



Complete genome characterization of cacao leafroll virus, a newly described cacao-infecting polerovirus

Raphael O. Adegbola¹ · Cory V. Keith¹ · Osman Gutierrez² · Ricardo Goenaga³ · Judith K. Brown¹

Received: 17 September 2023 / Accepted: 4 March 2024 / Published online: 23 March 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2024

Abstract

The complete genome sequence of cacao leafroll virus (CaLRV; family *Solemoviridae*, genus *Polerovirus*) was determined by high-throughput sequencing of total RNA isolated from symptomatic cacao *Theobroma cacao* L. plants (n=4). The CaLRV genome sequences ranged from 5,976 to 5,997 nucleotides (nt) in length and contained seven open reading frames (ORFs). Nucleotide and amino acid (aa) sequence comparisons showed that, among selected well-characterized poleroviruses, the CaLRV genome shared the highest nt sequence identity of 62% with that of potato leafroll virus (PLRV, NC_076505). A comparison of the predicted aa sequence of the CaLRV coat protein indicated that cotton leafroll dwarf virus (CLRDV, NC_014545) and melon aphid-borne yellows virus (MABYV, NC_010809) were the closest relatives, sharing 57% aa sequence identity. Bayesian phylogenetic analysis based on complete genome sequences showed that CaLRV grouped with well-characterized poleroviruses that cause diseases of cereal and vegetable crops. During the course of publishing this work, the nearly complete genome sequence of a member of the same polerovirus species, referred to as “cacao polerovirus” (OR605721), with which CaLRV shares 99% nt sequence identity, was reported.

Communicated by Elvira Fiallo-Olivé

✉ Judith K. Brown
jkbrown@arizona.edu
Raphael O. Adegbola
roadegbola@arizona.edu
Cory V. Keith
corykeith@arizona.edu
Osman Gutierrez
Osman.Gutierrez@usda.gov
Ricardo Goenaga
ricardo.goenaga@usda.gov

¹ School of Plant Sciences, The University of Arizona, Tucson, AZ 85721, USA

² USDA-ARS Subtropical Horticultural Research Station, Miami, FL 33158, USA

³ USDA-ARS Tropical Agriculture Research Station, Mayaguez, PR 00680, USA

The cacao tree, *Theobroma cacao* (L.) (family Malvaceae), is an economically important tree crop that is cultivated in tropical regions and from which the beans are harvested to produce high-value confectionaries and other products [1]. A number of plant viruses cause diseases in cacao trees that limit crop production where cacao is grown, and recently, at least three have been identified in cacao germplasm collections [2–7]. The family *Solemoviridae* includes four genera of plant viruses: *Enamovirus*, *Polemavirus*, *Polerovirus*, and *Sobemovirus* [8]. Poleroviruses have non-enveloped icosahedral virions that are 26–30 nm in diameter and a polycistronic positive-sense single-stranded RNA (+ssRNA) genome of approximately 5.6–6.2 kilobases (kb) with 5'- and 3'-terminal noncoding regions and 7–10 open reading frames (ORFs). They are transmitted in a persistent, circulative manner by aphids (s.o. Homoptera; o. Hemiptera) [9] and are economically important pathogens of cereal grains, fruits, and vegetable crops. Infected plants exhibit symptoms that consist of leaf rolling, foliar chlorosis, and overall stunting [10]. To safeguard the transport of virus-free cacao germplasm and ensure that cacao seedlings planted

on commercial cacao farms are virus-free, high-throughput sequencing has been used for screening of germplasm and planting materials for the presence of plant viruses to prevent their introduction. This has led to the characterization of numerous badnaviruses and, recently, to the detection of a previously uncharacterized polerovirus [3]. In this study, a previously uncharacterized polerovirus, which we have named "cacao leafroll virus" (CaLRV; family *Solemoviridae*; genus *Polerovirus*), was found to be associated with symptomatic cacao seedlings (n=4) held in quarantine, making it only the second polerovirus known to infect cacao, [11]. The virus was initially detected by discovery Illumina RNAseq and *de novo* assembly of the complete genome sequence, and it was identified as a polerovirus based on phylogenetic and pairwise distance analysis. The complete CaLRV genome sequence was obtained for two of the four isolates (MITC2028 and MITC2039) by rapid amplification of cDNA ends (RACE). The coding and non-coding regions of all four genome sequences were annotated based on previously published polerovirus genome sequences available in the NCBI GenBank database.

