

Article

Cotton Leafroll Dwarf Virus US Genomes Comprise Divergent Subpopulations and Harbor Extensive Variability

Roberto Ramos-Sobrinho ¹, Raphael O. Adegbola ¹, Kathy Lawrence ², Drew W. Schrimsher ², Thomas Isakeit ³, Olufemi J. Alabi ^{4,*} and Judith K. Brown ^{1,*}

¹ School of Plant Sciences, The University of Arizona, Tucson, AZ 85721, USA; ramosrs@email.arizona.edu (R.R.-S.); roadebola@arizona.edu (R.O.A.)

² Department of Entomology and Plant Pathology, Auburn University, Auburn, AL 36849, USA; lawrekk@auburn.edu (K.L.); drew.schrimsher@greenpointag.com (D.W.S.)

³ Department of Plant Pathology and Microbiology, Texas A&M University, College Station, TX 77843, USA; t-isakeit@tamu.edu

⁴ Department of Plant Pathology and Microbiology, Texas A&M AgriLife Research and Extension Center, Weslaco, TX 78596, USA

* Correspondence: olufemi.alabi@ag.tamu.edu (O.J.A.); jbrown@ag.arizona.edu (J.K.B.); Tel.: +1-520-621-1402 (J.K.B.)

Abstract: Cotton leafroll dwarf virus (CLRDV) was first reported in the United States (US) in 2017 from cotton plants in Alabama (AL) and has become widespread in cotton-growing states of the southern US. To investigate the genomic variability among CLRDV isolates in the US, complete genomes of the virus were obtained from infected cotton plants displaying mild to severe symptoms from AL, Florida, and Texas. Eight CLRDV genomes were determined, ranging in size from 5865 to 5867 bp, and shared highest nucleotide identity with other CLRDV isolates in the US, at 95.9–98.7%. Open reading frame (ORF) 0, encoding the P0 silencing suppressor, was the most variable gene, sharing 88.5–99.6% and 81.2–89.3% amino acid similarity with CLRDV isolates reported in cotton growing states in the US and in Argentina and Brazil in South America, respectively. Based on Bayesian analysis, the complete CLRDV genomes from cotton in the US formed a monophyletic group comprising three relatively divergent sister clades, whereas CLRDV genotypes from South America clustered as closely related sister-groups, separate from US isolates, patterns reminiscent of phylogeographical structuring. The CLRDV isolates exhibited a complex pattern of recombination, with most breakpoints evident in ORFs 2 and 3, and ORF5. Despite extensive nucleotide diversity among all available CLRDV genomes, purifying selection ($dN/dS < 1$) was implicated as the primary selective force acting on viral protein evolution.

Keywords: cotton viruses; genetic diversity; plant virus evolution; *Polerovirus*; *Solemoviridae*



Citation: Ramos-Sobrinho, R.; Adegbola, R.O.; Lawrence, K.; Schrimsher, D.W.; Isakeit, T.; Alabi, O.J.; Brown, J.K. Cotton Leafroll Dwarf Virus US Genomes Comprise Divergent Subpopulations and Harbor Extensive Variability. *Viruses* **2021**, *13*, 2230. <https://doi.org/10.3390/v13112230>

Academic Editor: K. Andrew White

Received: 28 September 2021

Accepted: 2 November 2021

Published: 5 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Cotton (*Gossypium* spp.) is an economically important fiber crop grown in over 80 countries worldwide, mainly for use in the textile industry [1–3]. In 2019, global production of cotton exceeded 82 million tons, harvested from a >38-million-hectare cultivation area, with an estimated value of \$74.4 billion dollars [4]. In the United States (US), 12 million tons of cotton were produced, representing a market value of \$5.8 billion US dollars [5].

Bacteria, fungi, nematodes, and viruses are important limiting factors to cotton production globally [6]. The cotton blue disease (CBD) was first described in Central African Republic (CAR) in 1949, where cotton plants exhibiting virus-like symptoms were associated with infestations of the cotton aphid, which was implicated as the vector of the suspected viral pathogen [7,8]. Virus-like diseases were also described in aphid-infested cotton-growing fields in Brazil in 1938 and 1962 [9,10], and in the Misiones Province of Argentina during 1982–1983 [11]. Based on the similar foliar symptoms and transmission by

the cotton aphid, the former diseases were hypothesized to share a similar undetermined virus-like etiology [12].

During 2003–2010, symptoms reminiscent of CBD in CAR were observed in commercially grown cotton in Brazil [12], East Timor [13,14], India [15], and Thailand [16]. Molecular characterization of partial and complete genome sequences from similarly symptomatic cotton plants revealed the presence of an aphid-transmitted polerovirus named cotton leafroll dwarf virus (CLRVD) [12,14,15,17,18]. Since the first report of CLRVD infecting cotton in Alabama, US, during 2016–2017 [19], CLRVD isolates have been reported from symptomatic cotton in Georgia, Mississippi, North Carolina, Texas, Arkansas, Florida, Louisiana, Oklahoma, and South Carolina [20–30]. Thus, CLRVD has become recognized as an emerging polerovirus [31,32]. Plants infected with CLRVD may develop shortened internodes, and thus the extent of overall stunting may range from extensive to none. Infected cotton leaves may develop reddening of leaf blades and petioles, curling and downward cupping, blue green, intense green, or red rust-coloration, vein-yellowing, or blistering, with yield loss attributed to reduced boll set [14,18,19,32,33]. Perhaps unexpectedly, CLRVD has been found to infect non-cotton hosts, including several uncultivated plant species [34].

Cotton leafroll dwarf virus belongs to the genus *Polerovirus* (family *Solemoviridae*). The virus has a linear, positive-sense, single-stranded, monopartite RNA genome of ~5.8 kb in size, encapsidated in a spherical virion of approximately 23 nm in diameter [35,36]. As for other poleroviruses, CLRVD is phloem limited. To date, the only insect vector identified is the cotton aphid (*Aphis gossypii* Glover), which transmits the virus in a persistent, circulative, non-propagative manner [8,37,38]. The CLRVD genome organization resembles that of other poleroviruses [12,17,32,39], consisting of seven open reading frames (ORFs) partitioned into two regions separated by a 200-nucleotide (nt) intergenic region (IR) that functions in the synthesis of sub-genomic RNAs [40–42]. The 5'-proximal ORF0 encodes a 28.9 kDa RNA silencing suppressor protein (P0), which is considered to function as an avirulence (AVr) determinant [40,43–45]. ORF1 encodes a 70.1 kDa protein (P1) expressed from genomic RNA [46] and ORF1-2 encodes the 118.7 kDa viral replication-associated protein P1-P2, including the RNA-dependent RNA polymerase (RdRp), which is expressed through a ribosomal frameshift [42,46]. The 3'-end comprises structural genes associated with viral encapsidation (ORF3), long-distance movement (ORF3a), cell-to-cell movement (ORF4), and aphid transmission (ORF5), all of which are expressed as subgenomic RNAs [46]. The viral ORF3 encodes the 22.4 kDa coat protein (CP; P3), while ORF4, which is nested within ORF3, encodes the 19.4 kDa movement protein (MP; P4) and is expressed as a frameshift. The ORF3a that encodes the P3a protein (5.2 kDa) is expressed through leaky scanning of the subgenomic RNA. The ORF5 is expressed through in-frame suppression of the ORF3 stop codon to yield a 77.2 kDa P3-P5 read-through domain (RTD). The CP-RTD is required for aphid transmission and in planta viral accumulation [38,47,48]. Finally, the 5'-terminal end is covalently bound to a genome-linked viral protein (VPg), but no poly(A)-tail or tRNA-like structures have been associated with the viral 3'-terminus [49,50].

The rapid genomic evolution observed among RNA plant viruses has been attributable to high rates of mutation, short generation times, and large population sizes [51,52]. The RdRp protein utilized for virus replication lacks proof-reading activity that can potentially result in a high mutation rate [53] and lead to recombination between viral genomes [54]. These evolutionary features of RNA viruses have been associated with enhanced virulence, differences in infectivity, transmission rate, symptom severity, and host range, including host-shifts [55,56].

The P0 silencing suppressor, encoded by the CLRVD ORF0, is the most variable coding region at both nucleotide and amino acid levels [32,40,57–59]. Despite the recent increased spread and emergence of CLRVD disease and its detrimental effects on cotton productivity worldwide, there is little understanding of the genetic structure of CLRVD populations. Increased knowledge of genomic diversity and population structure of extant

CLRVD isolates associated with cotton and non-cultivated host plants would help inform the selection of resistant cotton varieties while also improving the reliability of molecular detection and identification of variants required to elucidate the epidemiology locally and globally.

The objective of this study was to characterize the full-length genome of CLRVD isolates associated with cotton plants exhibiting disease symptoms observed in three US states affected by the 2017–2019 CLRVD outbreak. Here, eight new CLRVD genome sequences were determined by Sanger DNA sequencing and/or by high throughput sequencing (HTS) using Illumina technology. Analyses of the phylogenetic relationship, extent of recombination, and population structure were carried out for the 8 genomes and all available CLRVD sequences in the GenBank database. The goal was to inform local (US) and global perspectives of the dynamics surrounding the emergence of CLRVD and epidemiological implications.

2. Materials and Methods

2.1. The Plant Samples

Leaf samples were collected from cotton plants exhibiting mild to severe virus-like symptoms in Alabama (AL), Florida (FL), and Texas (TX) during 2019. The symptoms observed in some of the plants were similar to (while others were distinct from) those observed in CLRVD-infected cotton plants from AL during 2016–2017, an isolate referred to as CLRVD-AL [19,32]. The leaves were frozen immediately in liquid nitrogen and shipped by courier on dry ice (USDA-APHIS permit issued to the JK Brown Lab), to The University of Arizona, and stored at -80°C .

2.2. Total RNA Isolation and RT-PCR Diagnostics

Total RNA was isolated from 100 mg of cotton leaves and petioles using a silica protocol adapted from [60] or with the Spectrum™ Plant Total RNA Kit (Sigma-Aldrich, St. Louis, MO, USA). The RNA was used as a template for reverse transcription polymerase chain reaction (RT-PCR) with primers designed to target the CLRVD ORF0/P0 or ORF4/P4 genomic regions [14,32]. The expected size amplicons were cloned into the pGEM-T easy plasmid vector (Promega Corp. Madison, WI, USA) used to transform competent cells of *Escherichia coli* DH5 α and sent for Sanger DNA sequencing at Eton Biosciences (San Diego, CA, USA).

