strain, RCAs -30, 33, and 34 (*S. rhombifolia*), cluster together with the other SimMV sequences (100% bootstrap value).

Fig. 2 Neighbor-joining trees of DNA-A and DNA-B for the isolates detected in this study (RCA-30, -31, -32, -33, -34, -35.1, -35.2, -36) compared to the sequences with highest sequence identity from the international database GenBank (January 15, 2010) and using African cassava mosaic virus (ACMV) as an outgroup. Numbers next to the branch points indicate bootstrap values (1,000 replicates) above 50% (0.5). Newly described isolates are highlighted and marked with asterisks if they represent new virus species. The corresponding identifier numbers, names, abbreviations and accession numbers are listed in Supplementary Table 1
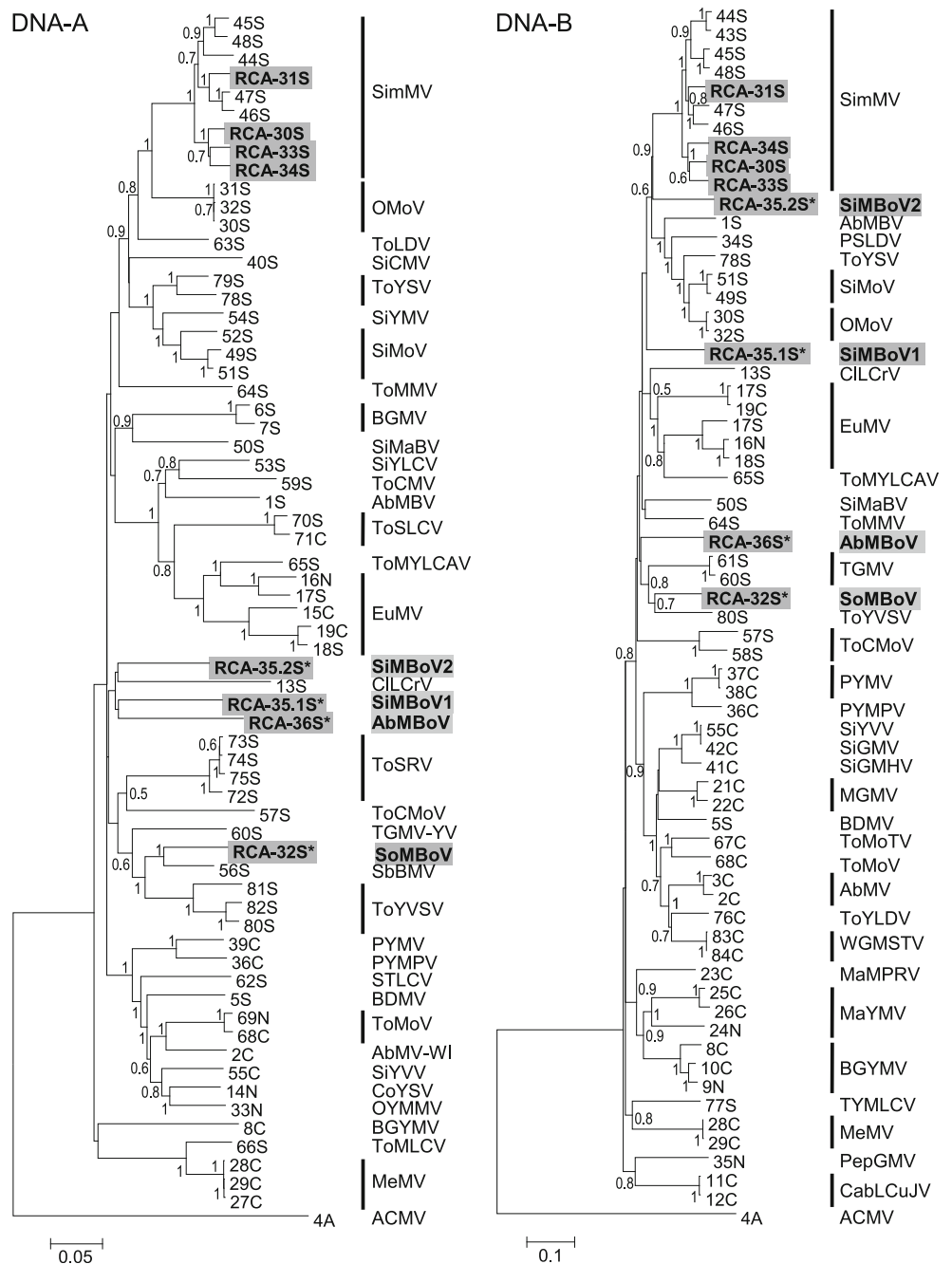
The isolate from sample RCA-31 (*S. micrantha*), classified by pairwise comparison analysis as a variant of SimMV-MGS2:07, clustered with SimMV-MGS1:07 and SimMV-MGS2:07 (100% bootstrap value), as expected. Similarly, the DNA-B components of samples RCA-35 (*S. micrantha*) and RCA-36 (*Abutilon sp*) did not group with any other sequence, while that of RCA-31 (*S. micrantha*) and those of RCAs-30, -33 and -34 (*S. rhombifolia*) were located within the SimMV cluster (Fig. 2). In contrast, the DNA-B sequence from sample RCA-32 (*Solanum*) seems most closely related to tomato yellow vein streak virus (ToY-VSV), but this may be because there is no SbBMV DNA-B sequence available in the database, and ToYVSV is a close relative of SbBMV (Fig. 2). The results of RCA-RFLP fragment pattern analysis (Fig. 1h) were in agreement with what was expected from *in silico Hpa*II restriction of the determined sequences [27], with only a few unexpected bands, which may have resulted from polymorphic restriction sites or amplification of additional small circular DNA components present in the sample, such as mito-chondrial DNA or unidentified begomovirus components.

As we can see from the sequence and phylogenetic analysis, there is a large amount of diversity amongst begomoviruses infecting plants all over America. Sequences from South America tend to group together, showing a closer relationship to each other than to sequences from Central and North America, which also show closer relationship to each other than to South America sequences, suggesting that the viruses from one geographical area are evolving from other viruses present in that area rather than being spread over long distances.

Considering their sequences, phylogenetic relationships, symptoms and the hosts from which the new viruses were isolated (Fig. 1a-g), we propose the name Solanum mosaic Bolivia virus (SoMBoV) for the species represented by RCA-32 (*Solanum*), Sida mosaic Bolivia virus 1 (SiMBoV1) for the species represented by RCA-35.1 (*S. micrantha*), Sida mosaic Bolivia virus 2 (SiMBoV2) for the species represented by RCA-35.2 (*S. micrantha*) and Abutilon mosaic Bolivia virus (AbMBoV) for the species represented by RCA-36 (*Abutilon sp*). Although we have not transmitted the viral isolates back to the original host plants and therefore cannot assign any symptom name with confidence, we are following the current policy of the ICTV study group (pers. communication) with these names. For the new SimMV strain isolates from RCAs-30, -33 and -34 (*S. rhombifolia*), we suggest the name Sida micrantha mosaic virus - Rhombifolia for the same reasons, although our previous research results did not support the symptom name [15].

References

1. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402
2. Anonymous (2009) The World Factbook, vol 2010
3. Arguello-Astorga G, Ascencio-Ibáñez JT, Dallas MB, Orozco BM, Hanley-Bowdoin L (2007) High-frequency reversion of geminivirus replication protein mutants during infection. J Virol 81:11005–11015
4. Argüello-Astorga GR, Guevara-González RG, Herrera-Estrella LR, Rivera-Bustamante RF (1994) Geminivirus replication origins have a group-specific organization of iterative elements: a model for replication. Virology 203:90–100
5. Argüello-Astorga GR, Ruiz-Medrano R (2001) An iteron-related domain is associated to Motif 1 in the replication proteins of geminiviruses: identification of potential interacting amino acid-base pairs by a comparative approach. Arch Virol 146:1465–1485
6. Assunção IP, Listik AF, Barros MCS, Amorim EPR, Silva SJC, Izael OS, Ramalho-Neto CE, Lima GSA (2006) Genetic diversity of begomoviruses infecting weeds in Northeastern Brazil. Planta Daninha 24:239–244
7. Blanco L, Bernad A, Lazaro JM, Martin G, Garmendia C, Salas M (1989) Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. Symmetrical mode of DNA replication. J Biol Chem 264:8935–8940
8. Dean FB, Nelson JR, Giesler TL, Lasken RS (2001) Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. Genome Res 11:1095–1099
9. Fauquet CM, Briddon RW, Brown JK, Moriones E, Stanley J, Zerbini M, Zhou X (2008) Geminivirus strain demarcation and nomenclature. Arch Virol 153:783–821
10. Fernandes FR, Cruz ARR, Faria JC, Zerbini FM, Aragao JL (2009) Three distinct begomoviruses associated with soybean in central Brazil. Arch Virol 154:1567–1570
11. Frischmuth T, Engel M, Lauster S, Jeske H (1997) Nucleotide sequence evidence for the occurrence of three distinct whitefly-transmitted, Sida-infecting bipartite geminiviruses in Central America. J Gen Virol 78:2675–2682
12. Haible D, Kober S, Jeske H (2006) Rolling circle amplification revolutionizes diagnosis and genomics of geminiviruses. J Virol Methods 135:9–16
13. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98
14. Hellens RP, Edwards EA, Leyland NR, Bean S, Mullineaux PM (2000) pGreen: a versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation. Plant Mol Biol 42:819–832
15. Jeske H, Gotthardt D, Kober S (2010) In planta cloning of geminiviral DNA: the true Sida micrantha mosaic virus. J Virol Methods 163:301–308
16. Jovel J, Preiß W, Jeske H (2007) Characterization of DNA intermediates of an arising geminivirus. Virus Res 130:63–70
17. Kleinow T, Nischang M, Beck A, Kratzer U, Tanwir F, Preiss W, Kepp G, Jeske H (2009) Three C-terminal phosphorylation sites

in the Abutilon mosaic virus movement protein affect symptom development and viral DNA accumulation. Virology 390:89–101

18. Lizardi PM, Huang X, Zhu Z, Bray-Ward P, Thomas DC, Ward DC (1998) Mutation detection and single-molecule counting using isothermal rolling-circle amplification. Nat Genet 19:225–232

19. Mansoor S, Briddon RW, Zafar Y, Stanley J (2003) Geminivirus disease complexes: an emerging threat. Trends Plant Sci 8:128–134

20. Milne I, Wright F, Rowe G, Marshal DF, Husmeier D, McGuire G (2004) TOPALi: software for automatic identification of recombinant sequences within DNA multiple alignments. Bioinformatics 20:1806–1807

21. Moffat AS (1999) Geminiviruses emerge as serious crop threat. Science 286:1835

22. Morales FJ, Anderson PK (2001) The emergence and dissemination of whitefly-transmitted geminiviruses in Latin America. Arch Virol 146:415–441

23. Paprotka T, Boiteux LS, Fonseca MEN, Resende RO, Jeske H, Faria JC, Ribeiro SG (2010) Genomic diversity of sweet potato geminiviruses in a Brazilian germplasm bank. Virus Res 149:224–233

24. Stanley J, Bisaro DM, Briddon RW, Brown JK, Fauquet CM, Harrison BD, Rybicki EP, Stenger DC (2005) Geminiviridae. In: Fauquet CM, Mayo MA, Maniloff J, Desselberger U, Ball LA (eds) Virus taxonomy. VIIIth report of the international committee on taxonomy of viruses. Elsevier/Academic Press, London, pp 301–326

25. Stocsits RR, Hofacker IL, Fried C, Stadler PF (2005) Multiple sequence alignments of partially coding nucleic acid sequences. BMC Bioinformatics 6:160

26. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

27. Vincze T, Posfai J, Roberts RJ (2003) NEBcutter: a program to cleave DNA with restriction enzymes. Nucleic Acids Res 31:3688–3691

28. Zhang DY, Brandwein M, Hsuih TC, Li H (1998) Amplification of target-specific, ligation-dependent circular probe. Gene 211:277–285

29. Zhang DY, Zhang W, Li X, Konomi Y (2001) Detection of rare DNA targets by isothermal ramification amplification. Gene 274:209–216

30. Zhang Z, Schwartz S, Lukas W, Web M (2000) A greedy algorithm for aligning DNA sequences. J Comput Biol 7:203–214

Supplementary Table 1. Sequences retrieved from GenBank: Virus names, abbreviations and accession numbers of sequences used for the phylogenetic analysis. A – Africa; C – Central America; S – South America; N – North America.

ID	Virus name	Abbreviation	DNA A	DNA B	Origin
1S	Abutilon mosaic Brazil Virus	AbMBV	FN434438	FN434439	Brazil
2C	Abutilon mosaic virus - [Germany]	AbMV-[DE]	X15983.2	X15984.3	West India
3C	Abutilon mosaic virus - South African isolate	AbMV-SA		AM886130.1	South Africa
4A	African cassava mosaic virus	ACMV	NC_001467.1	NC_001468.1	Africa
5S	Bean dwarf mosaic virus - [Colombia:1987]	BDMV-[CO:87]	M88179.1	M88180.1	Colombia
6S	Bean golden mosaic Virus-[Brazil:Campinas1:1978]	BGMV-[BR:Cam1:78]	M88686.1		Brazil
7S	Bean golden mosaic virus-[Brazil:SAdG 7:soybean:2008]	BGMV-[BR:SAdG7:soybean:2008]	FJ665283.1		Brazil
8C	Bean golden yellow mosaic virus	BGYMV	NC_001439.1	NC_001438.1	Puerto Rico - Japan
9N	Bean golden yellow mosaic virus [United States of America:Homestead:2005]	BGYMV-[US:Hom:05]		DQ119825.1	USA
10C	Bean golden yellow mosaic virus-[Dominican Republic:1987]	BGYMV-[DO:87]		L01636.1	Dominican Republic
11C	Cabbage leaf curl Jamaica virus - [Jamaica:CUc3:2005]	CabLCuJV-[JM:Cuc3:05]		DQ178609.1	Jamaica
12C	Cabbage leaf curl Jamaica virus - [Jamaica:CUc32:2005]	CabLCuJV- [JM:Cuc32:05]		DQ178611.1	Jamaica
13S	Cleome leaf crumple virus	ClLCrV	FN435999	FN436000	Brazil
14N	Corchorus yellow spot virus - [Mexico:Yucatan:2005]	CoYSV-[MX:Yuc:05]	DQ875868.1		Mexico
15C	Euphorbia mosaic virus - Peru	EuMV-Peru	AM886131.1		Peru
16N	Euphorbia mosaic virus - [Jamaica:Wissadula]	EuMV-Wis	FJ407052.1	EU740969.1	Jamaica
17S	Euphorbia mosaic virus - A [Mexico:Yucatan:2004]	EuMV-[MX:Yuc:04]	DQ318937.1	DQ318938.1	Mexico
18S	Euphorbia mosaic virus [Mato Grosso do Sul1:2007]	EuMV- [MGS1:07]	FN435995	FN435996	Brazil
19C	Euphorbia mosaic virus [Mato Grosso do Sul2:2007]	EuMV- [MGS2:07]	FN435997	FN435998	Brazil
20S	Euphorbia mosaic virus clone EF2H	EuMV-EF2H		FJ807783.1	Cuba
21C	Macroptilium golden mosaic virus-[Jamaica:Wissadula:August Town]	MGMV-JM:W:AT		EU158097.1	Jamaica
22C	Macroptilium golden mosaic virus-[Jamaica1:SpanishTown]	MGMV-JM1:ST		FJ981727.1	Jamaica
23C	Macroptilium mosaic Puerto Rico virus - [Puerto Rico:1990]	MaMPRV-[PR:90]		AY044134.1	Puerto Rico
24N	Macroptilium yellow mosaic Florida virus - [United States of America:Florida:1985]	MaYMV-[US:Flo:85]		AY044136.1	USA
25C	Macroptilium yellow mosaic virus isolate Hope Pastures	MaYMV-HP		EF585289.1	Jamaica
26C	Macroptilium yellow mosaic virus isolate pmrsrfb14	MaYMV-pmrsrfb14		EF585292.1	Jamaica
27C	Merremia mosaic virus - Puerto Rico [Puerto Rico:]	MeMV-PR[PR:]	AF068636.2		Puerto Rico

28C	Merremia mosaic virus isolate PR4	MeMV-PR4	DQ644558.1	DQ644559.1	Puerto Rico
29C	Merremia mosaic virus isolate PR80	MeMV-PR80	DQ644557.1	DQ644560.1	Puerto Rico
30S	Okra mottle virus - [Brazil:okra:6319]	OMoV-[BR:Ok:6319]	EU914817.1	EU914818.1	Brazil
31S	Okra mottle virus - [Brazil:okra:6328]	OMoV-[BR:Ok:6328]	EU914819.1		Brazil
32S	Okra mottle virus strain - [Brazil:Sag8:Soy:08]	OMoV-[BR:Sag8:Soy:08]	FJ686695.1	FJ686696.1	Brazil
33N	Okra yellow mosaic Mexico virus - [Mexico:Santa Teresa 2 :3abg]	OYMMV-[MX:ST:3abg]	EF591629.1		Mexico
34S	Passionfruit severe leaf distortion virus - [Brazil:LNS2:Pas:01]	PSLDV-[BR:LNS2:Pas:01]		FJ972768.1	Brazil
35N	Pepper golden mosaic virus – United States of America [Mexico:Tamaulipas]	PepGMV-US[MX:Tam]		AF499442.1	Mexico
36C	Potato yellow mosaic Panama virus - [Panama:Divisa:Tomato]	PYMPV-[PA:Div:Tom]	Y15034.1	Y15033.1	Panama
37C	Potato yellow mosaic virus – Tomato [Puerto Rico:Tomato:2004]	PYMV-To[PR:Tom:04]		AY965898.1	Puerto Rico
38C	Potato yellow mosaic virus – Tomato[Guadeloupe:Tomato]	PYMV-To[GP:Tom]		AY120883.1	Caribbean
39C	Potato yellow mosaic virus - Trinidad [Trinidad & Tobago:Tomato]	PYMV-TT[TT:Tom]	AF039031.1		Trinidad & Tobago
40S	Sida common mosaic virus isolate BR:Coi4:07	SiCMV-BR:Coi4:07	EU710751.1		Brazil
41C	Sida golden mosaic Honduras virus - [Honduras]	SiGMHV-[HN]		Y11098.1	Honduras
42C	Sida golden mosaic virus	SiGMV		AJ250731.1	Honduras
43S	Sida micrantha mosaic virus - [Brazil:A1B3]	SimMV-[BR:B1]		AJ557452.1	Brazil
44S	Sida micrantha mosaic virus - [Brazil:A2B2]	SimMV-[BR:A2B2]	AJ557451.1	AJ557453.1	Brazil
45S	Sida micrantha mosaic virus - [Brazil:okra] isolate 5157	SimMV-[5157]	EU908733.1	EU908734.1	Brazil
46S	Sida micrantha mosaic virus [Mato Grosso do Sul1:2007]	SimMV-MGS1:07	FN436003	FN436004	Brazil
47S	Sida micrantha mosaic virus [Mato Grosso do Sul2:2007]	SimMV-MGS2:07	FN436005	FN436006	Brazil
48S	Sida micrantha mosaic virus strain BR:Sag3:Soy:08	SimMV-BR:Sag3:Soy:08	FJ686693.1	FJ686694.1	Brazil
49S	Sida micrantha mosaic virus, isolate SimMV-[SP77]	SimMV-[SP77]	FN557522.1	FN557523.1	Brazil
50S	Sida mosaic Brazil virus	SiMBV	FN436001	FN436002	Brazil
51S	Sida mottle virus - Micrantha [Brazil:A1B3]	SiMoV-mic[BR:A1B3]	AJ557450.1	AJ557454.1	Brazil
52S	Sida mottle virus - Rhombifolia [Brazil:Vicosa1:1999]	SiMoV-rho[BR:Vic1:99]	AY090555.1		Brazil
53S	Sida yellow leaf curl virus isolate BR:Coi3:07	SiYLCV-BR:Coi3:07	EU710750.1		Brazil
54S	Sida yellow mosaic virus-[Brazil:Vicosa2:1999]	SiYMV-[BR:Vic2:99]	AY090558.1		Brazil
55S	Sida yellow vein virus - [Honduras]	SiYVV-[HN]	Y11099.1	Y11101.1	Honduras
56S	Soybean blistering mosaic virus - [Argentina:NOA:2005]	SbBMV-[AR:NOA:05]	EF016486.1		Argentina
57S	Tomato chlorotic mottle virus - Minas Gerais [Brazil:Igorape1:1996]	ToCMoV-MG[BR:Iga1:96]	DQ336353.1	DQ336354.1	Brazil

58S	Tomato chlorotic mottle virus-Bahia[Brazil:Seabra1:1996]	ToCMoV-BA[BR:Sea1:96]		AF491306.1	Brazil
59S	Tomato common mosaic virus isolate BR:Coi22:07	ToCMV-BR:Coi22:07	EU710754.1		Brazil
60S	Tomato golden mosaic virus-[Brazil:Common:1984]	TGMV-[BR:Com:84]	K02029.1	K02030.1	Brazil
61S	Tomato golden mosaic virus-[Brazil:Yellow Vein]	TGMV-[BR:YV]		M73794.1	Brazil
62S	Tomato leaf curl Sinaloa virus – Nicaragua:Santa Lucia]	ToLCSinV-[NI:SL]	AJ608286.1		Nicaragua
63S	Tomato leaf distortion virus isolate BR:Pda4:05	ToLDV-BR:Pda4:05	EU710749.1		Brazil
64S	Tomato mild mosaic virus isolate BR:Pda58:05	ToMMV-BR:Pda58:05	EU710752.1	EU710753.1	Brazil
65S	Tomato mild yellow leaf curl Aragua virus - [Venezuela:10]	ToMYLCAV-[VE:10]	AY927277.1	EF547938.1	Venezuela
66S	Tomato mosaic leaf curl virus	ToMLCV	AY508991.1		Venezuela
67C	Tomato mottle Taino virus - [Cuba]	ToMoTV-[CU]		AF012301.1	Cuba
68C	Tomato mottle virus - [Puerto Rico:2004]	ToMoV-[PR:04]	AY965900.1	AY965901.2	Puerto Rico
69N	Tomato mottle virus-[United States of America:Florida:1989]	ToMoV-[US:Flo:89]	L14460.1		USA
70S	Tomato severe leaf curl virus - Guatemala [Guatemala:Sansirisay:1996]	ToSLCV-GT[GT:San:96]	AF130415.2		Guatemala
71C	Tomato severe leaf curl virus - Guatemala[Mexico:Rioverde1:2005]	ToSLCV-GT[MX:Rio1:05]	DQ347946.1		Brazil
72S	Tomato severe rugose virus - [Brazil:Uberlandia2:2000]	ToSRV-[BR:Ube2:2000]	AY029750.1		Brazil
73S	Tomato severe rugose virus strain Petrolina de Goias	ToSRV-PetrolinadeGoias	DQ207749.1		Brazil
74S	Tomato severe rugose virus strain PJU-Sao Paulo	ToSRV-PJU-SP	FJ824808.1		Brazil
75S	Tomato severe rugose virus strain Sumare	ToSRV-Sumare	EU086591.2		Brazil
76C	Tomato yellow leaf distortion virus clone EF5N	ToYLDV		FJ999999.1	Cuba
77S	Tomato yellow margin leaf curl virus - [Venezuela:Merida:57]	TYMLCV-[VE:Mer:57]		AY508994.1	Venezuela
78S	Tomato yellow spot virus - [Brazil:Bicas2:1999]	ToYSV-[BR:Bic2:99]	DQ336350.1	DQ336351.1	Brazil
79S	Tomato yellow spot virus from Argentina	ToYSV-[AR]	FJ538207.1		Argentina
80S	Tomato yellow vein streak virus isolate Ba-3	ToYVSV-Ba-3	EF417915.1	EF417916.1	Brazil
81S	Tomato yellow vein streak virus isolate G-22 from Brazil	ToYVSV-G22	EF459696.1		Brazil
82S	Tomato yellow vein streak virus isolate T06	ToYVSV-T06	GQ387369.1		Argentina
83C	Wissadula golden mosaic St Thomas Virus	WGMSTV		EU158095.1	Jamaica
84C	Wissadula golden mosaic St Thomas Virus clone W132BFL5	WGMSTV-W132BFL5		GQ355487.1	Jamaica

Supplementary Table 2. Pairwise comparisons of the nucleotide sequences obtained in this study and those with the highest nucleotide identities retrieved from GenBank.

	RCA-30	RCA-31	RCA-32	RCA-33	RCA-34	RCA-35 1	RCA-35 2	RCA-36
SimMV-MGS1:07	92	94	73	91	91	78	78	76
SimMV-MGS2:07	93	95	74	92	92	79	78	76
SimMV-5157	94	93	74	93	93	79	79	77
SbBMV	78	77	86	76	76	80	78	75
ToLDV-BR:Pda4:05	82	83	77	83	82	79	79	77
SiYLCV-BR:Coi3:07	77	77	72	77	77	74	76	73
SiMoV-rho[BR:Vic1:99]	82	82	74	81	81	79	79	76
ToSRV-Sumare	76	76	78	76	76	76	75	74
TMoV	74	74	76	73	74	74	76	71
RCA-30	x	92	74	96	96	80	79	76
RCA-31	x	x	74	92	92	79	79	76
RCA-32	x	x	x	74	74	76	77	73
RCA-33	x	x	x	x	96	80	79	76
RCA-34	x	x	x	x	x	80	79	76
RCA-35 1	x	x	x	x	x	x	80	78
RCA-35 2	x	x	x	x	x	x	x	77

Supplementary Table 3. Pairwise comparisons of the amino acid sequences for the *in silico* translated open reading frames (Fig. 1).