CaLRV genomic RNA was recovered from the cacao accessions CATIE-R4, CCN 51, CC 137, MX C67, and CATIE-R1, which were identified by genetic analysis as crosses between different genetic groups of cacao [12]. These plants were held in the USDA-ARS-SHRS, Miami, FL, quarantine greenhouse facility for mandatory observation prior to release. The plants exhibited foliar symptoms consisting of bluish-greenish discoloration, foliar malformation, and downward leaf rolling. Leaves with the petioles attached were collected from young cacao seedlings at the 12- to 14-leaf stage and shipped by courier to The University of Arizona, Tucson, AZ, under a USDA-APHIS permit (issued to J.K. Brown). Total RNA was isolated from leaf samples using a modified silica RNA isolation protocol [13] and submitted to Novogene Corp., Sacramento, CA, for Illumina RNAseq sequencing. Following ribosomal RNA depletion, cDNA libraries were constructed and sequenced (150 base-paired ends) on an Illumina NovaSeq 6000 platform (Illumina, San Diego, CA). Adapter sequences and low-quality reads were removed using BBDuk (Department of Energy, Joint Genome Institute, Berkeley, CA), and the read quality was assessed using the FastQC quality control tool, which is available online at <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. Quality-filtered reads were assembled *de novo* using rnaviralSPAdes, implemented in SPAdes ver. 3.15.4 [14]. Virus-specific contigs were initially identified using BLASTn [15] to query a local database (UA Plant Sciences Diagnostic Lab) containing sequences downloaded from the NCBI GenBank Virus Refseq database, available at <https://ftp.ncbi.nlm.nih.gov/refseq/release/viral/>, including the previously published partial

CaLRV genome sequences [11]. Bioinformatic analysis was carried out on The University of Arizona High Performance Computing (HPC) cluster (UA Research Data Center, Tucson, AZ). Viral coding regions were predicted using NCBI ORF Finder (<https://www.ncbi.nlm.nih.gov/orffinder/>), implemented in Geneious Prime v2023.1.2 (Biomatters Inc., San Diego, CA). The resultant viral nucleotide and amino acid sequences were aligned in MUSCLE [16], available in Geneious Prime v2023.1.2 (Biomatters Inc., San Diego, CA). Pairwise distance analysis of the viral nt sequences and predicted aa sequences was carried out using Sequence Demarcation Tool (SDT) v1.2 [17]. Phylogenetic analysis of the aligned sequences was carried out using MrBayes software, version 3.2.7a [18], implemented on The University of Arizona High Performance Computing (HPC) cluster (UA Research Data Center, Tucson, AZ). The phylogenetic tree was drawn using FigTree v1.4.4 (<https://github.com/rambaut/figtree/releases>) and edited in Adobe Illustrator 2019. The amino acid sequence protein family (pfam) matches were determined using the InterProScan tool [19] and InterPro protein signature databases. The putative viral fusion protein was predicted using KnotInFrame software (Turner model 2004) [20] to identify the location of the -1 ribosomal frameshift site at the ORF1 and ORF2 overlap. The 5'- and 3'-terminal ends of the genome were determined by rapid amplification of cDNA ends (RACE) as described previously [21]. Briefly, to determine the 5'-terminal sequence, first-strand complementary DNA (cDNA) was synthesized using a SuperScript IV reverse transcription (RT) system (Invitrogen, Carlsbad, CA) with the virus-specific antisense primer CaPol_1335R (5'- GGTCGCATCGAGCCATTC TA -3'). The cDNA was purified using a PureLink PCR Micro Kit (Invitrogen) and C-tailed using recombinant terminal deoxynucleotide transferase (rTdT) in 10X TdT reaction buffer, 2 mM deoxycytidine triphosphate (dCTP), and 2.5 mM cobalt(II) chloride (COCl₂) solution (New England Biolabs, Ipswich, MA). The 5'-terminal end was amplified from the C-tailed cDNA template by PCR amplification. The reaction mixture contained the sense primer P5-poly-G-F (5'- GGGGGGGGGGGGACAAAAGA- 3'), for which the conserved polerovirus sequence [21] is shown underlined, the virus-specific antisense primer CaPol_653R (5'- CGTCCACCATATCCATGCAAAC -3'), and JumpStart REDTaq ReadyMix (Sigma-Aldrich, St. Louis, MO). To determine the sequence of the 3'-terminal end of the genome, the total RNA was poly-A tailed using *Escherichia coli* poly(A) polymerase in 10X *E. coli* poly(A) polymerase reaction buffer containing 10 mM ATP (New England Biolabs). The poly-A-tailed RNA was reverse transcribed using the SuperScript IV RT system (Invitrogen) with the primer poly-T (5'- TCCCGGGTTTTTTTTTTTTTTTTTTTTTTTTTTT -3') [21]. The cDNA template was amplified by PCR using

the virus-specific primer CaPol_5225F (5'-AGATGAAAC TACCTCTGTGGCG-3') and poly-T as the antisense primer [21]. The amplicons were separated by agarose gel (1.0%) electrophoresis in 1X TAE buffer, pH 8.0, and excised from the gel with a razor (VWR, Radnor, PA), followed by gel-purification using a Wizard SV Gel and PCR Clean-up System (Promega, Madison, WI). Gel-purified amplicons were ligated into the pGEM-T Easy plasmid vector (Promega) and used to transform chemically competent *E. coli* strain DH5 α cells. Bacterial colonies containing a plasmid bearing a fragment of the expected size were screened by colony PCR amplification as described previously [22]. The DNA sequence of each insert was determined by Sanger sequencing (Eurofins Genomics LLC, Louisville, KY).