2.3. Determination of CLRVD Genomes

To determine the complete CLRVD genome sequences associated with mild to severe symptoms, total RNA was purified from eight representative cotton samples (Table 1), and genomic DNA was removed by treatment with RNase-free DNase I (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. The CLRVD genome sequence was obtained from four of the eight samples using high throughput sequencing (HTS) while the remaining four were determined by Sanger sequencing. For HTS, the cDNA library was constructed from total RNA following ribosomal RNA depletion and samples were subjected to Illumina RNAseq sequencing at Novogene Co. (Beijing, China). Removal of adapter sequences was carried out at Novogene Co., and low-quality bases were removed based on a sliding window size of 4 and Q score < 20 . The de novo assembly of quality-trimmed raw Illumina reads was carried out using SeqMan NGen v.12 software (DNASTAR Inc. Madison, WI, USA) with a kmer parameter of 21, and maximum of 2 mismatches. To assess the assembled contig sequence quality, trimmed reads were mapped against the apparently full-length genomes of CLRVD using the Bowtie2 software [61] implemented in Geneious v.8.1.9 (Biomatters, San Diego, CA, USA). The assembled contigs were subjected to a BLASTn [62] search of the GenBank nucleotide database to establish preliminary virus identification. For Sanger sequencing, cDNA was synthesized with random hexamers using the PrimeScript 1st strand cDNA Synthesis Kit (Takara Bio USA Inc., Mountain View, CA, USA). A 2- μL cDNA aliquot was used in 25 μL PCR reaction with PrimeSTAR

GXL DNA Polymerase (Takara Bio), as per manufacturer's recommendation, to obtain the near complete CLRDV genome as a single DNA fragment with a primer pair designed based on aligned sequences of available CLRDV genomes in GenBank (Table S1). The obtained ~5.8 Kb sample-specific DNA bands were gel eluted with the Zymoclean Gel DNA Recovery Kit (Zymo Research Corporation, Irvine, CA, USA) and blunt-end cloned individually into the pJET1.2 plasmid vector (ThermoFisher Scientific, Waltham, MA, USA). Plasmid DNA was obtained from two positive recombinant clones per insert using the GenElute Five-Minute Plasmid Miniprep Kit (Sigma-Aldrich, St. Louis, MO, USA), as per manufacturer's instructions, and these were Sanger sequenced and genome-walked until completion with additional genome-walking primers listed in Table S1. The overlapping genome sequence fragments were assembled with BioEdit v.7.2.5 program [63].

Table 1. Cotton leafroll dwarf virus (CLRDV) isolates recovered from symptomatic cotton plants using high throughput sequencing or Sanger sequencing.

Isolate	Location	Symptoms	PCR-ORF0	# of Reads	Coverage	Reference
CLRDV-USA-AL-MC2 ¹	Alabama	Severe	Positive	33,141	850	This study
CLRDV-USA-AL-SC4 ¹	Alabama	Mild	Positive	16,964	435	This study
CLRDV-USA-TX-CT2	Texas	Severe	Positive	13,043	336	This study
CLRDV-USA-TX-CT3 ¹	Texas	Severe	Positive	147,639	3,780	This study
CLRDV-USA-TXa	Texas	Severe	Positive	Sanger-derived		This study
CLRDV-USA-TXb	Texas	Severe	Positive	Sanger-derived		This study
CLRDV-USA-TXc	Texas	Severe	Positive	Sanger-derived		This study
CLRDV-USA-TXd	Texas	Severe	Positive	Sanger-derived		This study
CLRDV-USA-AL ¹	Alabama	Severe	Positive	-	-	Avelar et al., 2020

¹ Representative CLRDV isolates for which viral 5' and 3' UTRs were determined.

2.4. Sequencing of CLRDV 5'- and 3'-Untranslated Ends

To recover the 5'-untranslated region (UTR) of representative CLRDV isolates, cDNA synthesis was carried out for total RNA preparations using the reverse primer GSP1-CLRDV-5UTR-Rev (Table S1), according to the manufacturer's protocol for the SuperScript IV First-Strand Synthesis System (Invitrogen, Carlsbad, CA, USA). Aliquots of cDNA were used as template for PCR amplification of CLRDV 5'-UTR with the primer pair CLRDV-P20-For/GSP2-CLRDV-5UTR-Rev (Table S1). Additionally, to recover the CLRDV 3'-UTR, total RNA was poly(A) tailed using the Poly(A) Tailing kit (Invitrogen, Carlsbad, CA, USA), and cDNA synthesis was carried out using oligo(dT) primer according to the manufacturer's instructions accompanying the 3' RACE kit (Invitrogen, Carlsbad, CA, USA). The cDNA was used as a template for PCR amplification of CLRDV 3'-UTR with primers GSP2-CLRDV-3UTR-For/anchor (3' RACE kit) (Table S1).

The PCR amplification reactions were carried out using LongAmp Hot Start *Taq* 2X Master Mix (New England Biolabs, Ipswich, MA, USA), in a final volume of 25 µL: 12.5 µL of 2X LongAmp master mix, 0.2 µM of each primer, 1 µL of cDNA (template), and 10.5 µL of nuclease-free water. Cycling parameters were: initial denaturation for 2 min at 94 °C, followed by 35 cycles of denaturation at 94 °C for 20 s, annealing at 55 °C for 60 s, and extension at 65 °C for 30 s, and a final extension at 65 °C for 10 min. The amplicons were gel-purified and ligated to pGEM-T easy plasmid vector (Promega Corp. Madison, WI, USA) and transformed into competent cells of *E. coli* DH5α. Clones harboring inserts of the expected size were confirmed by colony PCR using M13 universal primers, and two or more clones per sample were sequenced bidirectionally (Eton Biosciences, San Diego, CA, USA).

2.5. Sequence Comparisons

Pairwise nucleotide sequence comparisons of the CLRDV complete genomes determined in this study and CLRDV sequences available in the GenBank database (Table S2), and CLRDV ORFs 0–5 were carried out using the Sequence Demarcation Tool (SDT) v.1.2 [64].

The pairwise amino acid (aa) similarity for the P0–P5 predicted protein sequences was calculated using the SDT software [64]. The species identification criterion of >10% divergence at the aa level for any poliovirus protein has been established for classification into the genus *Poliovirus* [36]. However, for CLRDV variants specifically, it has been proposed that isolates sharing <90% amino acid identity for ORF0, and >90% amino acid identity across the remaining proteins, represent different genotypes or variants of this species [32,58,59].

2.6. Phylogenetic Analysis

The CLRDV complete genome and ORF0 nucleotide (nt) sequences were aligned with the CLRDV sequences available in the GenBank database using MUSCLE [65]. The Bayesian phylogenetic trees were reconstructed using MrBayes v.3.2 [66] through the CIPRES web portal [67], and the GTR+I+G evolutionary model-of-best-fit. The analyses were carried out using two replicates of four chains each, and 10 million generations with sampling every 1000 generations. The first 2500 trees were discarded (burn-in), and the posterior probabilities [68] were determined based on a majority-rule consensus tree for the 15,000 remaining trees. Trees were edited in FigTree v.1.4 (tree.bio.ed.ac.uk/software/figtree) and Inkscape (<https://inkscape.org/pt/>). Furthermore, subpopulation structuring was tested by calculating Wright's *F* fixation index (*Fst*) [69] using DnaSP v.6.10 [70]. The phylogenetic trees were midpoint-rooted because it was not possible to identify a poliovirus outgroup species that did not significantly confound the alignment and lead to a poorly resolved phylogeny.

2.7. Recombination Analysis

The eight newly determined CLRDV genomes were aligned with isolates previously reported in the US, and CLRDV-typical and -atypical sequences from South America. Non-treelike evolution analysis was carried out using the Neighbor-Net method implemented in SplitsTree v.4.10 [71]. The putative parental sequences and recombination breakpoints were predicted using the methods RDP, Geneconv, Bootscan, MAXCHI, Chimaera, SISCAN and 3Seq implemented in the RDP v.4.0 program [72]. The default settings were implemented for each different method, except that the sequences were considered linear instead of circular. The statistical significance threshold was set as a *P*-value lower than a Bonferroni-corrected cut-off of 0.05. Recombination events detected by at least five of the seven algorithms available in the RDP software with robust statistical support were accepted as reliable predictions.

2.8. Genetic Variability and Selection

The mean pairwise number of nucleotide differences per site (nucleotide diversity, π) was estimated using a 100-nt sliding window with a step size of 10 nt across the CLRDV genome subpopulation sequences using DnaSP v.6.10 [70]. Statistically significant differences amongst π values for each dataset/subpopulation were calculated by estimating the 95% bootstrap confidence intervals from 1000 nonparametric simulations using the Simpleboot package implemented in R (R Core Team) [73]. Additionally, the per-site nucleotide diversity was calculated for ORFs 0–5 sequences determined for each CLRDV subpopulation. The fixation index or F_{ST} and N_{ST} test statistics were calculated using DnaSP v.6.10 to compare the extent of genetic differentiation between populations/subpopulations and to characterize the biogeographical population structure, with potential F_{ST} and N_{ST} values ranging from 0 (no genetic differentiation) to 1 (complete differentiation) [69].

The aa sites evolving under positive or negative selection in ORFs 0–5 were predicted using the Fixed-Effect Likelihood (FEL) algorithm [74] implemented in DataMonkey (www.datamonkey.org; accessed on 15 July 2021). To reduce the likelihood of spurious selection estimates caused by recombination, subpopulation datasets were constructed for each recombination pattern predicted by both RDP4 and SplitsTree4, consisting of only non-recombinant sequences. The FEL *P*-values of <0.1 were considered significant. The mean ratio of non-synonymous to synonymous substitutions (dN/dS) was calculated using

the Single-Likelihood Ancestor Counting (SLAC) method [74], where $dN/dS = 1$ predicts neutral evolution (no selection), and $dN/dS > 1$ or $dN/dS < 1$ indicates positive or negative selection, respectively [75].

