

SHORT REPORT

Open Access

Molecular diversity of *Cotton leaf curl Gezira virus* isolates and their satellite DNAs associated with okra leaf curl disease in Burkina Faso

Fidèle Tiendrébéogo^{1,2*}, Pierre Lefeuvre², Murielle Hoareau², Julie Villemot², Gnissa Konaté³, Alfred S Traoré¹, Nicolas Barro¹, Valentin S Traoré³, Bernard Reynaud², Oumar Traoré³, Jean-Michel Lett^{2*}

Abstract

Okra leaf curl disease (OLCD) is a major constraint on okra (*Abelmoschus esculentus*) production and is widespread in Africa. Using a large number of samples representative of the major growing regions in Burkina Faso (BF), we show that the disease is associated with a monopartite begomovirus and satellite DNA complexes. Twenty-three complete genomic sequences of *Cotton leaf curl Gezira virus* (CLCuGV) isolates associated with OLCD, sharing 95 to 99% nucleotide identity, were cloned and sequenced. Six betasatellite and four alphasatellite (DNA-1) molecules were also characterized. The six isolates of betasatellite associated with CLCuGV isolates correspond to *Cotton leaf curl Gezira betasatellite* (CLCuGB) (88 to 98% nucleotide identity). One isolate of alphasatellite is a variant of Cotton leaf curl Gezira alphasatellite (CLCuGA) (89% nucleotide identity), whereas the three others isolates appear to correspond to a new species of alphasatellite (CLCuGA most similar sequence present 52 to 60% nucleotide identity), provisionally named Okra leaf curl Burkina Faso alphasatellite (OLCBFA). Recombination analysis of the viruses demonstrated the interspecies recombinant origin of all CLCuGV isolates, with parents being close to *Hollyhock leaf crumple virus* (AY036009) and *Tomato leaf curl Diana virus* (AM701765). Combined with the presence of satellites DNA, these results highlight the complexity of begomoviruses associated with OLCD.

Findings

Okra leaf curl disease (OLCD) is commonly observed among okra (*Abelmoschus esculentus*) crops in Burkina Faso (BF) and several African countries [1-5]. Affected plants are severely stunted with apical leaf curl (upward or downward), distortion and thickening of the veins. In BF, okra is widely grown in both rainy and dry seasons. It is a major source of income particularly for small-scale farming. Viral diseases are important constraints in the production of this crop [6]. Recently, it was shown that OLCD in Africa is associated with a complex of begomoviruses: *Cotton leaf curl Gezira virus* (CLCuGV; [7,4,5]), *Okra yellow crinkle virus* (OYCrV; [8]) and *Hollyhock leaf crumple virus* (HoLCrV; [9,10]).

Viruses of the genus *Begomovirus* belong to the family *Geminiviridae* and are transmitted by the whitefly vector

Bemisia tabaci to dicotyledonous plants [11]. They have emerged as a major constraint for many vegetable and fibre crops throughout the world [12]. Begomoviruses are either bipartite with two genomic components, designated as DNA-A and DNA-B or monopartite with only DNA-A like components [13]. Some of the monopartite begomoviruses are also associated with additional circular ssDNA molecules, such as betasatellite or alphasatellite (previously known as DNA-1) that are nearly half the size of DNA-A. Betasatellites have been involved in pathogenicity but alphasatellites have no known function and are certainly not involved in symptom induction [14-16]. Alphasatellites have only been shown to be present in plants infected with monopartite begomoviruses in association with betasatellites [17].

The aim of our study was to characterize at the molecular level the complex of viruses involved in OLCD in BF and their relationship with other begomoviruses. In association with a single Old World begomovirus, we describe their associated satellite DNAs.

* Correspondence: fidelet@gmail.com; lett@cirad.fr

¹Laboratoire de Biochimie & Biologie Moléculaire, CRSBAN/UFR/SVT, Université de Ouagadougou 03 BP 7021 Ouagadougou 03, Burkina Faso
²CIRAD, UMR 53 PVBMT CIRAD-Université de la Réunion, Pôle de Protection des Plantes, 7 Chemin de l'IRAT, 97410 Saint Pierre, La Réunion, France

During May 2008 to April 2009, 74 leaf samples exhibiting typical OLCD symptoms were collected from okra fields in the major growing regions of BF around Tiébélé, Kampala, Pô, Kamboinsé, Bazèga and Bama (Kou valley) localities. Total DNA was extracted using DNeasy® Plant Minikit (Qiagen) before detection of begomoviruses using polymerase chain reaction (PCR) with specific primers of either the DNA-A [18] or betasatellite and alphasatellite [19,20]. Full-length viral genomes were amplified from the PCR-positive samples by rolling-circle amplification (RCA) [21]. The amplified DNAs were digested with endonucleases *Bam*HI or *Pst*I, and the DNA fragments of the expected size (~2.8 kb for DNA-A and ~1.4 kb for satellites) were cloned into pGEM®-3Zf (+) vector (Promega Biotech). Cloned genome components were sequenced by MacroGen Inc. (South Korea). Contigs were assembled with the DNAMAN software (Lynnon, Quebec, Canada) and subsequently aligned using the ClustalW tool [22] implemented in MEGA 4 [23]. Sequence comparisons were performed in MEGA 4 with pairwise deletion of gaps. The optimal model of sequence evolution, defined with ModelTest [24], was used for maximum likelihood (ML) phylogenetic reconstruction using PHYML v2.4.4 [25]. The degree of support for individual branches within the resulting phylogenetic trees was assessed with 1000 full ML bootstrap iterations. The trees were visualized using FigTree v1.1.1 software.

Recombination was analyzed using our sequences and a set of sequences representing the whole African begomovirus diversity (representing an alignment of 121 sequences). Detection of potential recombinant sequences, identification of likely parental sequences, and localization of possible recombination breakpoints was carried out using RDP [26], GENECONV [27], BOOTSCAN [28], MAXIMUM CHI SQUARE [29], CHIMAERA [28], SISTER SCAN [30] and 3Seq [31] recombination detection methods as implemented in RDP3 [32]. The analysis was performed with default settings for the different detection methods and a Bonferroni corrected *P*-value cut-off of 0.05. Only events detected with 3 methods or more were accepted.