% identity			RCA-30	RCA-31	RCA-32	RCA-33	RCA-34	RCA-35.1	RCA-35.2	RCA-36
AV1		RCA-30	100	97	80	97	99	84	92	85
		RCA-31	-	100	80	97	98	84	90	84
		RCA-32	-	-	100	80	80	77	83	79
		RCA-33	-	-	-	100	99	84	93	86
		RCA-34	-	-	-	-	100	85	93	86
		RCA-35.1	-	-	-	-	-	100	86	85
		RCA-35.2	-	-	-	-	-	-	100	88
		RCA-36	-	-	-	-	-	-	-	100
AC1		RCA-30	100	91	72	96	96	77	74	76
		RCA-31	-	100	71	91	91	75	73	77
		RCA-32	-	-	100	72	73	69	80	71
		RCA-33	-	-	-	100	96	77	75	77
		RCA-34	-	-	-	-	100	77	75	77
		RCA-35.1	-	-	-	-	-	100	74	75
		RCA-35.2	-	-	-	-	-	-	100	77
		RCA-36	-	-	-	-	-	-	-	100
AC2		RCA-30	100	88	64	94	89	66	68	69
		RCA-31	-	100	66	85	83	66	64	68
		RCA-32	-	-	100	62	67	81	71	72
		RCA-33	-	-	-	100	86	66	70	69
		RCA-34	-	-	-	-	100	72	72	73
		RCA-35.1	-	-	-	-	-	100	77	78
		RCA-35.2	-	-	-	-	-	-	100	76
		RCA-36	-	-	-	-	-	-	-	100
AC3		RCA-30	100	87	81	95	95	76	82	77
		RCA-31	-	100	83	89	89	77	78	79
		RCA-32	-	-	100	81	80	83	81	80
		RCA-33	-	-	-	100	95	78	82	77
		RCA-34	-	-	-	-	100	77	80	76
		RCA-35.1	-	-	-	-	-	100	77	76
		RCA-35.2	-	-	-	-	-	-	100	75
		RCA-36	-	-	-	-	-	-	-	100
AC4		RCA-30	100	70	45	94	91	65	66	58
		RCA-31	-	100	44	69	67	69	56	51
		RCA-32	-	-	100	47	46	42	44	40
		RCA-33	-	-	-	100	94	84	63	56
		RCA-34	-	-	-	-	100	83	66	56
		RCA-35.1	-	-	-	-	-	100	63	62
		RCA-35.2	-	-	-	-	-	-	100	51
		RCA-36	-	-	-	-	-	-	-	100
BV1		RCA-30	100	91	76	93	92	79	74	77
		RCA-31	-	100	75	90	91	80	74	76
		RCA-32	-	-	100	76	75	75	72	77
		RCA-33	-	-	-	100	90	81	73	77
		RCA-34	-	-	-	-	100	79	73	76
		RCA-35.1	-	-	-	-	-	100	79	78
		RCA-35.2	-	-	-	-	-	-	100	77
		RCA-36	-	-	-	-	-	-	-	100
BC1		RCA-30	100	98	83	98	97	87	91	87
		RCA-31	-	100	84	97	97	88	92	87
		RCA-32	-	-	100	83	83	87	83	86
		RCA-33	-	-	-	100	97	87	90	86
		RCA-34	-	-	-	-	100	86	91	86
		RCA-35.1	-	-	-	-	-	100	85	86
		RCA-35.2	-	-	-	-	-	-	100	86
		RCA-36	-	-	-	-	-	-	-	100

Circomics of geminiviruses: begomoviruses in bean crops and weeds of Northeastern Brazil

Patricia Soares Wyant[1], Stephan Strohmeier[1], Benjamin Schäfer[1], Björn Krenz[1], Iraildes Pereira Assunção[2], Gaus Silvestre de Andrade Lima[2] and Holger Jeske[1]

[1]Institute of Biology, Department of Molecular Biology and Plant Virology, University of Stuttgart, Pfaffenwaldring 57, 70550 Stuttgart, Germany.

[2]Setor de Proteção de Plantas do Centro de Ciências Agrárias da Universidade Federal de Alagoas. BR 104 Norte, Km 85, Rio Largo, AL. CEP 57100-000

Abstract

Begomovirus is the largest genus of the *Geminiviridae* family. Its members are transmitted by whiteflies, able to infect dicotyledonous plants and are responsible for several diseases in economically important crops worldwide, such as cassava, cotton, bean, pepper and tomato. In Brazil, the most affected crops are beans and tomatoes. In order to accelerate their diagnosis and identification, we investigated the genome structures of small circular DNAs by combining rolling circle amplification (RCA) with restriction fragment length polymorphism (RFLP) and pyrosequencing, collectively called "circomics". Here, we exemplify the usefulness of this strategy for 22 symptomatic plant samples collected in the Northeast of Brazil from beans and weeds samples, which act as perennial reservoirs of viruses. In 21 virus-positive samples, the completely determined sequences revealed a predominance of a new strain of Bean golden mosaic virus (BGMV), which was found in 13 samples, followed by a variant of Cleome leaf curl virus (ClLCrV), detected in four samples. Two samples contained Blainvillea yellow spot virus (BlYSV) variants, one *Sida rhombifolia* revealed infection with a variant of Abutilon mosaic Brazil virus (AbMBV), and one sample infecting a *Sida spp.*, a new virus showing 83% homology with the AbMBV, for which we suggest the name of Sida mosaic Alagoas virus (SiMAlV).

Keywords: Brazil, geminivirus, begomovirus, Bean golden mosaic virus, rolling circle amplification (RCA).

Introduction

The *Geminiviridae* family consists of a large and diverse group of plant viruses that are composed by one or two circular, single-stranded DNA (ssDNA) genomes, which are encapsidated in incomplete twinned icosahedral virions. The members of the *Geminiviridae* can be classified into four genera according to their host-range properties, insect vector and genome organization (Stanley et al., 2005). The *Begomovirus* is the largest genus of this family. Begomoviruses are transmitted only by whitefly (*Bemisia tabaci*) and infect dicotyledonous plants. New World begomoviruses have bipartite genomes referred to as DNA A, involved with virus replication, regulation of gene expression, and particle encapsidation, and DNA B, that encodes proteins responsible for cell-to-cell and long-distance movement within the plant (Jeske, 2009; Rojas et al., 2005).

Begomoviruses are one of the most economically important group of plant viruses due to their high incidence and disease severity in field crops such as cassava, cotton, bean, pepper and tomato in tropical and subtropical regions (Mansoor et al., 2003; Moffat, 1999; Morales and Anderson, 2001). In Brazil, the most severely affected crops are beans and tomatoes (Castillo-Urquiza et al., 2008; Faria and Maxwell, 1999). It has been suggested that the emergence of tomato-infecting begomoviruses was the result of horizontal transfer of begomoviruses that infect wild hosts or weed by a new biotype of the whitefly vector (*Bemisia tabaci* subtype B), which emerged on the early 1990s and is capable to transmit the viruses for new plant hosts. Once present in the new host, these indigenous viruses would have rapidly evolved via recombination and pseudo recombination, giving rise to the species currently detected in the field (Castillo-Urquiza et al., 2008).

With the aim of accelerate diagnosis and identification of begomoviruses, we have combined rolling circle amplification (RCA) with restriction fragment length polymorphism (RFLP) and pyrosequencing. As the combination of these methods made the detection of all small circular DNAs within a sample feasible, we called it "circomics". In this study, we examined 22 symptomatic plant samples, including beans (6 *Phaseolus lunatus* and 3 *Phaseolus vulgaris*) and weeds (4 *Cleome affinis*, 3 *Macroptilium lathyroides*, 2 *Blainvillea rhomboidea*, 1 *Sida rhombifolia*, 1 *Sida spp.* and 2 non specified *Leguminoseae*) samples collected in the Northeast of Brazil for

the presence of begomoviruses. The detected viruses were characterized by conventional and pyrosequencing for phylogenetics.

Material and methods

Samples

Twenty two Brazilian samples, eighteen from symptomatic plants of different sites of Alagoas, two from Pernambuco and two from Bahia (Supplementary Table 1) were collected on FTA Cards® (Whatman), from which DNA was recovered for RCA according to Haible et al. (2006). Three disks (3 mm diameter each) were cut out from the cards, mixed with 300 µl of TE [10 mM Tris-HCl; 0.1 mM ethylenediamine tetraacetic acid (EDTA), pH8.0] and incubated at room temperature for 5 minutes. The supernatants were discarded and the disks were sequentially washed with 70% ethanol, TE, wash buffer (30% ethanol; 100 mM Tris-HCl, pH 8.0; 0.25% Nonidet P40), and 70% ethanol, dried at room temperature, and the nucleic acids were eluted in 10 mM Tris-HCl, pH8.0 with 9 µl per mm diameter of the card. The samples were collected and nucleic acid material was extracted by Drs. Iraildes Pereira Assunção and Gaus Silvestre de Andrade Lima, and the nucleic acid was placed on FTA Cards® (Whatman) by from the Universidade Federal de Alagoas, Brazil.

RCA

Amplification of circular DNA was performed using TempliPhi™ Kit (GE Healthcare, Munich, Germany) following the manufacturer's protocol. Briefly, 2 µl (10 ng to 20 ng) of total nucleic acids were dissolved in 5 µl of sample buffer, denatured for 3 min at 95 °C and cooled down for 1 min on ice followed by addition of 5 µl reaction buffer and 0.2 µl enzyme mix. The amplification was for 16-20 h at 30 °C and stopped for 10 min at 65 °C.

Diagnosis of infection

Aliquots of the RCA products corresponding to 300 ng DNA per 10 µl (final volume) were digested by the restriction enzymes *Hpa*II or *Bfu*CI (New England Biolabs,

Frankfurt, Germany) for 2 h, according to supplier's recommendation. Fragments were resolved by electrophoresis in 2% agarose gels following standard protocols (Sambrook and Russell, 2001) and stained with 0.5 µg/ml ethidium bromide. Fragment sizes were estimated by comparison to known reference fragments of a molecular weight marker or a parallel RCA product of Abutilon mosaic virus (AbMV) using an exponential fit between retention factor (Rf) and the logarithms of the molecular weights using Microsoft Excel for calculations.

Sequencing

The virus-positive Brazilian samples were pooled together with 28 further samples from all over the world, including the recently determined geminiviruses from Bolivia (Wyant et al., 2011) and sent for commercial deep sequencing at GATC Biotech (Constance, Germany).

Bacterial cloning

RCA products of each sample were digested with the enzyme *Bfu*CI (New England Biolabs, Frankfurt, Germany) and inserted into *Bam*HI-cut pGreen0029 plasmid (John Innes Centre, Norwich, United Kingdom), described by Hellens et al.(2000). The recombinant DNA were transformed into *E. coli* DH5α cells (Invitrogen, Darmstadt, Germany), and positive plasmids were sequenced partially using universal primers (M13F/R) by Macrogen (Seoul, South Korea) as described by Wyant et al. (2011).

Sequences analysis

Complete sequences of DNAs A and B components were assembled using Contig Assembly Program on Bioedit software (Hall, 1999) and analyzed with BLAST (Altschul et al., 1997). The contig sequences were named C followed by consecutive numbers given according to sequence assemblage. According to the ICTV criteria the contig sequences were grouped as components of a virus species when shared >89% sequence identity (SI) and provisionally named BgV (Begomovirus) followed by a species identification number, e.g. BgV01. For improved phylogenetic analysis, codaln software (Stocsits et al., 2005) was used in order to optimize the alignment.

Examples of the most closely related viral sequences were retrieved from the international databases in March 2011 (Zhang et al., 2000), and neighbor joining trees were calculated using the algorithms included in MEGA 4.0 (Tamura et al., 2007) with 1,000 bootstrap replications each. Viral names and accession numbers are listed in Table 1. The contig assembly of 32 of 35 sequences was carried out by Stephan Strohmeier.

Table 1. Sequences described in this study or retrieved from GenBank: Virus names, abbreviations and accession numbers used for the phylogenetic analysis.

Name	Abbreviation	DNA A		DNA B	
		BgV component	Accession number	BgV component	Accession number
Novel sequences					
Abutilon mosaic Brazil virus	AbMBV	A01.1.C21	JF694480	B01.1.C35	JF694483
		A01.1.C67	JF694481		
		A01.1.C22	JF694482		
Bean golden mosaic virus - Brazil:Alagoas	BGMV-BR:AL	A03.1.C3	JF694449	B03.1.C43	JF694455
		A03.1.C14	JF694450	B03.1.C44	JF694456
		A03.1.C60	JF694451	B03.1.C45	JF694457
		A03.1.C62	JF694452	B03.1.C46	JF694458
		A03.1.C69	JF694453	B03.1.C47	JF694459
		A03.1.C79	JF694454	B03.1.C48	JF694460
Blainvillea yellow spot virus	BlYSV	A06.1.C80	JF694468	B06.1.C82	JF694469
		A06.1.C81	JF694476	B06.1.C55	JF694470
				B06.1.C56	JF694477
				B06.1.C70	JF694478
				B06.1.C77	JF694479
Cleome leaf crumple virus	ClLCrV	A05.1.C75	JF694461	B05.1.C52	JF694462
				B05.1.C53	JF694463
				B05.1.C54	JF694464
				B05.1.C49	JF694465
				B05.1.C50	JF694466
				B05.1.C51	JF694467
Sida mosaic Alagoas virus	SiMAlV	A02.1.C59	JF694471	B02.1.C34	JF694473
		A02.1.C61	JF694472	B02.1.C63	JF694474
				B02.1.C66	JF694475
Sequences from databases					
Abutilon mosaic Bolivia virus	AbMBoV		HM585445		HM585446
Abutilon mosaic Brazil virus	AbMBV		NC_014138.1		NC_014139.1
Abutilon mosaic virus	AbMV		NC_001928.2		NC_001929.2
African cassava mosaic virus	ACMV		NC_001467.1		NC_001468.1
Bean golden mosaic virus-[Brazil:SAdG 7:soybean:2008]	BGMV		NC_004042		NC_004043
Blainvillea yellow spot virus isolate BR:Coi25:07	BlYSV - [BR:Coi25:07]		EU710756		EU710757
Cleome leaf crumple virus	ClLCrV		FN435999.1		FN436000.1
Sida micrantha mosaic virus - Rhombifolia [Bolivia:Boyuibe-Villamontes:2007]	SimMV - rho [BoVi07]		HM585431		HM585432
Sida micrantha mosaic virus - Rhombifolia [Bolivia:Cerro Fraile 1:2007]	SimMV - rho [Bo:CF1:07]		HM585437		HM585438
Sida micrantha mosaic virus – Rhombifolia [Bolivia:Cerro Fraile 2:2007]	SimMV - rho [Bo:CF2:07]		HM585439		HM585440
Sida micrantha mosaic virus [Brazil:Mato grosso do sul:2:2007]- [Bolivia]	SimMV - MGS2:07-Bo		HM585433		HM585434
Sida mosaic Bolivia virus 1	SiMBoV1		HM585441		HM585442
Sida mosaic Bolivia virus 2	SiMBoV2		HM585443		HM585444
Solanum mosaic Bolivia virus	SoMBoV		HM585435		HM585436

Results

Most of the samples (21 of 22) were tested virus-positive by rolling circle amplification (RCA) in which the sum of the fragments obtained by restriction fragment length polymorphism (RFLP) was at least 5 kb, which is consistent with an expected result from an infection with bipartite New World begomoviruses (Fig. 1).

Sequences analysis

The contig sequences generated by deep sequencing (accession numbers JF694449 to JF694483) revealed several variants with identities above 99%, which may represent the quasispecies population of the viruses within or between the samples. In order to assign the sequences to the original sample, fragments of RCA products were additionally cloned and sequenced in a conventional way and compared to the sequences obtained from deep sequencing. Furthermore, to compare expected with observed fragment patterns and sizes from *in silico* and biochemical digestion, respectively, samples were analyzed by RCA/RFLP (Fig. 1). The quality of this comparison was validated by using Pearson's correlation coefficient (R^2) which indicates the best fit by approaching to 1.00. The observed and expected patterns matched well with each other.

Comparison of the generated DNA A sequences with other geminiviruses in the database according to the International Committee on Taxonomy of Viruses (ICTV) criteria (Fauquet et al., 2008) showed that they were classified as new species, strains or variants with sequence identity (SI) of <89%, 90-93% and >93%, respectively (Table 2).

Five distinct begomovirus species were found amongst the twenty one samples (Fig.1, Table 2 and Supplementary Table 1). The most prevalent sequence (present in 13 of 21 samples), provisionally named BgV03, showed 90% SI with Bean golden mosaic virus [Brazil:SAdG7:soybean:2008] (BGMV; accession number: FJ665283.1), and is therefore classified as a new strain of BGMV, for which the name Bean golden mosaic virus - Brazil:Alagoas (BGMV-BR:AL) is suggested. Six of these samples stem from *Phaseolus lunatus* (lima bean), three from *Phaseolus vulgaris* (common bean), three from *Macropitilium lathyroides* (wild bushbean), a weed member of the

Leguminosae family, and one from *Cleome affinis* plant, which included a mix infection with Cleome leaf crumple virus (ClLCrV; HM195184.1).

Additionally, four *C. affinis* plant samples, including the one with mixed infection mentioned above, contained virus variants (provisionally named BgV05) with 98% SI to ClLCrV (HM195184.1). In one *Sida spp.* sample, a new virus species (BgV02) was detected, with Abutilon mosaic Brazil virus (AbMBV; FN434438.1) as the closest related virus from the database (83% SI), for which we propose the name Sida mosaic Alagoas virus (SiMAlV). *Blainvillea rhomboidea* plants from Alagoas and Bahia (BgV06) revealed variants of Blainvillea yellow spot virus (BlYSV; EU710756.1) with 94% and 96% SI, respectively. One *Sida rhombifolia* plant collected in Bahia (BgV01) was found to be infected with AbMBV (FN434438.1), however we could not precisely classify if it is a new strain or a variant of the sequence retrieved from GenBank (discussed below). Finally, for a *Leguminosae* plant, which could not be specified further, partial DNA A and B sequences were detected and the analysis of the ~2400 nt of DNA A revealed that it is a begomovirus related to BGMV (FJ665283), with 86% SI. Despite the SI <89% this cannot be considered a new virus species because only a partial sequence, comprising ~92% of the total size, was analyzed.

Within the common region (CR), component A/B pairs mostly shared >90% SI, with the exception of variants BgV06A.1.C81 and BgV06B.1.C56, C60 and C70 (80.5% SI), and identical iterons, which are species-specific in New World begomoviruses (Argüello-Astorga et al., 1994; Argüello-Astorga and Ruiz-Medrano, 2001).

Figure 1. Composite of ethidium bromide-stained agarose gels for RCA/RFLP diagnosis using restriction enzymes *Hpa*II (top gels) or *Bfu*CI (bottom gels). The lanes were rearranged according to the corresponding virus, adjusted to the same dimensions with reference to the AbMV standards, and inverted in grey scale for better resolution of faint bands. The expected fragments sizes [nts.] for DNA A (red) and DNA B (green) shown. at the right side of each group of samples were inferred by *in silico* digestion of the contig sequences. Underlined fragments indicate polymorphic fragments, "+" indicates that the sum of two predicted fragments fit to an observed band, "," two fragments with similar size which cannot be resolved in this gel system. Values in round brackets refer to bands that are too small to be resolved. Non-assigned fragments may result from further polymorphism within the viral quasispecies, from still unidentified begomovirus components or from other small circular DNAs, like viral defective DNA or mitochondrial plasmids. Pearson's coefficients of correlation (R^2) refer to the respective gel section for validation of the goodness of the fit between Rf values and expected fragment sizes (in logarithmic scale). The laboratory sample numbers (RCA) are shown for sample identification. For the sample RCA-58 the fragments refer to partial sequences analysis. Host plants were *Phaseolus lunatus* (*P.l.*), unspecified *Leguminosae* (*L.*), *Macroptilium lathyroides* (*M.l.*), *P. vulgaris* (*P.v.*), *Cleome affinis* (*C.a.*), *Blainvillea rhomboidea* (*B.r.*), *S. rhombifolia* (*S.r.*) and *S. spp.* (*S.*). M, molecular weight marker; u, unidentified. BgV, provisional name for the detected species (Begomovirus) followed by number given in order of detection. ⟶
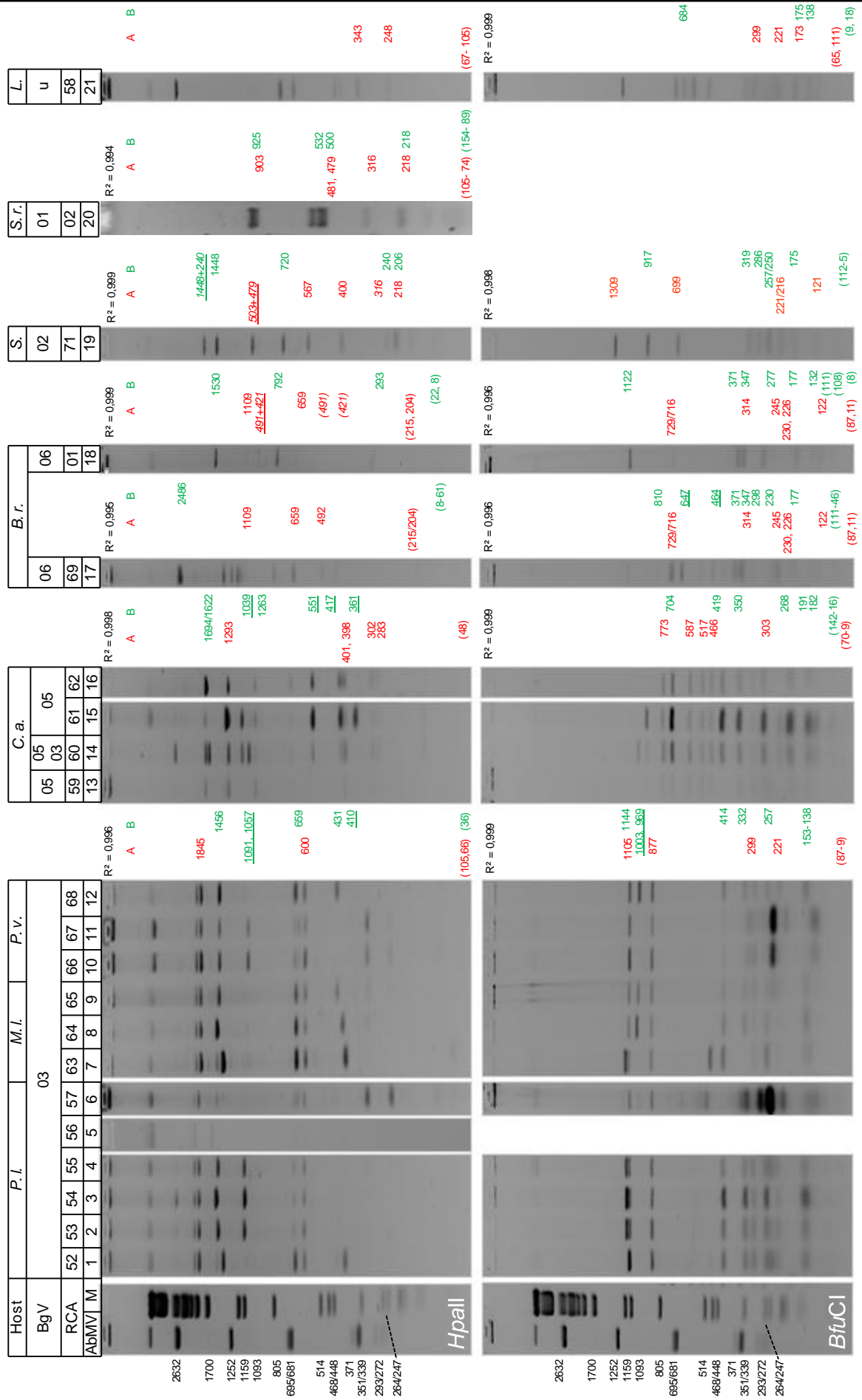
BgV03 was represented on 454 data set by six distinct sequences of the DNA A component with few polymorphisms (C 3, 14, 60, 62, 69 and 79; Table 2) which shared at least 99.4% SI. Six such variant sequences were also detected for DNA B (C 43, 44, 45, 46, 47 and 48) sharing more than 93.7% SI (data not shown). As expected from *in silico* predicted digestion, resulted these polymorphisms in slightly different enzymatic restriction patterns for *Hpa*II or *Bfu*CI (Fig. 1). The same holds true for six ClLCrV (BgV05) DNA B sequences (C 49, 50, 51, 52, 53 and 54; > 97.8% SI; data not shown) and three BIYSV (BgV06) DNA B sequences (C 56, 70 and 77; >97.1% SI; data not shown). The novel virus species (BgV02) was represented by two DNA A (C 59 and 61; 99.9% SI; Table 2) and three DNA B sequences (C 34, 63 and 66; >99.7% SI; data not shown). These data show that most differences in the viral DNA populations can be inferred from the 454 data base and validated by comparison with the results of RCA/RFLP diagnosis.