The number of reads used for *de novo* assembly ranged from 75,350,628 (isolate MITC2007) to 84,531,254 (isolate MITC2028), yielding 131,803 to 166,226 contigs, respectively (Supplementary Table S1A). The sequence coverage ranged from 1,883 (isolate MITC2007) to 30,103 (isolate MITC2039). A BLASTn analysis of the assembled reads (<https://ftp.ncbi.nlm.nih.gov/refseq/release/viral/>) identified contigs from the four samples, ranging in size from 5,973 bp to 6256 bp, that yielded matches to one of five previously characterized poleroviruses, melon aphid-borne yellows virus (MABYV, NC_010809), pepo aphid-borne yellows virus (PABYV, NC_030225), potato leafroll virus (PLRV, NC_0076505), pumpkin polerovirus (PuPV, NC_055513), or wheat leaf yellowing-associated virus (WLYaV, NC_035451) (Supplementary Table S1B) with 71.8–79.7% nucleotide sequence identity and e-values from $7.27e^{-31}$ (PABYV) to $1.13e^{-78}$ (Supplementary Table S1B).

Four CaLRV genome sequences were submitted to the GenBank database and assigned the accession numbers OR423047–50. The complete CaLRV genome sequences MITC2007, MITC2028, MITC2039, and MITC2064, which were assembled from Illumina RNAseq reads and/or RACE, were 5,994 to 5,997 nt in length. Each genome contained seven ORFs (ORFs 0, 1, 1–2, 3a, 3, 4, and 3–5), 5'

and 3' untranslated regions (UTRs), and an intergenic region (IGR) and exhibited a genome organization like that of other well-characterized poleroviruses (Fig. 1). The 5'-UTR (nt 1–101) is 101 nt in length and contains a polerovirus-like conserved terminal “ACAAAA” sequence [8]. ORF0 (nt 102–791) is 690 nt in length and encodes a predicted protein (P0; 229 aa) with a predicted molecular weight (MW) of 26.7 kDa. The polerovirus P0 protein, which is believed to be a suppressor of host gene silencing, contains the F-box motif FPFLX₁₄P_{64–83}, which is conserved among the well-characterized poleroviruses [23, 24]. ORF1 (nt 229–2289) is 2,061 nt in length and encodes a predicted multifunctional polyprotein (P1) of 686 aa with a predicted MW of 76.3 kDa that is predicted to be involved in virus replication [25]. P1 contains a conserved protease motif with the predicted catalytic triad of a chymotrypsin-like serine protease (HX₃₁DX₆₃TAAGFSG_{276–378}) [25–27] and includes the viral genome-linked protein (VPg), which is released following cleavage by the serine protease portion of P1. ORF2 (nt 1857–3566) is 1,710 nt in length and encodes an RNA-dependent RNA polymerase, RdRp (P2), of 569 aa, with an MW of 64.3 kDa. It is expressed as an overlapping protein from the C-terminal region of ORF1, which, with P1, forms a putative fusion protein, RdRp (P1-P2), due to a -1 ribosomal frameshift at nt 1,863 during translation. This is mediated by a slippery heptanucleotide sequence (GGGAAA AA) located at nt 1,857–1,863 within the ORF1-ORF2 overlapping region. The RdRp consists of the conserved ‘GDD’ aa core GSYNTSSSNX₁₉GDD_{1220–1250}, which is present in other viral RdRps [28]. The ORF1-ORF2 region (nt 229–3,566) is 3,338 nt in length and encodes a P1-P2 protein of 1,112 aa, with an estimated MW of 124.3 kDa. The viral ORF2 is followed by an intergenic region (IGR; nt 3567–3660) of 94 nt. The ORF3a sequence (nt 3661–3783) is 123 nt in length and begins with a non-canonical ‘CUG’ initiation codon [29]. It encodes the 40-aa P3a protein, which has an MW of 4.4 kDa. In PLRV, the P3a is involved in long-distance movement [30]. ORF3 (nt 3790–4395) is 606 nt long

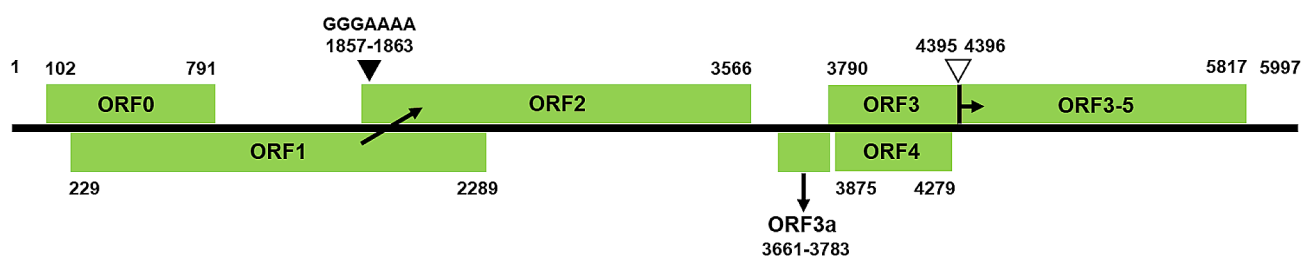


Fig. 1 Genome organization of cacao leafroll virus (CaLRV) showing an illustration of the open reading frames. Numbers indicate the nucleotide positions of ORFs within the genome. Boxes show open reading frames (ORF). The shaded triangle (▲) indicates the location of the slippery heptamer ‘GGGAAA’ that is responsible for the -1