Despite a very poor preservation of samples (high necrosis), 48 samples of the 74 were detected as being infected with begomovirus using PCR amplifications with the universal primer pair VD360-CD1266 recovering the conserved CP ORF [18]. From the positive samples, 23 begomovirus genome sequences with length between 2761 to 2773 nucleotides (nt) were successfully obtained using RCA. Pairwise sequence comparison demonstrated that the 23 new genome sequences of monopartite begomoviruses from BF are genetically related to the same strain (94.7 to 100% identity amongst themselves). A BLAST search identify the most similar virus sequences as being *Cotton leaf curl Gezira*

virus (CLCuGV) isolates, a monopartite *Begomovirus* first identified in Sudan [33]. Further pairwise sequence analyses showed that the 23 sequences shared between 94.8 to 98.8% nt identity with CLCuGV isolates from Niger (FJ469626, EU432373, EU432374), 93.7 to 96.2% with CLCuGV from Sudan (AY036007, AY036008), 92.4 to 96.1% with CLCuGV from Egypt (AY036006, AY036010) and 89.3 to 91.4% with CLCuGV from Cameroon (FM164726). According to the ICTV guidelines, these results of nucleotide identity <93% between isolates of CLCuGV suggest the existence of several strains within this begomovirus species. Similar comparisons performed with the other two *Begomovirus* species infecting *Malvaceae* in Africa showed low nucleotide sequence identity: 71.7 to 72.6% obtained with *Okra yellow crinkle virus* (DQ902715, EU024118, FM164724) and 83.4 to 84.2 with *Hollyhock leaf crumple virus* (AY036009, AF014881).

All 23 isolates of begomovirus infecting okra in BF have the typical genome organization of Old World monopartite begomoviruses. This organization consisted of the presence of six open reading frames (ORFs) on the DNA-A corresponding to V1 and V2 on the virion strand and C1, C2, C3 and C4 on the complementary strand [34]. The IR sequences located between the start codons of the C1 and V2 are 289 to 300 nt. In this region, they present a typical replication origin (↓), including an inverted repeat sequence containing the highly conserved nanonuclotide sequence TAATAT-T↓AC [35,5].

Based on the presently applicable species demarcation threshold of 89% for begomoviruses [36], we conclude that the 23 begomovirus isolates isolated from okra in BF belong to the species *Cotton leaf curl Gezira virus* and the Niger strain (See Table 1 for percentage of similarities and Table 2 for isolates description and accession numbers). In addition, a maximum-likelihood phylogenetic tree constructed using PHYML and the GTR+I+G model of sequence evolution (ModelTest), confirms that okra begomoviruses reported here cluster with the isolates of *Cotton leaf curl Gezira virus* (CLCuGV) (Figure 1). A clear phylogeographic separation is observed between the diversity of CLCuGV isolates of okra: West Africa (Niger strain), Central Africa (Cameroon strain), East Africa (Sudan strain) and north-east of the Africa (Egypt strain).

Betasatellites were found associated to all isolates from BF except CLCuGV-NE[BF:Baz:Ok:09] and CLCuGV-NE[BF:Bam:Ok:09] while alphasatellites were detected only in association with the seven following isolates: CLCuGV-NE[BF:Kap:Ok7:08], CLCuGV-NE[BF:Pô:Ok1:08], CLCuGV-NE[BF:Pô:Ok2:08], CLCuGV-NE[BF:Pô:Ok4:08], CLCuGV-NE[BF:Pô:Ok5:08], CLCuGV-NE[BF:Pô:Ok6:08] and CLCuGV-NE[BF:Pô:Ok7:08].

Table 1 Matrix of pairwise identity percentages between the complete DNA-A nucleotide sequences of the twenty-three isolates of okra-infecting Cotton leaf curl Gezira virus (CLCuGV) isolates from Burkina Faso and representatives of some begomoviruses infecting okra in Africa.

	CLCuGV-NE [NE:Sad: NG1:07]*	CLCuGV-NE [NE:Sad: AF1:07]*	CLCuGV-NE [NE:Sad: NG2:07]*	CLCuGV-NE [NE:Sad: OK]	CLCuGV-EG [EG:Cai: OK]	CLCuGV-SD [SD:Sha: OK]*	CLCuGV-SD [SD:Gez: OK]*	CLCuGV-SD [SD:Gez: OK]*	CLCuGV-SD [SD:Gez: S]	CLCuGV- CM	CLCuGV-CM [CM:Lys: Ok:08]*	OYCrV- [ML:06]	OYCrV- [CM:No: Ok:08]*	OYCrV- [ML:01: 05]*	HoLCrV- [EG: Cai:97]	HoLCrV- [EG: Cai:97]
NE [BF:TieOk1:08]*	96.3	96.8	98.7	93.3	95	94.6	94.8	89.6	72.4	72.6	72.4	72.4	72.4	84.1	83.9	
NE [BF:TieOk2:08]*	96.3	96.8	98.7	93.3	95	94.6	94.8	89.6	72.4	72.4	72.4	72.4	72.4	84.1	83.9	
NE [BF:KapOk1:08]*	97.3	96.8	96.8	94.4	96.1	95.9	96.0	91.4	72.0	72.4	72.0	72.4	72.0	83.7	83.7	
NE [BF:KapOk2:08]*	96.4	96.4	98.8	93.3	94.9	94.6	94.8	89.3	72.2	72.3	72.2	72.3	72.2	83.9	83.7	
NE [BF:KapOk3:08]*	96.8	97.1	96.7	94.4	96.2	95.9	96.0	91.4	72.1	72.6	72.0	72.6	72.0	84.1	84.1	
NE [BF:KapOk4:08]*	96.3	96.8	98.7	93.3	95.0	94.7	94.8	89.6	72.4	72.6	72.4	72.6	72.4	84	83.8	
NE [BF:KapOk5:08]*	96.6	96.2	96.8	94.2	96.1	95.9	96.1	91.2	72	72.3	71.9	72.3	71.9	83.5	83.5	
NE [BF:KapOk6:08]*	95.9	96.2	96.2	94.6	96.2	95.8	96.1	91.5	71.7	72.4	71.7	72.4	71.7	84.0	84.1	
NE [BF:KapOk7:08]*	96.3	96.8	98.6	93.2	95	94.6	94.8	89.6	72.3	72.5	72.3	72.5	72.3	83.9	83.7	
NE [BF:KapOk8:08]*	96.8	96.4	96.7	94.3	96.1	95.9	96.0	91.3	72.0	72.4	71.9	72.4	71.9	83.8	83.7	
NE [BF:KapOk9:08]*	96.3	96.8	98.7	93.3	95.0	94.6	94.8	89.6	72.4	72.6	72.3	72.6	72.3	83.9	83.8	
NE [BF:KapOk10:08]*	95.0	94.8	97.3	93.2	94.8	94.6	94.7	89.3	72.2	72.0	72.2	72.0	72.2	83.6	83.4	
NE [BF:P6:Ok1:08]*	95.4	95.1	95.9	94.0	95.7	95.4	95.6	90.7	71.9	72.3	71.9	72.3	71.9	83.4	83.4	
NE [BF:P6:Ok2:08]*	95.3	95.1	95.8	94.0	95.7	95.4	95.6	90.6	71.9	72.3	71.9	72.3	71.9	83.4	83.4	
NE [BF:P6:Ok3:08]*	96.6	96.5	96.5	94.4	95.9	95.7	95.8	91.1	72	72.3	72	72.3	71.9	83.8	83.8	
NE [BF:P6:Ok4:08]*	96.6	96.5	96.5	94.4	95.9	95.7	95.8	91.1	72	72.3	72	72.3	71.9	83.8	83.8	
NE [BF:P6:Ok5:08]*	96.6	96.5	96.5	94.4	95.9	95.7	95.8	91.1	72	72.3	72	72.3	71.9	83.8	83.8	
NE [BF:P6:Ok6:08]*	96.6	96.5	96.5	94.4	95.9	95.7	95.8	91.1	72	72.3	72	72.3	71.9	83.8	83.8	
NE [BF:P6:Ok7:08]*	96.6	96.5	96.5	94.4	95.9	95.7	95.8	91.1	72	72.3	72	72.3	71.9	83.8	83.8	
NE [BF:KabOk1:08]*	96.6	96.5	96.5	94.4	95.9	95.7	95.8	91.1	72	72.3	72	72.3	71.9	83.8	83.8	
NE [BF:KabOk2:08]*	96.6	96.5	96.5	94.4	95.9	95.7	95.8	91.1	72	72.3	72	72.3	71.9	83.8	83.8	
NE [BF:BazOk09]*	96.4	96.7	98.8	93.4	95.1	94.7	95	89.6	72.4	72.4	72.4	72.4	72.4	83.9	83.7	
NE [BF:BamOk09]*	98.7	98.2	95.9	92.4	93.9	93.7	93.8	89.3	72.4	72.7	72.4	72.7	72.3	84.2	84.0	

