Complete genome sequence of a novel fusarivirus from the phytopathogenic fungus *Corynespora cassiicola*

Mingming Liu (lmm1503827@126.com)
Henan Academy of Agricultural Sciences https://orcid.org/0000-0003-0580-0341
Xintao Liu
Henan Agricultural Academy of Science: Henan Academy of Agricultural Sciences
Hui Zhao
Henan Agricultural Academy of Science: Henan Academy of Agricultural Sciences
Yunxia Ni
Henan Agricultural Academy of Science: Henan Academy of Agricultural Sciences
Min Jia
Henan Agricultural Academy of Science: Henan Academy of Agricultural Sciences
Peilin Hu
Henan Agricultural Academy of Science: Henan Academy of Agricultural Sciences
Hongyan Liu
Henan Agricultural Academy of Science: Henan Academy of Agricultural Sciences
Baoming Tian
Zhengzhou University

Research Article

Keywords: Mycovirus, Fusariviridae, Corynespora cassiicola

Posted Date: November 29th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1094125/v1

Version of Record: A version of this preprint was published at Archives of Virology on April 9th, 2022. See the published version at https://doi.org/10.1007/s00705-022-05428-4.

Abstract

Corynespora cassiicola is an important phytopathogenic fungus and it has severely impaired the production of crops. In this study, we report on the molecular characterization of a novel (+) ssRNA mycovirus, Corynespora cassiicola fusarivirus 1 (CcFV1) isolated from *C. cassiicola* strain 20200826-3-1. Excluding a poly (A) tail, the genome of the virus is 6491 nt containing three putative open reading frames. The large ORF1 encodes a polypeptide of 1524aa with a conserved RNA-dependent RNA polymerase (RdRp) domain, a helicase (Hel) domain, and a Phage-holin-3-6 (Phage-holin) domain. ORF2 encodes a polypeptide with a conserved Chromosome segregation ATPase (Smc) domain. The smallest ORF3 encodes a putative protein with an unknown function. Phylogenetic analysis based on the ORF1 and ORF2 of CcFV1 encoded polypeptide showed that CcFV1 is phylogenetically related to the newly proposed family *Fusariviridae*. Thus, we suggest that CcFV1 might be a novel member of the family *Fusariviridae* and is also the first discovered in *C. cassiicola*.

Introduction

Mycoviruses are ubiquitous in all major fungi, which are transmitted intracellularly by cell division, sporogenesis, and cell-to-cell fusion [1]. Most fungal virus genomes consist of double-stranded RNA (dsRNA) or positive single-stranded RNA (+ssRNA) [1]. In addition, single-stranded circular DNA (ssDNA) viruses and negative-stranded RNA (-ssRNA) viruses have recently been reported in filamentous fungi [2, 3]. Mycoviruses with +ssRNA genomes are classified into eight families, including *Alphaflexiviridae*, *Barnaviridae*, *Botourmiaviridae*, *Deltaflexiviridae*, *Endornaviridae*, *Gammaflexiviridae*, *Hypoviridae*, and *Narnaviridae* (https://talk.ictvonline.org/taxonomy). In 2014, a new family named "*Fusariviridae*" is proposed in which the genomes in the family are +ssRNA with a size of 6-10 kb [4]. The viruses of *Fusariviridae* infect various phytopathogenic fungi, including *Nigrospora oryzae* [5], *Neofusicoccum luteum* [6], *Botryosphaeria dothidea* [7], *Setosphaeria turcica* [8], *Alternaria solani* [9].

Corynespora cassiicola is the phytopathogenic fungus of the most devastating leaf disease[10], which has severely impaired the production of important crops. It has been found on plant leaves, stems, and roots; nematode cysts; and human skin [11–16]. The phytopathogenic fungus *C. cassiicola* has a wide host range which has been reported to be growing on at least 530 plant species from 380 genera [17]. However, so far, no fusarivirus has been reported to infect this important phytopathogenic fungus. This is also the first report of the complete genome sequence of a fusarivirus infecting *C. cassiicola*, so the mycovirus is tentatively named "Corynespora cassiicola fusarivirus 1" (CcFV1).

The Provenance Of The Virus Material

Seven isolated strains of *C. cassiicola* used in this study were preserved at the Institute of Plant Protection, Henan Academy of Agricultural Sciences, China. For total RNA extraction, *C. cassiicola* strains were inoculated on potato dextrose agar (PDA) plates at 28℃ in the dark for eight days. The total RNA of each strain was extracted from 0.2g mycelium using an RNAiso Kit (Takara, Dalian, China) following the

manufacturer's instructions. The RNA concentrations of the 7 strains were adjusted to 200ng/ul and then 15ul of each strain was pooled together.