ribosomal frameshift (indicated by a slanting arrow) that results in the synthesis of the P1-P2 fusion protein. The unshaded triangle (Δ) indicates the position of suppression of the amber stop codon of ORF3 that results in a translational readthrough from ORF3 to ORF5

and encodes a coat protein (P3) that has 201 aa and an MW of 22.3 kDa. ORF4 (nt 3875–4279) is 405 nt long, is nested within ORF3, and encodes the viral cell-to-cell movement protein (P4) of 134 aa, with an MW of 15.2 kDa [28]. ORF5 (nt 4396–5817) is 1,422 nt long and is expressed through a translational readthrough domain (RTD) due to suppression of the amber stop codon of ORF3. This results in an ORF referred to as ORF 3–5 (nt 3790–5817), which encodes the P3-P5 fusion protein, or readthrough protein (RTP), consisting of 675 aa, with an MW of 76.1 kDa. This protein is essential for aphid-mediated polerovirus transmission [9]. The viral 3'UTR (nt 5818–5997) is 180 nt in length. The 3'-proximal ORFs ORF3a, ORF3, ORF4, and ORF5 of poleroviruses are expressed from subgenomic mRNA (sgRNA) to synthesize the capsid protein (CP) and CP-RTD readthrough protein (RTP) and the P4 and P3a movement proteins (see above) [31].

Pairwise comparisons of the four CaLRV genome sequences from this study with those of other well-characterized poleroviruses indicated that they shared the lowest and highest nt identity of 58.3% and 62.2% with beet chlorosis virus (NC_002766) and potato leafroll virus (NC_076505), respectively (Table 1).

The P0 proteins of the four CaLRV isolates from this study shared the least aa sequence similarity (17.9% identity) with wheat leaf yellowing-associated virus (YP_009407906) and the most (28.4% identity) with cowpea polerovirus 1 (YP_009352244) (Table 1). The CaLRV P1 protein shared the least aa sequence similarity (27.4% identity) with the P1 protein of pumpkin polerovirus 1 (YP_010087204) and the most similarity (36.8% identity) with that of beet mild yellowing virus (NP_612214). The CaLRV P1-P2 protein shared the least aa sequence similarity (39.4% identity) with that of suakwa aphid-borne yellows virus (YP_006666506) and the most (48.5% identity) with that of beet mild yellowing virus (NP_620479). The CaLRV P3 protein is 36.7% identical to that of strawberry polerovirus-1 (YP_009100306) and 57.3% identical to that of melon aphid-borne yellows virus (YP_001949873). The CaLRV P4 protein shared the least aa sequence similarity (27.8% identity) with that of luffa aphid-borne yellows virus (YP_009162337) and the most aa sequence similarity (45.3% identity) with those of potato leafroll virus (NP_056750) and tobacco virus 2 (YP_009352891). The CaLRV P3-P5 protein shared the least aa sequence similarity (17.1% identity) with that of sugarcane yellow leaf virus (NP_050010) and the least (37.5% identity) with that of cereal yellow dwarf virus-RPV (NP_840025).

Based on the polerovirus species threshold of <75% nt sequence identity in the complete genome sequence [8] and >10% aa sequence divergence from any polerovirus protein [32], CLRaV should be considered a member of a new

species in the genus *Polerovirus*. The P3 protein of CaLRV shares no more than 57% aa sequence identity with any of the P3 proteins of ICTV-recognized poleroviruses, while the most divergent CaLRV protein, P3-P5, shares no more than 17% aa sequence identity with any of the other polerovirus proteins.

Bayesian phylogenetic analysis based on the complete genome sequence of CLRaV and reference polerovirus genome sequences obtained from the GenBank database revealed that the closest relatives of CLRaV are poleroviruses belonging to clade I, one of three well-supported sister groups (90–100% posterior probability), which exclude the highly divergent strawberry polerovirus 1 (88%) (Fig. 2).