*Begomovirus DNA-A isolated on Okra (*Abelmoschus esculentus*) from Burkina Faso
 Acronyms: Niger: NE; Burkina Faso: BF; Tiébélé: Tie; Kampala: Kap; Kamboinsé: Kab; Bazéga: Baz; Bama: Bam
 Percentages of identity over the 89% taxonomic threshold of begomovirus are presented in bold

Table 2 Name, acronym and accession numbers of begomovirus reported in this study.

Virus name	Acronym	Isolates groups	DNA-A length (nt)	Accession numbers
Cotton leaf curl Gezira virus-Niger [Burkina Faso:Tiébélé:Okra1:2008]	CLCuGV-NE [BF:Te:Ok1:08]	G1	2762	FN554519
Cotton leaf curl Gezira virus-Niger [Burkina Faso:Tiébélé:Okra2:2008]	CLCuGV-NE [BF:Te:Ok2:08]	G1	2762	FN554520
Cotton leaf curl Gezira virus-Niger [Burkina Faso:Kampala:Okra1:2008]	CLCuGV-NE [BF:Kap:Ok1:08]	G1	2761	FN554521
Cotton leaf curl Gezira virus-Niger [Burkina Faso:Kampala:Okra2:2008]	CLCuGV-NE [BF:Kap:Ok2:08]	G1	2762	FN554522
Cotton leaf curl Gezira virus- Niger [Burkina Faso:Kampala:Okra3:2008]	CLCuGV-NE [BF:Kap:Ok3:08]	G2	2763	FN554523
Cotton leaf curl Gezira virus-Niger [Burkina Faso:Kampala:Okra4:2008]	CLCuGV-NE [BF:Kap:Ok4:08]	G1	2762	FN554524
Cotton leaf curl Gezira virus-Niger [Burkina Faso:Kampala:Okra5:2008]	CLCuGV-NE [BF:Kap:Ok5:08]	G2	2762	FN554525
Cotton leaf curl Gezira virus-Niger [Burkina Faso:Kampala:Okra6:2008]	CLCuGV-NE [BF:Kap:Ok6:08]	G2	2763	FN554526
Cotton leaf curl Gezira virus-Niger [Burkina Faso:Kampala:Okra7:2008]	CLCuGV-NE [BF:Kap:Ok7:08]	G1	2762	FN554527
Cotton leaf curl Gezira virus-Niger [Burkina Faso:Kampala:Okra8:2008]	CLCuGV-NE [BF:Kap:Ok8:08]	G2	2762	FN554528
Cotton leaf curl Gezira virus-Niger [Burkina Faso:Kampala:Okra9:2008]	CLCuGV-NE [BF:Kap:Ok9:08]	G1	2765	FN554529
Cotton leaf curl Gezira virus-Niger [Burkina Faso:Kampala:Okra10:2008]	CLCuGV-NE [BF:Kap:Ok10:08]	G2	2773	FN554530
Cotton leaf curl Gezira virus-Niger [Burkina Faso:Po:Okra1:2008]	CLCuGV-NE [BF:Po:Ok1:08]	G2	2770	FN554531
Cotton leaf curl Gezira virus-Niger [Burkina Faso:Po:Okra2:2008]	CLCuGV-NE [BF:Po:Ok2:08]	G2	2770	FN554532
Cotton leaf curl Gezira virus-Niger [Burkina Faso:Po:Okra3:2008]	CLCuGV-NE [BF:Po:Ok3:08]	G2	2761	FN554533
Cotton leaf curl Gezira virus-Niger [Burkina Faso:Po:Okra4:2008]	CLCuGV-NE [BF:Po:Ok4:08]	G2	2761	FN554534
Cotton leaf curl Gezira virus-Niger [Burkina Faso:Po:Okra5:2008]	CLCuGV-NE [BF:Po:Ok5:08]	G2	2761	FN554535
Cotton leaf curl Gezira virus-Niger [Burkina Faso:Po:Okra6:2008]	CLCuGV-NE [BF:Po:Ok6:08]	G2	2761	FN554536
Cotton leaf curl Gezira virus-Niger [Burkina Faso:Po:Okra7:2008]	CLCuGV-NE [BF:Po:Ok7:08]	G2	2761	FN554537
Cotton leaf curl Gezira virus-Niger [Burkina Faso:Kamboinsé:Okra1:2008]	CLCuGV-NE [BF:Kab:Ok1:08]	G2	2761	FN554538
Cotton leaf curl Gezira virus-Niger [Burkina Faso:Kamboinsé:Okra2:2008]	CLCuGV-NE [BF:Kab:Ok2:08]	G2	2761	FN554539
Cotton leaf curl Gezira virus-Niger [Burkina Faso:Bazega:Okra:2009]	CLCuGV-NE [BF:Baz:Ok:09]	G1	2762	FN554540
Cotton leaf curl Gezira virus-Niger [Burkina Faso:Bama:Okra:2009]	CLCuGV-NE [BF:Bam:Ok:09]	G3	2771	FN554541