Table 2. Pairwise comparisons of the nucleotide sequences obtained in this study (assigned BgV) and the closest sequences retrieved from GenBank (ClLCrV, BGMV, AbMBV and BIYSV). Values are showed in percentage of sequence identity (SI).

	BgV05 C75	BgV03 C3	BgV03 C14	BgV03 C60	BgV03 C62	BgV03 C69	BgV03 C79	BgV02 C59	BgV02 C61	BgV06 C80	BgV06 C81	BgV01 C21	BgV01 C67	BgV01 C22
CILCrV	**98.9**	69.6	69.6	69.6	69.6	67.4	69.6	72.0	72.1	68.5	68.6	67.3	66.5	67.6
BGMV	70.6	**90.3**	**90.5**	**90.5**	**90.5**	**90.4**	**90.5**	75.4	75.5	75.4	75.5	72.2	72.2	72.4
AbMBV	68.9	73.2	73.2	73.2	73.2	73.0	73.1	<u>83.0</u>	<u>83.0</u>	70.9	71.0	**91.6**	**90.9**	**94.8**
BIYSV	69.6	75.1	75.1	75.0	75.1	75.1	75.0	72.3	72.4	**94.4**	**95.6**	70.6	71.9	70.8
BgV05 C75	-	70.7	70.6	70.6	70.6	70.8	70.6	72.9	72.9	69.1	69.4	68.2	69.1	67.6
BgV03 C3	-	-	**99.7**	**99.9**	**99.9**	**99.4**	**99.6**	75.0	75.0	74.9	74.9	72.4	72.9	72.3
BgV03 C14	-	-	-	**99.7**	**99.7**	**99.7**	**99.9**	74.9	75.0	74.9	75.0	72.3	72.8	72.3
BgV03 C60	-	-	-	-	**99.8**	**99.4**	**99.5**	74.9	74.9	74.8	74.9	72.4	72.9	72.3
BgV03 C62	-	-	-	-	-	**99.4**	**99.5**	74.9	74.9	74.9	74.9	72.4	72.9	72.3
BgV03 C69	-	-	-	-	-	-	**99.7**	75.2	75.3	74.9	75.0	72.2	72.6	72.2
BgV03 C79	-	-	-	-	-	-	-	74.9	74.9	74.8	74.9	72.2	72.8	72.3
BgV02 C59	-	-	-	-	-	-	-	-	**99.9**	72.8	72.8	79.5	79.4	85.0
BgV02 C61	-	-	-	-	-	-	-	-	-	72.9	72.9	79.6	79.4	85.0
BgV06 C80	-	-	-	-	-	-	-	-	-	-	**96.4**	70.8	71.9	70.2
BgV06 C81	-	-	-	-	-	-	-	-	-	-	-	70.5	71.7	71.0
BgV01 C21	-	-	-	-	-	-	-	-	-	-	-	-	**97.7**	**94.2**
BgV01 C67	-	-	-	-	-	-	-	-	-	-	-	-	-	**93.9**

Numbers in bold: variants or strains, sequence identity (SI) >89%;
Numbers underlined: New virus species SI<89%.

Three BgV01 DNA A sequences (C 21, 22 and 67) were recovered from the 454 data set. They are variants of each other sharing at least 93.9% SI (Table 2). However when compared with the AbMBV (FN434438.1) from the GenBank, two of them, C 21 and 67, can be classified as new strains, with 91.6% and 90.9% SI, respectively, whereas C 22, with 94.8% SI, would be considered as a variant. In addition, the expected *Hpa*II RFLP pattern of the three sequences fits with the observed (Fig. 1).

Validation of the fidelity of sequences generated by circomics

RCA products of samples originally collected in Bolivia, described in a previous study (Wyant et al., 2011) were part of the pool of samples sequenced together with the samples analyzed in this study. The clone-based sequences from seven samples detected on the preceding work were compared to the 454 data set, and three samples were represented amongst the generated contig sequences. Abutilon mosaic Bolivia virus (AbMBoV;) was represented by BgV04, which contains a DNA A (C 33), sharing 100% SI with AbMBoV DNA A (accession number HM585445) and three DNA B (C 29, 99.9% SI; C 30, 100% SI and C 57, 99.9% SI with AbMBoV DNA B, HM585446) sequences. Solanum mosaic Bolivia virus (SoMBoV; HM585436), BgV07, and Sida micrantha mosaic virus – Rhombifolia [Bolivia:Cerro Fraile 2:2007] (SimMV-rho[Bo:CF2:07]; HM585440), BgV08, were represented on the 454 data set only by DNA B sequences, C 18 and C 98 with 100% and 99.7% SI with SoMBoV, respectively, and C 82 with 99.9% SI with SimMV-rho[Bo:CF2:07]. These results validate the sequence fidelity of the generated 454 data set sequences and show a limitation of the method in detecting all sample sequences included on the pool.

Phylogenetic analysis

The phylogenetic trees confirmed the nucleotide sequence investigation with all sequences assigned as belonging to the same species clustering together (Fig. 2). The BgV02 sequences clustered with AbMBV, the closest related virus sequence, but in a different branch in both DNA A and DNA B trees, validating the classification as a new virus species. Whereas the sequences of BgV01, 03, 05 and 06 grouped with the respective virus species sequences from which they belong to, AbMBV, BGMV, CILCrV and BIYSV, respectively, in both trees with bootstrap value of 100%.

The same result was observed with the sequences of samples from Bolivia, AbMBoV, SoMBoV and SimMV-rho[Bo:CF2:07], that clustered with the sequences recovered from the 454 data set BgV04, BgV07 and BgV08, respectively, with 100% bootstrap value, what confirms the fidelity of the sequences generated in the present study.
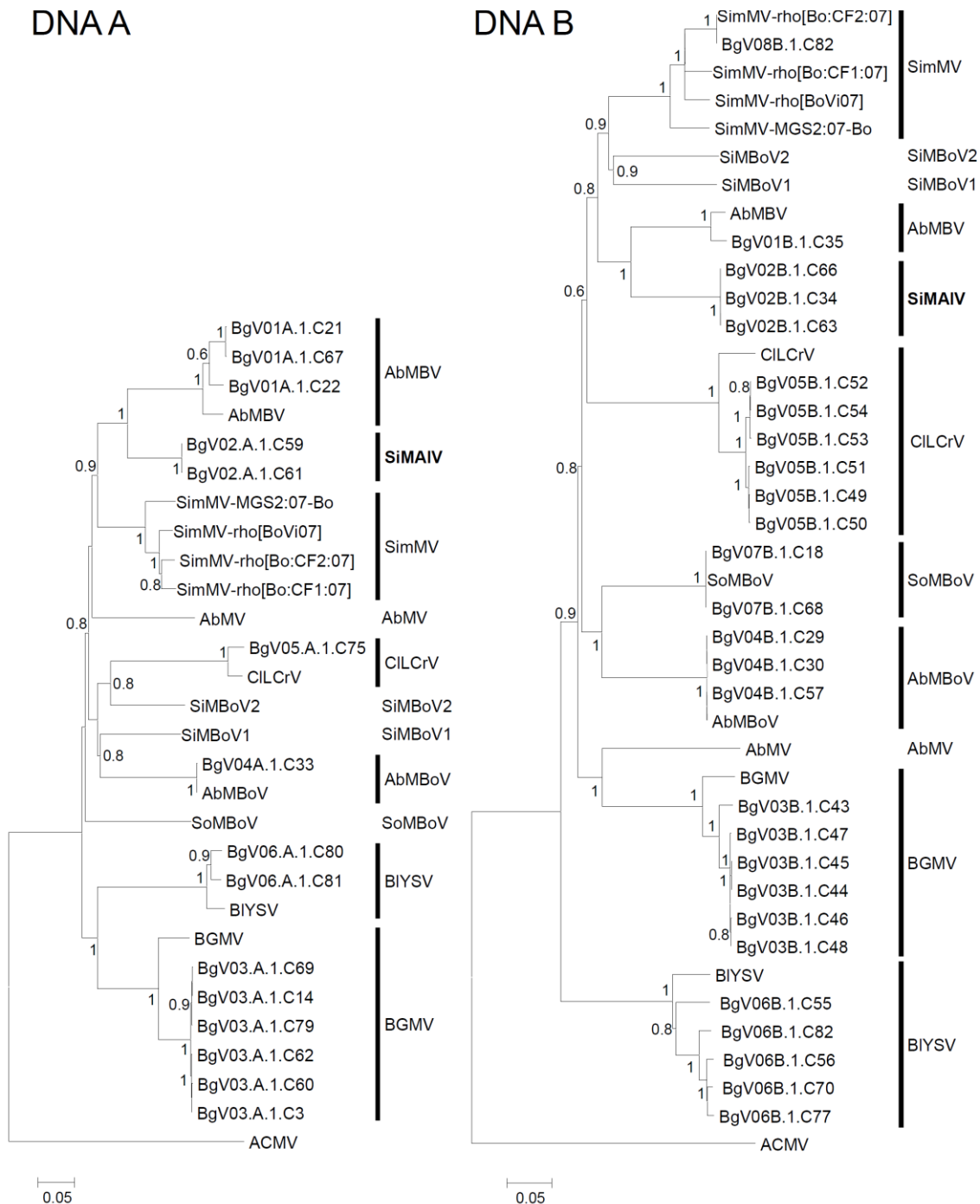
Figure 2. Neighbor Joining trees of DNA A and DNA B for the isolates detected in this study compared to selected sequences from the Americas retrieved from international databases, and using African cassava mosaic virus (ACMV) as outgroup. Numbers next to the nodes indicate bootstrap values (1,000 replicates) above 50% (0.5). The corresponding names, abbreviations and accession numbers are listed in Table 1. In bold, new virus species described in the present work, for which the name Sida mosaic Alagoas virus (SiMAIV) is suggested.

Relevance of weeds for epidemics

Eight weed samples were virus-positive, including the crop-infecting BGMV (*Leguminoseae* and *C. affinis* sample). These results stress the importance to control viral infections also in weeds, since they act as perennial reservoirs of viruses and provide a durable breeding ground for DNA recombination and creation of new viral genomes as discussed before (Castillo-Urquiza et al., 2008; Jovel et al., 2007; Jovel et al., 2004).

Discussion

Circomics has proven to be a significant step forward in the rapid and reliable identification of geminiviruses. Furthermore, it extends the view on the viruses because it is able to represent the population structure of viral sequences in one or several plants. In contrast to the bacterially cloned DNA, which represents only a small sample size, the potential variability already present in a host is elucidated. This information is extremely important for sustainable resistance breeding, irrespectively of whether this is done by classic crossing or molecular engineering. Therefore, it was worthy to note, that even minor changes in the viral DNA population above 99% SI can be monitored and verified by comparison with RFLPs. Now, it becomes possible for the first time to use the quasispecies concept originally set up by Manfred Eigen for bacteriophages and human retroviruses (Biebricher and Eigen, 2006; Eigen, 1993; Eigen, 1996) on an agricultural scale.

The quality of the pyrosequencing results was similar if not superior to conventional Sanger sequencing as reflected by the comparison of viral genomes from Bolivia and Brazil. Particular notorious errors of this techniques, e.g. with oligoA or oligoT stretches, were compensated by the high coverage of sequences of redundant entries, and must be considered only if few entries were obtained for a certain genome component. The limitation of the method in detecting all sample sequences included on the pool might be due to the large number of samples submitted to deep sequencing simultaneously.

The two used restriction enzymes are ideal in combination for the analysis of geminivirus genomes. In addition, this RCA/RFLP-based identification tool might be also valuable to ensure the release of only virus-free material for plantation.

In total, circomics rings in a new age for geminivirus investigation and a novel paradigm for understanding population genetics of these important plant pathogens.

Acknowledgements

Supplementary Table 1. Summary of information about plants samples evaluated in this study and the detected viruses. Lane# refers to lane numbers as in Figure 2.

Lane #	RCA	Virus DNA Contig	Collection site	Plant sample
1	52	BgV03	Belém de Maria, Pernambuco	*Phaseolus lunatus*
2	53		Pilar, Alagoas	
3	54		Atalaia, Alagoas	
4	55		Pilar, Alagoas	
5	56		Recife, Pernambuco	
6	57		Rio Largo, Alagoas	
7	63		Palmeira dos Índios, Alagoas	*Macropitilium lathyroides*
8	64		Rio Largo, Alagoas	
9	65		União dos Palmares, Alagoas	
10	66		Rio Largo, Alagoas	*Phaseolus vulgaris*
11	67		Rio Largo, Alagoas	
12	68		Rio Largo, Alagoas	
13	59	BgV05	Rio Largo, Alagoas	*Cleome affinis*
14	60	BgV05 & 03	Atalaia, Alagoas	
15	61	BgV05	Maceió, Alagoas	
16	62		Rio Largo, Alagoas	
17	69	BgV06	Rio Largo, Alagoas	*Blainvillea rhomboidea*
18	1		Salvador, Bahia	
19	71	BgV02	Rio Largo, Alagoas	*Sida spp.*
20	2	BgV01	Salvador, Bahia	*Sida rhombifolia*
21	58	unidentified	Barra de São Miguel, Alagoas	*Leguminoseae* (not specified)
-	70	-	Rio Largo, Alagoas	

References

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W.and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25** (17)**,** 3389-3402.

Argüello-Astorga, G. R., Guevara-González, R. G., Herrera-Estrella, L. R.and Rivera-Bustamante, R. F. (1994). Geminivirus replication origins have a group-

specific organization of iterative elements: a model for replication. *Virology* **203** (1)**,** 90-100.

Argüello-Astorga, G. R.and Ruiz-Medrano, R. (2001). An iteron-related domain is associated to Motif 1 in the replication proteins of geminiviruses: identification of potential interacting amino acid-base pairs by a comparative approach. *Arch. Virol.* **146,** 1465-1485.

Biebricher, C. K.and Eigen, M. (2006). What is a quasispecies? *Curr. Top. Microbiol. Immunol.* **299,** 1-31.

Castillo-Urquiza, G. P., Junior, J. E. A. B., Bruckner, F. P., Lima, A. T. M., Varsani, A., Alfenas-Zerbini, P.and Zerbini, F. M. (2008). Six novel begomoviruses infecting tomato and associated weeds in Southeastern Brazil. *Arch. Virol.* **153,** 1985-1989.

Eigen, M. (1993). Viral quasispecies. *Sci. Am.* **269** (1)**,** 42-49.

Eigen, M. (1996). On the nature of virus quasispecies. *Trends Microbiol.* **4** (6)**,** 216-218.

Faria, J. C.and Maxwell, D. P. (1999). Variability in geminivirus isolates associated with *Phaseolus* spp. in Brazil. *Phytopathol.* **89,** 262-268.

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

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

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41,** 95-98.

Hellens, R. P., Edwards, E. A., Leyland, N. R., Bean, S.and Mullineaux, P. M. (2000). pGreen: a versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation. *Plant Mol. Biol.* **42** (6)**,** 819-832.

Jeske, H. (2009). Geminiviruses. *In* "TT viruses: the still elusive human pathogens". E.-M. de Villiers, and H. zur Hausen, Eds.). Springer, Berlin Heidelberg.

Jovel, J., Preiß, W.and Jeske, H. (2007). Characterization of DNA-intermediates of an arising geminivirus. *Virus Res.* **130,** 63-70.

Jovel, J., Reski, G., Rothenstein, D., Ringel, M., Frischmuth, T.and Jeske, H. (2004). Sida micrantha mosaic is associated with a complex infection of begomoviruses different from Abutilon mosaic virus. *Arch. Virol.* **149** (4)**,** 829-841.

Mansoor, S., Briddon, R. W., Zafar, Y.and Stanley, J. (2003). Geminivirus disease complexes: an emerging threat. *Trends Plant. Sci.* **8** (3)**,** 128-134.

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

Morales, F. J.and Anderson, P. K. (2001). The emergence and dissemination of whitefly-transmitted geminiviruses in Latin America. *Arch. Virol.* **146** (3)**,** 415-441.

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

Sambrook, J.and Russell, D. W. (2001). "Molecular cloning: a laboratory manual." Cold Spring Harbor Press, Cold Spring Harbor, N Y.

Stanley, J., Bisaro, D. M., Briddon, R. W., Brown, J. K., Fauquet, C. M., Harrison, B. D., Rybicki, E. P.and Stenger, D. C. (2005). Geminiviridae. *In* "Virus

taxonomy. VIIIth report of the international committee on taxonomy of viruses". C. M. Fauquet, M. A. Mayo, J. Maniloff, U. Desselberger, and L. A. Ball, Eds.), pp. 301-326. Elsevier, Academic Press, London.

Stocsits, R. R., Hofacker, I. L., Fried, C.and Stadler, P. F. (2005). Multiple sequence alignments of partially coding nucleic acid sequences. *BMC Bioinformatics* **6,** 160.

Tamura, K., Dudley, J., Nei, M.and Kumar, S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24** (8)**,** 1596-1599.

Wyant, P. S., Gotthardt, D., Schafer, B., Krenz, B.and Jeske, H. (2011). The genomes of four novel begomoviruses and a new Sida micrantha mosaic virus strain from Bolivian weeds. *Arch. Virol.* **156** (2)**,** 347-352.

Zhang, Z., Schwartz, S., Lukas, W.and Web, M. (2000). A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.* **7** (1-2)**,** 203-214.

Tomato golden mosaic virus back in tomatoes

Patricia Soares Wyant, Sigrid Kober, Holger Jeske, Christina Wege

Institute of Biology, Department of Molecular Biology and Plant Virology, University of Stuttgart, Pfaffenwaldring 57, 70550 Stuttgart, Germany.

Abstract

Tomato golden mosaic virus (TGMV) is one of the standard geminiviruses used as a model in several laboratories providing information about its symptom expression, replication and transcription strategies. Despite its widespread laboratory use, the cloned virus has never been transmitted back to tomatoes, and was never shown to induce a golden mosaic in this or in any other hosts. Here we show for the first time an extremely low infection rate and very mild symptoms induced in tomatoes. We also analyzed the genomes sequences recovered from the infected tomato plants and thus the complete genome of DNA A of csTGMV (common strain TGMV) and that of DNA B of yvTGMV (yellow vein TGMV) with a non-truncated BC1 open reading frame (ORF) could be described for the first time.

Keywords: csTGMV, yvTGMV, Tomato golden mosaic virus.

Introduction

Tomato golden mosaic virus (TGMV) is a plant virus member of the *Geminiviridae* family, genus *Begomovirus*. It presents a bipartite genome and was the first tomato infecting geminivirus identified and assumed to be the causative agent of tomato mosaic disease in the seventies (Matyis et al., 1975).

TGMV is one of the standard geminiviruses analyzed in several laboratories. It has been used for more than 20 years now, and has revealed key information about geminiviral replication, transcription strategies, and symptom expression (Egelkrout et al., 2001; Gardiner et al., 1988; Hanley-Bowdoin et al., 1988; Hanley-Bowdoin et al., 1989; Orozco and Hanley-Bowdoin, 1996; Thommes et al., 1993; von Arnim and Stanley, 1992).

Since its identification, viral samples of TGMV were propagated in different laboratories worldwide in several host plants, mainly in *Nicotiana,* as *N.*

benthamiana, *N. tabacum* and *N. glutinosa*. All initial inocula originated from plant samples which were kindly provided by Dr. A. S. Costa (†) to various researchers. The first infectious cloned sequence of TGMV was obtained in 1984 from extractions of *N. benthamiana,* and is now known as yellow vein (yv) TGMV (Hamilton et al., 1984). Later on, in 1992, the full sequence of a DNA B of another more severe "strain" was published, which induced extensive chlorosis throughout systemically infected leaves in *N. benthamiana,* in contrast to the chlorotic or yellowing symptoms largely confined to veins associated with yvTGMV. Therefore the new and more severe virus was named common strain (cs) TGMV (von Arnim and Stanley, 1992). We have now sequenced DNA A and DNA B of both cs and yvTGMV, kindly provided to us by Dr. J. Stanley (Norwich, United Kingdom) as agroinfectious clones. Here, we report the complete genome sequences retrieved from agroinoculated, systemically invaded tomato plants. For the first time, the full sequence of csTGMV DNA A, and yvTGMV DNA B with a corrected BC1 ORF are published. Furthermore, we have compared all genomic TGMV sequences available in the GenBank database originating from other host plants with the sequences we were able to obtain from TGMV re-transmission to tomato.

Despite all information and progress in virus research provided by the use of TGMV as a model, it has never been transmitted back to tomato before, and therefore was never shown to induce a golden mosaic in this or other hosts. We have re-investigated the issue of TGMV transmission to tomato plants systematically and show, for the first time, very mild symptoms in this original host species, with an extremely low infection rate. Although the phenotype did not resemble a mosaic, we propose that the well-established name of TGMV should be maintained, to avoid any future confusion.

Materials and Methods

Sample collection

Samples of systemically yvTGMV- or csTGMV-infected leaf tissue were collected from tomato plants (*Solanum lycopersicum L.)* previously infected by agroinoculation (Klinkenberg et al., 1989) using partial repeats of each genomic component of csTGMV and yvTGMV: clones csTA1.6, csTB1.4, yvTA1.3 and yvTB1.5 in pBIN19

(von Arnim and Stanley, 1992) in *Agrobacterium tumefaciens* LBA 4404 (Hoekema et al., 1983). The collection of the samples was carried out by Sigrid Kober.

Nucleic acid extraction

Total nucleic acids from *S. lycopersicum* and, as controls, *N. benthamiana* plant samples were extracted using a CTAB-based method (Kleinow et al., 2009). The total nucleic acid extraction was performed by Sigrid Kober.

RCA/RFLP

Amplification of small circular DNA was performed by rolling circle amplification (RCA) using a TempliPhi™ Kit (GE Healthcare) following the manufacturer's protocol.

RCA products (~300 ng DNA) were digested by the restriction enzyme *Hpa*II (New England Biolabs, Frankfurt, Germany), in 10 µl final volume for 2 h at 37 °C, following the manufacturer's protocol, separated in 2 % agarose gels in TBE, and stained with 0.5 µg/ml ethidium bromide. The resulting restriction fragment length polymorphism (RFLP) patterns were compared to the fragment sizes predicted for csTGMV DNA B (GenBank accession number M73794) and yvTGMV DNA A and -B (accession numbers K02029, K02030).

Sequencing and sequence analysis

RCA products were directly used as templates for sequencing without further purification, using either a GenomeLab™ DTCS (Dye Termination Cycle Sequencing) Quick Start Kit (Beckman Coulter GmbH, Krefeld, Germany), and capillary electrophoresis (CEQ 8000 Genetic Analysis System, Beckman-Coulter GmbH, Krefeld, Germany), or a BigDye® Terminator Cycle Sequencing Kit (Applied Biosystems GmbH, Darmstadt, Germany) and capillary electrophoresis (3730xl DNA Analyzer, Applied Biosystems GmbH, Darmstadt, Germany) performed by a custom sequencing service (Macrogen Inc., South Korea). For both methods, unlabelled TGMV-specific primers (Table 1) were used. For the first method, 50 fmol DNA, 10 pmol primer, 4 µl Quick Start Master Mix were combined with water to a volume of 10 µl and subjected to cycle sequencing (40 or 50 cycles 96 °C 20 s, 50 °C 20 s, 60 °C

4 min). The second method was conducted by Macrogen Inc. Sequencing of samples carried out in our lab were performed by either Gabi Kepp or Benjamin Schäfer.

Sequences were compared to the nucleotide sequences of yvTGMV DNA A (GenBank accession number K02029), yvTGMV DNA B (K02030) and csTGMV DNA B (M73794), using the BLAST algorithm (Altschul et al., 1997) followed by alignment using CLUSTALW (Thompson et al., 1994). Open reading frames (ORFs) were identified using BioEdit software (Hall, 1999).

Table 1. Primers used for direct sequencing of RCA products: * specific for yvTGMV; $ degenerated primer: M stands for A (yvTGMV) or C (csTGMV). § There might be small differences of ±2nt in the annealing positions among different strains.