The only known host of CLRaV is *T. cacao*, (family Malvaceae). However, CaLRV grouped with poleroviruses that infect plants of various families, including the Apiaceae, Brassicaceae, Fabaceae, Poaceae, Solanaceae, and others. This is consistent with the observation that many poleroviruses infect a wide variety of plants and indicates that host range does not closely correlate with genus-level evolutionary relationships. This is the first report of the complete genome sequence of the newly discovered polerovirus CLRaV. A recently published partial genome sequence of cotton leafroll dwarf virus (CLRVDV) recovered from commercial *T. cacao* trees in Brazil [3] and the nearly complete genome sequence of a polerovirus isolate from cacao, referred to as 'cacao polerovirus', which is nearly identical to that of CLRaV [33] and was assembled from transcriptome libraries of several *T. cacao* accessions originating from the germplasm collection maintained at CATIE (Centro Agronómico Tropical de Investigación y Enseñanza, Costa Rica), and CLRaV, are the first sequences from poleroviruses associated with symptomatic cacao trees.

Table 1 Comparison of the percent nucleotide and amino acid sequence identity values for the complete genome and predicted proteins, respectively, of cacao leafroll virus (CaLRV) and selected members of the genus *Polerovirus*

Virus	Accession number	Genome size (nt)	Nucleotide identity (%)	Amino acid similarity (%)						
				P0	P1	P1-P2 (RdRp)	P3 (CP)	P3a	P4 (MP)	P3-P5
Barley virus G	NC_029906.1	5,620	60.1	23.7	34.6	44.7	52.3	48.7	39.4	33.7
Maize yellow dwarf virus-RMV	NC_021484.1	5,612	60.1	24.3	33.3	44.5	49.7	37.8	36.4	29.7
Maize yellow dwarf virus-RMV2	NC_029990.1	5,625	59.5	24.3	31.5	-	50.3	40.0	40.9	33.3
Sauropus yellowing virus	NC_025837.1	5,820	58.9	22.2	30.9	40.7	46.8	42.5	36.6	29.6
African eggplant yellowing virus	NC_034207.1	5,953	61.3	24.6	35.1	44.4	51.6	46.2	40.3	31.7
Tobacco vein distorting virus	NC_010732.1	5,920	60.5	23.7	31.2	41.8	49.3	47.5	41.8	30.7
Pepper vein yellows virus	NC_015050.1	6,244	60.7	25.9	32.3	44.5	49.5	50.0	41.8	30.2
Pepper vein yellows virus 2	NC_055129.1	6,028	60.3	22.9	33.1	-	51.8	35.3	40.9	30.6
Pepper vein yellows virus 5	NC_036803.1	6,125	60.6	25.7	33.9	44.8	49.0	50.0	41.0	30.5
Cowpea polerovirus 1	NC_034246.1	5,845	59.8	28.4	30.7	44.0	50.5	45.0	42.0	32.7
Cowpea polerovirus 2	NC_034247.1	5,945	61.4	24.8	30.3	41.2	52.3	47.5	38.2	35.3
Chickpea chlorotic stunt virus	NC_008249.1	5,900	60.3	22.0	31.8	42.2	55.2	55.0	36.8	31.3
White clover mottle virus	NC_031747.1	6,205	59.7	27.9	33.6	45.0	53.6	52.5	35.1	31.0
Phasey bean mild yellows virus	NC_028793.2	5,851	60.7	23.4	29.6	39.9	56.0	52.5	36.8	32.8
Luffa aphid-borne yellows virus	NC_027703.1	5,961	60.5	20.8	32.6	46.0	43.9	53.1	27.8	26.2
Cotton leafroll dwarf virus	NC_014545.1	5,866	59.7	19.6	32.5	45.5	57.1	50.0	40.2	34.5
Pepo aphid-borne yellows virus	NC_030225.1	5,813	60.2	24.4	31.8	43.0	54.5	50.0	41.4	32.8
Pumpkin polerovirus 1	NC_055513.1	5,810	60.2	24.6	27.4	41.5	53.5	50.0	41.4	32.0
Suakwa aphid-borne yellows virus	NC_018571.2	5,843	60.2	26.3	28.5	39.4	55.2	50.0	39.6	32.2
Cucurbit aphid-borne yellows virus	NC_003688.1	5,669	60.1	22.4	34.9	45.0	55.1	50.0	43.4	31.6
Melon aphid-borne yellows virus	NC_010809.1	5,674	61.3	23.8	30.4	45.1	57.3	50.0	39.8	32.8
Wheat yellow dwarf virus-GPV	NC_012931.1	5,673	59.0	22.3	34.2	44.1	55.8	50.0	44.6	35.1
Cereal yellow dwarf virus-RPS	NC_002198.2	5,662	59.1	24.5	33.0	43.4	55.9	37.5	43.1	35.7
Cereal yellow dwarf virus-RPV	NC_004751.1	5,723	59.6	25.7	33.3	44.6	56.1	42.5	42.3	37.5
Potato leafroll virus	NC_001747.1	5,987	62.2	24.0	33.9	45.6	53.6	45.0	45.3	31.6
Tobacco virus 2	NC_034265.1	5,979	61.7	25.0	30.8	-	53.6	45.0	45.3	-
Carrot red leaf virus	NC_006265.1	5,723	60.8	20.8	35.9	45.0	49.2	42.5	40.3	32.4
Beet chlorosis virus	NC_002766.1	5,776	58.3	25.5	29.9	43.1	54.9	43.8	38.1	32.4
Beet mild yellowing virus	NC_003491.1	5,722	59.9	24.2	36.8	48.5	55.2	43.8	38.1	30.8
Beet western yellows virus	NC_004756.1	5,666	60.0	23.1	32.1	44.4	55.7	32.5	38.8	30.9
Faba bean polerovirus 1	NC_055495.1	5,631	60.7	25.1	33.6	44.8	56.7	35.0	35.8	32.2
Brassica yellows virus	NC_016038.2	5,666	60.3	24.3	34.2	46.6	56.5	47.5	39.6	19.7
Turnip yellows virus	NC_003743.1	5,641	60.0	22.2	34.5	45.7	56.0	47.5	38.8	31.9
Strawberry polerovirus 1	NC_025435.1	5,986	59.2	21.4	29.8	42.3	36.7	-	30.6	25.4
Sugarcane yellow leaf virus	NC_000874.1	5,899	59.1	23.2	29.9	42.0	40.3	59.0	38.5	17.1
Wheat leaf yellowing-associated virus	NC_035451.1	5,772	59.1	17.9	32.4	44.3	44.6	53.8	33.1	27.0