*GenBank-EMBL-DBJ data bases



Figure 1 Maximum likelihood tree based on the complete DNA-A sequences of twenty-three *Cotton leaf curl Gezira virus* isolates from Burkina Faso (in bold; see Table 2 for isolates name and acronyms), plus additional sequences from African and Asian monopartite and bipartite begomoviruses. Begomovirus acronyms used are *Cotton leaf curl Gezira virus* (CLCuGV), *Hollyhock leaf crumple virus* (HoLCrV), *Cotton leaf curl Bangalore virus* (CLCuBV), *Okra yellow vein mosaic virus* (OYVMV), *Tomato leaf curl Togo virus* (ToLCTGV), *Tomato leaf curl Ghana virus* (ToLCGHV), *Tomato leaf curl Nigeria virus* (ToLCNGV), *South African cassava mosaic virus* (SACMV), *Tomato yellow leaf curl Sardinia virus* (TYLCSV), *Tomato leaf curl Sudan virus* (ToLCSDV), *Tomato yellow leaf curl Mali virus* (TYLCMLV), *Tomato yellow leaf curl virus-Mild* (TYLCV-Mld), *Pepper yellow vein Mali virus* (PepYVMV), *Tomato curly stunt virus* (ToCSV), *Tobacco leaf curl Zimbabwe virus* (TbLCZV), *Tomato leaf curl Mali virus* (ToLCMV), *Okra yellow crinkle virus* (OYCrV) and *Malvastrum leaf curl virus* (MaLCV). For the complete description of isolate descriptors refer to Fauquet et al. (2008). Four genetic groups (G1 to G4) have been defined on the presence or absence of recombination events (Figure 4), and are represented here.

Betasatellites associated with CLCuGV-NE[BF:Tie:Ok2:08], CLCuGV-NE[BF:Kap:Ok1:08], CLCuGV-NE[BF:Kap:Ok3:08], CLCuGV-NE[BF:Kap:Ok5:08], CLCuGV-NE[BF:Kap:Ok6:08] and CLCuGV-NE[BF:Pô:Ok6:08] consisted of 1348, 1347, 1349, 1348, 1347 and 1347 nucleotides, respectively. All betasatellites showed typical features consisting of the presence of a single ORF β C1 in the complementary-sense, a region of sequence rich in adenine (A) (nt 703-892 with 58.4 to 58.7% A residues) and a satellite conserved region (SCR) with a predicted stem-loop structure containing the geminivirus nonanucleotide sequence (TAATATTAC) [37]. The nucleotide sequence comparison showed that our sequences had nucleotide identities ranging from 88.1 to 98.7% with betasatellites from Cameroon, Egypt, Mali, Niger and Sudan. In a phylogenetic analysis based upon alignments of the complete betasatellites sequences, the BF betasatellite sequences segregated with

betasatellites associated with okra begomoviruses from Africa (Figure 2). Based on the recently established species demarcation threshold for betasatellites (78% nucleotide sequence identity; [38]), the betasatellites reported in this study belong to the same species *Cotton leaf curl Gezira betasatellite* (see Table 3 for betasatellites isolates description and accession numbers). Interestingly and under our knowledge, this species represent the only known betasatellite described in Africa on malvaceous and tomato plants. Associated to the absence of betasatellites in the New World and the existence of a high diversity of betasatellites in Asia, this result confirms that the centre of diversity appears to be in southern Asia [39].

The complete nucleotide sequences of alphasatellites associated with CLCuGV-NE[BF:Kap:Ok7:08], CLCuGV-NE[BF:Pô:Ok1:08], CLCuGV-NE[BF:Pô:Ok4:08] and CLCuGV-NE[BF:Pô:Ok5:08] were determined to be