DNA	Name	Sequence	Genomic annealing position[§]
DNA A	AV1 revc	5' ttt gaa tta aag atc cac gaa acg 3'	0188-0165
	AV1 av	5' atg cct aag cgg gat gcc cc 3'	0190-0209
	AV1 rev	5' acg cag aaa cgc tta cca ac 3'	0494-0475
	AC1R	5' cca gac gag atg aga caa tg 3'	1884-1865
	AIR-F	5' agc cac agt tca ggt caa ttc g 3'	1383-1362
	AC1/AC2	5' ctc aac tcc ccc ctc tat ca 3'	1447-1428
	AV1F	5' cta caa cca cca gga ggc ag 3'	0792-0811
	AC1ac	5' atg cca tcg cat cca aaa cgg 3'	2764-2444
	AC1F3	5' gca acc tcc tct agc act tcg 3'	2081-2101
	AC1F4	5' ccc ata caa gag tat ctc cgt 3'	2122-2142
DNA B	CR/BV1*	5' tgg tgt ggt ccc ctt gtg tta 3'	0066-0086
	BV1 revc	5' agt taa tga gtg cag aaa taa c 3'	0136-0115
	BV1v	5' ggt ccc ttt aat ttg aaa tat 3'	0137-0157
	BV1rev	5' gct tcc acg cca tca atg taa 3'	0247-0227
	BV1-IR	5' ctt gtt gtt gta gtt gat cgg ca 3'	0708-0730
	BV1for	5' gct ata agt tct gcg ttg aag gac 3'	0810-0833
	IR-BV1	5' ctc cgt cac agc cca aaa 3'	1217-1200
	IR-BC1	5' ctg ggt cta aga tgg tgg ttc c 3'	1335-1356
	CR-BC1[$]	5' gct acc cat ttc cag ttc tmc gc 3'	2368-2346

Biolistic inoculation of plant tissues

Tomato plants (*S. lycopersicum* cv. 'Moneymaker', special breeding cultivar of Hild Samen GmbH, Marbach, Germany) in the cotyledon stage were inoculated biolistically as described previously (Unseld et al., 2001).

Results

Sequencing

Sequences of DNA A and DNA B of common strain Tomato golden mosaic virus (csTGMV) and of yellow vein (yv) TGMV (accession numbers JF694488 to JF694491), were determined completely by direct sequencing of rolling circle amplification (RCA) products. These were derived from systemically infected tomato

plants by use of TGMV-specific primers and a primer walking strategy (Table 1). The resulting genome organization is shown in Fig. 1. For csTGMV, the sequence of its accompanying DNA A cloned from the same source material was obtained and is described in this study for the first time. The molecule of 2589 bases in length was compared to the reference sequence of the first cloned yvTGMV DNA A (K02029; 2588 bases). We detected 38 nucleotide exchanges and a nucleotide insertion, resulting in an overall sequence identity of 98.5%. Of these nucleotide variations between csTGMV and yvTGMV, 30 were located within open reading frames (ORFs), eight in AV1, seventeen in AC1, four in AC2, five in AC3 and one in AC4, with five of the variations in regions shared between two ORFs each. Twelve of them led to amino acid exchange affecting all DNA A-encoded proteins. In addition to conserved exchanges in AV1, AC1, and AC3 (two in every protein), the following non-conserved exchanges were found: four in AC1, two in AC2, and one each in AC3 and in AC4. Single-site substitutions within both the overlapping coding regions for AC1 and AC4, and AC2 and AC3, respectively, caused amino acid exchange in all four of these proteins (Table 2). The sequence of the cloned csTGMV DNA B applied in this study (M73794) had been published before (von Arnim and Stanley, 1992). Our sequence re-amplified from tomato plants harbored a single T deletion in comparison to the reference sequence at genome position 1137, within the intergenic region (IR) between BV1 and BC1, and thus exhibited a total length of 2523 in contrast to the originally published 2524 bases.

For yvTGMV, both DNA components of clones established by von Arnim and Stanley (1992) were sequenced completely for the first time. As detailed in the introduction, these clones were derived from plant samples directly obtained from Dr. A. S. Costa, and thus were independent of the first sequences published, K02029 and K02030 (Bisaro et al., 1982; Hamilton et al., 1983; Hamilton et al., 1984). Strikingly, the newly determined DNA A sequence, after re-isolation of the agroinoculated virus from tomato, was nearly identical to that of the first yvTGMV DNA A clone. With the exception of a single silent nucleotide C to A exchange at genome position 1732 within ORF AC1, no further differences of the 2588 bases molecule were found in comparison to the reference sequence (K02029). In contrast, DNA B in relation to the reference (K02030) was more divergent. Its length was increased from 2508 bases denoted in K02030 to 2524 bases (see below), which however, paralleled findings of

Fontes et al. (1994) not submitted to the GenBank database. Several further differences affected exclusively the cell-to-cell movement protein encoded by ORF BC1, and the IR. Overall, in comparison to the first clone (K02030), the independent yvTGMV DNA B clone exhibited four nucleotide substitutions: a silent one located within BV1 ORF, another one at position 1418, considered to be in the intergenic region (IR) on the first clone (K02030) sequence and now described within ORF BC1, and two in the IR (Table 2). In addition, two deletions, one in the IR and one in BC1, and eighteen insertions were detected, seventeen of them in BC1 (including a two nucleotide- and an eleven nucleotide-insertion event) and one in the IR, totalizing 16 nucleotides more than the K02030 sequence. The insertion of an A between positions 1538 and 1539 led to the exchange of eleven amino acids, eight of them being non-conserved. Another insertion of an A between positions 1498 and 1499 led to the exchange of a stop codon into a lysine and, as a consequence thereof, to the C-terminal extension of the ORF BC1 to 293 amino acids, which is 109 amino acids longer than in the reference sequence for yvTGMV (K02030) as shown on Table 2. Amino acids differences are represented in Figure 1.

▲ single-site aa differences between csTGMV and yvTGMV (this study and M73794)

aa differences between yvTGMV (this study) and yvTGMV reference sequences (K02029 and K02030), all of which were located in DNA-B:
★ 11 substitutions resulting from an insertion (see Table S1 for details);
⚡ exchange of a stop codon for a lysine, resulting in
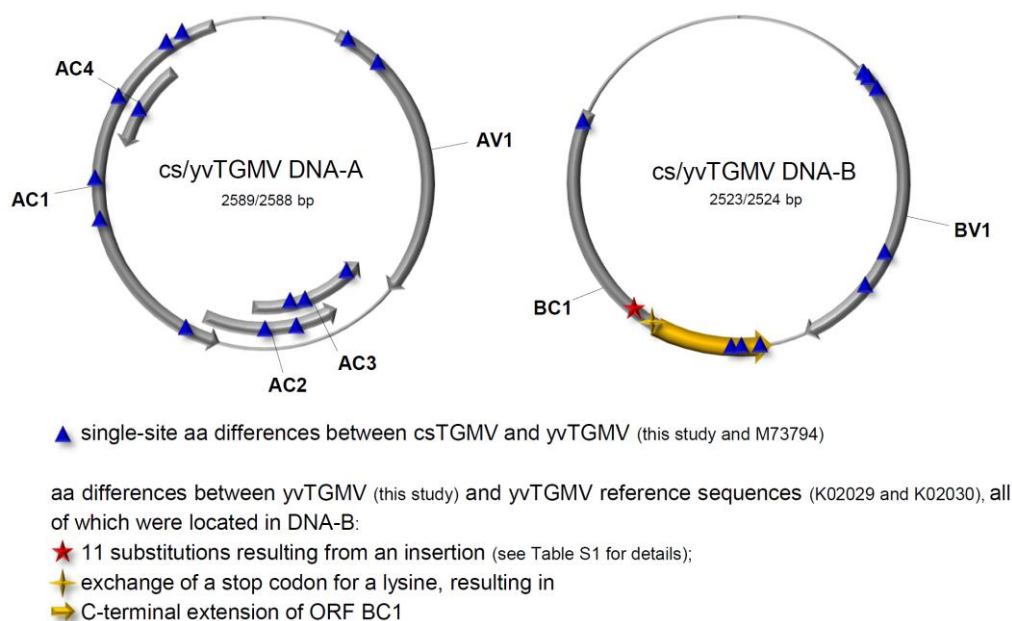➡ C-terminal extension of ORF BC1

Figure 1. Genomic organization of Tomato golden mosaic virus (TGMV). Arrows show the open reading frames (ORFs) positions and directions of translation. DNA A encodes proteins involved in viral replication, DNA B proteins associated with viral transport. AV1 codes for the coat protein (CP); AC1 the replication-associated "Rep"-protein, responsible for the initiation of viral DNA replication; AC2 for the transcription activator of AV1 and BV1 (TrAP); AC3 for the replication enhancer (REn); the function of AC4 is still unclear. BV1 encodes the nuclear shuttle protein (NSP) and BC1 the movement protein (MP), involved in cell to cell transport. Triangles represent amino acid (aa) differences between common strain (cs) TGMV and yellow vein (yv) TGMV analyzed in this study, red star represent eleven aa differences and yellow star represents exchange of a stop codon by a lysine, leading to the C-terminal extension of the ORF BC1 (yellow arrow), between yvTGMV sequenced in this study and the reference sequence of an independent yvTGMV clone (GenBank accession numbers K02029 and K02030 for DNA A and DNA B, respectively).

We also sought to assess whether certain amino acid variations identified within the different TGMV gene products may reflect advantageous alterations prevailing in several begomoviruses. Therefore we analyzed whether codons exchanged in the newly sequenced molecules in relation to those at the same positions in the reference sequences, or in the variant being compared to, conform better to the respective protein sequences of other members in the *Geminiviridae* or not. To this aim, all amino acid divergences between csTGMV and yvTGMV found in this study, or between these clones and the reference sequences of independent clones as deposited in the database (K02029 or K02030), were compared with distinct closely related geminiviral amino acid sequences retrieved from the GenBank database (Table 2). Comparing csTGMV and the reference sequence of yvTGMV (K02029), five out of fourteen amino acid variations observed on DNA A of csTGMV conform better to the sequences of other geminiviral proteins, six are better conformed in the original reference sequence of yvTGMV, and three variations were located at positions with variable amino acid composition (details are listed in Table 2). For the comparison of the newly obtained DNA B of yvTGMV with the reference yvTGMV (K02030), the result is very uniform: All amino acid variations detected conform better to the consensus sequences of other geminiviral proteins.

Overall the DNA As of the csTGMV and yvTGMV clones retrieved in this study were 98.5% identical with 38 nucleotides difference, and the yvTGMV sequence described here exhibited only a single silent nucleotide exchange in comparison to the formerly published clone (GenBank accession number K02029). The DNA Bs were more divergent, showing 97% SI between that of csTGMV and the novel yvTGMV sequence, with 69 divergent nucleotides leading to five amino acid variations within the BV1 ORF encoding the nuclear shuttle protein, four of them even non-conserved, and four amino acid exchanges within ORF BC1 coding for the cell-to-cell movement protein, with one of them non-conserved (Table 2).

Biolistic inoculation of TGMV from tomato

In order to evaluate the infectivity of the viral DNAs retrieved from tomato plants under high inoculum pressure, their RCA products were inoculated biolistically onto *Solanum lycopersicum* cv. 'Moneymaker' seedlings. At two weeks post inoculation (wpi), one out of thirteen plants inoculated with yvTGMV showed yellowing of the

veins in newly expanded leaves. With time, in subsequent leaf generations, the symptoms ameliorated and were completely absent from leaves developed after six wpi (Fig. 2). RCA-amplified csTGMV DNA bombarded into thirteen tomato seedlings of the same variety in parallel did not induce symptoms in any of the plants, when compared to mock-inoculated seedlings. RCA/RFLP-based diagnosis of all plants treated with TGMV inocula showed the expected fragment pattern only for the single symptomatic plant infected with yvTGMV. This pattern was detectable even in newly formed leaves at six wpi, developing symptomless (data not shown).
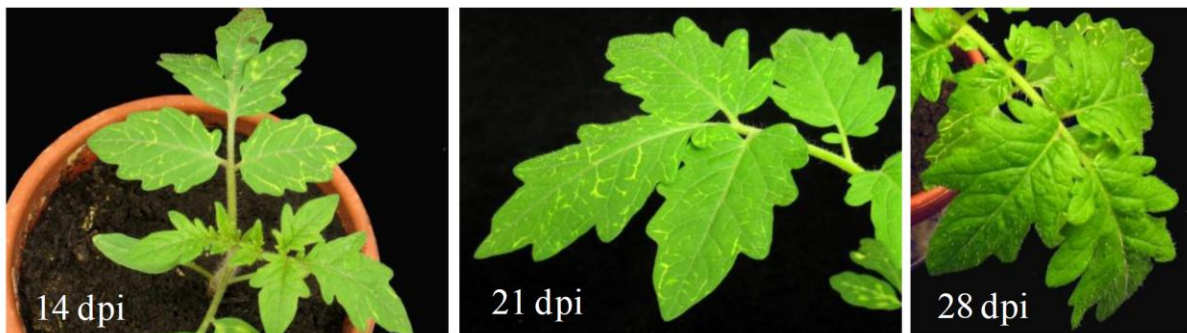

Figure 2. Tomato plants (*Solanum lycopersicum*, cv. 'Monemaker', special breeding variety) 14, 21 or 28 days post inoculation (dpi), respectively, with yellow vein Tomato golden mosaic virus (yvTGMV) rolling circle amplification (RCA) products by biolistics, showing initial symptomatology, and amelioration and complete disappearance of symptoms in subsequent leaf generations.

Discussion

Viral DNA samples were retrieved from systemically infected tomato plants agroinoculated with either csTGMV or yvTGMV, respectively, *via* clones of von Arnim et al. (1992). Using a primer walking strategy directly on RCA products, we determined the complete sequences of the DNA A and DNA B components of both so-called strains after their multiplication and spread in tomato tissues. Whereas a yvTGMV DNA B reference sequence obtained with an independent plasmid clone (deposited in the GenBank database under accession number K02030) shows a truncated movement protein ORF BC1, the yvTGMV sequence described in this study presents an undisrupted ORF BC1 for the first time. These data confirm and specify observations of Fontes et al. (1994), who also indicated the need of certain corrections in the yvTGMV DNA B sequence, but have not published the respective details so far.

Here the first complete sequence available for csTGMV DNA A is reported. Interestingly, this isolate harbors 98.5 % SI of its DNA A with that of yvTGMV. Hence,

csTGMV and yvTGMV should be classified as different TGMV variants according to the current ICTV rules (Fauquet and Stanley, 2005), instead of the so far prevailing designation as "strains". According to Fauquet et al (2008), the correct name of these variants within the Tomato golden mosaic virus species are Tomato golden mosaic virus-[Brazil:Common;1984] (TGMV-[BR:Com:84]) and Tomato golden mosaic virus-[Brazil:Yellow Vein] (TGMV-[BR:YV]).

Even though csTGMV and yvTGMV show high sequence identity, relatively few differences are enough to cause prominent discrepancies between their phenotypes in several *Nicotiana* species, *Datura stramonium* (von Arnim and Stanley, 1992) and, as shown in this study for the first time, the original host tomato. While csTGMV induces local up to extensive chlorosis in leaves of distinct host plants, veinal yellowing and chlorosis in close association with the vascular bundles is a typical feature of yvTGMV in several susceptible species. Notably, symptoms were less severe with yvTGMV in most virus-host combinations analyzed so far (von Arnim and Stanley, 1992), which differs from our observations with significantly more pronounced chlorotic phenotypes induced by yvTGMV on tomato, in comparison to the scarce and mild patchy chlorosis accompanying csTGMV invasion (data not shown). It has been suggested that phenotypic differences are primarily based on divergences in the DNA B components of the viruses, more specifically due to the exchanges of V272I and Q288K in the C-terminal portion of BC1, which encodes the viral movement protein, MP (Saunders et al., 2001). Our sequence data confirm the above-mentioned amino acid substitutions defining two major symptom determinants within the movement proteins of yvTGMV versus csTGMV, respectively.

To again ensure infectivity and symptomatology of the nucleic acids enriched for sequencing by RCA amplification after their multiplication in tomato, we biolistically re-inoculated a portion of the RCA products into tomato plants again. One out of thirteen seedlings bombarded with yvTGMV DNA, but none of thirteen plants treated with csTGMV DNA developed symptoms and accumulated viral DNA. The thus obtained symptoms were identical to those of the initially agroinfected tomato plants, from which we extracted the viral samples analyzed in this study. As for the directly agroinoculated tomatoes (data not shown), biolistically infected plants retained the virus and revealed a latent infection after symptoms declined and were completely lost in newly expanded leaves a few weeks post inoculation.

It is surprising that TGMV, originally extracted from tomato plants during the seventies and assumed to be the causative agent of tomato golden mosaic disease, does not induce the respective strong symptoms when re-transmitted back to tomato. This observation raises the question if the virus has suffered mutations during passages within *Nicotiana* species and became less virulent to tomato plants, or if the real causative agent of the "mosaico dourado" first observed on tomato in Brazil was a different, probably co-infecting virus.

The observation of two phenotypically different TGMV "strains" was pointed out in the literature after passage of the viruses in different *Nicotiana* species for several plant generations (von Arnim and Stanley, 1992), and it is not clear if there was such segregation in tomato, too. One can speculate that therefore, it is likely that different variants, initially named "strains", were generated due to mutations caused by the passage of the virus in *Nicotiana* plants, since all TGMV accessions studied so far have been provided by Dr. A. S. Costa, but were propagated in different laboratories in several plants henceforth.

Table 2. Comparison between csTGMV and yvTGMV sequences. *csTGMV[92/11]* or *yvTGMV[92/11]*, respectively, correspond to clones of von Arnim and Stanley [1992], analyzed in this study [2011]; reference (ref.) M73794 is the first published sequence of this csTGMV DNA B clone. K02029, K02030: ref. sequences of independent former yvTGMV DNA A or DNA B clones, respectively. Divergence (Div.) of nucleotides (nt) is represented by the nt in the ref. sequence, followed by genome position (pos.) and nt identified in the clones of this study (ref→[92/11]). When comparing these clones [92/11] to each other, the order is csTGMV[92/11]→(pos.)→yvTGMV[92/11]. Amino acid (aa) substitutions are shown the same way; grey boxes: non-conserved aa exchanges. Deletions: "d" followed by nt and genomic pos. Insertions: genomic pos. before insertion followed by "↓" and nt(s) inserted. *: stop codon. Consensus column: comparison between aa detected in this study and closely related proteins recovered from the GenBank. "yes" - aa change conforms with consensus in other geminiviruses, "no" - ref. sequence conforms better, "var": variable sequence with no decision possible. Genomic regions: ORFs (nc: non-coding).

	yvTGMV DNA A (K02029)			
	Genomic region	Div. nt pos. (ref→[92/11])	aa pos. exchange	Consensus
	AV1	T240A	S18T	no
		A338G	-	-
		C347T	-	-
		T494C	-	-
		T551C	-	-
		C698G	I170M	yes
		G773A	-	-
		G821A	-	-
	AC1	G1490A	T325I	yes
		T1534A	-	-
		T1594C	-	-
		A1603G	-	-
		A1612G	-	-
		A1672G	-	-
		G1678A	-	-
		C1711T	-	-
		C1732A	-	-
		G1795A	-	-
		T1844C	K207R	yes
		T1867C	-	-
		A1930C	-	-
		T1940G	D175A	var
		G2156T	T103K	no
		A2345G	I40T	yes
		A2379C	S29A	yes
	AC2	C1161T	-	-
		A1189C	L92R	var
		T1286C	R60G	no
		T1313C	-	-
csTGMV[92/11] DNA A	AC3	G981A	S116L	no
		C1161T	R56K	no
		A1189C	L47V	no
		T1286C	-	-
		T1313C	-	-
	AC4	G2156T	R52S	var
	nc	A45G	-	-
		C60A	-	-
		T90C	-	-
		A124G	-	-
		T135C	-	-
		135↓C136	-	-
		T149A	-	-
		T169A	-	-
		A175C	-	-
	csTGMV DNA B (M73794)			
csTGMV[92/11] DNA B	Genomic region	Div. nt pos. (ref→[92/11])	aa pos. exchange	Consensus
	nc	dT1137	-	-

(Table 2 continued on following page)

(Table 2 continued)

yvTGMV[92/11] DNA A	yvTGMV DNA A (K02029)			
	Genomic region	Div. nt pos. (ref→[92/11])	aa pos. exchange	Consensus
	AC1	C1732A	-	-
	yvTGMV DNA B (K02030)			
	Genomic region	Div. nt pos. (ref→[92/11])	aa pos. exchange	Consensus
yvTGMV[92/11] DNA B	BV1	C558T	-	-
	BC1	1367↓G	na	na
		1377↓G	na	na
		1416↓GC	na	na
		G1418T	na	na
		dG1454	na	na
		1481↓CCCAATAGTAG	na	na
		1498↓A	*185K	yes
		1538↓A	K184R	yes
			E183D	yes
			I182-	yes
			G181W	yes
			N180K	yes
			M178Y	yes
			D177G	yes
			W176V	yes
			T175H	yes
			A174S	yes
			S173F	yes
	nc	2069↓G	-	-
		A200T	-	-
		dC1135	-	-
		T2088C	-	-
	yvTGMV DNA B (K02030)			
	Genomic region	Div. nt pos. (ref→[92/11])	aa pos. exchange	Consensus
csTGMV[92/11] DNA B (continued next page)	BV1	A336G	-	-
		T351A	F10L	var
		G364C	G15R	yes
		G404A	R28H	yes
		C468T	-	-
		T759C	-	-
		G811A	V164I	yes
		A843T	-	-
		C870T	-	-
		C900T	-	-
		T908C	M196T	yes
	BC1 (to be continued next page)	T1208G	na	na
		T1256C	na	na
		A1277T	na	na
		C1296T	na	na
		1366↓G	na	na
		1377↓G	na	na
		1416↓GC	na	na
		G1418T	na	na
		C1421T	na	na
		dG1458	na	na
		1478↓GAGCCCAATAG	na	na
		A1480G	na	na
		1498↓A	* 185M	yes
		1536↓A	E183R	yes
			I182D	yes
			G181W	yes
			N180K	yes
			M178Y	yes
			D177G	yes
			W176V	yes
			T175H	yes
			A174S	yes
			S173F	yes

(Table 2 continued; ref. seq. yvTGMV DNA B (K02030)

	BC1 (continued)	C1772T	-	-
		A1823G	-	-
		A1957G	-	-
		G2022A	A11V	yes
		G2064A	-	-
		C2066T	-	-
csTGMV[92/11] DNA B (continued)	nc	T40C	-	-
		T69C	-	-
		T79A	-	-
		A100C	-	-
		C184T	-	-
		G188A	-	-
		A191G	-	-
		G192A	-	-
		T193C	-	-
		T258C	-	-
		A302G	-	-
		T1100C	-	-
		dC1135	-	-
		2069↓G	-	-
		C2089A	-	-
		C2140A	-	-
		A2207G	-	-
		G2209A	-	-
		A2231T	-	-
		A2262G	-	-
		C2269A	-	-
		C2287G	-	-
		G2296T	-	-
		A2297T	-	-
		C2300A	-	-
		T2301A	-	-
		T2312G	-	-
		G2324A	-	-
		G2327T	-	-
		T2331G	-	-
		C2365T	-	-
		T2374C	-	-
		G2386T	-	-
		G2394A	-	-
		A2395G	-	-
		G2399C	-	-
		T2409A	-	-
		G2413A	-	-
		T2414G	-	-
		T2415G	-	yes
		T2416A	-	-
		A2423C	-	-
		A2437T	-	-
		dT2502	-	-

(Table 2 continued on following page)

(Table 2 continued)

	yvTGMV[92/11] DNA A			
	Genomic region	Div. nt pos. (yv→cs)	aa pos. exchange	Consensus
	AV1	T240A	S18T	no
		A338G	-	-
		C347T	-	-
		T494C	-	-
		T551C	-	-
		C698G	I170M	yes
		G773A	-	-
		G821A	-	-
	AC1	G1490A	T325I	yes
		T1534A	-	-
		T1594C	-	-
		A1603G	-	-
		A1612G	-	-
		A1672G	-	-
		G1678A	-	-
		C1711T	-	-
		C1732A	-	-
		G1795A	-	-
		T1844C	K207R	yes
		T1867C	-	-
		A1930C	-	-
		T1940G	D175A	var
		G2156T	T103K	no
		A2345G	I40T	yes
		A2379C	S29A	yes
	AC2	C1161T	-	-
		A1189C	L92R	var
		T1286C	R60G	no
		T1313C	-	-
	AC3	G981A	S116L	no
		C1161T	R56K	no
		A1189C	L47V	no
		T1286C	-	-
		T1313C	-	-
	AC4	G2156T	R52S	var
csTGMV[92/11] DNA A	nc	A45G	-	-
		C60A	-	-
		T90C	-	-
		A124G	-	-
		T135C	-	-
		135↓C	-	-
		T149A	-	-
		T169A	-	-
		A175C	-	-
	yvTGMV[92/11] DNA B			
	Genomic region	Div. nt pos. (yv→cs)	aa pos. exchange	Consensus
yvTGMV[92/11] DNA B (continued next page)	BV1	A336G	-	-
		T351A	F10L	var
		G364C	G15R	yes
		G404A	R28H	yes
		C468T	-	-
		T558C	-	-
		T759C	-	-
		G811A	V164I	yes
		A843G	-	-
		C870T	-	-
		C900T	-	-
		T908C	M196T	yes
	BC1 (to be continued next page)	T1207C	K288Q	no
		T1255C	I272V	var
		A1276T	L265M	yes
		C1295T	-	-
		C1424T	-	-

(Table 2 continued; ref. seq. csTGMV[92/11] DNA B)

	BC1 (continued)	T1481G	-	-
		A1493G	-	-
		C1787T	-	-
		A1838G	-	-
		A1972G	-	-
		C2037A	A11V	yes
csTGMV[92/11] DNA B (continued)	nc	T40C	-	-
		T69C	-	-
		T79A	-	-
		A100C	-	-
		C184T	-	-
		G188A	-	-
		A191G	-	-
		G192A	-	-
		T193C	-	-
		T200A	-	-
		T258C	-	-
		A302G	-	-
		T1100C	-	-
		G2079A	-	-
		C2081T	-	-
		C2104T	-	-
		C2105A	-	-
		C2156A	-	-
		A2223G	-	-
		G2225A	-	-
		A2247T	-	-
		A2278G	-	-
		C2285A	-	-
		C2303G	-	-
		G2312T	-	-
		A2313C	-	-
		C2316A	-	-
		T2317A	-	-
		T2328G	-	-
		G2340A	-	-
		G2343T	-	-
		T2347G	-	-
		C2381T	-	-
		T2390C	-	-
		G2402T	-	-
		G2410A	-	-
		A2411G	-	-
		G2415C	-	-
		T2425A	-	-
		G2429A	-	-
		T2430G	-	-
		T2431G	-	-
		T2432A	-	-
		A2439C	-	-
		A2453T	-	-
		2517↓T	-	-

"-": no aa exchange (in non-coding regions); "na": not applicable.