3. Results

3.1. CLRDV Complete Genome Sequences

The eight apparently full-length newly determined CLRDV genome sequences were de novo and guided-assembled or Sanger-sequenced from representative samples displaying mild to severe symptoms. The Illumina sequences of 5865 to 5867 bp in length were derived from 13,043 to 147,639 reads from four isolates, having a depth of coverage of 336 to 3780 times (Table 1). The Sanger-derived sequences from the other four isolates were each 5579-nt in length, putatively missing 27-nt 5'-UTR and 79-nt 3'-UTR sequences. The 5'- and 3'-UTRs of isolates CLRDV-AL [32], CLRDV-USA-AL-MC2, CLRDV-USA-FL-SC4, and CLRDV-USA-TX-CT3 were amplified by RT-PCR amplification (RACE), cloned, and Sanger sequenced, revealing fragments of 480 and 335 bp in length, respectively. The Sanger and Illumina-derived 5'- and 3'-UTRs shared >99.0% nt identity, indicating the two approaches used for sequencing and assembling the genome sequence data produced nearly identical and/or highly comparable genome sequences.

3.2. Pairwise Sequence Comparisons

The novel CLRDV complete genomes ($n = 8$) shared 96.5–100.0% nt identity with one another, and 94.5–8.1% nt identity with the next most closely related isolates for which sequences were available in GenBank, i.e., the CLRDV-AL variants previously reported in the US. The ORF0 has been recognized as the most variable coding region among the CLRDV isolates, with the new isolates sharing 93.3–100.0% nt identity among each other, and 89.4–99.7% identity with isolates from GenBank. Compared to the CLRDV-typical and -atypical genotypes reported from South America, the ORF0 nt sequences of the US isolates were 89.4–92.1% identical. As predicted, the P0 aa sequences exhibited the highest molecular variability among all of the viral coding regions, with the US CLRDV isolates sharing 90.4–100.0% aa similarity among each other, and 81.2–89.3% aa similarity with South American isolates. The three isolates collected in TX [CLRDV-USA-TX-CT3, CLRDV-USA-TXa, and CLRDV-USA-TXb (GenBank Accession Nos. OK185942-OK185944)] shared the highest aa similarity with CLRDV isolates previously identified in cotton samples from TX (Accession MN872302), at 95.0–96.2%, whereas five isolates from TX, FL, and AL [CLRDV-USA-TXc, CLRDV-USA-TXd, CLRDV-USA-TX-CT2, CLRDV-USA-AL-MC2, and CLRDV-USA-FL-SC4 (GenBank Accession Nos. OK185945-OK185946, OK185941, OK185939, and OK185940)] were more closely related to the CLRDV-AL genotype previously reported from AL and Georgia (Accession Nos. MN071395 and MT800932), at 95.0–99.6% aa similarity. For ORFs 2–5, the nt and aa sequence pairwise distance estimates exceeded 92.0% and 90.0% identity, respectively, except for isolate CLRDV-GA40 from Georgia (GA), which showed aa similarities between 87.5–91.5 and 87.8–88.5% for ORF2, compared to CLRDV-US and CLRDV-SA isolates, respectively (Table 2). Thus, the newly characterized genomes most closely resembled the first CLRDV genome sequences discovered in the US, the CLRDV-AL genotypes [18,32].

Table 2. Nucleotide and amino acid identities for the complete genome and open reading frame (ORF) sequences of cotton leafroll dwarf virus (CLRDV) isolates from the US and South America (SA).

Genomes—Nucleotide/Amino Acid			ORF0—Nucleotide/Amino Acid		
	CLRDV-SA	CLRDV-US		CLRDV-SA	CLRDV-US
CLRDV-SA	95.9–99.5/- ¹		CLRDV-SA	91.1–99.5/87.0–98.1	
CLRDV-US	94.0–96.2/-	94.5–100.0/-	CLRDV-US	89.4–94.3/81.6–90.4	92.1–100.0/90.8–100.0
ORF1—Nucleotide/Amino Acid			ORF1-ORF2—Nucleotide/Amino Acid		
	CLRDV-SA	CLRDV-US		CLRDV-SA	CLRDV-US
CLRDV-SA	94.3–99.5/93.8–99.5		CLRDV-SA	95.5–99.4/93.3–99.1	
CLRDV-US	92.4–95.5/90.3–95.0	93.8–100.0/91.6–100.0	CLRDV-US	92.9–96.2/87.8–94.7	92.5–100.0/87.5–100.0
ORF3—Nucleotide/Amino Acid			ORF3a—nucleotide/amino acid		
	CLRDV-SA	CLRDV-US		CLRDV-SA	CLRDV-US
CLRDV-SA	97.9–100.0/98.5–100.0		CLRDV-SA	97.8–100.0/100.0	
CLRDV-US	95.0–98.3/95.5–100.0	96.2–100.0/95.5–100.0	CLRDV-US	97.1–100.0/100.0	97.8–100.0/100.0
ORF4—Nucleotide/Amino Acid			ORF5—Nucleotide/Amino Acid		
	CLRDV-SA	CLRDV-US		CLRDV-SA	CLRDV-US
CLRDV-SA	98.1–100.0/95.4–100.0		CLRDV-SA	96.7–99.7/96.8–100.0	
CLRDV-US	94.9–98.9/89.7–97.1	96.0–100.0/92.5–100.0	CLRDV-US	94.8–96.8/94.1–98.3	96.5–100.0/95.7–100.0

¹ No amino acid sequence comparisons were carried out for CLRDV complete genomes.

3.3. Phylogenetic Relationship

Bayesian phylogenetic trees were reconstructed from the apparently full-length genome and ORF0 nt sequences of the new CLRDV isolates ($n = 8$) and sequences previously reported in the US ($n = 9$) and South America ($n = 6$). The CLRDV-US isolates formed a monophyletic group consisting of three sister clades. The subclade 1-US harbored sequences from TX, AL, and FL ($n = 6$), subclade 2-US isolates found only in TX ($n = 4$), while the subclade 3-US was comprised of sequences in GA and AL ($n = 7$) (Figure 1a). Furthermore, the CLRDV sequences from the US (CLRDV-AL) and South America (CLRDV-typical and -atypical) formed distinct phylogenetic groups, reinforcing the hypothesis that extant CLRDV-US genotypes or subpopulations are divergent and geographically structured in relation to other isolates recognized so far. The CLRDV-typical and -atypical variants grouped into two subclades (Figure 1a). The CLRDV ORF0 and complete genome trees were incongruent, with some isolates from GA being more closely related to CLRDV reported in SA (Figure 1b).

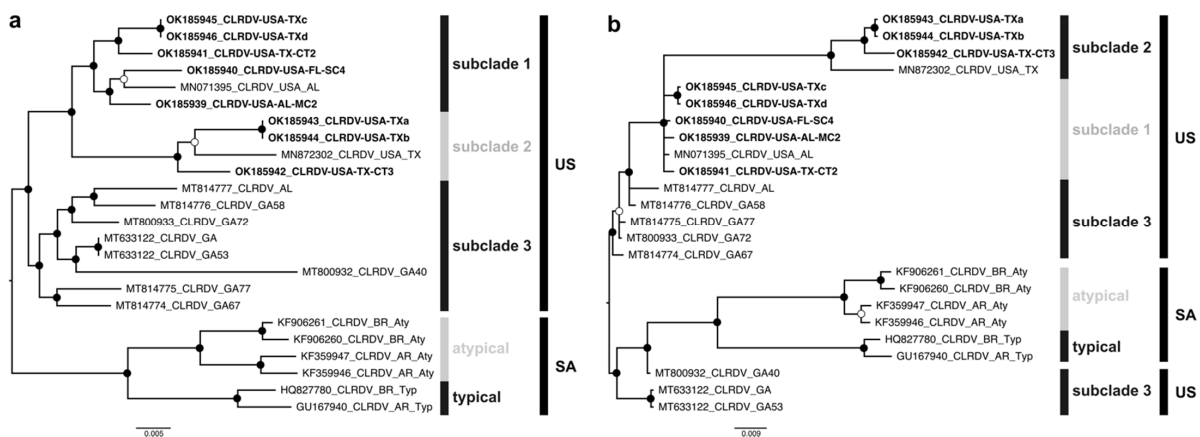


Figure 1. Midpoint-rooted Bayesian phylogenetic trees reconstructed from complete genomes (a) and ORF0 nucleotide sequences (b) of cotton leafroll dwarf virus (CLRDV). Posterior probability values are represented by filled (0.95–1.00) and empty (0.80–0.94) circles near the branch nodes. Sequences determined in this study are highlighted in bold font. 3.4. Recombination.

Reticulate (non-treelike) branching patterns were detected for the CLRDV complete genomes using the Neighbor-Net distance-based method. Several parallel/boxes paths were observed in the network, indicating conflicting signals probably caused by recombination events. The reticulation was pronounced for the different subclades from the US (subclades 1–3) and South America (subclades typical and atypical), suggesting that the evolution of these CLRDV sequences may have been shaped by genetic recombination. Additionally, the long branches associated with the CLRDV isolates indicated that these genomes harbored more mutations than isolates grouping in clades having shorter branch lengths (Figure 2a).