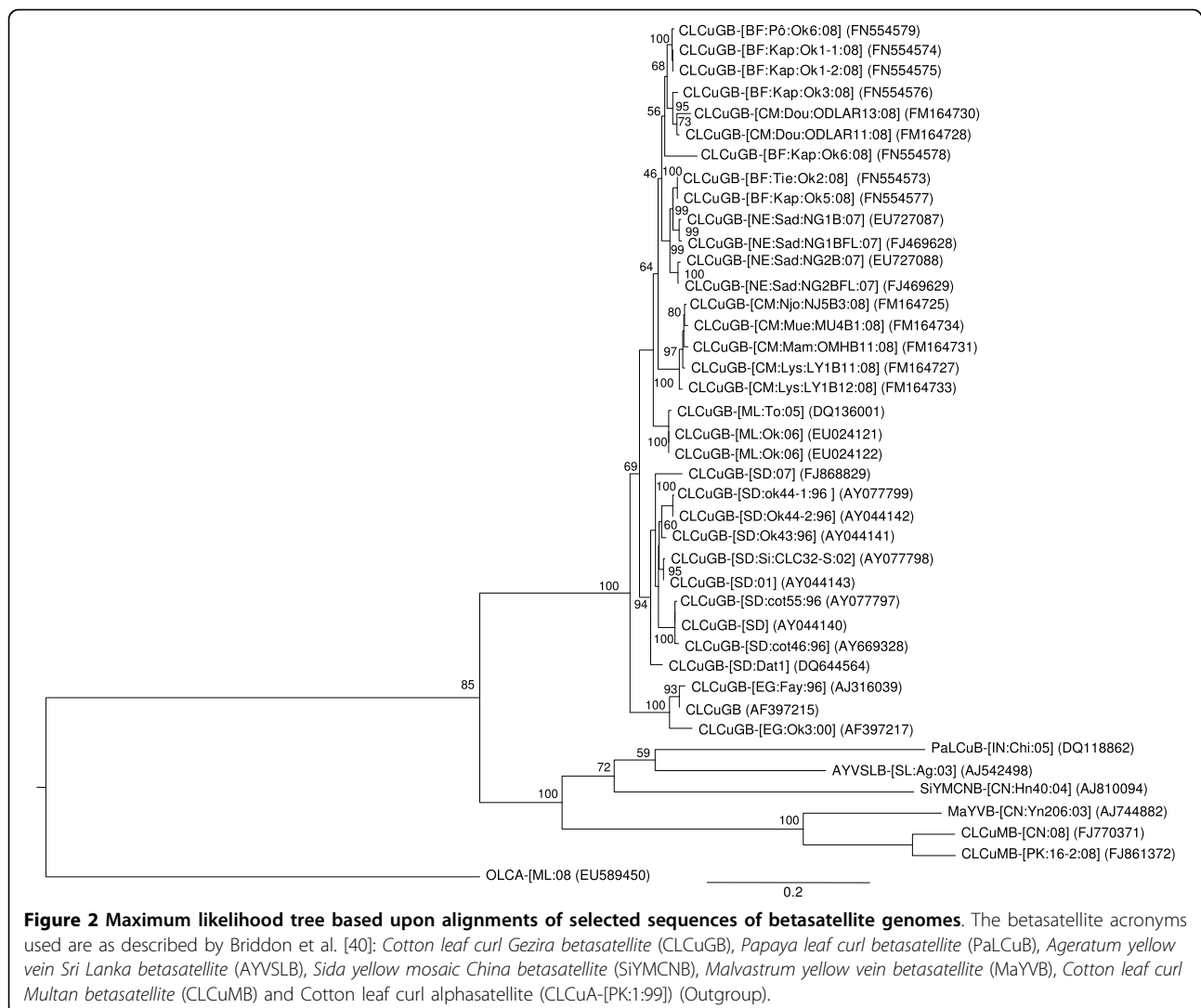


Table 3 Betasatellites and alphasatellites characterized in this study.

Betasatellite and alphasatellite names	Acronym	DNA molecule and size (nt)		Accession numbers
		Betasatellite	Alphasatellite	
<i>Cotton leaf curl Gezira betasatellite</i> -[Burkina Faso:Tiébélé:Okra2:2008]	CLCuGB- [BF:Tie:Ok2:08]	1348		FN554573
<i>Cotton leaf curl Gezira betasatellite</i> -[Burkina Faso:Kampala:Okra1-1:2008]	CLCuGB- [BF:Kap:Ok1-1:08]	1347		FN554574
<i>Cotton leaf curl Gezira betasatellite</i> -[Burkina Faso:Kampala:Okra1-2:2008]	CLCuGB- [BF:Kap:Ok1-2:08]	1347		FN554575
<i>Cotton leaf curl Gezira betasatellite</i> -[Burkina Faso:Kampala:Okra3:2008]	CLCuGB-[BF:Kap:Ok3:08]	1349		FN554576
<i>Cotton leaf curl Gezira betasatellite</i> -[Burkina Faso:Kampala:Okra5:2008]	CLCuGB- [BF:Kap:Ok5:08]	1348		FN554577
<i>Cotton leaf curl Gezira betasatellite</i> -[Burkina Faso:Kampala:Okra6:2008]	CLCuGB- [BF:Kap:Ok6:08]	1347		FN554578
<i>Cotton leaf curl Gezira betasatellite</i> -[Burkina Faso:Pô:Okra6:2008]	CLCuGB- [BF:Pô:Ok6:08]	1347		FN554579
<i>Cotton leaf curl Gezira alphasatellite</i> -[Burkina Faso:Kampala:Okra7:2008]	CLCuGA- [BF:Kap:Ok7:08]		1387	FN554580
Okra leaf curl Burkina Faso alphasatellite-[Burkina Faso:Pô:Okra1:2008]	OLCBFA- [BF:Pô:Ok1:08]		1353	FN554581
Okra leaf curl Burkina Faso alphasatellite-[Burkina Faso:Pô:Okra4:2008]	OLCBFA- [BF:Pô:Ok4:08]		1299	FN554582
Okra leaf curl Burkina Faso alphasatellite-[Burkina Faso:Pô:Okra5:2008]	OLCBFA- [BF:Pô:Ok5:08]		1353	FN554583