The major phenotypic difference between cs and yvTGMV has been attributed to not more than two amino acid differences in the MP (Saunders et al., 2001), which might be a result of serial passaging *in planta*. However, in early publications on the tomato golden mosaic disease in Brazil, Costa et al. (1975) mention the existence of several distinct viruses in affected tomato. It is therefore possible as well that in the original

tomato samples, both TGMV variants, yv and cs, were present and were segregated through their passage in different laboratory host plants.

Acknowledgements

References

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W.and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25** (17)**,** 3389-3402.

Bisaro, D. M., Hamilton, W. D., Coutts, R. H.and Buck, K. W. (1982). Molecular cloning and characterisation of the two DNA components of tomato golden mosaic virus. *Nucleic Acids Res.* **10** (16)**,** 4913-4922.

Costa, A. S., Oliveira, A. R.and Silva, D. M. (1975). Transmissão mecânica do mosaico dourado do tomateiro. *Fitopatol. Bras.* **6,** 147.

Egelkrout, E. M., Robertson, D.and Hanley-Bowdoin, L. (2001). Proliferating cell nuclear antigen transcription is repressed through an E2F consensus element and activated by geminivirus infection in mature leaves. *Plant Cell* **13** (6)**,** 1437-1452.

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

Fauquet, C. M.and Stanley, J. (2005). Revising the way we conceive and name viruses below the species level: a review of geminivirus taxonomy calls for new standardized isolate descriptors. *Arch. Virol.* **150** (10)**,** 2151-2179.

Fontes, E. P., Gladfelter, H. J., Schaffer, R. L., Petty, I. T.and Hanley-Bowdoin, L. (1994). Geminivirus replication origins have a modular organization. *Plant Cell* **6** (3)**,** 405-416.

Gardiner, W. E., Sunter, G., Brand, L., Elmer, J. S., Rogers, S. G.and Bisaro, D. M. (1988). Genetic analysis of tomato golden mosaic virus: the coat protein is not required for systemic spread or symptom development. *EMBO J.* **7** (4)**,** 899-904.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41,** 95-98.

Hamilton, W. D., Bisaro, D. M., Coutts, R. H.and Buck, K. W. (1983). Demonstration of the bipartite nature of the genome of a single-stranded DNA plant virus by infection with the cloned DNA components. *Nucleic Acids Res.* **11** (21)**,** 7387-7396.

Hamilton, W. D., Stein, V. E., Coutts, R. H.and Buck, K. W. (1984). Complete nucleotide sequence of the infectious cloned DNA components of tomato

golden mosaic virus: potential coding regions and regulatory sequences. *EMBO J.* **3** (9)**,** 2197-2205.

Hanley-Bowdoin, L., Elmer, J. S.and Rogers, S. G. (1988). Transient expression of heterologous RNAs using tomato golden mosaic virus. *Nucleic Acids Res.* **16** (22)**,** 10511-10528.

Hanley-Bowdoin, L., Elmer, J. S.and Rogers, S. G. (1989). Functional expression of the leftward open reading frames of the A component of tomato golden mosaic virus in transgenic tobacco plants. *Plant Cell* **1** (11)**,** 1057-1067.

Hoekema, A., Hirsch, P. R., Hooykaas, P. J. J.and Schilperoort, R. A. (1983). A binary plant vector strategy based on separation of vir- and T-region of the Agrobacterium tumefaciens Ti-plasmid **303** (5913)**,** 179-180.

Kleinow, T., Nischang, M., Beck, A., Kratzer, U., Tanwir, F., Preiss, W., Kepp, G.and Jeske, H. (2009). Three C-terminal phosphorylation sites in the Abutilon mosaic virus movement protein affect symptom development and viral DNA accumulation. *Virology* **390** (1)**,** 89-101.

Klinkenberg, F. A., Ellwood, S.and Stanley, J. (1989). Fate of African cassava mosaic virus coat protein deletion mutants after agroinoculation. *J. Gen. Virol.* **70,** 1837-1844.

Matyis, J. C., Silva, D. M., Oliveira, A. R.and Costa, A. S. (1975). Purificacao e morfologia do virus do mosaico dourado do tomateiro. *Summa Phytopathol.* **1,** 267–273.

Orozco, B. M.and Hanley-Bowdoin, L. (1996). A DNA structure is required for geminivirus replication origin function. *J. Virol.* **70** (1)**,** 148-158.

Saunders, K., Wege, C., Veluthambi, K., Jeske, H.and Stanley, J. (2001). The distinct disease phenotypes of the common and yellow vein strains of Tomato golden mosaic virus are determined by nucleotide differences in the 3'-terminal region of the gene encoding the movement protein. *J. Gen. Virol.* **82** (Pt 1)**,** 45-51.

Thommes, P., Osman, T. A., Hayes, R. J.and Buck, K. W. (1993). TGMV replication protein AL1 preferentially binds to single-stranded DNA from the common region. *FEBS Letters* **319** (1-2)**,** 95-99.

Thompson, J. D., Higgins, D. G.and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22** (22)**,** 4673-4680.

Unseld, S., Hohnle, M., Ringel, M.and Frischmuth, T. (2001). Subcellular targeting of the coat protein of African cassava mosaic geminivirus. *Virology* **286** (2)**,** 373-383.

von Arnim, A.and Stanley, J. (1992). Determinants of tomato golden mosaic virus symptom development located on DNA B. *Virology* **186** (1)**,** 286-293.

Segregating geminiviruses in *Asystasia gangetica* induce distinct symptoms

Patricia Soares Wyant[1], Stephan Strohmeier[1], Benjamin Schäfer[1], Annika Allinger[1], Björn Krenz[1], Rob W Briddon[2] and Holger Jeske[1]

[1]Institute of Biology, Department of Molecular Biology and Plant Virology, University of Stuttgart, Pfaffenwaldring 57, 70550 Stuttgart, Germany.

[2]National Institute for Biotechnology and Genetic Engineering, Jhang Road, P. O. Box 577, Faisalabad, Pakistan.

Abstract

Asystasia gangetica of the family *Acanthaceae* originates from tropical Africa and Asia, but was introduced into many other tropical regions. It is used as source of food and for medical applications. Plants with typical geminivirus symptoms were collected in West Africa in the 1980s, and propagated since then vegetatively in Norwich, United Kingdom and in Stuttgart, Germany. Interestingly, these plants showed a curious segregation of symptoms such as yellow veining, curling and mosaic, which can be present simultaneously or separately in different leaves of the same plant or different cuttings. By the use of rolling circle amplification (RCA) in combination with restriction fragment length polymorphism (RFLP) followed by deep sequencing of the RCA products, we identified two geminiviruses in these plants. One with a bipartite genome, Asystasia begomovirus 1, and the other with a monopartite genome together with its defective DNA, Asystasia begomovirus 2. The interrelatedness of leaf symptoms and virus segregation under different light regimes has been investigated, and showed for the first time a remarkable segregation of geminiviruses in a single plant.

Key words: *Asystasia gangetica*, RCA, segregation of symptoms, geminivirus.

Introduction

Asystasia gangetica (Linnaeus) Thomas Anderson, also known as *Asystasia coromandeliana* Nees, and by the popular names, chinese violet, coromandel,

ganges primrose, or philippine violet, is a member of the family *Acanthaceae* (Integrated taxonomic information system, ITIS, available on: http://www.itis.gov). It is indigenous in tropical Africa and Asia, but has been also introduced into many other tropical regions (Plant resources of tropical Africa, PROTA, 2010, available at www.prota.org). *A. gangetica* is a source for food and medical application in several countries. A study regarding food value of tropical plants consumed as vegetables by ethnic communities in Singapore, Malaysia and Indonesia showed that their leaves contain high amounts of proteins, amino acids, minerals, sugars, lipids, and fiber (Yeoh and Wong, 1993). In India it is used as stomachic, astringent, diaphoretic and antiasthmatic, flower extracts are used to treat pyrexia and conjunctivitis, and the whole plant is used as a medicine for spasm of bladder, and the roots in the treatment of edema (Sudhakar et al., 2006). Sudhakar et al. (2006) investigated the antimicrobial activity of plants used in India as traditional medicine including *A. gangetica* and have found that ethanolic extract exhibited antimicrobial activity against *Escherichia coli*, *Proteus vulgaris*, *Pseudomonas aeruginosa* and *Staphylococcus aureus*. In Nigeria, extracts of the leaves are used to treat asthma. The anti-asthmatic property of the leaf extracts was evaluated, and it was suggested, that they were indeed able to inhibit trachea contractions induced by histamine, serotonin and acetylcholine, agents implicated in the pathogenesis of asthma (Akah et al., 2003).

Geminivirus infected *A. gangetica* plants collected originally by Dr. Peter Markham (Norwich, United Kingdom) in West-Africa during the 1980s were propagated vegetatively in Norwich, United Kingdom and in Stuttgart, Germany. The symptoms yellow-green mosaic, yellow vein and leaf curling are observed all together or separately in different leaves of the same plant or different cuttings. A geminivirus detected infecting the plants was tentatively named Asystasia golden mosaic virus, AGMV (Briddon and Markham, 1994; Briddon et al., 1993) and a first draft sequence of one of the viral DNA components was determined (Rob Briddon, unpublished). The virus has been shown to be whitefly-transmitted specifically by the *Bemisia tabaci* biotype E (Brown et al., 1995) or an *Asystasia B. tabaci* population (Bedford et al., 1994; Maruthi et al., 2002).

Rolling circle amplification (RCA) combined with restriction fragment length polymorphism (RFLP) has been proven to be a useful tool for the detection of

geminiviruses (Haible et al., 2006; Inoue-Nagata et al., 2004; Jeske et al., 2010; Kumar et al., 2008; Paprotka et al., 2010a; Wyant et al., 2011). It utilizes the DNA polymerase of the *Bacillus subtilis* bacteriophage phi (Φ) 29 (Blanco et al., 1989; Dean et al., 2001), which possesses polymerase and strand-displacement activities, allowing circular templates to be amplified preferentially (Lizardi et al., 1998; Zhang et al., 1998; Zhang et al., 2001). The technique uses random hexamer primers, allowing the amplification of viral DNA without initial knowledge of the sequence (Dean et al., 2001; Haible et al., 2006; Jeske et al., 2010; Lizardi et al., 1998; Paprotka et al., 2010a; Paprotka et al., 2010c; Wyant et al., 2011). Combining RCA/RFLP-based diagnosis and deep sequencing we identified two geminiviruses within the *Asystasia* plants and investigated their fate in correlation to symptom segregation in leaves.

Material and methods

Nucleic acid extraction

Total DNA of samples from *A. gangetica* plants, which were kindly provided by Dr. Peter Markham (Norwich, UK) were extracted using a CTAB-based method modified from described by Kleinow et al.(2009). Briefly, leaf disks (10 - 50 mg) were ground in liquid nitrogen and added to 500 µl extraction buffer (100 mM Tris–HCl pH 8.0; 20 mM EDTA; 1.4 M NaCl; 2% cetyltrimethylammonium bromide, CTAB; 0.5 M glucose). After shaking at 60 °C for one hour, the suspension was mixed with one volume of chloroform/isoamylalcohol (24:1) and centrifuged for 5 min at maximum speed at 4 °C. Nucleic acids in the aqueous phase were pelleted by adding 0.8 volumes of isopropanol and centrifugation (10 min, at 14,000x*g* at 4 °C). DNA was washed with 70% ethanol and resuspended in 50 µl of sterile water.

RCA

Amplification of circular DNA was performed using TempliPhi™ Kit (GE Healthcare, Munich, Germany) following the manufacturer's protocol. Briefly, 2 µl (10 ng to 20 ng) of total DNA were dissolved in 5 µl of sample buffer, denatured for 3 min at 95 °C and cooled down on ice for 1 minute followed by addition of 5 µl reaction buffer and 0.2 µl

enzyme mix. Amplification was performed for 16-20 h at 30 °C and the reaction was stopped by heating for 10 min at 65 °C to inactivate the enzyme.

RFLP

The RFLP analysis was performed by digesting 1 µl (~300 ng DNA) of the RCA product with the restriction enzymes *Hpa*II or *Bfu*CI (New England Biolabs, Frankfurt, Germany). The reaction was carried out at 37 °C for 2 h, followed by treatment for 20 min at 65 °C for enzyme inactivation, according to supplier's recommendation. DNA fragments were separated on 2% agarose gel following standard protocols (Sambrook and Russell, 2001) and visualized by ethidium bromide staining. Two molecular weight markers with known reference fragments were used to the estimation of fragment sizes, bacteriophage Lambda DNA (Fermentas, St. Leon-Roth, Germany) digested by *Pst*I (New England Biolabs, Frankfurt, Germany) and RCA products of Abutilon mosaic virus (AbMV) digested in parallel by *Hpa*II or *Pst*I (New England Biolabs, Frankfurt, Germany). Fragments sizes were estimated by using an exponential fit between retention factor (Rf) and the logarithms of the molecular weights using Microsoft Excel for calculations.

Sequencing

Samples positive for the amplification of a small circular DNA as monitored by the RFLP analysis were pooled with RCA products of AbMV generated from samples of the *Abutilon sellovianum* used to fulfill Koch's postulates for AbMV (Wege et al., 2000) and from Tobacco mosaic virus (TMV) recombinant plasmid pTMV (Kadri et al., 2011) and sent for commercial pyrosequencing at GATC Biotech (Constance, Germany). Complete sequences of DNAs A and B components as well as a defective DNA A were assembled using an own alpha version of software (unpublished). The assembly of the DNAsA and B sequences was carried out by Stephan Strohmeier.

Sequences analysis

The DNA sequences were compared with the international databases using BLAST (Altschul et al., 1997). The analysis of open reading frames (ORFs) integrity, sequence similarity and *in silico* enzymatic restriction was made using the Geneious

Pro[TM] software; Biomatters Ltd., Auckland, New Zealand (Drummond et al., 2010). For the phylogenetic analysis, sequences of the ten most closely related viral sequences were retrieved from the GenBank (March, 2011). Sequence alignments were performed using codaln software (Stocsits et al., 2005), and neighbor joining trees were calculated using the algorithms included in MEGA 4.0 (Tamura et al., 2007) with 1,000 bootstrap replications each. Viral names and accession numbers are listed in Table 1.

Table1. Sequences retrieved from GenBank: Virus names, abbreviations and accession numbers of sequences used for the phylogenetic analysis.

| Name | abbreviation | accession number | | origin |
		DNA A	DNA B	
Abutilon mosaic virus	AbMV	NC_014138	NC_014139	Caribbean region
Ageratum leaf curl Cameroon virus	ALCuCMV	FR717144	-	Cameroon
Ageratum yellow vein Sri Lanka virus	AYVSLV	AF314144	-	Sri Lanka
Chilli leaf curl virus-India	ChiLCuV	DQ989326	-	India
Croton yellow vein mosaic virus	CYVMV	FN645915	-	India
East African cassava mosaic Cameroon virus	EACMCV	-	AF112355	Cameroon
East African cassava mosaic Kenya virus	EACMKV	-	AJ704970	Kenya
East African cassava mosaic virus	EACMV	-	AJ704936	Kenya
East African cassava mosaic Zanzibar virus	EACMZV	-	AJ704943	Kenya
Horsegram yellow mosaic virus	HYMV	-	AM932430	India
Kudzu mosaic virus	KuMV	-	FJ539015	China
Mimosa yellow leaf curl virus	MiYLCV	DQ641695	-	Vietnam
Mungbean yellow mosaic virus	MYMV	-	DQ400849	India
Papaya leaf curl China virus	PaLCuCNV	AJ704604	-	China
Pepper yellow vein Mali virus	PepYVMV	AM691547	-	China
Rhynchosia yellow mosaic virus	RhYMV	-	AM999982	Pakistan
Tobacco curly shoot virus-[Y41]	TbCSV-[Y41]	AJ457986	-	China
Tobacco leaf curl Zimbabwe virus	TbLCZV	AM701756	-	Comoros
Tomato curly stunt virus - [South Africa]	ToCSV-[ZA]	AF261885	-	South Africa
Tomato leaf curl Anjouan virus	ToLCAnjV	AM701759	-	Comoros
Tomato leaf curl Antsiranana virus	ToLCAntV	AM701761	-	Comoros
Tomato leaf curl Cameroon virus	ToLCCMV	FM210278	-	Cameroon
Tomato leaf curl Comoros virus	ToLCKMV	FN600540	-	Comoros
Tomato leaf curl Diana virus	TYLCDiV	AM701765	-	Madagascar
Tomato leaf curl Hainan virus	ToLCHnV	FN434083	-	China
Tomato leaf curl Kerala virus	ToLCKeV	EU910140	-	India
Tomato leaf curl Madagascar virus	ToLCMGV	AJ865338	-	Madagascar
Tomato leaf curl virus-[Buea]	TYLCV	FM210277	-	Cameroon
Tomato yellow leaf curl virus-[Jiroft:Iran]	TYLCV-OM[IR:Ji31:06]	GU076441	-	Iran
Watermelon chlorotic stunt virus-[SD]	WmCSV-SD	-	AJ245651	Iran

DNAs A are from monopartite begomoviruses with exception of AbMV;
DNAs B are from bipartite begomoviruses used for the DNA B phylogenetic analysis.

Plant propagation under different light regimes

A. gangetica plants were propagated via stem cuttings. Two months after propagation and after the cuttings had rooted, half of them were grown under high light or low light conditions for six months, May to November, in the greenhouse with 16 hours of photoperiod, temperature of 14 °C during the night and 22 °C during the day and humidity of at least 52% and maximum of 86%. High light conditions

correspond to a photosynthetically active radiation (PAR) of 75.6 µmol photons/m^2/s, i.e. spectral range from 400 to 700 nm that photosynthetic organisms are able to use in the process of photosynthesis and is measured as the photosynthetic photon flux density per m^2 of area per seconds to which the plant is exposed, whereas low light/shade conditions represent a PAR of 1.6 µmol photons/m^2/s. Plant propagation was performed by Annika Allinger.

Biolistic inoculation

Datura stramonium and *Nicotiana benthamiana* in the 2-leaf stage were biolistically inoculated using the particle gun PDS1000/HE (Bio-Rad, München, Germany). Gold particles (1 µm in diameter) were coated with RCA products and bombarded using 900 psi rupture discs under a vacuum pressure of 27 inch. Hg as described by Unseld et al. (2001).

Results and discussion

Sequence analysis

For investigation of the geminiviruses present in the *Asystasia gangetica* plants, small circular DNAs were amplified from total nucleic acid extracts of fresh leaf material and analyzed by rolling circle amplification (RCA) followed by restriction fragment length polymorphism (RFLP). RCA products of positive samples were pooled and examined by direct pyrosequencing. The assembly of sequence data clearly revealed the presence of four begomoviral components. A cognate DNA A and B pair (DNA A1 and B1) sharing 76.6% sequence identity (SI) within their common region (CR), which consists of 184 nt, and identical iterons, that are the predicted Rep-binding DNA motifs (Argüello-Astorga et al., 1994; Argüello-Astorga and Ruiz-Medrano, 2001) as well as a second DNA A (DNA A2), and its defective DNA were identified.

DNA A1 and DNA A2 shared 67.4% overall nucleotide SI and therefore classified as different begomovirus species according to ICTV criteria (Fauquet et al., 2005). Comparing their sequences with those from the international databases (March, 2011), DNA A1 showed the closest similarity to Ageratum yellow vein Sri Lanka virus

(AYVSLV, accession number AF314144.1; 72% SI for 83% coverage), DNA A2 to Tomato yellow leaf curl virus-[Roodan:Iran] isolate TYLCV-OM[IR:Ro25:06] (GU076451.1; 76% SI for 97% coverage) and the DNA B1 sequence was distinct from any other virus sequence available on GenBank. The closest sequence to DNA B was from East African cassava mosaic Zanzibar virus (AJ704943.1) and showed 73% SI with coverage of only 25% of the total genome. Therefore, the viruses found in this study can be classified as novel species and the names Asystasia begomovirus 1 (ABgV1) for DNA A1 (accession number JF694484) and DNA B1 (JF694485), and Asystasia begomovirus 2 (ABgV2) for DNA A2 (JF694486) and its defective DNA (JF694487) is given tentatively.

DNA A1 and DNA B1 genomic organization resemble bipartite Old World begomoviruses. DNA A1 contained six putative open reading frames (ORFs), which encode pre-coat protein (PCP), coat protein (CP), replication-associated protein (Rep), transcriptional activator protein (TrAP), replication enhancer protein (REn) and AC4 protein, and DNA B1 contains two ORFs, one encoding the movement protein (MP) while the other the nuclear shuttle protein (NSP) (Jeske, 2009, and references therein). The DNA A2 genome also harbors six putative ORFs including PCP, CP, Rep, TrAP, REn and C4 (Fig. 1). The defective DNA A contains an incomplete Rep ORF. Since we have not found any cognate DNA B for DNA A2 in the large data set, DNA A2 may represent a monopartite Old World begomovirus, which does not need DNA B to be infectious (Jeske, 2009).
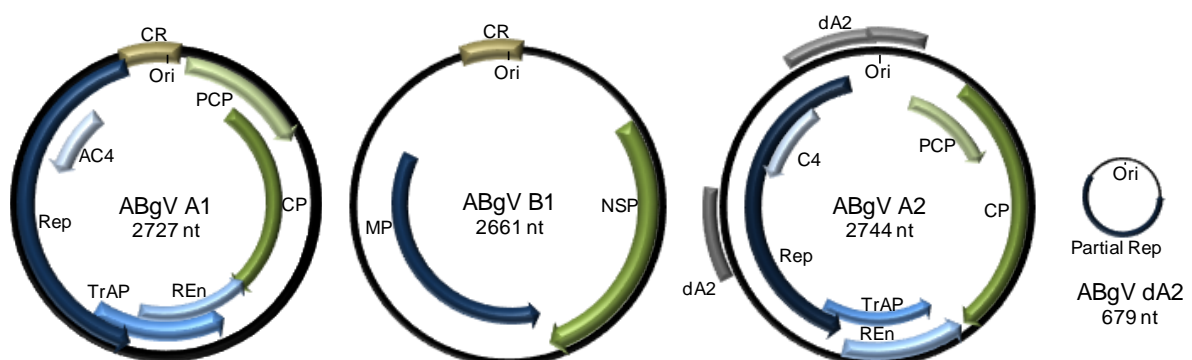

Figure 1. Genomic organization predicted by sequence data analysis: open reading frame (ORF) positions and directions of translation are indicated by arrows in DNA A and DNA B, PCP: Pre-coat protein, CP: coat protein, Rep: Rep-protein, TrAP: transcriptional activator protein, REn: replication enhancer, AC4/C4: protein with silencing suppressor and symptoms inducer function reported for some Old World begomoviruses, NSP: nuclear shuttle protein, MP: movement protein, CR: common region, Ori : origin of replication. The defective DNA A2 (ABgV dA2) is represented by grey boxes outside the map of ABgV A2 (dA2).