P0, silencing suppressor protein; P1, P1 protein; P1-P2, RNA-dependent RNA polymerase fusion protein; P3a, P3a protein; P3, coat protein (CP); P4, movement protein (MP); P3-P5; coat protein readthrough (CP-RTD) protein

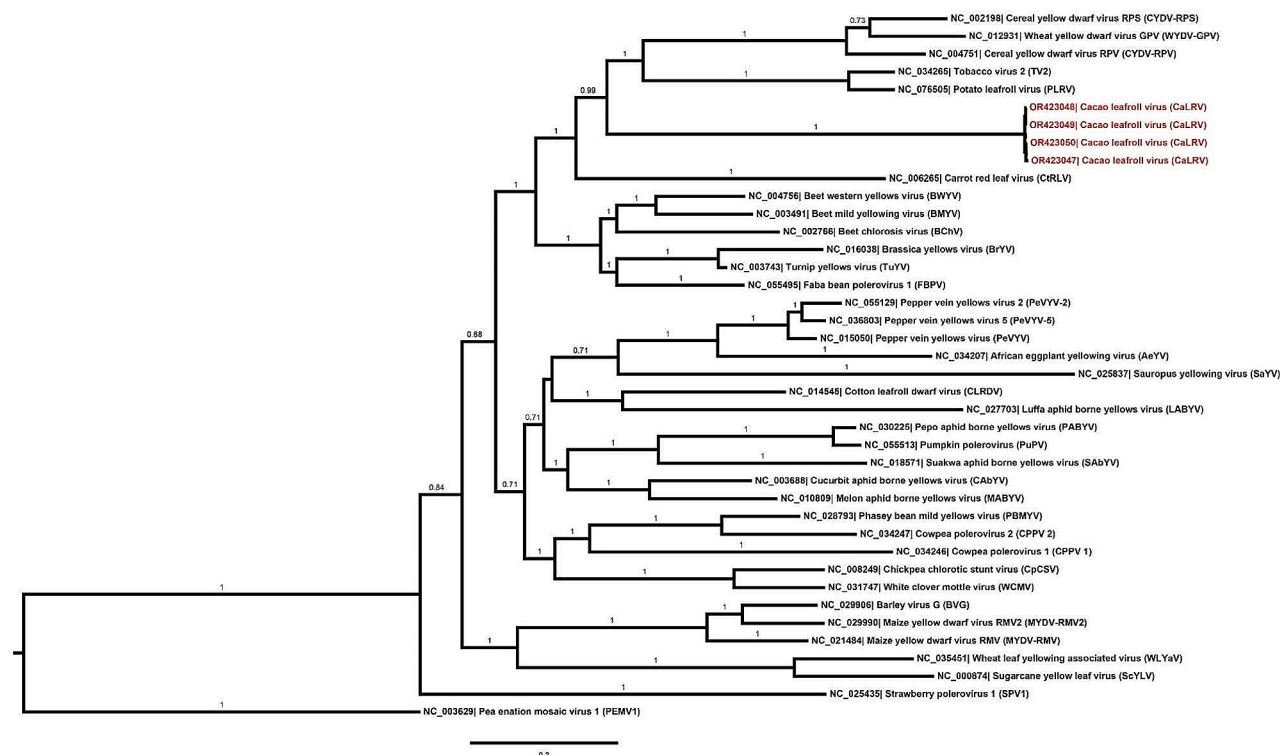


Fig. 2 Bayesian phylogenetic analysis of the genome sequence of cacao leafroll virus (CaLRV) and reference genome sequences of poleroviruses from the GenBank database. The tree was rooted using pea enation mosaic virus (PEMV, family *Solemoviridae*, genus *Enamovirus*). The evolutionary history was inferred using MrBayes 3.2.7a [18], with 50,000,000 generations (nruns=2) and a burn-in of 2,500 generations. The evolutionary distances were computed using

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00705-024-06013-7>.