*GenBank-EMBL-DDBJ data bases

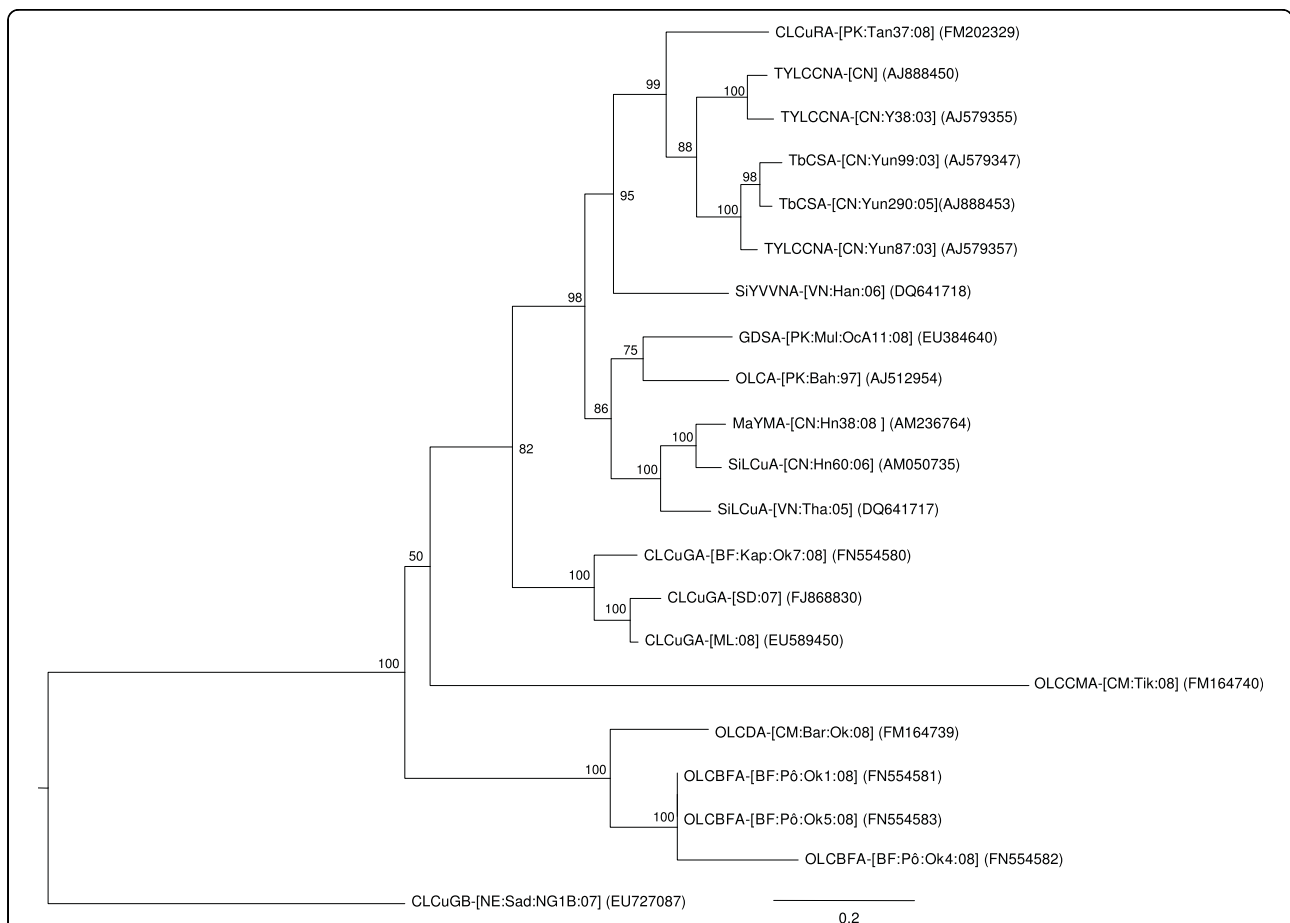


Figure 3 Maximum likelihood tree based on selected alphasatellite sequences. Acronyms used are as described by Mubin et al. [17]: Malvastrum yellow mosaic alphasatellite (MaYMA), Sida leaf curl alphasatellite (SiLCuA), Gossypium darwinii symptomless alphasatellite (GDSA), Okra leaf curl alphasatellite (OLCA), Cotton leaf curl Rajasthan alphasatellite (CLCuRA), Tomato yellow leaf curl China alphasatellite (TYLCCNA), Tobacco curly shoot alphasatellite (TbCSA), Sida yellow vein Vietnam alphasatellite (SiYVVNA), Okra leaf curl alphasatellite (OLCA), Okra leaf curl Cameroon alphasatellite (OLCCMA) and *Cotton leaf curl Gezira betasatellite* (CLCuGB) (outgroup).

1382, 1353, 1299 and 1353 nt respectively. The alphasatellite sequence associated with CLCuGV-NE[BF:Kap:Ok7:08] display the highest level of nucleotide sequence identity (88.9%) with Cotton leaf curl Gezira alphasatellite from Mali (CLCuGA-[Mali:Bamako]; EU589450). The phylogenetic analysis showed that the alphasatellite associated with CLCuGV-NE[BF:Kap:Ok7:08] segregate with CLCuGA-[Mali:Bamako] and OLCA-[Sudan:2007] (Figure 3) and has an arrangement typical of characterized alphasatellites [40], containing a single ORF in the virion sense, an A-rich region with 51% adenine and a hairpin structure with the loop sequence TAGTATTAC. The alphasatellites associated with CLCuGV-NE[BF:Pô:Ok1:08], CLCuGV-NE[BF:Pô:Ok4:08] and CLCuGV-NE [BF:Pô:Ok5:08] shared between 84.8 to 100% nucleotide sequence identity amongst themselves and only 52.4 to 60.1% with the alphasatellites associated with CLCuGV-NE[BF:Kap:Ok7:08] and those characterized in Mali and Sudan (respectively, CLCuGA-[Mali:Bamako] and OLCD1-[Sudan:2007]). Considering the suggested species demarcation threshold of 83% nucleotide sequence identity for alphasatellites [17], these alphasatellites represent isolates of a new species provisionally named Okra leaf curl Burkina Faso alphasatellite, clustering together in the phylogenetic tree (Figure 3; see Table 3 for alphasatellites accession numbers). These particular

alphasatellite isolates contain a single ORF in the virion sense and a predicted hairpin structure with the loop sequence CAGTATTAC.

Further to the sequence description of the viral isolates, we were interested in their possible recombinant origin. Three distinct recombination events (a, b and c) were detected within the full genome sequences of CLCuGV isolates (Figure 4), using a large sequence alignment of geminiviruses [41]. The presence or absence of these recombination events has identified four genetic groups of viruses (G1 to G4; Figures 1 and 4). Recombination event b present in all CLCuGV isolates involves a major parent being related to the HoLCrV described in north Africa (Egypt; [9]) and a minor parent related to ToLCDiaV described in the south-west Indian Ocean Islands (Madagascar; [41]). Compared to events a and c based on intra-strain recombination, event b seems to be more ancient. The recombination events a and c specific to isolates G1, G3 and G4 have been characterized in Burkina Faso and in Niger and appear to represent a specific geographic signature. The distribution of the recombination breakpoints observed here confirm the existence of recombination hot spots over the intergenic region (IR) and the centre of C1 ORF (Figure 4) as described by Lefeuvre et al. [41]. The recombination event c of