The phylogenetic analyses confirmed that both DNA A1 and DNA A2 are distantly related with each other and also with any other virus (Fig. 2). Consistent with the location of sample collection, DNA A2 and DNA B1 nucleotide sequences grouped mainly with sequences obtained from African virus isolates, with bootstraps values of 90% and 100%, respectively. Interestingly, DNA A1 did not cluster with any other virus sequence, showing to be distantly related with all viruses used for the analysis. This could be explained by the fact that all sequences used for the DNA A phylogeny were from monopartite begomoviruses whereas DNA A1 belongs to a bipartite begomovirus. Curiously, the one hundred first hits on BLAST alignment for DNA A1 using the international sequence databank, from where the ten first were used for phylogenetics, were sequences of monopartite begomoviruses. The phylogenetic analyses results support the classification based on nucleotide sequence of DNA A1 and DNA A2 analysis as new virus species.
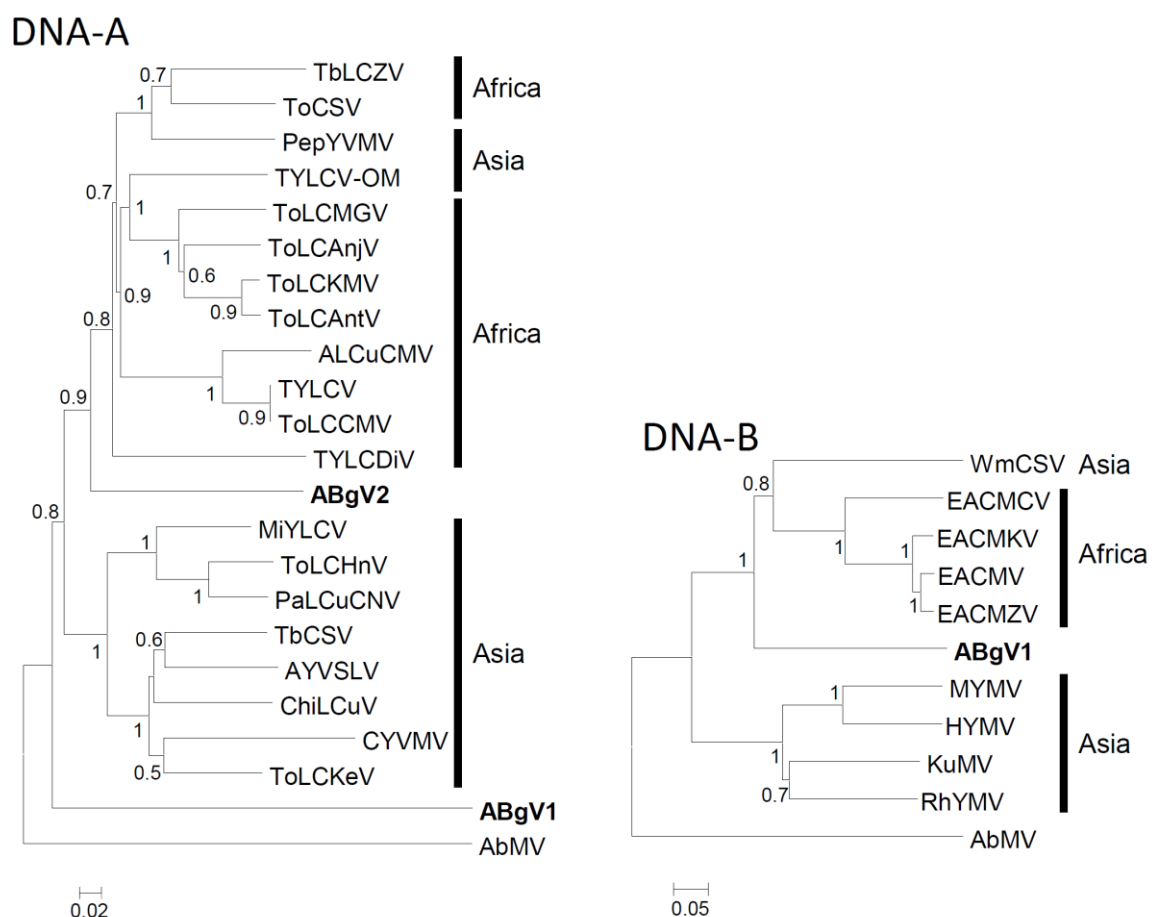
Figure 2. Neighbor Joining trees of DNA A and DNA B for the isolates detected in this study compared to the most closely related sequences retrieved from international databases (March, 2011), and using Abutilon mosaic virus (AbMV) as outgroup. Numbers next to the nodes indicate bootstrap values (1,000 replicates) above 50% (0.5). The novel viruses are highlighted. The corresponding names, abbreviations and accession numbers are listed in Table 1.

Symptom appearance and virus segregation

An example of symptom segregation of mosaic, yellow vein and curling in leaves of an *A. gangetica* plant is shown in Figure 3a. For analyses of symptom segregation, twenty stem cuttings from a single *Asystasia* plant, named Ag1, were prepared, consecutively numbered (Ag1.1 - Ag1.20.) and further cultivated. After six weeks of propagation, the offshoots showed separation of different symptoms depending on which branch of the original plant they were originated from. The most predominant symptoms observed on leaves were a yellow-green mosaic, yellow veins or yellow veins combined with curling. Some offshoots showed no symptoms at all. Plantlets with predominant leaf mosaic symptoms exhibited also stunting as compared to those possessing only yellow veins. To evaluate the influence of light on the symptom development, plant populations were divided into two groups and each was exposed to different light regimes (Table 2). Four plants showing predominantly leaf mosaic (Ag1.1, 1.3, 1.5 and 1.9), four with yellow veins (Ag1.7, 1.15, 1.18 and 1.20) and two without symptoms (Ag1.10 and 1.12) were placed under high light conditions. The group grown under low light conditions comprised four plants exhibiting a leaf mosaic (Ag1.2, 1.4, 1.6 and 1.8), five with yellow veins (Ag1.13, 1.14, 1.16, 1.17 and 1.19), from which one (Ag1.13) showed additional curling in some leaves, and one non-symptomatic plant (Ag1.11). For both groups the photoperiod, temperature and humidity for cultivation were kept identical.

Table 2. Summary of symptoms segregation observed on new stem cuttings of *Asystasia* plants after 40 days and light regime, under which the plants were submitted to for growing 40 days after cutting. High and low refer to the different light intensity regime. High: plants growing under high light, with photosynthetically active radiation of 75.6 μmol photon/m^2/s. Low: low light with PAR of 1.6 μmol photon/m^2/s.

High	Symptom	Low
Ag1.1		Ag1.2
Ag1.3	mosaic	Ag1.4
Ag1.5		Ag1.6
Ag1.9		Ag1.8
Ag1.7		Ag1.14
Ag1.15	yellow vein	Ag1.16
Ag1.18		Ag1.17
Ag1.20		Ag1.19
-	yellow vein + curling	Ag1.13
Ag1.10	none	Ag1.11
Ag1.12		-

First, a RCA/RFLP diagnosis using *Hpa*II or *Bfu*CI (Fig. 3b) was performed from total DNA extracted from the offshoots Ag1.4 and Ag1.5 (only leaf mosaic), Ag1.13 (yellow veins and curling) as well as Ag1.15 and Ag1.19 (only yellow veins). Different band patterns were detected for samples with varying symptoms. No products were detected for non-symptomatic plants (data not shown), indicating a clearance of viruses in those stem cuttings. Fragments were compared to the *in silico* predicted enzymatic restriction patterns of DNA A1, B1, A2 and dA2 revealing that in plants with mosaic symptoms or with yellow veins in combination with leaf curling all components were present. However, samples of plants showing yellow vein and curling showed a higher concentration of DNA A1 and B1 in comparison to those exhibiting mosaic-specific symptoms.

On Figure 3b, when fragment intensity is compared on samples Ag1.19 and Ag1.15, a clear predominance of DNA A2, either alone or together with DNA dA2, was found for samples derived from plants with only yellow vein symptoms. The same result was obtained when this comparison was made using leaves with different symptoms in their halves; one half displaying a mosaic combined with either yellow vein or no symptom on the other half (data not shown). The outcome suggests that the yellow vein symptom correlates with the presence of DNA A2, irrespectively of whether DNA dA2 was present. However, further investigation is needed to confirm this observation. Moreover, the data support the hypothesis that DNA A2 belongs to a monopartite begomovirus which is capable to induce symptoms and move in the absence of a DNA B component (Jeske, 2009, and references therein).
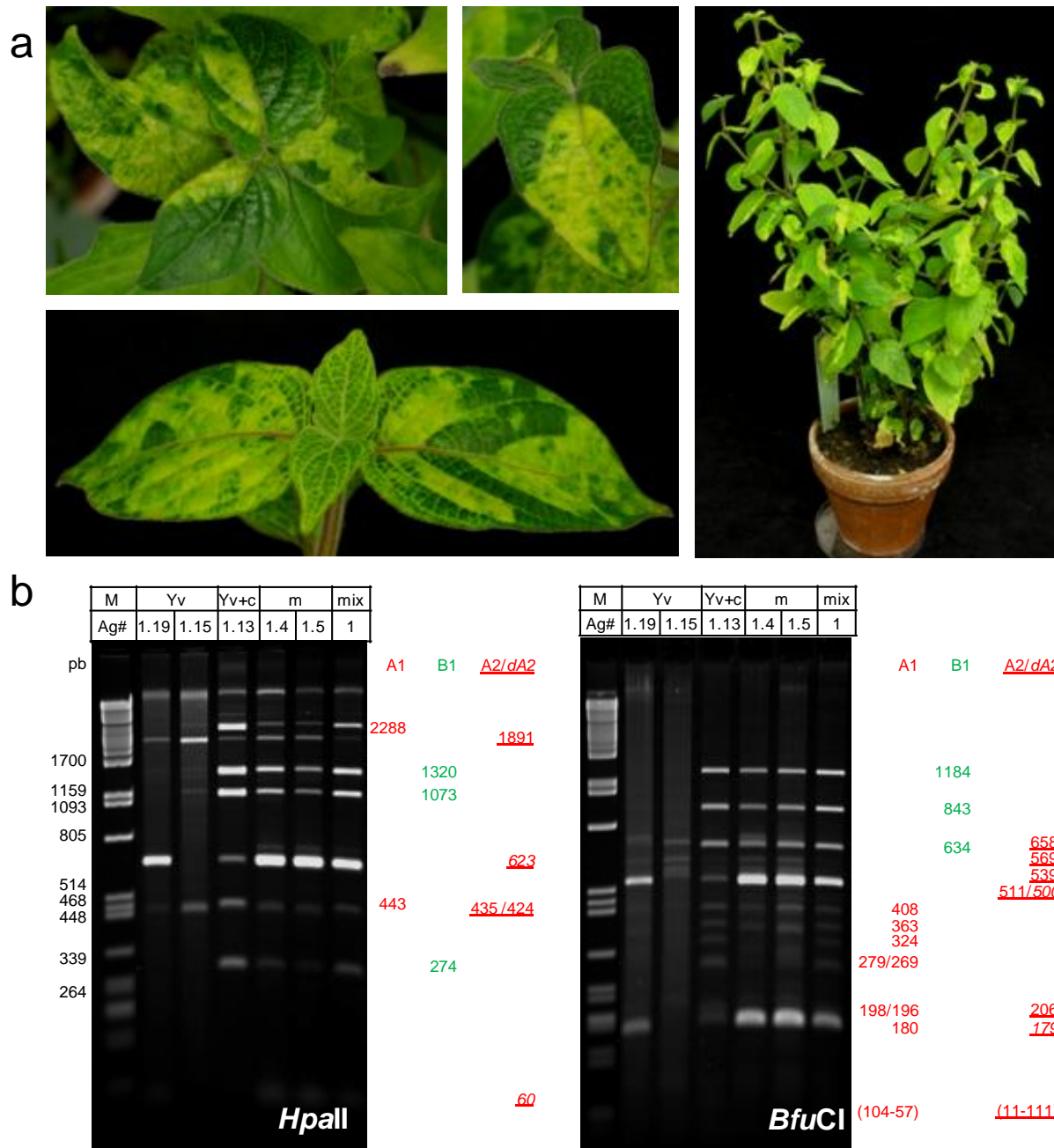
Figure 3. (a) Symptoms segregation in *Asystasia gangetica* plant and leaves. (b) Gel electrophoresis of rolling circle amplification (RCA) followed by restriction fragment length polymorphism (RFLP) products using *Hpa*II (left gel) or *Bfu*CI (right gel). The expected fragments sizes obtained by *in silico* digestion of the retrieved sequences are shown at the right side of each gel with DNA A1 fragments in red and DNA B1 fragments in green. Underlined fragments in red indicate DNA A2 and in italic are defective DNA dA2. Values in brackets refer to bands that are too small to be resolved in this gel system. Non-assigned fragments may result from polymorphism within the viral quasispecies, from other defective DNAs, from still unidentified begomovirus components or from other small circular DNAs from the plant, like mitochondrial DNA. M, molecular weight marker; yv, yellow vein; c, curling; m, mosaic; mix, positive control, sample of the plant Ag1 (mother plant, from which the cuttings were taken) with mixed symptoms. Ag, *A. gangetica*.

Six months after the plants have grown under different light conditions, the symptoms and the composition of the viruses were analyzed with the same techniques. The plants that had presented leaf mosaic symptoms and grown under low light

conditions died all. Whereas plants showing leaf mosaic and grown under the high light regime survived. The plants were severely stunted (Fig. 4a; Ag1.1, 1.3, 1.5, 1.9) and chlorotic (Fig. 4c; Ag1.1, 1.3, 1.5, 1.9). Plants that showed originally yellow vein symptoms (Ag1.7, 1.13 - 1.20) developed a yellow-green leaf mosaic, independent on the light condition (Fig. 4). Two initially asymptomatic plants (Ag1.11, grown under low light and Ag1.10 under high light) remained symptomless, whereas Ag1.12 (growing under high light) developed mosaic and yellow vein symptoms (Fig. 4c and d).

All plants showing only mosaic symptoms submitted to low light intensity died, whereas the plants submitted to high light conditions survived. In agreement with the observed result, several studies suggest that light plays an important role in plant defense against viral, bacterial and fungal pathogens with attenuated responses often been observed in the dark (Chandra-Shekara et al., 2006; Genoud et al., 2002; Griebel and Zeier, 2008). On the other hand, all plants with predominance of mosaic symptoms grown under high light conditions showed chlorosis, in agreement with Jeske and Werz (1978), that investigated light intensity in plants infected by Abutilon mosaic virus (AbMV) and found that plants submitted to high light conditions show increased mosaic symptoms and chlorosis. The appearance of mosaic symptoms in plants that initially showed only yellow veining and the emergence of symptoms in asymptomatic plants after six months might be related with the circadian clock, seasonal changes with difference in the length of the day (Roden and Ingle, 2009). These differences might interfere with the defense response of the plant and with viral replication processes, favoring replication of distinct components responsible for the induction of different symptoms.

A second round of RCA/RFLP-based diagnosis was performed on total nucleic acid extracted from plants after six months treatment with different light conditions. In plants that had showed only yellow vein and developed mosaic symptoms after the six months (Ag1.7, 1.13 - 1.20), DNA A1 and DNA B1 were detected, although initially those plants showed more predominantly detectable DNA A2 (Fig. 3b and 5).
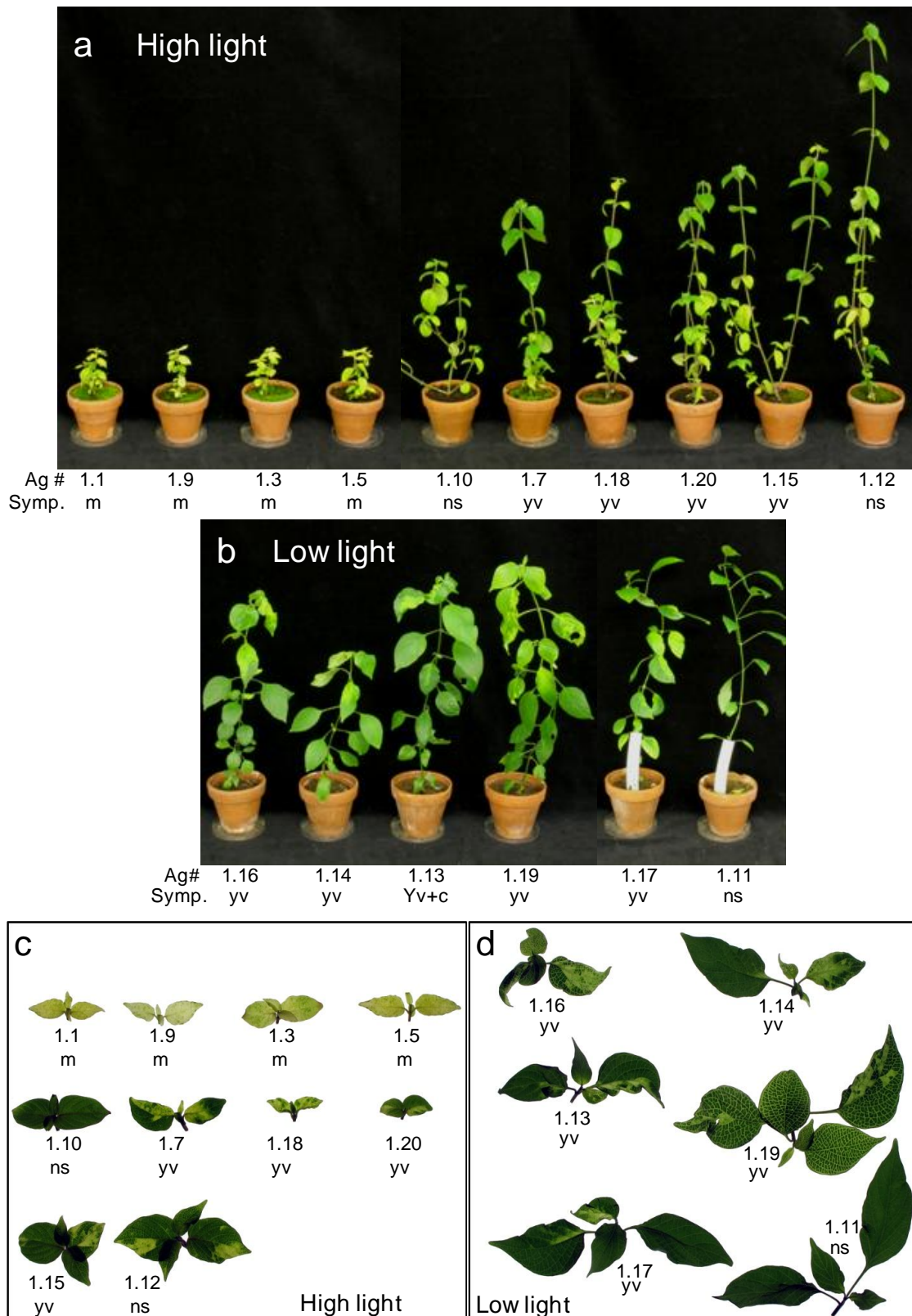
Figure 4. Comparison of symptom development between *Asystasia* plants growing under high light regime, with photosynthetically active radiation (PAR), i.e. measurement of photosynthetic photon flux density per m^2 of area per seconds to which the plant is exposed, of 75.6 μmol photons/m^2/s (a and c) and under low light regime with PAR of 1.6 μmol photons/m^2/s (b and d). Ag, *A. gangetica*; Symp., original predominant symptoms shown by the plantlets before separation in different light conditions; m, mosaic: yv, yellow vein; c, curling; ns, no symptoms. (a) and (b) Plants grown for six months under high light and low light regime, respectively; (c) and (d) Top leaves of the plants grown under high and low light regimes respectively.
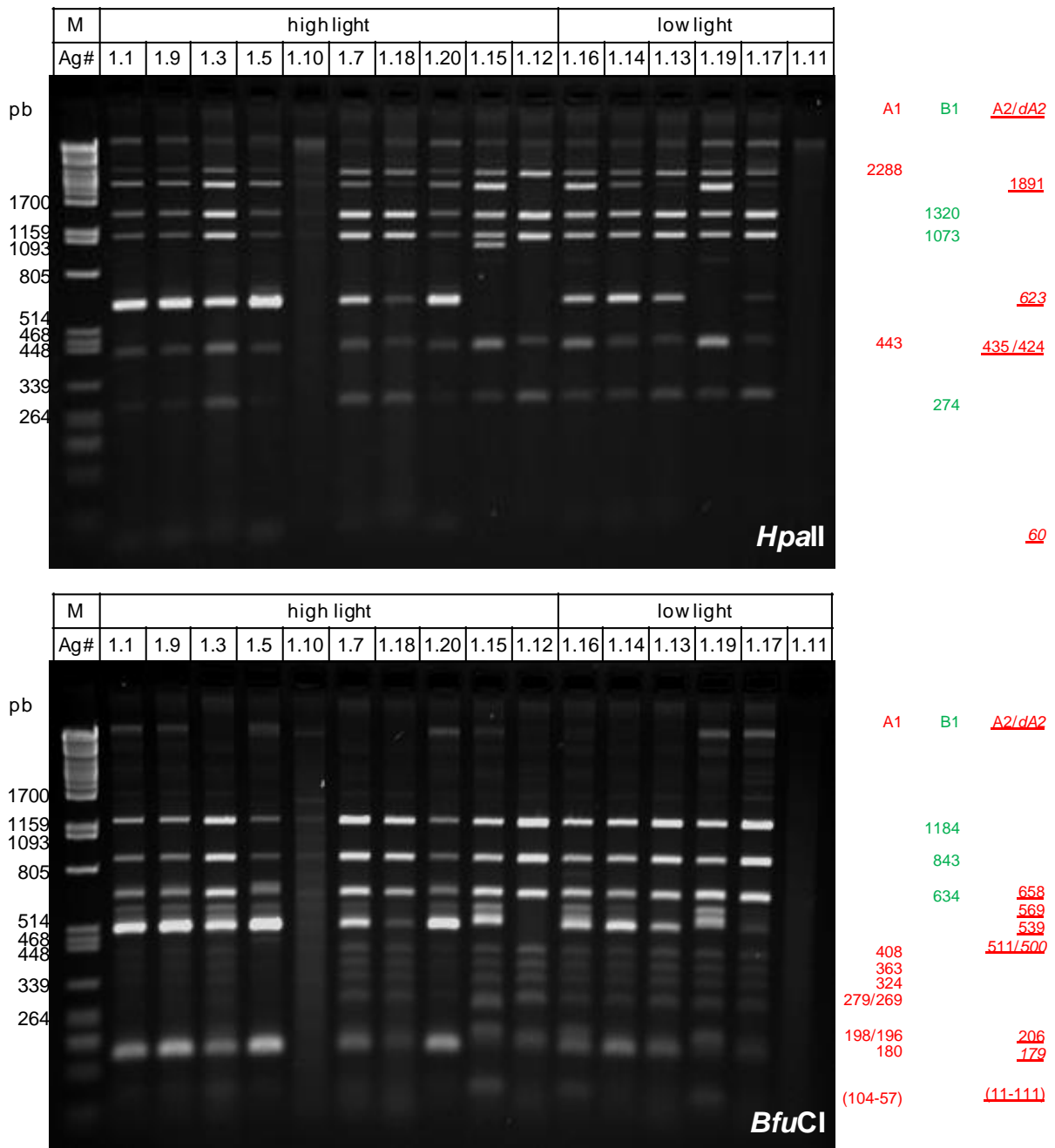
Figure 5. Gel electrophoresis of RCA/RFLP products of *Asystasia* stem cuttings six months after grown under distinct light conditions, high light regime with photosynthetically active radiation (PAR) of 75.6 µmol/m^2/s or low light regime, PAR of 1.6 µmol/m^2/s using *Hpa*II (top gel) or *Bfu*CI (bottom gel). The expected fragment sizes obtained by *in silico* digestion of the retrieved sequences are shown at the right side of each gel with DNA A1 fragments in red and DNA B1 fragments in green. Underlined red fragments indicate DNA A2 and in italic are defective DNA dA2. Values in brackets refer to bands that are too small to be resolved in this gel system. Fragments not assigned might be result of polymorphism within the viral quasispecies, new defective DNAs or plant circular DNAs, like mini mitochondrial plasmid. M, molecular weight marker.

The presence of DNA A1 and B1 but no DNA A2 was confirmed for Ag1.12, a plant that was asymptomatic and developed a leaf mosaic and yellow vein, suggesting that the DNA A2, although apparently related to yellow vein symptoms induction as discussed before, is not essential for yellow vein symptoms development. As the

RCA/RFP was performed using the same nucleic acid extraction, further experiments are needed to ensure the absence of DNA A2 in this plant. The sample Ag1.10, which after six months still did not develop symptoms showed to be infected with both viruses, with very low titers, on RFLP analysis (Fig. 5).

Retransmission of viruses to *Datura stramonium* and *Nicotiana benthamiana*

In order to evaluate the infectivity of the viruses retrieved from *A. gangetica* in other plant species, the RCA products of DNA A1, A2 and B1, from plant samples with mixed symptoms, including yellow-green mosaic, yellow veins and curling, were inoculated biolistically onto *D. stramonium* and *N. benthamiana*. The presence of the DNA components in the RCA products to be bombarded was confirmed by RCA/RFLP (mix on Fig. 3b). At four and six weeks post inoculation (wpi) D*. stramonium* and *N. benthamiana* plants, respectively, were scored for their symptom phenotype and analyzed by RCA/RFLP. Four out of ten inoculated *D. stramonium* showed stunting and leaf curling symptoms, whereas for *N. benthamiana* only one out of ten plants showed stunting symptom when compared to non-inoculated controls (Fig. 6a). RCA/RFLP-based diagnosis of all bombarded plants showed the expected fragment pattern for DNAs A1, B1 and A2 only for the single symptomatic *N. benthamiana* and for the four symptomatic *D. stramonium* plants. Additionally, three non-symptomatic *D. stramonium* plants showed to be infected by RCA/RFLP (Fig. 6b). These results confirm the infectivity of the described viruses in *N. benthamiana* and *D. stramonium* and suggest that *D. stramonium* is a better model to study these specific viruses since it seems to be more susceptible to the viral infection. In addition the results suggest that the defective DNA dA2 is not essential for systemic infection. Defective DNA molecules have been detected in plants infected with begomoviruses (Horn et al., 2011; Liu et al., 1998; Paximadis and Rey, 2001; Stanley et al., 1997) and to play role in symptom induction by attenuating (Stanley et al., 1990; Stanley et al., 1997) or increasing symptoms severity (Horn et al., 2011). Further analyses are needed to clarify the role of the defective DNA dA2 detected in this study in symptoms induction and viral infection.