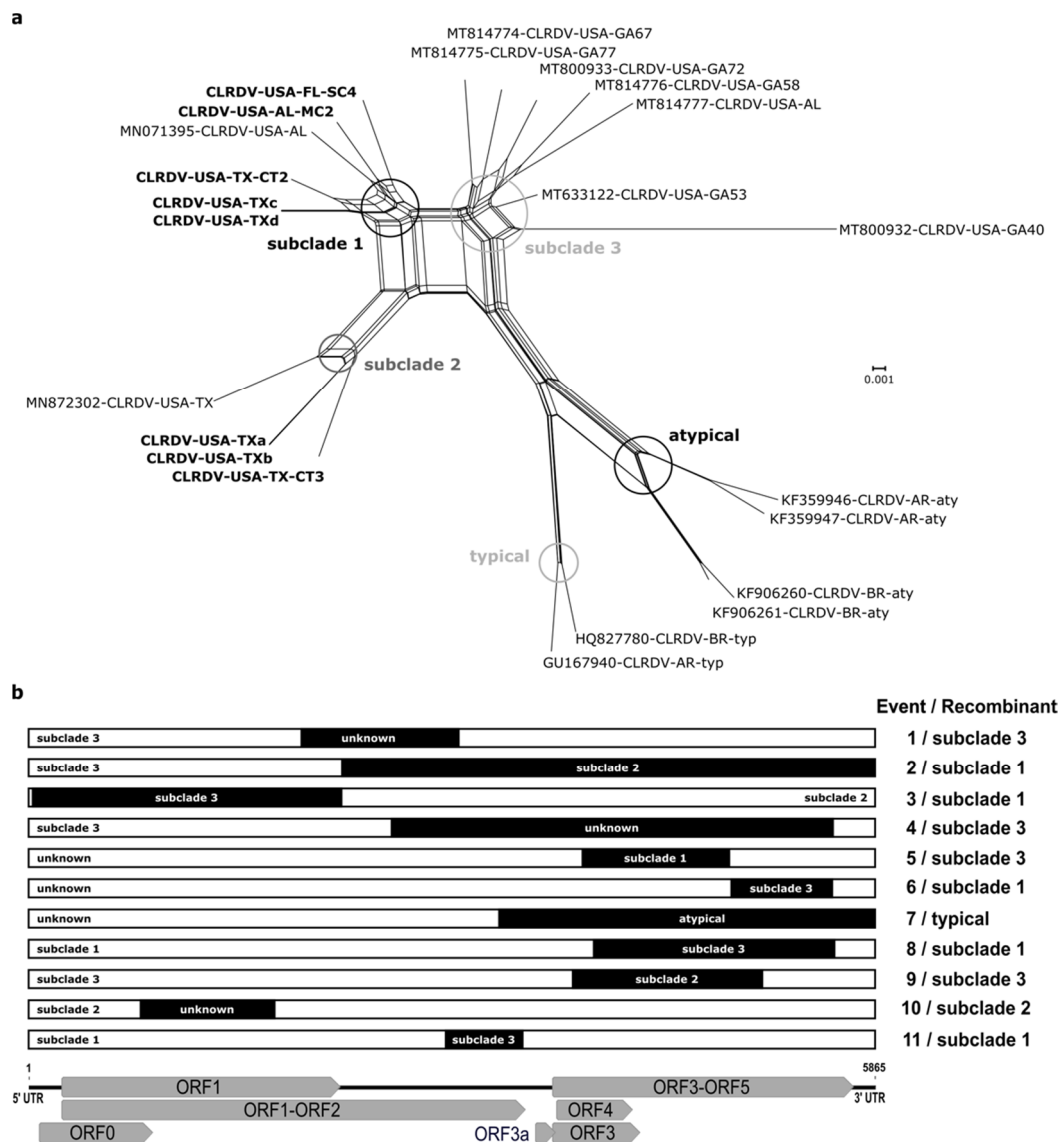


Figure 2. Reticulate phylogenetic network constructed using the Neighbor-Net method (a) and recombination events detected in RDP4 (b) based on analysis of aligned complete genomes of cotton leafroll dwarf virus (CLRDV). (a) Reticulation among the viral isolates is shown by parallel paths instead of a bifurcating evolutionary tree indicative of putative recombination. The branch internodes for the phylogenetic subclades are circled. (b) The regions highlighted in black correspond to the portion donated by the predicted minor parent, while the remaining portions represent the major parent of each independent recombination event. The CLRDV genome organization is shown in relation to the aligned sequences.

To predict possible recombination among CLRDV isolates, and identify the respective breakpoints and parents, all available CLRDV complete genomes were aligned and analyzed using the RDP software package. At least 11 independent recombination events

were predicted among the CLRDV isolates, with most recombination breakpoints having occurred in the second half of ORF2 and ORF5, and in the first half of ORF3 (events 1–9 and 11) (Figure 2b; Table S3). The CLRDV-GA isolates (subclade 3-US) were identified as putative recombinants in three events (1, 4, and 9) and as possible parental sequences in eight recombination events (1–4, 6, 8, 9, and 11). Further, CLRDV isolates in subclade 1-US were identified as recombinants, with sequences of isolates from TX (subclade 2-US) or GA (subclade 3-US) implicated as possible minor and major parents (events 2, 3, 6, 8, and 11). The CLRDV-typical isolates from Brazil (BR) and Argentina (AR) were also identified as putative recombinants, with CLRDV-atypical (BR) as the predicted minor parent, but the major parent was unidentified (Figure 2b; Table S3).

3.4. Population Genetics

The mean pairwise number of nucleotide differences per site (nucleotide diversity, π) was estimated for the CLRDV complete genomes using DnaSP v.6. The CLRDV sequences were grouped as seven subpopulations, representing isolates extant in South America (SA; $n = 6$), typical-SA ($n = 2$), atypical-SA ($n = 4$) and the US ($n = 17$), subclade 1-US ($n = 6$), subclade 2-US ($n = 4$), and subclade 3-US ($n = 7$). Based on F_{st} and N_{st} values, extensive population structure was evident between most of the subpopulations, indicative of genetic divergence. In contrast, relatively low F_{st} and N_{st} indices were obtained when subclades 1-US and 3-US were compared with the entire US subpopulation, at 0.07 and 0.06, respectively (Table 3). Finally, a comparison of atypical-SA with SA subpopulations revealed negative F_{st} and N_{st} values of -0.07 (Table 3).

Table 3. Genetic structure among cotton leafroll dwarf virus (CLRDV) subpopulations.

Subpopulations		N_{st} ¹	F_{st} ¹
subclade 1-US	subclade 2-US	0.52	0.51
subclade 1-US	Subclade 3-US	0.31	0.31
subclade 1-US	US	0.07	0.07
subclade 1-US	typical-SA	0.72	0.72
subclade 1-US	atypical-SA	0.67	0.67
subclade 1-US	SA	0.56	0.56
subclade 2-US	subclade 3-US	0.51	0.50
subclade 2-US	US	0.27	0.27
subclade 2-US	typical-SA	0.71	0.71
subclade 2-US	atypical-SA	0.66	0.66
subclade 2-US	SA	0.56	0.55
subclade 3-US	US	0.06	0.06
subclade 3-US	typical-SA	0.62	0.62
subclade 3-US	atypical-SA	0.56	0.55
subclade 3-US	SA	0.45	0.45
US	typical-SA	0.58	0.57
US	atypical-SA	0.53	0.52
US	SA	0.42	0.41
typical-SA	atypical-SA	0.60	0.60
typical-SA	SA	0.26	0.26
atypical-SA	SA	-0.07	-0.07

¹ The F_{ST} and N_{ST} values range from 0 for no genetic differentiation, to 1 for complete differentiation between populations.

Statistically significant differences for pairwise comparisons of π values calculated from full-length genomes of CLRDV subpopulations of different sample sizes were analyzed with a 95% bootstrap confidence interval (CIs) using the approach described by [73], where no statistically significant differences were observed for CI values including zero (grey line) (Figure 3). The nucleotide diversity analysis of the complete CLRDV genomes revealed that genetic variability was unevenly distributed across the viral genomes (Figure 4). The lowest nucleotide diversity for isolates affiliated with the subclades 1-US and 3-US resided within the 5'-portion of the genome, which encodes ORF0, whereas ORF0 was

the most variable coding region among all of the CLRDV-SA ($n = 6$) and CLRDV-US ($n = 17$) populations (Figure 4). The second half of ORFs 2 and 5, and the first half of ORF3 harbored the greatest nucleotide diversity among subclade 3-US isolates (Figure 4). Finally, the diversity was highest among US and SA populations, compared to individual subpopulations, an observation indicative of interpopulation variation. Although CLRDV genome sequences from South America were under-represented among the total sequences available, the genomic variability among the South American ($n = 6$) and US ($n = 17$) CLRDV isolates was comparable ($\pi_{SA} = 0.02867$ and $\pi_{US} = 0.02831$; Figures 3 and 4).

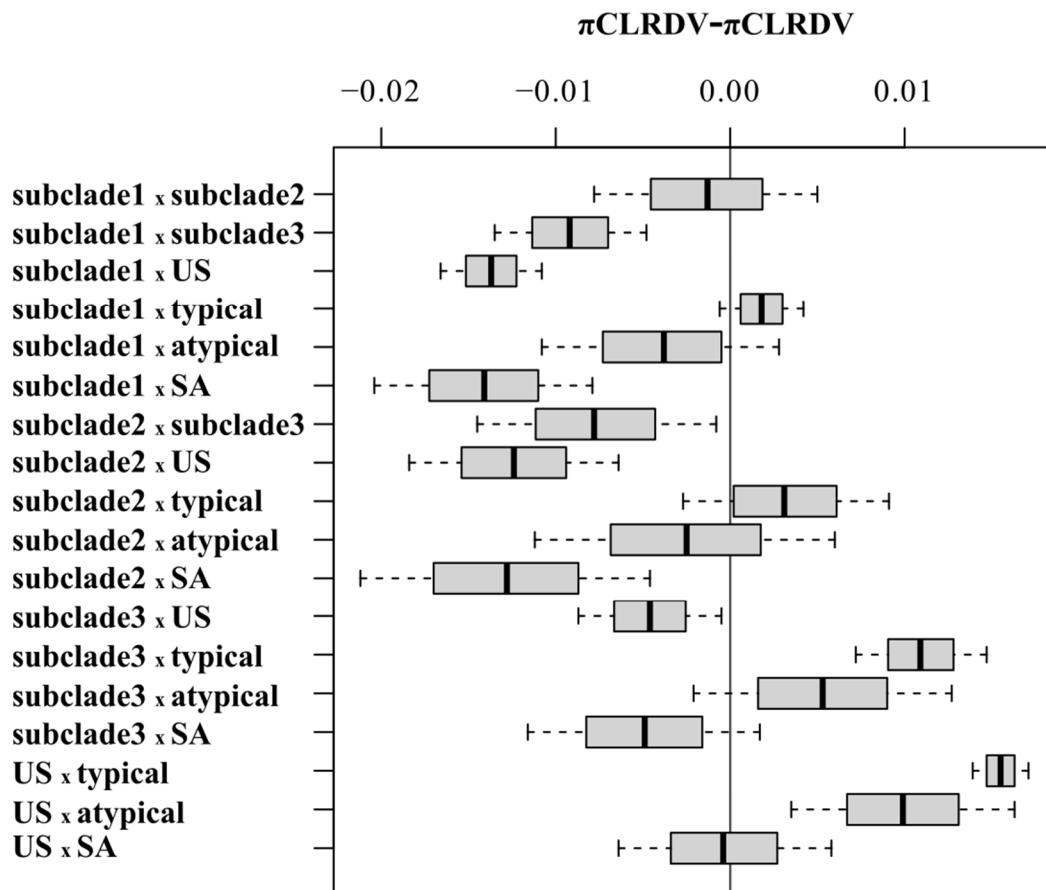


Figure 3. The statistical significance of the differences amongst the mean pairwise number of nucleotide differences per site (π), as evaluated by 95% bootstrap confidence intervals (CIs) calculated using 1000 nonparametric simulations. No statistically significant differences observed for CI values including zero (grey line) for pairwise comparisons based on the complete genome of cotton leafroll dwarf virus (CLR DV) subpopulations.