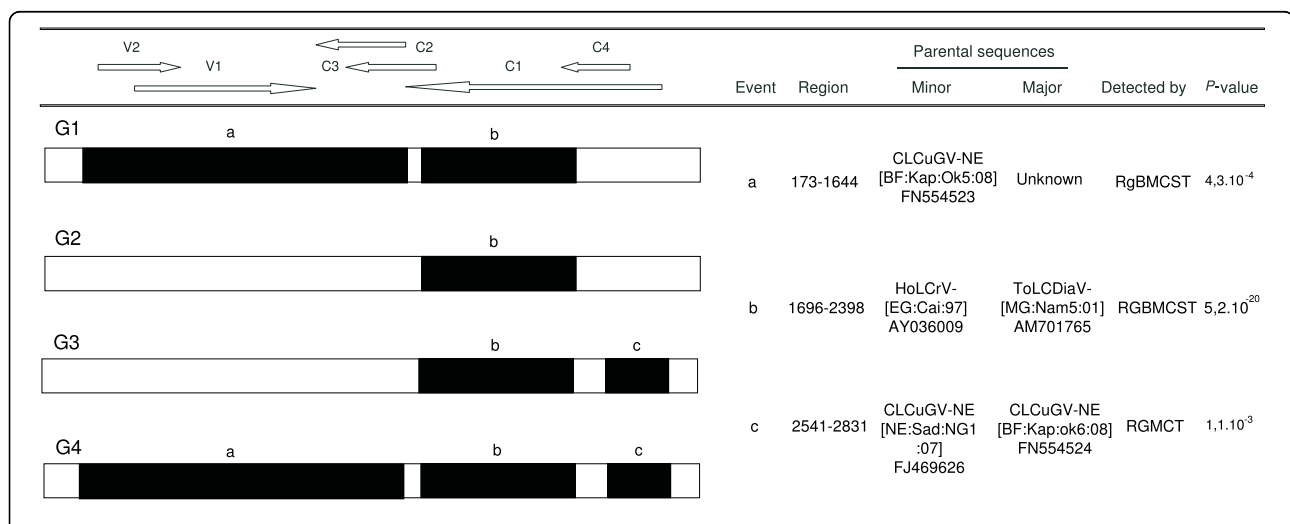


Figure 4 Recombinant regions (a, b and c) detected within the African isolates of CLCuGV sequences using RDP3. Four genetic groups (G1 to G4) have been defined on the presence or absence of recombination events. The genome at the top of the figure corresponds to the schematic representation of sequences below. Region coordinates are nucleotide positions of detected recombination breakpoints in the multiple sequence alignment used to detect recombination. Wherever possible, parental sequences are identified. "Major" and "Minor" parents are sequences that were used, along with the indicated recombinant sequence, to identify recombination. Whereas for each identified event the minor parent is apparently the contributor of the sequence within the indicated region, the major parent is the apparent contributor of the rest of the sequence. Note that the identified "parental sequences" are not the actual parents but are simply those sequences most similar to the actual parents in the analysed dataset. Recombinant regions and parental viruses were identified using the RDP (R), GENECONV (G), BOOTSCAN (B), MAXIMUM CHI SQUARE (M), CHIMAERA (C), SISTER SCAN (S) and 3Seq (T) methods. Whereas upper case letters imply a method detected recombination with a multiple comparison corrected P-value < 0.01, lower case letters imply the method detected recombination with a multiple comparison corrected P-value < 0.05 but > = 0.01.

isolates G3 and G4 covers the N terminus of the replication associated protein (Rep) which contains the iteron-related domain (IRD) [42]. This domain is involved in the specificity of interaction with iterated DNA motifs (iterons) of the geminivirus origin of replication (*ori*), functioning as essential elements for specific virus replication. Since the IRD domain of G3 and G4 isolates (MAPTKKFRINSKKNYFL) is different from the IRD domains of G1 and G2 isolates (MPPSKRFLINA-KKNYFL or MPFGTHYLSTDILER), the biological aspects of recombination events should be investigated in the future.

In conclusion, in Burkina Faso OLCB is mainly caused by a single begomovirus species and a complex of beta and alpha satellite species, contrary to what happens in the neighbouring countries Mali and Niger (respectively, [5,4]). Taken together, the current molecular results highlight the complex aetiology of the OLCB in Africa and the need for further investigations.

Acknowledgements

This study was supported by the following institutions: International Foundation for Science (IFS) fellowship N°C/4472-1 to F. Tiendrébéogo, AIRES-Sud: a programme from the French Ministry of Foreign and European Affairs implemented by the Institut de Recherche pour le Développement (IRD-DSF), CRSBAN/UFR-SVT (University of Ouagadougou), CIRAD, Conseil Régional de La Réunion, European Union (FEDER) and GIS Centre de recherche et de veille sanitaire sur les maladies émergentes dans l'océan Indien (N°PRAO/AIRD/CRVOI/08/03). The authors wish to thank the anonymous reviewer for excellent comments and Ben Warren for revising the English version of the manuscript. FT completed this research as part of his PhD Degree.

Author details

¹Laboratoire de Biochimie & Biologie Moléculaire, CRSBAN/UFR/SVT, Université de Ouagadougou 03 BP 7021 Ouagadougou 03, Burkina Faso.
²CIRAD, UMR 53 PVBMT CIRAD-Université de la Réunion, Pôle de Protection des Plantes, 7 Chemin de l'IRAT, 97410 Saint Pierre, La Réunion, France.
³Institut de l'Environnement et de Recherches Agricoles (INERA) 01 BP 476 Ouagadougou 01, Burkina Faso.

Authors' contributions

FT, VSET and OT collected samples; FT, MH, JV cloned and sequenced the viruses and satellites; FT, PL and JML analysed the data and prepared the manuscript. JML, OT, NB, GK, AST, VSET and BR secured funding for the project's execution, and provided ideas and comments during manuscript preparation. All authors have read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 20 November 2009

Accepted: 23 February 2010 Published: 23 February 2010

References

1. N'Guessan KP, Fargette D, Fauquet C, Thouvenel JC: **Aspects of the epidemiology of okra leaf curl virus in Côte d'Ivoire.** *Trop Pest Manag* 1992, **38**:122-126.
2. Konaté G, Barro N, Fargette D, Swanson MM, Harrison BD: **Occurrence of whitefly-transmitted geminiviruses in crops in Burkina Faso, and their serological detection and differentiation.** *Ann Appl Biol* 1995, **126**:121-129.
3. Swanson MM, Harrison BD: **Serological relationships and epitope profiles of isolates of okra leaf curl geminivirus from Africa and the Middle-East.** *Biochimie* 1993, **75**:707-711.
4. Shih SL, Kumar S, Tsai WS, Lee LM, Green SK: **Complete nucleotide sequences of okra isolates of Cotton leaf curl Gezira virus and their associated DNA-β from Niger.** *Arch Virol* 2009, **154**:369-372.
5. Kon T, Rojas MR, Abdourhamane IK, Gilbertson RL: **Roles and interactions of begomoviruses and satellite DNAs associated with okra leaf curl disease in Mali, West Africa.** *J Gen Virol* 2009, **90**:1001-1013.
6. Ndunguru J, Rajabu AC: **Effect of okra mosaic virus disease on the above-ground morphological yield components of okra in Tanzania.** *Sci Hortic* 2004, **99**:225-235.
7. Idris AM, Brown JK: **Molecular Analysis of Cotton Leaf Curl Virus-Sudan Reveals an evolutionary history of recombination.** *Virus Genes* 2002, **24**:249-256.
8. Shih SL, Green SK, Tsai WS, Lee LM, Levasseur V: **First report of a distinct begomovirus associated with okra yellow crinkle disease in Mali.** *Plant Pathol* 2007, **56**:718.
9. Bigarré L, Chazly M, Salah M, Ibrahim M, Padidam M, Nicole M, Peterschmitt M, Fauquet C, Thouvenel JC: **Characterization of a new begomovirus from Egypt infecting hollyhock (*Althea rosea*).** *Eur J Plant Pathol* 2001, **107**:701-711.
10. Idris AM, Hussein MH, Abdel-Salam AM, Brown JK: **Phylogenetic relationships for okra leaf curl- and hollyhock leaf crumple-associated begomoviruses and first report of associated satellite DNAs.** *Arab J Biotechnol* 2002, **5**:67-82.
11. Fauquet CM, Stanley J: **Geminivirus classification and nomenclature: progress and problems.** *Ann Appl Biol* 2003, **142**:165-189.
12. Mansoor S, Briddon RW, Bull SE, Bedford ID, Bashir A, Hussain M, Saeed M, Zafar Y, Malik KA, Fauquet C, Markham PG: **Cotton leaf curl disease is associated with multiple monopartite begomoviruses supported by single DNAβ.** *Arch Virol* 2003, **148**:1969-1986.
13. Lazarowitz SG: **Geminiviruses: genome structure and gene function.** *Crit Rev Plant Sci* 1992, **11**:327-349.
14. Mansoor S, Khan SH, Bashir A, Saeed M, Zafar Y, Malik KA, Briddon R, Stanley J, Markham PG: **Identification of a novel circular single stranded DNA associated with cotton leaf lurl disease in Pakistan.** *Virology* 1999, **259**:190-199.
15. Briddon RW, Stanley J: **Sub-viral agents associated with plant-infecting single-stranded DNA viruses.** *Virology* 2006, **344**:198-210.
16. Saunders K, Briddon RW, Stanley J: **Replication promiscuity of DNAβ satellites associated with monopartite begomoviruses: deletion mutagenesis of the Ageratum yellow vein virus DNAβ satellite localises sequences involved in replication.** *J Gen Virol* 2008, **89**:3165-3172.
17. Mubin M, Briddon RW, Mansoor S: **Complete nucleotide sequence of chili leaf curl virus and its associated satellites naturally infecting potato in Pakistan.** *Arch Virol* 2009, **154**:365-368.
18. Delatte H, Martin DP, Naze F, Golbach RW, Reynaud B, Peterschmitt M, Lett JM: **South West Indian Ocean islands tomato begomovirus populations represent a new major monopartite begomovirus group.** *J Gen Virol* 2005, **86**:1533-1542.
19. Briddon RW, Bull SE, Mansoor S, Amin I, Markham PG: **Universal primers for the PCR-mediated amplification of DNA β: a molecule associated with some monopartite begomoviruses.** *Mol Biotechnol* 2002, **20**:315-318.
20. Bull SE, Briddon RW, Markham PG: **Universal primers for the PCR-mediated amplification of DNA1: a satellite-like molecule associated with begomovirus-DNA β complexes.** *Mol Biotechnol* 2003, **23**:83-86.
21. Inoue-Nagata AK, Albuquerque LC, Rocha WB, Nagata T: **A simple method for cloning the complete begomovirus genome using the bacteriophage φ29 DNA polymerase.** *J Virol Methods* 2004, **116**:209-211.
22. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22**:4673-4680.
23. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0.** *Mol Biol Evol* 2007, **24**:1596-1599.
24. Posada D: **ModelTest Server: a web-based tool for the statistical selection of models of nucleotide substitution online.** *Nucleic Acids Res* 2006, **34**:W700-3.
25. Guindon S, Lethiec F, Duroux P, Gascuel O: **PHYML Online-a web server for fast maximum likelihood-based phylogenetic inference.** *Nucleic Acids Res* 2005, **33**:557-559.
26. Martin D, Rybicki E: **RDP: detection of recombination amongst aligned sequences.** *Bioinformatics* 2000, **16**:562-563.