Mixed viral infection on a same plant is often described in the literature (Renteria-Canett et al., 2011; Sanchez-Campos et al., 1999; Sikora et al., 1998; Wyant et al., 2011). Several studies showed that geminiviruses involved in mixed infection on a

plant frequently co-infect the same cell (Arguello-Astorga et al., 2007; Morilla et al., 2004; Patil et al., 2007; Saunders and Stanley, 1995).
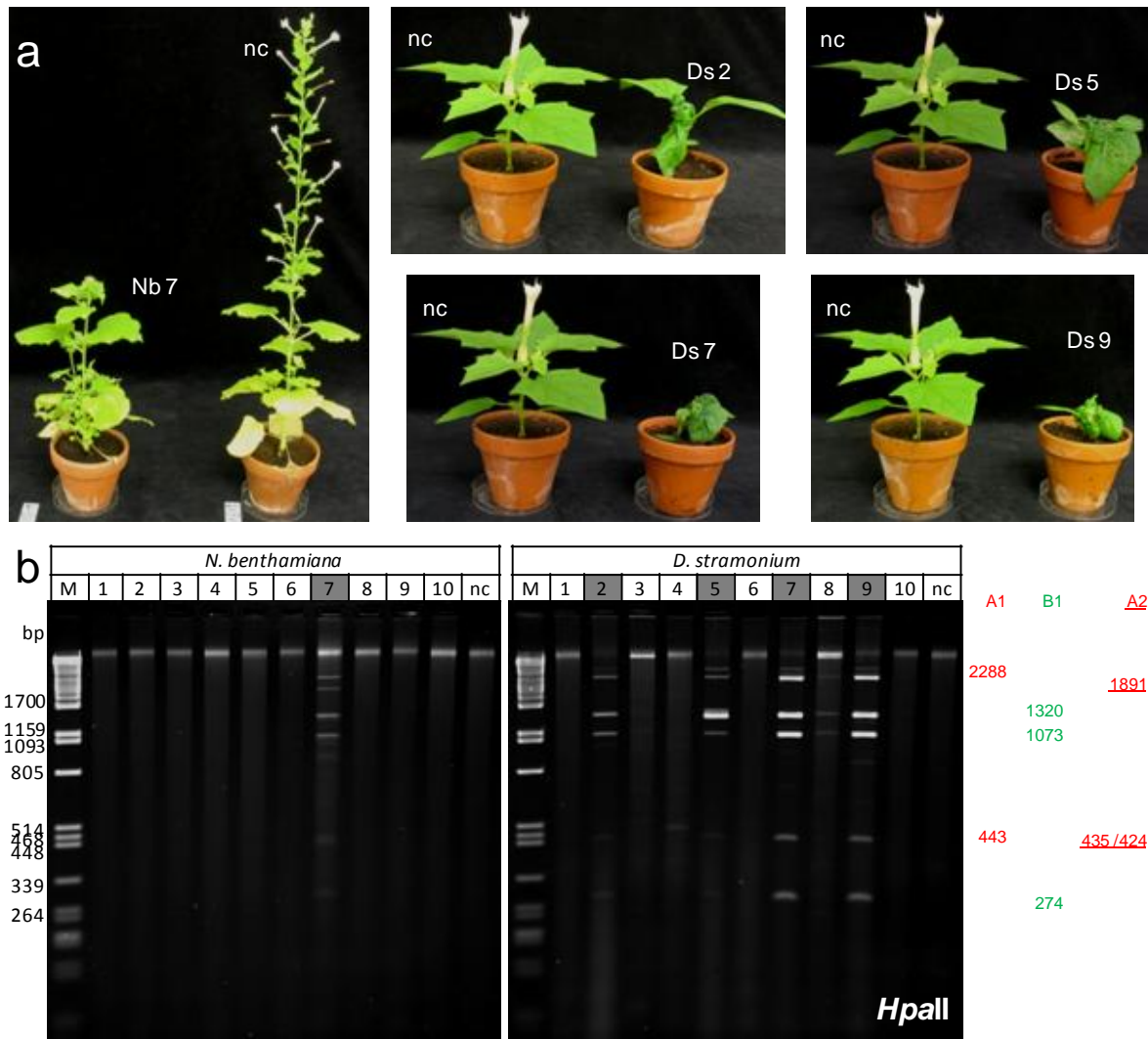

Figure 6. (a) Symptomatic *Nicotiana benthamiana* (Nb) and *Datura stramonium* (Ds) six and four weeks post biolistic inoculation (wpi) with RCA products of mixed symptom infected samples. nc, negative control. Numbers represent different plant samples. (b) Gel electrophoresis of RCA/RFLP products from bombarded plants, six wpi for *N. benthamiana* (left gel) and four wpi for *D. stramonium* plants using *Hpa*II. The expected fragment sizes of components DNA A1 B1 and A2 (underline fragments) are shown at the right side of each gel with DNA A1 fragments in red and DNA B1 fragments in green. M, molecular weight marker; nc, negative control. In gray, symptomatic plants.

This is the first report on segregation of symptoms within the same plant caused by infection by two distinct geminiviruses. Neither *N. benthamiana* nor *D. stramonium* showed symptoms related to the ones initially observed on *A. gangetica*. They did not present any kind of symptom segregation in different leaves or branches, suggesting that the symptom induction and segregation is determined by host factors.

Acknowledgments

References

Akah, P. A., Ezike, A. C., Nwafor, S. V., Okoli, C. O.and Enwerem, N. M. (2003). Evaluation of the anti-asthmatic property of *Asystasia gangetica* leaf extracts. *J. Ethnopharmacol.* **89,** 25-36.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W.and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25** (17)**,** 3389-3402.

Arguello-Astorga, G., Ascencio-Ibáñez, J. T., Dallas, M. B., Orozco, B. M.and Hanley-Bowdoin, L. (2007). High-frequency reversion of geminivirus replication protein mutants during infection. *J. Virol.* **81,** 11005-11015.

Argüello-Astorga, G. R., Guevara-González, R. G., Herrera-Estrella, L. R.and Rivera-Bustamante, R. F. (1994). Geminivirus replication origins have a group-specific organization of iterative elements: a model for replication. *Virology* **203** (1)**,** 90-100.

Argüello-Astorga, G. R.and Ruiz-Medrano, R. (2001). An iteron-related domain is associated to Motif 1 in the replication proteins of geminiviruses: identification of potential interacting amino acid-base pairs by a comparative approach. *Arch. Virol.* **146,** 1465-1485.

Bedford, I. D., Briddon, R. W., Brown, J. K., Rosell, R. C.and Markham, P. G. (1994). Geminivirus transmission and biological characterisation of *Bemisia tabaci* (Gennadius) biotypes from different geographic regions. *Ann. Appl. Biol.* **125** (2)**,** 311-325.

Blanco, L., Bernad, A., Lazaro, J. M., Martin, G., Garmendia, C.and Salas, M. (1989). Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. Symmetrical mode of DNA replication. *J. Biol. Chem.* **264** (15)**,** 8935-8940.

Briddon, R. W.and Markham, P. G. (1994). Universal primers for the PCR amplification of dicot-infecting geminiviruses. *Mol. Biotechnol.* **1** (2)**,** 202-205.

Briddon, R. W., Prescott, A. G., Lunness, P., Chamberlin, L. C.and Markham, P. G. (1993). Rapid production of full-length, infectious geminivirus clones by abutting primer PCR (AbP-PCR). *J. Virol. Methods* **43** (1)**,** 7-20.

Brown, J. K., Frohlich, D. R.and Rosell, R. C. (1995). The sweetpotato or silverleaf whiteflies: Biotypes of *Bemisia tabaci* or a species complex? *Annu. Rev. Entomol.* **40,** 511-534.

Chandra-Shekara, A. C., Gupte, M., Navarre, D., Raina, S., Raina, R., Klessig, D.and Kachroo, P. (2006). Light-dependent hypersensitive response and resistance signaling against Turnip crinkle Virus in Arabidopsis. *Plant J.* **45** (3)**,** 320-334.

Dean, F. B., Nelson, J. R., Giesler, T. L.and Lasken, R. S. (2001). Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res.* **11** (6)**,** 1095-1099.

Drummond, A. J., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Heled, J., Kearse, M., Moir, R., Stones-Havas, S., Sturrock, S., Thierer, T.and Wilson, A. Geneious 5.0. Auckland, New Zealand: Biomatters Ltd.

Fauquet, C. M., Mayo, M. A., Maniloff, J., Desselberger, U.and Ball, L. A. (2005). "Virus taxonomy, VIIIth report of the ICTV." Elsevier/Academic Press, London.

Genoud, T., Buchala, A. J., Chua, N. H.and Metraux, J. P. (2002). Phytochrome signalling modulates the SA-perceptive pathway in Arabidopsis. *Plant J.* **31** (1)**,** 87-95.

Griebel, T.and Zeier, J. (2008). Light regulation and daytime dependency of inducible plant defenses in Arabidopsis: phytochrome signaling controls systemic acquired resistance rather than local defense. *Plant Physiol.* **147** (2)**,** 790-801.

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

Horn, J., Lauster, S., Krenz, B., Kraus, J., Frischmuth, T.and Jeske, H. (2011). Ambivalent effects of defective DNA in beet curly top virus-infected transgenic sugarbeet plants. *Virus Res.* **158** (1-2)**,** 169-178.

Inoue-Nagata, A. K., Albuquerque, L. C., Rocha, W. B.and Nagata, T. (2004). A simple method for cloning the complete begomovirus genome using the bacteriophage phi 29 DNA polymerase. *J. Virol. Methods* **116,** 209-211.

ITIS Integrated taxonomic information system. *In* "Integrated taxonomic information system". [online]. cited 08.02.2011. Available from World Wide Web http://www.itis.gov

Jeske, H. (2009). Geminiviruses. *In* "TT viruses: the still elusive human pathogens". E.-M. de Villiers, and H. zur Hausen, Eds.). Springer, Berlin Heidelberg.

Jeske, H., Gotthardt, D.and Kober, S. (2010). In planta cloning of geminiviral DNA: The true Sida micrantha mosaic virus. *J. Virol. Methods* **163,** 301-308.

Jeske, H.and Werz, G. (1978). The influence of light intensity on pigment composition and ultrastructure of plastids in leaves of diseased Abutilon sellowianum Reg. (Malvaceae). *J. Phytopath.* **91,** 1-13.

Kadri, A., Maiss, E., Amsharov, N., Bittner, A. M., Balci, S., Kern, K., Jeske, H.and Wege, C. (2011). Engineered *Tobacco mosaic virus* mutants with distinct physical characteristics *in planta* and enhanced metallization properties. *Virus Res.* **157** (1)**,** 35-46.

Kleinow, T., Nischang, M., Beck, A., Kratzer, U., Tanwir, F., Preiss, W., Kepp, G.and Jeske, H. (2009). Three C-terminal phosphorylation sites in the Abutilon mosaic virus movement protein affect symptom development and viral DNA accumulation. *Virology* **390** (1)**,** 89-101.

Kumar, Y., Hallan, V.and Zaidi, A. A. (2008). Molecular characterization of a distinct bipartite begomovirus species infecting tomato in India. *Virus Genes* **37** (3)**,** 425-431.

Liu, Y., Robinson, D. J.and Harrison, B. D. (1998). Defective forms of cotton leaf curl virus DNA-A that have different combinations of sequence deletion, duplication, inversion and rearrangement. *J. Gen. Virol.* **79 (Pt 6),** 1501-1508.

Lizardi, P. M., Huang, X., Zhu, Z., Bray-Ward, P., Thomas, D. C.and Ward, D. C. (1998). Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nat. Genet.* **19** (3)**,** 225-232.

Maruthi, M. N., Colvin, J., Seal, S., Gibson, G.and Cooper, J. (2002). Co-adaptation between cassava mosaic geminiviruses and their local vector populations. *Virus Res.* **86** (1-2)*,* 71-85.

Morilla, G., Krenz, B., Jeske, H., Bejarano, E. R.and Wege, C. (2004). Tete a tete of tomato yellow leaf curl virus and tomato yellow leaf curl sardinia virus in single nuclei. *J. Virol.* **78** (19)*,* 10715-10723.

Paprotka, T., Boiteux, L. S., Fonseca, M. E. N., Resende, R. O., Jeske, H., Faria, J. C.and Ribeiro, S. G. (2010a). Genomic diversity of sweet potato geminiviruses in a Brazilian germplasm bank. *Virus Res.* **149** (2)*,* 224-233.

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

Patil, B. L., Dutt, N., Briddon, R. W., Bull, S. E., Rothenstein, D., Borah, B. K., Dasgupta, I., Stanley, J.and Jeske, H. (2007). Deletion and recombination events between the DNA-A and DNA-B components of Indian cassava-infecting geminiviruses generate defective molecules in Nicotiana benthamiana. *Virus Res.* **124** (1-2)*,* 59-67.

Paximadis, M.and Rey, M. E. (2001). Genome organization of Tobacco leaf curl Zimbabwe virus, a new, distinct monopartite begomovirus associated with subgenomic defective DNA molecules. *J. Gen. Virol.* **82** (Pt 12)*,* 3091-3097.

PROTA (2010). Plant resources of tropical Africa. *In* "Plant resources of tropical Africa". [online]. cited 18/04/2011. Available from World Wide Web www.prota.org

Renteria-Canett, I., Xoconostle-Cazares, B., Ruiz-Medrano, R.and Rivera-Bustamante, R. F. (2011). Geminivirus mixed infection on pepper plants: Synergistic interaction between PHYVV and PepGMV. *Virol. J.* **8,** 104-117.

Roden, L. C.and Ingle, R. A. (2009). Lights, rhythms, infection: the role of light and the circadian clock in determining the outcome of plant-pathogen interactions. *Plant Cell* **21** (9)*,* 2546-2552.

Sambrook, J.and Russell, D. W. (2001). "Molecular cloning: a laboratory manual." Cold Spring Harbor Press, Cold Spring Harbor, N Y.

Sanchez-Campos, S., Navas-Castillo, J., Camero, R., Soria, C., Diaz, J. A.and Moriones, E. (1999). Displacement of Tomato yellow leaf curl virus (TYLCV)-Sr by TYLCV-Is in tomato epidemics in Spain. *Phytopathol.* **89** (11)*,* 1038-1043.

Saunders, K.and Stanley, J. (1995). Complementation of African cassava mosaic virus AC2 gene function in a mixed bipartite geminivirus infection. *J. Gen. Virol.* **76 (Pt 9),** 2287-2292.

Sikora, E. J., Gudauskas, R. T., Murphy, J. F., Porch, D. W., Andrianifahanana, M., Zehnder, G. W., Bauske, E. M., Kemble, J. M.and Lester, D. (1998). A multivirus epidemic of tomatoes in Alabama. *Plant Dis.* **82,** 117-120.

Stanley, J., Frischmuth, T.and Ellwood, S. (1990). Defective viral DNA ameliorates symptoms of geminivirus infection in transgenic plants. *Proc. Natl. Acad. Sci. U S A* **87** (16)*,* 6291-6295.

Stanley, J., Saunders, K., Pinner, M. S.and Wong, S. M. (1997). Novel defective interfering DNAs associated with ageratum yellow vein geminivirus infection of Ageratum conyzoides. *Virology* **239** (1)*,* 87-96.

Stocsits, R. R., Hofacker, I. L., Fried, C.and Stadler, P. F. (2005). Multiple sequence alignments of partially coding nucleic acid sequences. *BMC Bioinformatics* **6,** 160.

Sudhakar, M., Rao, C. V., Rao, P. M., Raju, D. B.and Venkateswarlu, Y. (2006). Antimicrobial activity of *Caesalpinia pulcherrima*, *Euphorbia hirta* and *Asystasia gangeticum*. *Fitoterapia* **77,** 378-380.

Tamura, K., Dudley, J., Nei, M.and Kumar, S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24** (8)**,** 1596-1599.

Unseld, S., Hohnle, M., Ringel, M.and Frischmuth, T. (2001). Subcellular targeting of the coat protein of African cassava mosaic geminivirus. *Virology* **286** (2)**,** 373-383.

Wege, C., Gotthardt, R. D., Frischmuth, T.and Jeske, H. (2000). Fulfilling Koch's postulates for Abutilon mosaic virus. *Arch. Virol.* **145** (10)**,** 2217-2225.

Wyant, P. S., Gotthardt, D., Schafer, B., Krenz, B.and Jeske, H. (2011). The genomes of four novel begomoviruses and a new Sida micrantha mosaic virus strain from Bolivian weeds. *Arch. Virol.* **156** (2)**,** 347-352.

Yeoh, H.-H.and Wong, P.-F. M. (1993). Food value of lesser utilised tropical plants. *Food Chem.* **46,** 239-241.

Zhang, D. Y., Brandwein, M., Hsuih, T. C.and Li, H. (1998). Amplification of target-specific, ligation-dependent circular probe. *Gene* **211** (2)**,** 277-285.

Zhang, D. Y., Zhang, W., Li, X.and Konomi, Y. (2001). Detection of rare DNA targets by isothermal ramification amplification. *Gene* **274** (1-2)**,** 209-216.

References

Akah, P. A., Ezike, A. C., Nwafor, S. V., Okoli, C. O.and Enwerem, N. M. (2003). Evaluation of the anti-asthmatic property of *Asystasia gangetica* leaf extracts. *J. Ethnopharmacol.* **89,** 25-36.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W.and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25** (17)**,** 3389-3402.

Anonymus (2004). Systematics and molecular phylogenetics. *In* "Just the facts: A basic introduction to the science underlying NCBI resources". NCBI [online]. cited 12-04-2011. Available from World Wide Web http://www.ncbi.nlm.nih.gov/About/primer/phylo.html

Archer, J. P. (2008). Phylogenetic trees. *In* "The diversity of HIV-1". PhD Thesis (pp. 75-80. The University of Manchester, Manchester.

Arguello-Astorga, G., Ascencio-Ibáñez, J. T., Dallas, M. B., Orozco, B. M.and Hanley-Bowdoin, L. (2007). High-frequency reversion of geminivirus replication protein mutants during infection. *J. Virol.* **81,** 11005-11015.

Argüello-Astorga, G. R., Guevara-González, R. G., Herrera-Estrella, L. R.and Rivera-Bustamante, R. F. (1994). Geminivirus replication origins have a group-specific organization of iterative elements: a model for replication. *Virology* **203** (1)**,** 90-100.

Argüello-Astorga, G. R.and Ruiz-Medrano, R. (2001). An iteron-related domain is associated to Motif 1 in the replication proteins of geminiviruses: identification of potential interacting amino acid-base pairs by a comparative approach. *Arch. Virol.* **146,** 1465-1485.

Assunção, I. P., Listik, A. F., Barros, M. C. S., Amorim, E. P. R., Silva, S. J. C., Izael, O. S., Ramalho-Neto, C. E.and Lima, G. S. A. (2006). Genetic diversity of begomovirus infecting weeds in Northeastern Brazil. *Planta Daninha* **24** (2)**,** 239-244.

Baldauf, S. L. (2003). Phylogeny for the faint of heart: a tutorial. *Trends Genet.* **19,** 345-351.

Bedford, I. D., Briddon, R. W., Brown, J. K., Rosell, R. C.and Markham, P. G. (1994). Geminivirus transmission and biological characterisation of *Bemisia tabaci* (Gennadius) biotypes from different geographic regions. *Ann. Appl. Biol.* **125** (2)**,** 311-325.

Biebricher, C. K.and Eigen, M. (2006). What is a quasispecies? *Curr. Top. Microbiol. Immunol.* **299,** 1-31.

Bisaro, D. M. (2006). Silencing suppression by geminivirus proteins. *Virology* **344** (1)**,** 158-168.

Bisaro, D. M., Hamilton, W. D., Coutts, R. H.and Buck, K. W. (1982). Molecular cloning and characterisation of the two DNA components of tomato golden mosaic virus. *Nucleic Acids Res.* **10** (16)**,** 4913-4922.

Blanco, L., Bernad, A., Lazaro, J. M., Martin, G., Garmendia, C.and Salas, M. (1989). Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. Symmetrical mode of DNA replication. *J. Biol. Chem.* **264** (15)**,** 8935-8940.

Blanco, L.and Salas, M. (1996). Relating structure to function in phi29 DNA polymerase. *J. Biol. Chem.* **271,** 8509-8512.

Böttcher, B., Unseld, S., Ceulemans, H., Russell, R. B.and Jeske, H. (2004). Geminate structures of African cassava mosaic virus. *J. Virol.* **78,** 6758-6765.

Briddon, R. W.and Markham, P. G. (1994). Universal primers for the PCR amplification of dicot-infecting geminiviruses. *Mol. Biotechnol.* **1** (2)**,** 202-205.

Briddon, R. W.and Markham, P. G. (1995). Use of PCR in the detection and characterization of geminiviruses. *EPPO* **25** (1-2)**,** 315-320.

Briddon, R. W., Pinner, M. S., Stanley, J.and Markham, P. G. (1990). Geminivirus coat protein gene replacement alters insect specificity. *Virology* **177,** 85-94.

Briddon, R. W., Prescott, A. G., Lunness, P., Chamberlin, L. C.and Markham, P. G. (1993). Rapid production of full-length, infectious geminivirus clones by abutting primer PCR (AbP-PCR). *J. Virol. Methods* **43** (1)**,** 7-20.

Brown, J. K., Frohlich, D. R.and Rosell, R. C. (1995). The sweetpotato or silverleaf whiteflies: Biotypes of *Bemisia tabaci* or a species complex? *Annu. Rev. Entomol.* **40,** 511-534.

Castillo-Urquiza, G. P., Junior, J. E. A. B., Bruckner, F. P., Lima, A. T. M., Varsani, A., Alfenas-Zerbini, P.and Zerbini, F. M. (2008). Six novel begomoviruses infecting tomato and associated weeds in Southeastern Brazil. *Arch. Virol.* **153,** 1985-1989.

Chandra-Shekara, A. C., Gupte, M., Navarre, D., Raina, S., Raina, R., Klessig, D.and Kachroo, P. (2006). Light-dependent hypersensitive response and resistance signaling against Turnip crinkle Virus in Arabidopsis. *Plant J.* **45** (3)**,** 320-334.

Choudhury, N. R., Malik, P. S., Singh, D. K., Islam, M. N., Kaliappan, K.and Kosalai, S. K. (2006). The oligomeric Rep protein of Mungbean yellow mosaic India virus (MYMIV) is a likely replicative helicase. *Nucleic Acids Res.* **34,** 6362-6377.

Clerot, D.and Bernardi, F. (2006). DNA helicase activity is associated with the replication initiator protein rep of tomato yellow leaf curl geminivirus. *J. Virol.* **80,** 11322-11330.

Costa, A. S., Oliveira, A. R.and Silva, D. M. (1975). Transmissão mecânica do mosaico dourado do tomateiro. *Fitopatol. Bras.* **6,** 147.

Dean, F. B., Nelson, J. R., Giesler, T. L.and Lasken, R. S. (2001). Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res.* **11** (6)**,** 1095-1099.

Drummond, A. J., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Heled, J., Kearse, M., Moir, R., Stones-Havas, S., Sturrock, S., Thierer, T.and Wilson, A. Geneious 5.0. Auckland, New Zealand: Biomatters Ltd.

Eckert, K. A.and Kunkel, T. A. (1990). High fidelity DNA synthesis by the *Therrmus aquaticus* DNA polymerase. *Nucleic Acids Res.* **18,** 3739-3744.

Egelkrout, E. M., Robertson, D.and Hanley-Bowdoin, L. (2001). Proliferating cell nuclear antigen transcription is repressed through an E2F consensus element and activated by geminivirus infection in mature leaves. *Plant Cell* **13** (6)**,** 1437-1452.

Eigen, M. (1993). Viral quasispecies. *Sci. Am.* **269** (1)**,** 42-49.

Eigen, M. (1996). On the nature of virus quasispecies. *Trends Microbiol.* **4** (6)**,** 216-218.

Faria, J. C.and Maxwell, D. P. (1999). Variability in geminivirus isolates associated with *Phaseolus* spp. in Brazil. *Phytopathol.* **89,** 262-268.

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

Fauquet, C. M., Mayo, M. A., Maniloff, J., Desselberger, U.and Ball, L. A. (2005). "Virus taxonomy, VIIIth report of the ICTV." Elsevier/Academic Press, London.

Fauquet, C. M.and Stanley, J. (2005). Revising the way we conceive and name viruses below the species level: a review of geminivirus taxonomy calls for new standardized isolate descriptors. *Arch. Virol.* **150** (10), 2151-2179.

Felsenstein, J. (2004). "Inferring phylogenies." Sinauer Associates, Inc., Sunderland, Massachusetts.

Fontes, E. P., Gladfelter, H. J., Schaffer, R. L., Petty, I. T.and Hanley-Bowdoin, L. (1994). Geminivirus replication origins have a modular organization. *Plant Cell* **6** (3), 405-416.