Author contributions All authors contributed to the conception and design of the study. Funding was procured by Judith K. Brown. Laboratory experiments, data collection, and analysis were performed by Raphael O. Adegbola and Cory V. Keith. The first draft of the manuscript was written by Raphael O. Adegbola. All authors read, edited, and approved the final manuscript.

Funding This study was supported by a USDA-ARS Non-Assistance Cooperative Agreement to JKB: Number: 6038-21000-025-001-S titled: Discovery Sequencing for Improved PCR/qPCR Diagnostic Test Development to Identify the Presence of Badnaviruses in Cacao Germplasm Collections.

Data availability Nucleotide sequences of the newly reported CaLRV isolates are available in the NCBI GenBank database and have been assigned the accession number, OR423047-50.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

the general time-reversible (GTR) model with invariable sites. The rate variation among sites was modeled with a gamma distribution (GTR + I + G). Evolutionary analysis was carried out on the University of Arizona High Performance Computing (HPC) cluster (UA Research Data Center, Tucson, AZ). The phylogenetic tree was drawn using Fig-Tree v1.4.4 and edited in Adobe Illustrator 2019

References

- de Souza PA, Moreira LF, Sarmiento DH, da Costa FB (2018) Cacao-*Theobroma cacao*. In: Rodrigues S, de Oliveira Silva E, de Brito ES (eds) Exotic Fruits. Academic Press, pp 69–76. <https://doi.org/10.1016/B978-0-12-803138-4.00010-1>
- Kandito A, Hartono S, Trisyono YA, Somowiyarjo S (2022) First Report of Cacao Mild Mosaic Virus Associated with Cacao Mosaic Disease in Indonesia. New Dis Rep 45:e12071. <https://doi.org/10.1002/ndr2.12071>
- Ramos-Sobrinho R, Ferro MMM, Lima GSA, Nagata T (2022) First report of cotton leafroll dwarf virus infecting cacao (*Theobroma cacao*) trees in Brazil. Plant Dis 107:1251. <https://doi.org/10.1094/PDIS-07-22-1570-PDN>
- Ramos-Sobrinho R, Ferro MMM, Nagata T, Puig AS, Keith CV, Britto DS, Gutierrez OA, Marelli J-P, Brown JK (2021) Complete genome sequences of three newly discovered cacao mild mosaic virus isolates from *Theobroma cacao* L. in Brazil and Puerto Rico and evidence for recombination. Arch Virol 166:2027–2031. <https://doi.org/10.1007/s00705-021-05063-5>
- Ullah I, Daymond AJ, Hadley P, End MJ, Umaharan P, Dunwell JM (2021) Identification of cacao mild mosaic virus (CaMMV) and cacao yellow vein-banding virus (CYVBV) in cocoa (*Theobroma cacao*) Germplasm. Viruses 13:2152
- Ramos-Sobrinho R, Chingandu N, Gutierrez OA, Marelli J-P, Brown JK (2020) A complex of badnavirus species infecting cacao reveals mixed infections, extensive genomic variability,