The potential recombination events among different CLR DV genomes were predicted using RDP4 and SplitsTree4 (Figure 2), and selective forces acting on CLR DV evolution were estimated based on subpopulation datasets consisting of only nonrecombinant sequences. Using the SLAC method, the dN/dS mean ratios for ORFs 0–3 and 5 were < 1 for nearly all datasets, which is indicative of purifying selection. The exception was ORF0 of the atypical-SA, with a dN/dS ratio of 1.63, indicative of positive selection. In contrast, the dN/dS values for ORF4 among all of the CLR DV isolates range from 0.98–1.26, revealing widespread positive selection on this viral ORF (Table 4). Further, most statistically significant aa sites identified in ORFs 0–5 were shown to be evolving under purifying selection. Only a small number of statistically significant aa sites exhibited positive selection, and these were distributed throughout ORFs 2–5 (Table 4). These results indicated that negative selection was the most important selective force acting on viral coding regions.

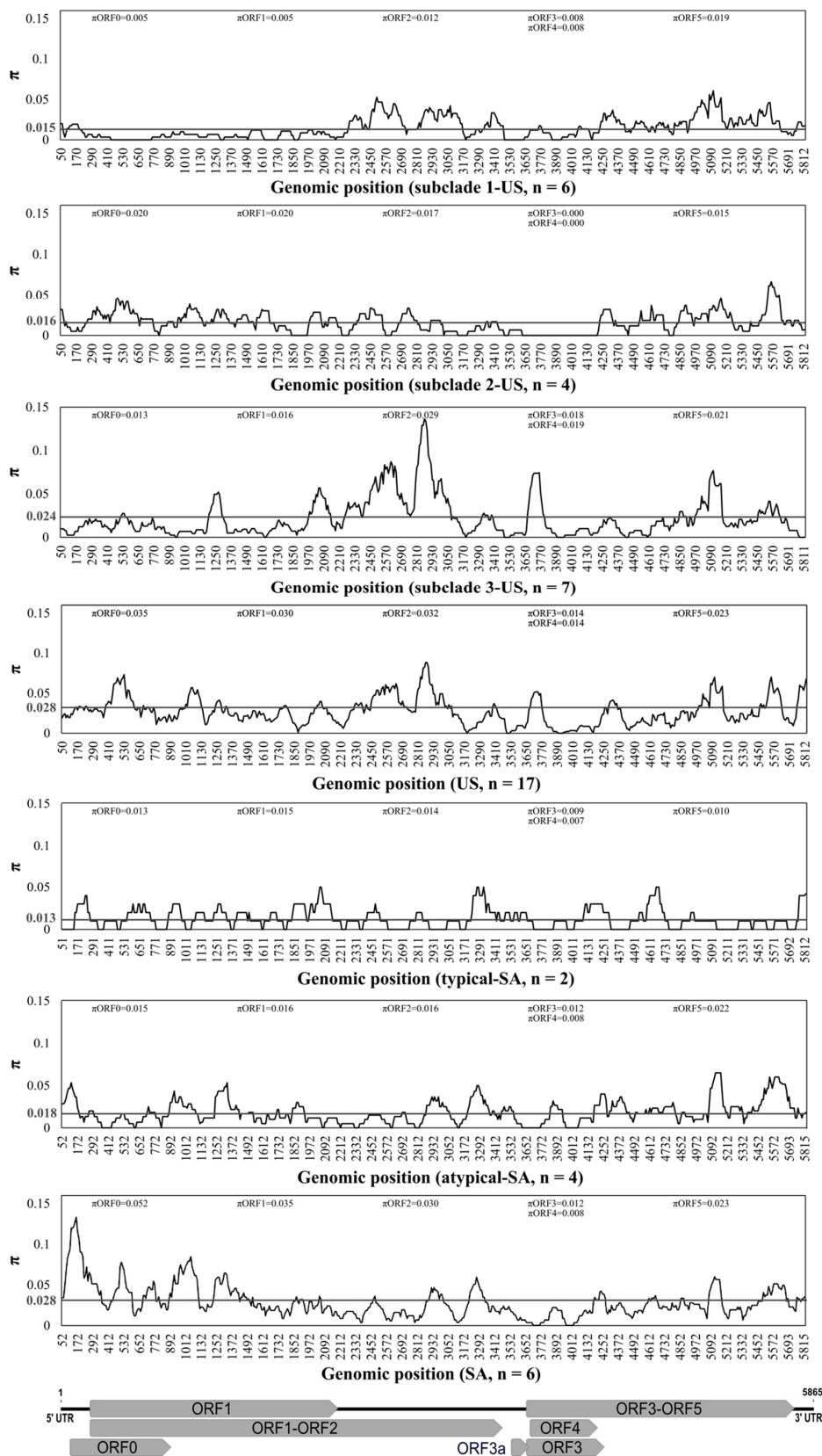


Figure 4. Nucleotide diversity across the complete genomes of cotton leafroll dwarf virus (CLRDV) isolates, calculated using DnaSP6 on a 100-nucleotide sliding window with a step size of 10 nucleotides. The horizontal lines (in grey) correspond to the genome-wide average nucleotide diversity for the full-length sequences within each subpopulation. The CLRDV genome organization is shown in relation to the alignment.

Table 4. Nonsynonymous to synonymous substitution rates (dN/dS) and number of positively and negatively selected sites in the ORFs 0–5 of cotton leafroll dwarf virus (CLRVDV) subpopulations using Single-Likelihood Ancestor Counting (SLAC) and Fixed-Effect Likelihood (FEL), respectively.

Dataset	Gene	SLAC (dN/dS)	FEL (# of Sites)	
			Negative	Positive
CLRVDV-subclade1-US	ORF0	0.52	2	0
	ORF1	0.36	3	0
	ORF2	0.15	33	0
	ORF3	0.16	3	0
	ORF4	1.19	1	0
	ORF5	0.19	30	0
CLRVDV-subclade2-US	ORF0	0.36	8	0
	ORF1	0.28	27	0
	ORF2	0.19	54	1
	ORF3	– ¹	–	–
	ORF4	–	–	–
	ORF5	0.24	20	0
CLRVDV-subclade3-US	ORF0	0.67	3	0
	ORF1	0.55	17	0
	ORF2	0.47	54	3
	ORF3	0.50	6	0
	ORF4	0.98	1	1
	ORF5	0.38	29	0
CLRVDV-atypical-SA ²	ORF0	1.63	1	0
	ORF1	0.33	13	0
	ORF2	0.25	32	0
	ORF3	0.16	5	1
	ORF4	1.26	1	0
	ORF5	0.15	16	1

¹ Analysis was not performed because the dataset was comprised by one only haplotype. ² The selection analysis was not carried out for the CLRVDV-typical-SA isolates because at least three different haplotypes are required.

4. Discussion

Cotton is one of the most economically important crops worldwide, and significant yield losses caused by cotton leafroll dwarf virus (the causal agent of cotton blue- and cotton blue-like diseases) have been reported in Brazil and Argentina [9,10,12,58,59], East Timor [13,14], India [15], Thailand [16], and, most recently, in the US [19,32]. Since 2016–2017, CLRVDV has emerged and spread quickly from those sites in cotton-growing states in the southern and south-central US [19–30,32]. Despite the potential for economic importance of CLRVDV, little is known about CLRVDV genome-level variability and the forces that may be acting upon it. In this study, the high nucleotide diversity documented among CLRVDV isolates was primarily influenced by mutation and recombination, leading to geographically structured viral populations, suggestive of admixture through the introduction of multiple variants and/or multiple introductions in or near the same foci, potentially followed by short and/or long-distance spread by the aphid vector.

Although cotton plants exhibiting CBD-like symptoms have been recognized in several cotton-growing areas of the world since the 1930's [7–10], the etiological agent was identified in 2005 when partial sequences of the polerovirus CLRVDV were associated with symptomatic cotton in Brazil [12]. The first CLRVDV complete genome sequence was reported five years later [17]. Most recently, several genomes of a new CLRVDV variant, CLRVDV-AL, were reported in the US [21,25,32,57]. In this study, eight CLRVDV full-length genome sequences were determined by HTS and Sanger sequencing of isolates collected from commercially grown cotton plants in AL, FL, and TX. The pairwise nt and aa comparisons for the eight CLRVDV genomes and the associated ORFs 0–5 indicated that all isolates

from the US were most closely related to the CLRDV-AL genotype, and therefore only distantly related to known isolates from South America. Previously, the CLRDV isolates from the US were reported to cluster into two divergent subclades, with the TX isolate (Accession MN872302) representing the most divergent among all genome sequences available thus far from US isolates [57]. However, with the addition of eight newly determined genome sequences representing isolates from three different states, the evolutionary relationship among CLRDV-AL isolates revealed at least three subclades, herein, referred to as subclades 1–3-US. The latter isolates clustered as a monophyletic group distinct from the genome sequences from South America. Based on the high F_{ST} (0.32–0.52) and N_{ST} (0.32–0.51) estimates, CLRDV isolates exhibited phylogenetic structure at the subclade level.

Based on phylogenetic analysis of the ORF0, several CLRDV isolates from Georgia grouped more closely with CLRDV-SA than with the other US isolates. This observation could suggest that gene flow may have occurred between CLRDV isolates extant in the two geographically distant locales. However, the F_{ST} (0.42) and N_{ST} (0.42) indices, based on complete genome sequences, revealed a high degree of genetic differentiation, suggesting that gene flow between South America and the US has been insufficient to group these isolates as a unified population. The relatively low F_{ST} and N_{ST} values observed between the subclades 1-US (0.07) and 3-US (0.06), compared to the US population, may suggest that the variability frequencies in these subpopulations are similar, while the negative F_{ST} and N_{ST} (−0.07) between atypical-SA and SA could indicate that most variability derives from within subpopulation-level differences [76]. Even so, additional CLRDV complete genome sequences are needed to better resolve the population structure and predict the recent origin(s) of CLRDV in both the US and South America. To accomplish this, additional isolates from cotton and from diverse wild host species from different locations throughout the southern US cotton growing states are needed. Additionally, a larger, more representative CLRDV genome database would aid in predicting the recent origin(s) of US isolates and their route(s) of spread within the US and their relationships to CLRDV from Asia and South America.