Frischmuth, T., Engel, M., Lauster, S.and Jeske, H. (1997). Nucleotide sequence evidence for the occurrence of three distinct whitefly-transmitted, Sida-infecting bipartite geminiviruses in Central America. *J. Gen. Virol.* **78,** 2675-2682.

Fujii, R., Kitaoka, M.and Hayashi, K. (2006). Error-prone rolling circle amplification: the simplest random mutagenesis protocol. *Nature Protocols* **1,** 2493 - 2497.

Gardiner, W. E., Sunter, G., Brand, L., Elmer, J. S., Rogers, S. G.and Bisaro, D. M. (1988). Genetic analysis of tomato golden mosaic virus: the coat protein is not required for systemic spread or symptom development. *EMBO J.* **7** (4), 899-904.

Garmendia, C., Bernad, A., Estebane, J. A., Blanco, L.and Salad, M. (1992). The bacteriophage phi29 DNA polymerase, a proofreading enzyme. *J. Biol. Chem.* **267,** 2594-2599.

Genoud, T., Buchala, A. J., Chua, N. H.and Metraux, J. P. (2002). Phytochrome signalling modulates the SA-perceptive pathway in Arabidopsis. *Plant J.* **31** (1), 87-95.

Givord, L., Fargette, D., Kounounguissa, B., Thouvenel, J. C., Walter, B.and Van Regenmortel, M. (1994). Detection of geminiviruses from tropical countries by a double monoclonal antibody ELISA using antibodies to African cassava mosaic virus. *Agronomie* **14,** 327-333.

Grafi, G., Burnett, R. J., Helentjaris, T., Larkins, B. A., Decaprio, J. A., Sellers, W. R.and Kaelin, W. G. J. (1996). A maize cDNA encoding a member of the retinoblastoma protein family: Involvement in endoreduplication. *Proc. Natl. Acad. Sci.* **93,** 8962-8967.

Griebel, T.and Zeier, J. (2008). Light regulation and daytime dependency of inducible plant defenses in Arabidopsis: phytochrome signaling controls systemic acquired resistance rather than local defense. *Plant Physiol.* **147** (2), 790-801.

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

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41,** 95-98.

Hamilton, W. D., Bisaro, D. M., Coutts, R. H.and Buck, K. W. (1983). Demonstration of the bipartite nature of the genome of a single-stranded DNA plant virus by infection with the cloned DNA components. *Nucleic Acids Res.* **11** (21), 7387-7396.

Hamilton, W. D., Stein, V. E., Coutts, R. H.and Buck, K. W. (1984). Complete nucleotide sequence of the infectious cloned DNA components of tomato golden mosaic virus: potential coding regions and regulatory sequences. *EMBO J.* **3** (9), 2197-2205.

Hanley-Bowdoin, L., Elmer, J. S.and Rogers, S. G. (1988). Transient expression of heterologous RNAs using tomato golden mosaic virus. *Nucleic Acids Res.* **16** (22)**,** 10511-10528.

Hanley-Bowdoin, L., Elmer, J. S.and Rogers, S. G. (1989). Functional expression of the leftward open reading frames of the A component of tomato golden mosaic virus in transgenic tobacco plants. *Plant Cell* **1** (11)**,** 1057-1067.

Harrison, B. D., Swanson, M. M.and Fargette, D. (2002). Begomovirus coat protein: serology, variation and functions. *Physiol. Mol. Plant Pathol.* **60,** 257-271.

Hehnle, S., Wege, C.and Jeske, H. (2004). Interaction of DNA with the movement proteins of geminiviruses revisited. *J. Virol.* **78** (14)**,** 7698-7706.

Hellens, R. P., Edwards, E. A., Leyland, N. R., Bean, S.and Mullineaux, P. M. (2000). pGreen: a versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation. *Plant Mol. Biol.* **42** (6)**,** 819-832.

Hoekema, A., Hirsch, P. R., Hooykaas, P. J. J.and Schilperoort, R. A. (1983). A binary plant vector strategy based on separation of vir- and T-region of the Agrobacterium tumefaciens Ti-plasmid **303** (5913)**,** 179-180.

Homs, M., Kober, S., Kepp, G.and Jeske, H. (2008). Mitochondrial plasmids of sugar beet amplified via rolling circle method detected during curtovirus screening. *Virus Res.* **136** (1-2)**,** 124-129.

Hormuzdi, S. G.and Bisaro, D. M. (1995). Genetic analysis of beet curly top virus: examination of the roles of L2 and L3 genes in viral pathogenesis. *Virology* **206** (2)**,** 1044-1054.

Horn, J., Lauster, S., Krenz, B., Kraus, J., Frischmuth, T.and Jeske, H. (2011). Ambivalent effects of defective DNA in beet curly top virus-infected transgenic sugarbeet plants. *Virus Res.* **158** (1-2)**,** 169-178.

Inoue-Nagata, A. K., Albuquerque, L. C., Rocha, W. B.and Nagata, T. (2004). A simple method for cloning the complete begomovirus genome using the bacteriophage phi 29 DNA polymerase. *J. Virol. Methods* **116,** 209-211.

ITIS Integrated taxonomic information system. *In* "Integrated taxonomic information system". [online]. cited 08.02.2011. Available from World Wide Web http://www.itis.gov

Jeske, H. (2009). Geminiviruses. *In* "TT viruses: the still elusive human pathogens". E.-M. de Villiers, and H. zur Hausen, Eds.). Springer, Berlin Heidelberg.

Jeske, H., Gotthardt, D.and Kober, S. (2010). In planta cloning of geminiviral DNA: The true Sida micrantha mosaic virus. *J. Virol. Methods* **163,** 301-308.

Jeske, H., Lütgemeier, M.and Preiss, W. (2001). Distinct DNA forms indicate rolling circle and recombination-dependent replication of Abutilon mosaic geminivirus. *EMBO J.* **20,** 6158-6167.

Jeske, H.and Werz, G. (1978). The influence of light intensity on pigment composition and ultrastructure of plastids in leaves of diseased Abutilon sellowianum Reg. (Malvaceae). *J. Phytopath.* **91,** 1-13.

Jovel, J., Preiß, W.and Jeske, H. (2007). Characterization of DNA-intermediates of an arising geminivirus. *Virus Res.* **130,** 63-70.

Jovel, J., Reski, G., Rothenstein, D., Ringel, M., Frischmuth, T.and Jeske, H. (2004). Sida micrantha mosaic is associated with a complex infection of begomoviruses different from Abutilon mosaic virus. *Arch. Virol.* **149** (4)**,** 829-841.

Kadri, A., Maiss, E., Amsharov, N., Bittner, A. M., Balci, S., Kern, K., Jeske, H.and Wege, C. (2011). Engineered *Tobacco mosaic virus* mutants with distinct

physical characteristics *in planta* and enhanced metallization properties. *Virus Res.* **157** (1)**,** 35-46.

Kleinow, T., Nischang, M., Beck, A., Kratzer, U., Tanwir, F., Preiss, W., Kepp, G.and Jeske, H. (2009). Three C-terminal phosphorylation sites in the Abutilon mosaic virus movement protein affect symptom development and viral DNA accumulation. *Virology* **390** (1)**,** 89-101.

Klinkenberg, F. A., Ellwood, S.and Stanley, J. (1989). Fate of African cassava mosaic virus coat protein deletion mutants after agroinoculation. *J. Gen. Virol.* **70,** 1837-1844.

Kumar, Y., Hallan, V.and Zaidi, A. A. (2008). Molecular characterization of a distinct bipartite begomovirus species infecting tomato in India. *Virus Genes* **37** (3)**,** 425-431.

Laufs, J., Traut, W., Heyraud, F., Matzeit, V., Rogers, S. G., Schell, J.and Gronenborn, B. (1995). In vitro cleavage and joining at the viral origin of replication by the replication initiator protein of tomato yellow leaf curl virus. *Proc. Natl. Acad. Sci.* **92,** 3879-3883.

Li, R., Salih, S.and Hurtt, S. (2004). Detection of geminiviruses in sweetpotato by polymerase chain reaction. *Plant Dis.* **88,** 1347-1351.

Liu, Y., Robinson, D. J.and Harrison, B. D. (1998). Defective forms of cotton leaf curl virus DNA-A that have different combinations of sequence deletion, duplication, inversion and rearrangement. *J. Gen. Virol.* **79 (Pt 6),** 1501-1508.

Lizardi, P. M., Huang, X., Zhu, Z., Bray-Ward, P., Thomas, D. C.and Ward, D. C. (1998). Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nat. Genet.* **19** (3)**,** 225-232.

Mansoor, S., Briddon, R. W., Zafar, Y.and Stanley, J. (2003). Geminivirus disease complexes: an emerging threat. *Trends Plant. Sci.* **8** (3)**,** 128-134.

Maruthi, M. N., Colvin, J., Seal, S., Gibson, G.and Cooper, J. (2002). Co-adaptation between cassava mosaic geminiviruses and their local vector populations. *Virus Res.* **86** (1-2)**,** 71-85.

Marziali, A.and Akeson, M. (2001). New DNA sequencing methods. *Annu. Rev. Biomed. Eng.* **3,** 195-223.

Matyis, J. C., Silva, D. M., Oliveira, A. R.and Costa, A. S. (1975). Purificacao e morfologia do virus do mosaico dourado do tomateiro. *Summa Phytopathol.* **1,** 267–273.

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

Morales, F. J.and Anderson, P. K. (2001). The emergence and dissemination of whitefly-transmitted geminiviruses in Latin America. *Arch. Virol.* **146** (3)**,** 415-441.

Morilla, G., Krenz, B., Jeske, H., Bejarano, E. R.and Wege, C. (2004). Tete a tete of tomato yellow leaf curl virus and tomato yellow leaf curl sardinia virus in single nuclei. *J. Virol.* **78** (19)**,** 10715-10723.

Nawaz-ul-Rehman, M. S.and Fauquet, C. M. (2009). Evolution of geminiviruses and their satellites. *FEBS Letters* **583,** 1825-1832.

Orozco, B. M.and Hanley-Bowdoin, L. (1996). A DNA structure is required for geminivirus replication origin function. *J. Virol.* **70** (1)**,** 148-158.

Paprotka, T., Boiteux, L. S., Fonseca, M. E. N., Resende, R. O., Jeske, H., Faria, J. C.and Ribeiro, S. G. (2010a). Genomic diversity of sweet potato geminiviruses in a Brazilian germplasm bank. *Virus Res.* **149** (2)**,** 224-233.

Paprotka, T., Metzler, V.and Jeske, H. (2010b). The complete nucleotide sequence of a new bipartite begomovirus from Brazil infecting Abutilon. *Arch Virol* **155,** 813-816.

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

Patil, B. L., Dutt, N., Briddon, R. W., Bull, S. E., Rothenstein, D., Borah, B. K., Dasgupta, I., Stanley, J.and Jeske, H. (2007). Deletion and recombination events between the DNA-A and DNA-B components of Indian cassava-infecting geminiviruses generate defective molecules in Nicotiana benthamiana. *Virus Res.* **124** (1-2)**,** 59-67.

Paximadis, M.and Rey, M. E. (2001). Genome organization of Tobacco leaf curl Zimbabwe virus, a new, distinct monopartite begomovirus associated with subgenomic defective DNA molecules. *J. Gen. Virol.* **82** (Pt 12)**,** 3091-3097.

Perveen, R., Fani, I., Rasheed, I., Chohan, S., Rehman, A.and Haider, S. (2010). Identification of Cotton leaf curl begomovirus in Pakistan in different symptomatic and asymptomatic plants though enzyme-linked immunosorbent assay (ELISA). *Eur. J. Soc. Sci.* **14,** 502-206.

Pooma, W.and Petty, I. T. (1996). Tomato golden mosaic virus open reading frame AL4 is genetically distinct from its C4 analogue in monopartite geminiviruses. *J. Gen. Virol.* **77,** 1947-1951.

Preiss, W.and Jeske, H. (2003). Multitasking in replication is common among geminiviruses. *J. Virol.* **77,** 2972-2980.

PROTA (2010). Plant resources of tropical Africa. *In* "Plant resources of tropical Africa". [online]. cited 18/04/2011. Available from World Wide Web www.prota.org

Qin, S., Ward, B. M.and Lazarowitz, S. G. (1998). The bipartite geminivirus coat protein aids BR1 function in viral movement by affecting the accumulation of viral single-stranded DNA. *J. Virol.* **72** (11)**,** 9247-9256.

Renteria-Canett, I., Xoconostle-Cazares, B., Ruiz-Medrano, R.and Rivera-Bustamante, R. F. (2011). Geminivirus mixed infection on pepper plants: Synergistic interaction between PHYVV and PepGMV. *Virol. J.* **8,** 104-117.

Rigden, J. E., Krake, L. R., Rezaian, M. A.and Dry, I. B. (1994). ORF C4 of tomato leaf curl geminivirus is a determinant of symptom severity. *Virology* **204** (2)**,** 847-850.

Roden, L. C.and Ingle, R. A. (2009). Lights, rhythms, infection: the role of light and the circadian clock in determining the outcome of plant-pathogen interactions. *Plant Cell* **21** (9)**,** 2546-2552.

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

Ronaghi, M. (2001). Pyrosequencing sheds light on DNA sequencing. *Genome Res.* **11,** 3-11.

Ronaghi, M., Uhlén, M.and Nyrén, P. (1998). A Sequencing Method Based on Real-Time Pyrophosphate. *Science* **17,** 363-365.

Rothberg, J. M.and Leamon, J. H. (2008). The development and impact of 454 sequencing. *Nat. Biotechnol.* **26** (10)**,** 1117-1124.

Rothenstein, D., Krenz, B., Selchow, O.and Jeske, H. (2007). Tissue and cell tropism of Indian cassava mosaic virus (ICMV) and its AV2 (precoat) gene product. *Virology* **359,** 137-145.

Rouhibakhsh, A., Priya, J., Periasamy, M., Haq, Q. M. I.and Malathi, V. G. (2008). An improved DNA isolation method and PCR protocol for efficient detection of multicomponents of begomovirus in legumes. *J. Virol. Methods* **147** (1)**,** 37-42.

Saitou, N.and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4,** 406-425.

Sambrook, J.and Russell, D. W. (2001). "Molecular cloning: a laboratory manual." Cold Spring Harbor Press, Cold Spring Harbor, N Y.

Sanchez-Campos, S., Navas-Castillo, J., Camero, R., Soria, C., Diaz, J. A.and Moriones, E. (1999). Displacement of Tomato yellow leaf curl virus (TYLCV)-Sr by TYLCV-Is in tomato epidemics in Spain. *Phytopathol.* **89** (11)**,** 1038-1043.

Saunders, K.and Stanley, J. (1995). Complementation of African cassava mosaic virus AC2 gene function in a mixed bipartite geminivirus infection. *J. Gen. Virol.* **76 (Pt 9),** 2287-2292.

Saunders, K., Wege, C., Veluthambi, K., Jeske, H.and Stanley, J. (2001). The distinct disease phenotypes of the common and yellow vein strains of Tomato golden mosaic virus are determined by nucleotide differences in the 3'-terminal region of the gene encoding the movement protein. *J. Gen. Virol.* **82** (Pt 1)**,** 45-51.

Settlage, S. B., See, R. G.and Hanley-Bowdoin, L. (2005). Geminivirus C3 protein: replication enhancement and protein interactions. *J. Virol.* **79,** 9885-9895.

Sikora, E. J., Gudauskas, R. T., Murphy, J. F., Porch, D. W., Andrianifahanana, M., Zehnder, G. W., Bauske, E. M., Kemble, J. M.and Lester, D. (1998). A multivirus epidemic of tomatoes in Alabama. *Plant Dis.* **82,** 117-120.

Stanley, J., Bisaro, D. M., Briddon, R. W., Brown, J. K., Fauquet, C. M., Harrison, B. D., Rybicki, E. P.and Stenger, D. C. (2005). Geminiviridae. *In* "Virus taxonomy. VIIIth report of the international committee on taxonomy of viruses". C. M. Fauquet, M. A. Mayo, J. Maniloff, U. Desselberger, and L. A. Ball, Eds.), pp. 301-326. Elsevier, Academic Press, London.

Stanley, J., Frischmuth, T.and Ellwood, S. (1990). Defective viral DNA ameliorates symptoms of geminivirus infection in transgenic plants. *Proc. Natl. Acad. Sci. U S A* **87** (16)**,** 6291-6295.

Stanley, J., Saunders, K., Pinner, M. S.and Wong, S. M. (1997). Novel defective interfering DNAs associated with ageratum yellow vein geminivirus infection of Ageratum conyzoides. *Virology* **239** (1)**,** 87-96.

Stocsits, R. R., Hofacker, I. L., Fried, C.and Stadler, P. F. (2005). Multiple sequence alignments of partially coding nucleic acid sequences. *BMC Bioinformatics* **6,** 160.

Sudhakar, M., Rao, C. V., Rao, P. M., Raju, D. B.and Venkateswarlu, Y. (2006). Antimicrobial activity of *Caesalpinia pulcherrima*, *Euphorbia hirta* and *Asystasia gangeticum. Fitoterapia* **77,** 378-380.

Sunter, G.and Bisaro, D. M. (1992). Transactivation of geminivirus AR1 and BR1 gene expression by the viral AL2 gene product occurs at the level of transcription. *Plant Cell* **4** (10)**,** 1321-1331.

Tamura, K., Dudley, J., Nei, M.and Kumar, S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24** (8)**,** 1596-1599.

Tamura, K., Nei, M.and Kumar, S. (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Natl. Acad. Sci.* **101,** 11030-11035.

Thommes, P., Osman, T. A., Hayes, R. J.and Buck, K. W. (1993). TGMV replication protein AL1 preferentially binds to single-stranded DNA from the common region. *FEBS Lett.* **319** (1-2)**,** 95-99.

Thompson, J. D., Higgins, D. G.and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22** (22)**,** 4673-4680.

Trinks, D., Rajeswaran, R., Shivaprasad, P. V., Akbergenov, R., Oakeley, E. J., Veluthambi, K., Hohn, T.and Pooggin, M. M. (2005). Suppression of RNA silencing by a geminivirus nuclear protein, AC2, correlates with transactivation of host genes. *J. Virol.* **79** (4)**,** 2517-2527.

Unseld, S., Hohnle, M., Ringel, M.and Frischmuth, T. (2001). Subcellular targeting of the coat protein of African cassava mosaic geminivirus. *Virology* **286** (2)**,** 373-383.

Vanitharani, R., Chellappan, P., Pita, J. S.and Fauquet, C. M. (2004). Differential roles of AC2 and AC4 of cassava geminiviruses in mediating synergism and suppression of posttranscriptional gene silencing. *J. Virol.* **78,** 9487-9498.

von Arnim, A.and Stanley, J. (1992). Determinants of tomato golden mosaic virus symptom development located on DNA B. *Virology* **186** (1)**,** 286-293.

Wege, C., Gotthardt, R. D., Frischmuth, T.and Jeske, H. (2000). Fulfilling Koch's postulates for Abutilon mosaic virus. *Arch. Virol.* **145** (10)**,** 2217-2225.

Wyant, P. S., Gotthardt, D., Schafer, B., Krenz, B.and Jeske, H. (2011). The genomes of four novel begomoviruses and a new Sida micrantha mosaic virus strain from Bolivian weeds. *Arch. Virol.* **156** (2)**,** 347-352.

Xie, Q., Sanz-Burgos, A. P., Hannon, G. J.and Gutierrez, C. (1996). Plant cells contain a novel member of the retinoblastoma family of growth regulatory proteins. *EMBO J.* **15,** 4900-4908.

Xie, Q., Suarez-Lopez, P.and Gutiérrez, C. (1995). Identification and analysis of a retinoblastoma binding motif in the replication protein of a plant DNA virus: requirement for efficient viral DNA replication. *EMBO J.* **14,** 4073-4082.

Yeoh, H.-H.and Wong, P.-F. M. (1993). Food value of lesser utilised tropical plants. *Food Chem.* **46,** 239-241.

Zhang, D. Y., Brandwein, M., Hsuih, T. C.and Li, H. (1998). Amplification of target-specific, ligation-dependent circular probe. *Gene* **211** (2)**,** 277-285.

Zhang, D. Y., Zhang, W., Li, X.and Konomi, Y. (2001). Detection of rare DNA targets by isothermal ramification amplification. *Gene* **274** (1-2)**,** 209-216.

Zhang, Z., Schwartz, S., Lukas, W.and Web, M. (2000). A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.* **7** (1-2)**,** 203-214.

Zhou, Y. C., Garrido-Ramirez, E. R., Sudarshana, M. R., Yendluri, S.and Gilbertson, R. L. (2007). The N-terminus of the Begomovirus nuclear shuttle protein (BV1) determines virulence or avirulence in Phaseolus vulgaris. *Mol. Plant Microbe Interact.* **20** (12)**,** 1523-1534.

Acknowledgments

I would like to express my most sincere thanks to:

Prof. Holger Jeske, for the opportunity of working in his group in this project, as well as for all the care and support, being always available for elucidation of questions and giving precious suggestions on the work.

Dr. Björn Krenz for all the assistance and for being always ready to help with any problems or doubts that arose during the promotion period.

Dr. Anan Kadri and PD Dr. Tatjana Kleinow for critically reading my thesis, giving me precious advices and suggestions.

PD Dr. Christina Wege for all the help, support and suggestions with my work.

Benjamin Schäfer, Annika Allinger, Judith Horn and Björn Krenz for the encouragement and friendship at the laboratory "upstairs". Thank you for all the fun and the great atmosphere.

Mrs. Pfafferott and Mrs. Entmooser, for the help with bureaucratic issues.

Annika Allinger and Diether Gotthardt for providing and taking care of the plants used for the projects during the promotion study.

The staff of the Institute of Biology, for the help during the project.

My colleagues, particularly Peter Rau, Jürgen Rac, René Lesnik, Alexander Sommer, Alexander Fischer, Tobias Paprotka, Fania Geiger, Katharina Hipp and Anan Kadri, for the good talks and exchange of experiences and knowledge.

My family, mainly my parents and my brother, for the support they provided me through my entire life, without which I could not have this work carried out.

My friends, especially Ricardo Alberti, Daniel Feldenheimer, Tainá Raiol and my boyfriend, Josh Halley, who, whether near or far away, were always helpful. Thank you for all the patience, support and encouraging words.

The European Research Area - Plant Genomics (ERA-PG), for the financial support to this project.

Also to anybody I forgot to mention who was involved in this work and contributed to its success.

Curriculum Vitae – PATRÍCIA SOARES WYANT

E-Mail: patriciasoaresw@gmail. com
Address: Allmandring 18A Zimmer 20 DE-70569, Germany
Telephone: (49) 01577-2718679
Date of Birth: 22. 07. 1981
Nationalities: Brazilian / British

Graduation:

2007-2010   Doctoral studies in Molecular Biology and Plant Virology at "University of Stuttgart", Germany

2005-2007   Master's degree in Molecular Biology at "University of Brasília", Brazil

2001-2004   Graduation in Biological Sciences at "University of Brasília", Brazil

Scientific Congresses:

02-04.06.2010   3rd RCA Genomics meeting – Malaga - Spain

14-16.04.2010   RNAi & the World of Small RNA Molecules – Wageningen - Nederland

11-12.03.2010   42nd Annual meeting of DPG working group "Virus diseases of plants" (42. Jahrestreffens des DGP-Arbeitskreises "Viruskrankheiten der Pflanzen") – Göttingen - Germany

08-09.04.2009   5th joint meeting of the DPG working group "Virus diseases of plants" and the "Dutch Society for Plant Virology" (5th joint meeting of the DPG working group "Viruskrankheiten der Pflanzen" and the "Nederlandse Kring voor Plantevirologie") – Hamburg - Germany

01-04.10.2008   2nd RCA genomics meeting – Gif sur Yvette - France

19-22.11.2006   XVII National Meeting of Virology - Campos do Jordão - Brazil

22-25.11.2005   XVI National Meeting of Virology – Salvador - Brazil

17-22.07.2005   57th Annual conference of Brazilian Society for the Progress of Science (SBPC) – Fortaleza - Brazil

25-27.10.2004   10th Congress of Scientific Initiation – Brasília - Brazil

14-17.09.2003   XIV National Meeting of Virology – Florianópolis - Brazil

20-22.08.2003   9th Congress of Scientific Initiation – Brasília - Brazil

13-18.07.2003   55th Annual conference of Brazilian Society for the Progress of Science(SBPC) – Recife - Brazil

Internships:

08.2005-12.2005   Supervised Apprenticeship in Biology Teaching

03.2005-07.2005   Supervised Apprenticeship in Science Teaching

03.2003-08.2003   Research in Vegetable Virology - "University of Brasília"

04.2002-06.2003   Molecular diagnosis of human and animal diseases - "University of Brasília"

03.2002-08.2002   Research in Molecular Biology - "University of Brasília"

08.2001-03.2002   Supervised Apprenticeship in Virology- "University of Brasília"

Languages:

Portuguese: native; English: fluent; German: basic; Spanish: intermediate.

Eidesstaatliche Erklärung

Hiermit versichere ich, dass ich diese Arbeit selbst angefertigt habe und nur die aufgeführten Quellen und Hilfsmittel verwendet habe.

Stuttgart, den 29. August 2011.

Patricia Soares Wyant