- and interspecific recombination. *Viruses* 12:443. <https://doi.org/10.3390/v12040443>
7. Muller E, Ravel S, Agret C, Abrokwah F, Dzahini-Obiatey H, Galyuon I, Kouakou K, Jeyaseelan EC, Allainguillaume J, Wetten A (2018) Next generation sequencing elucidates cacao badnavirus diversity and reveals the existence of more than ten viral species. *Virus Res* 244:235–251. <https://doi.org/10.1016/j.virusres.2017.11.019>
 8. Sömera M, Fargette D, Hébrard E, Sarmiento C (2021) ICTV Report Consortium. ICTV virus taxonomy profile: Solemoviridae 2021. *J Gen Virol* 102:12. <https://doi.org/10.1099/jgv.0.001707>
 9. Brault V, Bergdoll M, Mutterer J, Prasad V, Pfeffer S, Erdinger M, Richards KE, Ziegler-Graff V (2003) Effects of point mutations in the major capsid protein of beet western yellows virus on capsid formation, virus accumulation, and aphid transmission. *J Virol* 77:3247–3256. <https://doi.org/10.1128/jvi.77.5.3247-3256.2003>
 10. Delfosse VC, Barrios Barón MP, Distéfano AJ (2021) What we know about poleroviruses: Advances in understanding the functions of polerovirus proteins. *Plant Pathol* 70:1047–1061. <https://doi.org/10.1111/ppa.13336>
 11. Adegbola RO, Keith CV, Gutierrez OA, Goenaga R, Brown JK (2022) A Previously undescribed polerovirus (Solemoviridae) Infecting *Theobroma cacao* germplasm. *Plant Dis* 107:975. <https://doi.org/10.1094/PDIS-06-22-1449-PDN>
 12. Turnbull CJ, Hadley P (2021) International Cocoa Germplasm Database (ICGD). CRA Ltd./ICE Futures Europe/University of Reading, UK
 13. Rott ME, Jelkmann W (2001) 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* 107:411–420. <https://doi.org/10.1023/A:1011264400482>
 14. Pribelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A (2020) Using SPAdes de novo assembler. *Curr Protocols Bioinform* 70:e102. <https://doi.org/10.1002/cpbi.102>
 15. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
 16. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. <https://doi.org/10.1093/nar/gkh340>
 17. Muhire BM, Varsani A, Martin DP (2014) SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. *PLoS ONE* 9:e108277. <https://doi.org/10.1371/journal.pone.0108277>
 18. Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61:539–542. <https://doi.org/10.1093/sysbio/sys029>
 19. Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL et al (2023) InterPro in 2022. *Nucleic Acids Res* 51:D418–D427. <https://doi.org/10.1093/nar/gkac993>
 20. Theis C, Reeder J, Giegerich R (2008) KnotInFrame: prediction of -1 ribosomal frameshift events. *Nucleic Acids Res* 36:6013–6020. <https://doi.org/10.1093/nar/gkn578>
 21. Knierim D, Tsai WS, Deng TC, Green SK, Kenyon L (2013) Full-length genome sequences of four polerovirus isolates infecting cucurbits in Taiwan determined from total RNA extracted from field samples. *Plant Pathol* 62:633–641. <https://doi.org/10.1111/j.1365-3059.2012.02653.x>
 22. Bergkessel M, Guthrie C (2013) Colony PCR. *Methods in enzymology*, vol 529. Academic, Cambridge, MA, pp 299–309
 23. Pazhouhandeh M, Dieterle M, Marocco K, Lechner E, Berry B, Brault V, Hemmer O, Kretsch T, Richards KE, Genschik P, Ziegler-Graff V (2006) F-box-like domain in the polerovirus protein P0 is required for silencing suppressor function. *P Natl Acad Sci USA* 103:1994–1999. <https://doi.org/10.1073/pnas.0510784103>
 24. Pfeffer S, Dunoyer P, Heim F, Richards KE, Jonard G, Ziegler-Graff V (2002) P0 of beet Western yellows virus is a suppressor of posttranscriptional gene silencing. *J Virol* 76:6815–6824. <https://doi.org/10.1128/jvi.76.13.6815-6824.2002>
 25. Nickel H, Kawchuk L, Twyman RM, Zimmermann S, Junghans H, Winter S, Fischer R, Prüfer D (2008) Plantibody-mediated inhibition of the Potato leafroll virus P1 protein reduces virus accumulation. *Virus Res* 136:140–145. <https://doi.org/10.1016/j.virusres.2008.05.001>
 26. Miller WA, Liu S, Beckett R (2002) Barley yellow dwarf virus: Luteoviridae or Tombusviridae? *Mol Plant Pathol* 3:177–183. <https://doi.org/10.1046/j.1364-3703.2002.00112.x>
 27. Miller WA, Dinesh-Kumar SP, Paul CP (1995) Luteovirus gene expression. *CRC Crit Rev Plant Sci* 14:179–211. <https://doi.org/10.1080/07352689509701926>
 28. Kamer G, Argos P (1984) Primary structural comparison of RNA-dependent polymerases from plant, animal and bacterial viruses. *Nucleic Acids Res* 12:7269–7282. <https://doi.org/10.1093/nar/12.18.7269>
 29. Smirnova E, Firth AE, Miller WA, Scheidecker D, Brault V, Reinbold C, Rakotondrafara AM, Chung BY, Ziegler-Graff V (2015) Discovery of a small non-AUG-initiated ORF in poleroviruses and luteoviruses that is required for long-distance movement. *PLoS Pathog* 11:e1004868. <https://doi.org/10.1371/journal.ppat.1004868>
 30. Tacke E, Schmitz J, Prüfer D, Rohde W (1993) Mutational analysis of the nucleic acid-binding 17 kDa phosphoprotein of potato leaf-roll luteovirus identifies an amphipathic alpha-helix as the domain for protein/protein interactions. *Virology* 197:274–282. <https://doi.org/10.1006/viro.1993.1588>
 31. Hwang YT, Kalischuk M, Fusaro AF, Waterhouse PM, Kawchuk L (2013) Small RNA sequencing of Potato leafroll virus-infected plants reveals an additional subgenomic RNA encoding a sequence-specific RNA-binding protein. *Virology* 438:61–69. <https://doi.org/10.1016/j.virol.2012.12.012>
 32. Domier LL (2012) Family Luteoviridae. In: King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (eds) *Virus taxonomy*, ninth report of the international committee on taxonomy of viruses. Elsevier/Academic, London, pp 1045–1053
 33. Ullah I, Kamran M, Dunwell JM (2023) Identification of a novel polerovirus in cocoa (*Theobroma cacao*) germplasm and development of molecular methods for use in diagnostics. *Pathogens* 12:1284. <https://doi.org/10.3390/pathogens12111284>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.