Recombination provides a vital evolutionary mechanism shaping the molecular variability of plant viruses, including poleroviruses [77]. The intergenic regions (IR) between ORF2 and ORF3 of poleroviruses have been reported as a hotspot of recombination [18,78–81]. Preliminary analysis of partial CLRDV sequences suggested that the CLRDV-atypical isolates in South America emerged through recombination involving the 5'-block of the viral genome [18]. A complex recombination pattern among CLRDV-GA genomes was recently reported, where these isolates were detected as putative recombinant or parental sequences in seven of ten independent recombination events [57]. The results described here are consistent with previous findings, with at least 11 independent recombination events being predicted among CLRDV isolates, and CLRDV-GA involved as potential recombinant or parental genomes in eight of these events. Recombination breakpoints detected across the viral genomes were mainly localized between ORF2 and ORF3, and ORF5, underscoring the importance of recombination as an evolutionary mechanism underlying CLRDV diversification. In another study, CLRDV was detected in 23 weed species, representing 16 botanical families, naturally occurring in cotton fields in GA, US [34], pointing to the potential for a much greater-than-expected CLRDV host range, and widespread distribution of year-round virus reservoirs. Wild plant species are known to act as mixing vessels that can support high viral diversity, and mixed infections, facilitating virus diversification through recombination and emergence of new viral genotypes [82]. This may explain, in part, why most recombination events were observed among CLRDV in GA. Determining CLRDV genome sequences from wild reservoir hosts would aid in understanding their potential for driving evolution of recombinant genotypes.

Relatively low genome variability has been observed among CLRDV ORF2 and ORF3 [18]. In the context of cultivated cotton and breeding efforts to develop tolerance or resistance to CLRDV, it may not be surprising that CLRDV ORF0, which is known to function as a suppressor of plant host gene silencing, is the most variable coding

region among cotton-infecting isolates of CLRDV [32,34,40,43,44,57–59]. Here, high per site nucleotide diversity was observed among the CLRDV populations from the US and South America, with ORF0 exhibiting the greatest extent of diversification. Interestingly, the CLRDV isolates associated with the phylogenetic subclade 3-US ($\pi = 0.024$) had significantly higher nucleotide diversity indices compared to sequences in subclades 1-US ($\pi = 0.015$) and 2-US ($\pi = 0.016$), for which ORF2 ($\pi = 0.029$) was the most divergent coding region. These observations might be explained by recombination potentially having occurred among subclade 3-US isolates, which were predicted as putative recombinants or parents in nine of 11 independent recombination events. Several recombinants were also previously identified among the CLRDV sequences from GA (subclade 3-US) [57], adding further support for the hypothesis that both mutation and recombination influenced CLRDV evolution.

Selective forces acting on viral coding regions are expected to vary, particularly for multifunctional proteins [83]. Although high variability was observed among the CLRDV isolates, most prominently in ORF0/P0, negative selection ($dN/dS < 1.0$) appeared to have expunged most nonsynonymous mutational effects in the CLRDV ORFs, owing to deleterious effects [83]. Consequently, most mutations detected in viral coding regions showed no additional aa sequence modifications, a scenario that is consistent with conserved protein structure and function. The CLRDV 5' block of genes function in viral replication and in the suppression of host gene silencing, whereas the 3' block encodes proteins involved in genome encapsidation, cell-to-cell movement, and aphid vector transmission [38,40,42,46–48]. Conservation of the viral proteins involved in replication and aphid transmission is of particular importance for virus survival and spread. Notably, the viral movement protein (ORF4) of phylogenetic subclade 1-US and atypical-SA, and ORF0 of atypical-SA were found to be under positive or diversifying selection ($dN/dS > 1.0$). In contrast, ORF5 showed the lowest dN/dS value, a result consistent with its role in aphid-mediated transmission. Although these results suggested that either multiple CLRDV variants were introduced into the US at one time or that distinct isolates were introduced at different times, followed by subsequent diversification, it is not yet possible to determine the extent to which cotton genotype may have differentially influenced viral diversification. Likewise, the timespan(s) between the initial introduction(s) to the present cannot be determined without additional, broad-scale sampling of the southern US cotton-growing states. Overall, the small sample size represented by the two predominant CLRDV genome types extant in the Americas ($n = 6$), together with the lack of complete CLRDV genome sequences for Asian or African isolates precludes a more extensive review of population dynamics. Similarly, understanding the relationship between genomic variability and diversification in the context of local and global epidemiological patterns of spread will require an analysis of CLRDV sequences determined from all known host species and locations where the virus is now known to occur.

5. Conclusions

The introduction of CLRDV into previously uninfected commercial cotton-growing regions, and/or the emergence of new viral genotypes in the US and elsewhere, present economic threats to the cotton industry, with yield losses reported to be as great as 80% in susceptible varieties [18]. Although CLRDV-resistant cotton varieties have been developed that provide protection against the original CLRDV-typical isolates, the so-named resistant-breaking variant, CLRDV-atypical, has emerged in Brazil [59] and Argentina [58]. Considering that SA and US isolates represent two divergent, geographically distinct subpopulations, and interactions between cotton varieties and CLRDV variants of the US subclades 1–3 are unstudied, the movement of CLRDV isolates between these infection foci should be avoided to mitigate their spread to locations where they do not occur naturally. This new knowledge of genetic variability in CLRDV populations is expected to inform cotton breeding efforts to further develop CLRDV-resistant germplasm. Challenging the most promising cotton genotypes with the different CLRDV variants will aid in characterizing genetic resistance in cotton. Finally, additional research is required to better elucidate

CLRVDV epidemiology, including routes or pathways of short- and long-distance spread, understand the relationship between genome variation and recombination among CLRVDV isolates infecting different cotton genotypes and wild plant host species, as well as the role of cotton aphid transmission in CLRVDV diversification. Even so, this new knowledge is expected to advance the development of reliable tools for molecular detection, support resistance breeding efforts, and provide initial epidemiological clues about the distribution of the two main CLRVDV groups extant in the US or SA.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/v13112230/s1>, Table S1: Primers used for the genome amplification of complete cotton leafroll dwarf virus (CLRVDV) and its 5'/3' untranslated regions (UTRs) by random amplification of cDNA ends; Table S2: Cotton leafroll dwarf virus (CLRVDV) genome sequences retrieved from NCBI-GenBank; Table S3: Predicted recombination events detected for cotton leafroll dwarf virus (CLRVDV) isolates.

Author Contributions: Conceptualization, R.R.-S., K.L., D.W.S., O.J.A. and J.K.B.; methodology, R.R.-S., R.O.A., O.J.A. and J.K.B.; software, R.R.-S., R.O.A. and O.J.A.; validation, R.R.-S., R.O.A. and O.J.A.; formal analysis, R.R.-S., R.O.A. and O.J.A.; investigation, R.R.-S., R.O.A., T.I., O.J.A.; resources, K.L., D.W.S., T.I., O.J.A. and J.K.B.; data curation, R.R.-S., R.O.A. and O.J.A.; writing—original draft preparation, R.R.-S., R.O.A. and J.K.B.; writing—review and editing, R.R.-S., R.O.A., K.L., D.W.S., T.I., O.J.A. and J.K.B.; visualization, R.R.-S.; supervision, O.J.A. and J.K.B.; project administration, J.K.B.; funding acquisition, O.J.A. and J.K.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Cotton Incorporated grants 06-829 (to J.K.B.) and 20-391 (to O.J.A.).

Institutional Review Board Statement: This manuscript does not contain studies with human participants or animals.

Data Availability Statement: The new viral complete genome sequences are available NCBI-GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) under accession numbers OK185939-OK185946.

Acknowledgments: This research was supported by Cotton Incorporated grants 06-829 (to JKB) and 20-391 (to OJA). We thank Cecilia Villegas (Texas A&M University System) for technical support.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ahmad, S.; Iqbal, M.; Muhammad, T.; Mehmood, A.; Ahmad, S.; Hasanuzzaman, M. Cotton productivity enhanced through transplanting and early sowing. *Acta Sci. Biol. Sci.* **2018**, *40*, 34610. [CrossRef]
2. Ahmad, S.; Raza, I.; Ali, H.; Shahzad, A.N.; Rehman, A.U.; Sarwar, N. Response of cotton crop to exogenous application of glycinebetaine under sufficient and scarce water conditions. *Braz. J. Bot.* **2014**, *37*, 407–415. [CrossRef]
3. Sunilkumar, G.; Campbell, L.M.; Puckhaber, L.; Stipanovic, R.D.; Rathore, K.S. Engineering cottonseed for use in human nutrition by tissue-specific reduction of toxic gossypol. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 18054–18059. [CrossRef] [PubMed]
4. FAO. FAOSTAT. 2019. Available online: <http://www.fao.org/faostat/en/#home> (accessed on 26 December 2019).
5. USDA. Crop Values 2020 Summary. Available online: <https://quickstats.nass.usda.gov/results/B7E6307C-4344-387D-B041-5BA186B72D77> (accessed on 22 July 2021).
6. Chohan, S.; Perveen, R.; Abid, M.; Tahir, M.N.; Sajid, M. Cotton Diseases and Their Management. In *Cotton Production and Uses*; Springer: Dordrecht, The Netherlands, 2020; pp. 239–270.
7. Cauquil, J. Etudes sur une maladie d'origine virale du cotonnier, la maladie bleue. *Cot. Fib. Trop.* **1977**, *32*, 259–278.
8. Cauquil, J.; Vaissayre, M. La "Maladie Bleue" du Cotonnier en Afrique: Transmission de Cotonnier à Cotonnier par Aphis Gossypii Glover. Available online: <http://agritrop.cirad.fr/455715/1/ID455715.pdf> (accessed on 28 September 2021).
9. Costa, A.S.; Carvalho, A.M.B. Moléstias de vírus do algodoeiro. *Bragantia* **1962**, *21*, 50–51.
10. Costa, A.S.; Forster, R. Nota preliminar sobre una nova moléstia de vírus do algodoeiro: Mosaico das nervuras. *Revista de Agricultura* **1938**, *13*, 187–191.
11. Campagnac, N.A.; Bonacic Kresic, R.; Poisson, J. Mal de Misiones: Nueva enfermedad del algodón de probable origen virósico. *Jorn. Fitosanit. Argent.* **1986**, *6*, 503–511.
12. Corrêa, R.L.; Silva, T.F.; Simões-Araújo, J.L.; Barroso, P.A.V.; Vidal, M.S.; Vaslin, M.F.S. Molecular characterization of a virus from the family Luteoviridae associated with cotton blue disease. *Arch. Virol.* **2005**, *150*, 1357–1367. [CrossRef]