27. Padidam M, Sawyer S, Fauquet CM: **Possible emergence of new geminiviruses by frequent recombination.** *Virology* 1999, **265**:218-225.
28. Martin DP, Posada D, Crandall KA, Williamson C: **A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints.** *AIDS Res Hum Retroviruses* 2005, **21**:98-102.
29. Maynard SJ: **Analyzing the mosaic structure of genes.** *J Mol Evol* 1992, **34**:126-129.
30. Gibbs MJ, Armstrong JS, Gibbs AJ: **Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences.** *Bioinformatics* 2000, **16**:573-582.
31. Boni MF, Posada D, Feldman MW: **An Exact Nonparametric Method for Inferring Mosaic Structure in Sequence Triplets.** *Genetics* 2007, **176**:1035-1047.
32. Martin DP, Williamson C, Posada D: **RDP2: recombination detection and analysis from sequence alignments.** *Bioinformatics* 2005, **21**:260-262.
33. Idris AM, Brown JK: **Identification of a new, monopartite begomovirus associated with leaf curl disease of cotton in Gezira, Sudan.** *Plant Dis* 2000, **84**:809.
34. Rojas MR, Hagen C, Lucas WJ, Gilbertson RL: **Exploiting chinks in the plant's armor: evolution and emergence of geminiviruses.** *Annu Rev Phytopathol* 2005, **43**:361-394.
35. Zhou YC, Noussourou M, Kon T, Rojas M, Jiang H, Chen LF, Gamby K, Foster R, Gilbertson RL: **Evidence of local evolution of tomato-infecting begomovirus species in West Africa: characterization of tomato leaf curl Mali virus and tomato yellow leaf crumple virus from Mali.** *Arch Virol* 2008, **153**:693-706.
36. Fauquet CM, Briddon RW, Brown JK, Moriones E, Stanley J, Zerbini M, Zhou X: **Geminivirus strain demarcation and nomenclature.** *Arch Virol* 2008, **153**:783-821.
37. Briddon RW, Bull SE, Amin I, Idris AM, Mansoor S, Bedford ID, Dhawan P, Rishi N, Siwatch SS, Abdel-Salam AM, Brown JK, Zafar Y, Markham PG: **Diversity of DNA β : a satellite molecule associated with some monopartite begomoviruses.** *Virology* 2003, **312**:106-121.
38. Briddon RW, Brown JK, Moriones E, Stanley J, Zerbini M, Zhou X, Fauquet CM: **Recommendations for the classification and nomenclature of the DNA- β satellites of begomoviruses.** *Arch Virol* 2008, **153**:763-781.
39. Mansoor S, Zafar Y, Briddon RW: **Geminivirus disease complexes: the threat is spreading.** *Trends Plant Sci* 2006, **11**:209-212.
40. Briddon RW, Bull SE, Amin I, Mansoor S, Bedford ID, Rishi N, Siwatch SS, Zafar Y, Abdel-Salam AM, Markham PG: **Diversity of DNA 1: a satellite-like molecule associated with monopartite begomovirus-DNA beta complexes.** *Virology* 2004, **324**:462-474.
41. Lefevre P, Martin DP, Hoareau M, Naze F, Delatte H, Thierry M, Varsani A, Becker N, Reynaud B, Lett JM: **Begomovirus 'melting pot' in the south-west Indian Ocean islands: molecular diversity and evolution through recombination.** *J Gen Virol* 2007, **88**:3458-3468.
42. Argüello-Astorga GR, Ruiz-Medrano R: **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* 2001, **146**:1465-1485.

doi:10.1186/1743-422X-7-48

Cite this article as: Tiendrébéogo *et al.*: Molecular diversity of Cotton leaf curl Gezira virus isolates and their satellite DNAs associated with okra leaf curl disease in Burkina Faso. *Virology Journal* 2010 **7**:48.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