13. Ray, J.D.; Sharman, M.; Quintao, V.; Rossel, B.; Westaway, J.; Gambley, C. Cotton leafroll dwarf virus detected in Timor-Leste. *Australas. Plant Dis. Notes* **2016**, *11*, 29. [[CrossRef](#)]
14. Sharman, M.; Lapbanjob, S.; Seibunruang, P.; Belot, J.L.; Galbieri, R.; Giband, M.; Suassuna, N. First report of Cotton leafroll dwarf virus in Thailand using a species-specific PCR validated with isolates from Brazil. *Australas. Plant Dis. Notes* **2015**, *10*, 1–4. [[CrossRef](#)]
15. Mukherjee, A.K.; Chahande, P.R.; Meshram, M.K.; Kranthi, K.R. First report of Polerovirus of the family Luteoviridae infecting cotton in India. *New Dis. Rep.* **2012**, *25*, 22. [[CrossRef](#)]
16. Quyen, L.Q.; Hai, N.T.; Mai, T.A.H.; Hao, V.; Binh, N.T.T.; Bu’u, D.N.; Dieu, D.X.; Underwood, E. Cotton production in Vietnam. In *Environmental Risk Assessment of Genetically Modified Organisms: Challenges and Opportunities with bt Cotton in Vietnam*; Andow, D.A., Hilbeck, A., Eds.; CABI Publishing: Wallingford, UK, 2008; Volume 4, pp. 24–63.
17. Distéfano, A.J.; Kresic, I.B.; Hopp, H.E. The complete genome sequence of a virus associated with cotton blue disease, cotton leafroll dwarf virus, confirms that it is a new member of the genus Polerovirus. *Arch. Virol.* **2010**, *155*, 1849–1854. [[CrossRef](#)]
18. Silva, T.F.; Corrêa, R.L.; Castilho, Y.; Silvie, P.; Bêlot, J.-L.; Vaslin, M.F.S. Widespread distribution and a new re-combinant species of Brazilian virus associated with cotton blue disease. *Virol. J.* **2008**, *5*, 123. [[CrossRef](#)] [[PubMed](#)]
19. Avelar, S.; Schrimsher, D.W.; Lawrence, K.; Brown, J.K. First Report of Cotton leafroll dwarf virus Associated with Cotton Blue Disease Symptoms in Alabama. *Plant Dis.* **2019**, *103*, 592. [[CrossRef](#)]
20. Alabi, O.J.; Isakeit, T.; Vaughn, R.; Stelly, D.; Conner, K.N.; Gaytán, B.C.; Villegas, C.; Hitzelberger, C.; De Santiago, L.; Monclova-Santana, C.; et al. First Report of Cotton leafroll dwarf virus Infecting Upland Cotton (*Gossypium hirsutum*) in Texas. *Plant Dis.* **2020**, *104*, 998. [[CrossRef](#)]
21. Ali, A.; Mokhtari, S. Complete Genome Sequence of Cotton Leafroll Dwarf Virus Isolated from Cotton in Texas, USA. *Microbiol. Resour. Announc.* **2020**, *9*. [[CrossRef](#)]
22. Faske, T.R.; Stainton, D.; Ghanem-Sabanadzovic, N.A.; Allen, T.W. First Report of Cotton Leafroll Dwarf Virus from Upland Cotton (*Gossypium hirsutum*) in Arkansas. *Plant Dis.* **2020**, *104*. [[CrossRef](#)]
23. Iriarte, F.B.; Dey, K.K.; Small, I.M.; Conner, K.; O’Brien, G.K.; Johnson, L.; Savery, C.; Carter, E.; Sprague, D.; Nichols, R.L.; et al. First Report of Cotton Leafroll Dwarf Virus in Florida. *Plant Dis.* **2020**, *104*. [[CrossRef](#)]
24. Price, T.; Valverde, R.; Singh, R.; Davis, J.; Brown, S.; Jones, H. First Report of Cotton Leafroll Dwarf Virus in Louisiana. *Plant Heal. Prog.* **2020**, *21*, 142–143. [[CrossRef](#)]
25. Tabassum, A.; Roberts, P.M.; Bag, S. Genome Sequence of Cotton Leafroll Dwarf Virus Infecting Cotton in Georgia, USA. *Microbiol. Resour. Announc.* **2020**, *9*, 00812-20. [[CrossRef](#)]
26. Thiessen, L.D.; Schappe, T.L.; Zaccaron, M.; Conner, K.; Koebnick, J.; Jacobson, A.; Huseth, A. First Report of Cotton Leafroll Dwarf Virus in Cotton Plants Affected by Cotton Leafroll Dwarf Disease in North Carolina. *Plant Dis.* **2020**, *104*, 3275. [[CrossRef](#)]
27. Wang, H.; Greene, J.; Mueller, J.D.; Conner, K.; Jacobson, A. First Report of Cotton Leafroll Dwarf Virus in Cotton Fields of South Carolina. *Plant Dis.* **2020**, *104*, 2532. [[CrossRef](#)]
28. Aboughanem-Sabanadzovic, N.; Allen, T.W.; Wilkerson, T.H.; Conner, K.N.; Sikora, E.J.; Nichols, R.L.; Sabanadzovic, S. First Report of Cotton Leafroll Dwarf Virus in Upland Cotton (*Gossypium hirsutum*) in Mississippi. *Plant Dis.* **2019**, *103*, 1798. [[CrossRef](#)]
29. Huseth, A.; Reisig, D.; Collins, G.; Thiessen, L. Detection of Cotton Leafroll Dwarf Virus (CLR DV) in North Carolina. NC State Extension. 2019. Available online: <http://go.ncsu.edu/readext?639597> (accessed on 11 July 2021).
30. Tabassum, A.; Bag, S.; Roberts, P.; Suassuna, N.; Chee, P.; Whitaker, J.R.; Conner, K.N.; Brown, J.; Nichols, R.L.; Kemerait, R.C. First Report of Cotton Leafroll Dwarf Virus Infecting Cotton in Georgia, U.S.A. *Plant Dis.* **2019**, *103*, 1803. [[CrossRef](#)]
31. Mahas, J.W. Management of Aphis gossypii Populations and the Spread of Cotton leafroll dwarf virus in Southeastern Cotton Production Systems. Master’s Thesis, Department of Entomology and Plant pathology, Auburn University, Auburn, AL, USA, 2020.
32. Avelar, S.; Ramos-Sobrinho, R.; Conner, K.; Nichols, R.L.; Lawrence, K.; Brown, J.K. Characterization of the Complete Genome and P0 Protein for a Previously Unreported Genotype of Cotton Leafroll Dwarf Virus, an Introduced Polerovirus in the United States. *Plant Dis.* **2020**, *104*, 780–786. [[CrossRef](#)]
33. Bag, S.; Roberts, P.M.; Kemerait, R.C. Cotton Leafroll Dwarf Disease: An Emerging Virus Disease on Cotton in the U.S. *Crop. Soils* **2021**, *54*, 18–22. [[CrossRef](#)]
34. Sedhain, N.P.; Bag, S.; Morgan, K.; Carter, R.; Triana, P.; Whitaker, J.; Kemerait, R.C.; Roberts, P.M. Natural host range, incidence on overwintering cotton and diversity of cotton leafroll dwarf virus in Georgia USA. *Crop. Prot.* **2021**, *144*, 105604. [[CrossRef](#)]
35. ICTV.Taxonomy History: Cotton Leafroll Dwarf Virus. Available online: https://talk.ictvonline.org/taxonomy/p/taxonomy-history?taxnode_id=202003781 (accessed on 23 July 2021).
36. Domier, L.L.; King, A.M.Q.; Adams, M.J.; Carstens, E.B.; Lefkowitz, E.J. Family Luteoviridae. In *Virus Taxonomy*; Ninth report of the International Committee on Taxonomy of Viruses; Elsevier/Academic Press: London, UK, 2012; pp. 1045–1053. [[CrossRef](#)]
37. Gray, S.; Gildow, F.E. Luteovirus-aphid interactions. *Annu. Rev. Phytopathol.* **2003**, *41*, 539–566. [[CrossRef](#)] [[PubMed](#)]
38. Wang, J.; Chay, C.; Gildow, F.; Gray, S. Readthrough Protein Associated with Virions of Barley Yellow Dwarf Luteovirus and Its Potential Role in Regulating the Efficiency of Aphid Transmission. *Virology* **1995**, *206*, 954–962. [[CrossRef](#)]
39. Takimoto, J.K.; Benetti Queiroz-Voltan, R.; Caram de Souza-Dias, J.A.; Cia, E. Alterações anatômicas em algodoeiro infectado pelo vírus da doença azul. *Bragantia* **2009**, *68*, 109–116.

40. Cascardo, R.S.; Arantes, I.L.G.; Silva, T.F.; Sachetto-Martins, G.; Vaslin, M.F.S.; Corrêa, R.L. Function and diversity of P0 proteins among cotton leafroll dwarf virus isolates. *Virol. J.* **2015**, *12*, 1–10. [[CrossRef](#)]
41. Miller, W.; Brown, C.; Wang, S. New Punctuation for the Genetic Code: Luteovirus Gene Expression. *Semin. Virol.* **1997**, *8*, 3–13. [[CrossRef](#)]
42. Mayo, M.; Ziegler-Graff, V. Molecular Biology of Luteoviruses. *Adv. Virus Res.* **1996**, *46*, 413–460. [[CrossRef](#)] [[PubMed](#)]
43. Agrofoglio, Y.C.; Delfosse, V.C.; Casse, M.F.; Hopp, H.E.; Kresic, I.B.; Ziegler-Graff, V.; Distéfano, A.J. P0 protein of cotton leafroll dwarf virus-atypical isolate is a weak RNA silencing suppressor and the avirulence determinant that breaks the cotton Cbd gene-based resistance. *Plant Pathol.* **2019**, *68*, 1059–1071. [[CrossRef](#)]
44. Delfosse, V.C.; Agrofoglio, Y.C.; Casse, M.F.; Kresic, I.B.; Hopp, H.E.; Ziegler-Graff, V.; Distéfano, A.J. The P0 protein encoded by cotton leafroll dwarf virus (CLRDV) inhibits local but not systemic RNA silencing. *Virus Res.* **2014**, *180*, 70–75. [[CrossRef](#)] [[PubMed](#)]
45. Pfeffer, S.; Dunoyer, P.; Heim, F.; Richards, K.E.; Jonard, G.; Ziegler-Graff, V. P0 of beet Western yellows virus is a sup-pressor of posttranscriptional gene silencing. *J. Virol.* **2002**, *76*, 6815–6824. [[CrossRef](#)] [[PubMed](#)]
46. Smirnova, E.; Firth, A.; Miller, W.A.; Scheidecker, D.; Brault, V.; Reinbold, C.; Rakotondrafara, A.M.; Chung, B.Y.-W.; Ziegler-Graff, V. Discovery of a Small Non-AUG-Initiated ORF in Poloroviruses and Luteoviruses That Is Required for Long-Distance Movement. *PLoS Pathog.* **2015**, *11*, e1004868. [[CrossRef](#)]
47. Chay, C.; Gunasinge, U.; Kumarb, S.-; Miller, W.; Gray, S.M. Aphid Transmission and Systemic Plant Infection Determinants of Barley Yellow Dwarf Luteovirus-PAV are Contained in the Coat Protein Readthrough Domain and 17-kDa Protein, Respectively. *Virology* **1996**, *219*, 57–65. [[CrossRef](#)] [[PubMed](#)]
48. Brault, V.; Heuvel, J.V.D.; Verbeek, M.; Ziegler-Graff, V.; Reutenauer, A.; Herrbach, E.; Garaud, J.; Guilley, H.; Richards, K.; Jonard, G. Aphid transmission of beet western yellows luteovirus requires the minor capsid read-through protein P74. *EMBO J.* **1995**, *14*, 650–659. [[CrossRef](#)]
49. Miller, W.A.; Beckett, R.; Liu, S. Structure, Function, and Variation of the Barley Yellow Dwarf Virus and Cereal Yellow Dwarf Virus Genomes. In *Barley Yellow Dwarf Disease: Recent Advances and Future Strategies, Proceedings of the an International Symposium held at El Batán, Texcoco, Mexico, 1–5 September 2002*; Henry, M., McNab, A., Eds.; CIMMYT: Texcoco, Mexico, 2002.
50. van der Wilk, F.; Verbeek, M.; Dullemans, A.; Heuvel, J.V.D. The Genome-Linked Protein of Potato Leafroll Virus Is Located Downstream of the Putative Protease Domain of the ORF1 Product. *Virology* **1997**, *234*, 300–303. [[CrossRef](#)]
51. Pagán, I. The diversity, evolution and epidemiology of plant viruses: A phylogenetic view. *Infect. Genet. Evol.* **2018**, *65*, 187–199. [[CrossRef](#)] [[PubMed](#)]
52. Holmes, E.C. *The Evolution and Emergence of RNA Viruses*; Oxford University Press: Oxford, UK, 2009.
53. Drake, J.W.; Holland, J.J. Mutation rates among RNA viruses. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 13910–13913. [[CrossRef](#)] [[PubMed](#)]
54. Simon-Loriere, E.; Holmes, E. Why do RNA viruses recombine? *Nat. Rev. Genet.* **2011**, *9*, 617–626. [[CrossRef](#)]
55. Hall, G. Selective constraint and genetic differentiation in geo- graphically distant barley yellow dwarf virus popula-tions. *J. Gen. Virol.* **2006**, *87*, 3067–3075. [[CrossRef](#)]
56. Gibbs, M.J.; Weiller, G.F. Evidence that a plant virus switched hosts to infect a vertebrate and then recombined with a vertebrate-infecting virus. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 8022–8027. [[CrossRef](#)]
57. Tabassum, A.; Bag, S.; Suassuna, N.D.; Conner, K.N.; Chee, P.; Kemerait, R.C.; Roberts, P. Genome analysis of cotton leafroll dwarf virus reveals variability in the silencing suppressor protein, genotypes and genomic recombinants in the USA. *PLoS ONE* **2021**, *16*, e0252523. [[CrossRef](#)] [[PubMed](#)]
58. Agrofoglio, Y.C.; Delfosse, C.; Casse, F.; Hopp, H.E.; Kresic, B.; Distéfano, A.J. Identification of a new cotton disease caused by an atypical cotton leafroll dwarf virus in Argentina. *Phytopathology* **2017**, *107*, 1–8.
59. Da Silva, A.K.F.; Romanel, E.; da F Silva, T.; Castilhos, Y.; Schrago, C.; Galbieri, R.; Bélot, J.-L.; Vaslin, M.F.S. Complete genome sequences of two new virus isolates associated with cotton blue disease resistance breaking in Brazil. *Arch. Virol.* **2015**, *160*, 1371–1374. [[CrossRef](#)]
60. Rott, M.; Jelkmann, W. Characterization and Detection of Several Filamentous Viruses of Cherry: Adaptation of an Alternative Cloning Method (DOP-PCR), and Modification of an RNA Extraction Protocol. *Eur. J. Plant Pathol.* **2001**, *107*, 411–420. [[CrossRef](#)]
61. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [[CrossRef](#)] [[PubMed](#)]
62. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)]
63. Hall, T.A. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **1999**, *41*, 95–98.
64. Muhire, B.M.; Varsani, A.; Martin, D.P. SDT: A Virus Classification Tool Based on Pairwise Sequence Alignment and Identity Calculation. *PLoS ONE* **2014**, *9*, e108277. [[CrossRef](#)]
65. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [[CrossRef](#)]
66. NBSP; Huelsenbeck, J.P. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Syst. Biol.* **2012**, *61*, 539–542. [[CrossRef](#)]

67. Miller, M.A.; Pfeiffer, W.; Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In Proceedings of the 2010 Gateway Computing Environments Workshop (GCE), New Orleans, LA, USA, 14 November 2010; IEEE: New Orleans, LA, USA, 2010; pp. 1–8. [[CrossRef](#)]
68. Rannala, B.; Yang, Z. Probability distribution of molecular evolutionary trees: A new method of phylogenetic inference. *J Mol Evol* **1996**, *43*, 304–311.
69. Wright, S. The Genetical Structure of Populations. *Ann. Eugen.* **1949**, *15*, 323–354. [[CrossRef](#)] [[PubMed](#)]
70. Rozas, J.; Ferrer-Mata, A.; Sánchez-DelBarrio, J.C.; Guirao-Rico, S.; Librado, P.; Ramos-Onsins, S.; Sánchez-Gracia, A. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Mol. Biol. Evol.* **2017**, *34*, 3299–3302. [[CrossRef](#)]
71. Huson, D.H.; Bryant, D. Application of Phylogenetic Networks in Evolutionary Studies. *Mol. Biol. Evol.* **2005**, *23*, 254–267. [[CrossRef](#)] [[PubMed](#)]
72. Martin, D.P.; Murrell, B.; Golden, M.; Khoosal, A.; Muhire, B. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* **2015**, *1*, vev003. [[CrossRef](#)] [[PubMed](#)]
73. Lima, A.T.M.; Silva, J.C.F.; Silva, F.N.; Castillo-Urquiza, G.P.; Silva, F.F.; Seah, Y.M.; Mizubuti, E.S.G.; Duffy, S.; Zerbini, F.M. The diversification of begomovirus populations is predominantly driven by mutational dynamics. *Virus Evol.* **2017**, *3*, vex005. [[CrossRef](#)]
74. Pond, S.L.K.; Frost, S.D.W. Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. *Mol. Biol. Evol.* **2005**, *22*, 1208–1222. [[CrossRef](#)]
75. Yang, Z.; Nielsen, R. Estimating Synonymous and Nonsynonymous Substitution Rates Under Realistic Evolutionary Models. *Mol. Biol. Evol.* **2000**, *17*, 32–43. [[PubMed](#)]
76. Holsinger, K.E.; Weir, B.S. Genetics in geographically structured populations: Defining, estimating and interpreting *F_{ST}*. *Nat. Rev. Genet.* **2009**, *10*, 639–650. [[CrossRef](#)] [[PubMed](#)]
77. Knierim, D.; Deng, T.C.; Tsai, W.S.; Green, S.K.; Kenyon, L.C. Molecular identification of three distinct Polerovirus species and a recombinant Cucurbit aphid-borne yellows virus strain infecting cucurbit crops in Taiwan. *Plant Pathol.* **2010**, *59*, 991–1002. [[CrossRef](#)]
78. Shang, Q.-X.; Xiang, H.-Y.; Han, C.-G.; Li, D.; Yu, J.-L. Distribution and molecular diversity of three cucurbit-infecting poleroviruses in China. *Virus Res.* **2009**, *145*, 341–346. [[CrossRef](#)] [[PubMed](#)]
79. Lemaire, O.; Beuve, M.; Hauser, S.; Stevens, M.; Herrbach, E. Diversity and phylogeny of beet poleroviruses. In Proceedings of the 8. International Plant Virus Epidemiology Symposium, Ascherleben, Germany, 12–17 May 2002.
80. Moonan, F.; Molina, J.; Mirkov, T.E. Sugarcane yellow leaf virus: An emerging virus that has evolved by recombination between luteoviral and poleroviral ancestors. *Virology* **2000**, *269*, 156–171. [[PubMed](#)]
81. Guilley, H.; Richards, K.E.; Jonard, G. Nucleotide sequence of beet mild yellowing virus RNA. *Arch. Virol.* **1995**, *140*, 1109–1118. [[CrossRef](#)] [[PubMed](#)]
82. García-Arenal, F.; Zerbini, F.M. Life on the Edge: Geminiviruses at the Interface Between Crops and Wild Plant Hosts. *Annu. Rev. Virol.* **2019**, *6*, 411–433. [[CrossRef](#)]
83. Fraile, A.; García-Arenal, F.; Malpica, J.M. Variation and evolution of plant virus populations. *Int. Microbiol.* **2003**, *6*, 225–232. [[CrossRef](#)]