The mixed sample was sent to Shanghai Bohao Biotechnology Corporation for high-throughput sequencing. The sequencing library was prepared from rRNA-depleted total RNA of seven *C. cassiicola* isolates and the cDNA was sequenced using an Illumina Hiseq 2500 platform. The clean reads that can be used for data analysis were finally obtained by removing the unqualified raw reads based on default parameters. These clean reads were assembled *de novo* by the scaffolding contig algorithm in CLC Genomics Workbench (version: 6.0.4) to obtain the primary unigenes. The final unigenes sequences were obtained by applying the CAP3 EST splicing software for the second splicing of the primary unigenes. The final unigenes were screened by blastx against the NCBI GenBank database for homologous viral sequences. Through the above methods, we found that contig2308 was identified as a novel fusarivirus.

The cDNAs of seven isolated *C. cassiicola* strains were synthesized following the instruction with the PrimerScript II^{TM} 1st Strand cDNA Synthesis Kit (TaKaRa, Dalian, China). To confirm the newly discovered viruses in each isolated strain, the virus-specific primer designed based on contig2308 was used for reverse transcriptase PCR (RT-PCR) to detect specific amplicon. The specific amplicon of contig2308 was detected in the target *C. cassiicola* strain 20200826-3-1 which was isolated from the diseased stem of a sesame plant from Zhoukou county, Henan province, China. The full length of the contig2308 was verified through the assembling of 16 pairs of specific primer-amplified fragments (Supplementary Table S1).

To complete CcFV1 the 5'- and 3'-terminal genomic sequences, rapid amplification of cDNA ends (RACE) was performed using a SMARTer RACE 5'/3' Kit (TaKaRa, China). The 5'-RACE reaction was carried out with the primer R_GPS_Contig2308 (5'-CGGGTTGAAACCTGAACGGTTCGTGTA-3') to amplify the 5' cDNA end. The special primer F_GPS_Contig2308 (5'-TTCGATTGCGGAAACCTCCTGTTGAC-3') was used for the 3'-RACE reaction. The procedures were performed according to the user manual provided with the kit. All PCR products were purified by FastPure Gel DNA Extraction Kit (Vazyme, China) and cloned into the pMD^{TM}18-T vector, and then introduced into JM109 Chemically Competent Cell (TSINGKE, China) for propagation. At least three recombinant clone bacterial suspensions were selected for sending to Sangon Biotech for sequencing to verify the CcFV1 nucleotide sequence accuracy.

The ORF Finder and CD-search in the NCBI (https://www.ncbi.nlm.nih.gov/) were used to search for open reading frames and conserved proteins, respectively. The sequences obtained from clones were assembled using DNAMAN software. Multiple alignments of the amino acid sequence were performed with Clustal X and DNAMAN software. The phylogenetic trees were constructed by the maximum-likelihood (ML) method with 1000 bootstrap replicates and edited using MEGA (7.0 version) [18]. The GC content was determined by NoVopro (https://www.novopro.cn/tools/gc-content.html). The Expasy was used to calculate the protein molecular weight (Mw) and isoelectric point (pI) (https://web.expasy.org/compute_pi/). The schematic diagram of the genome organization of CcFV1 was drawn through the Illustrator for Biological Sequence (http://ibs.biocuckoo.org/dbvisualization.php).

Sequence Properties

The full-length cDNA sequence of CcFV1(GenBank accession number: OL456888)is 6491 nt long excluding the poly (A) tail and has a GC content of 43.54%. It comprises a positive-stranded RNA containing three disconnected two main ORFs (ORF1 and ORF2) and a smaller ORF3. The lengths of the 5'-UTR and 3'-UTR are 18nt and 47nt, respectively (Fig. 1A). A BLASTx search against the NCBI non-redundant protein sequence database showed that the three highest degree of viruses matching with CcFV1 are Setosphaeria turcica fusarivirus 1 (StFV1, 50.10% identity, E-value 0.0), Erysiphe necator associated fusarivirus 2 (EnFV2, 52.23% identity, E-value 0.0), and Plasmopara viticola lesion associated fusarivirus 1 (PvFV1, 50.10% identity, E-value 0.0). The three best-matched viruses with CcFV1 all belong to the newly proposed *Fusariviridae* family.

ORF1 (19-4593nt) encodes a polyprotein of 1524 amino acids (aa) with an approximate molecular weight (Mw) of 172.59kDa and an isoelectric point of 8.98. A BLASTp search against the NCBI non-redundant protein sequence database showed that this protein is 50.45%, 51.26%, and 49.55% identical to RdRp of StFV1, EnFV2, PvFV1 and the top 12 species with the highest degree of matching with this protein are all *Fusariviridae* viruses (Table 1).

Using the CD search program on the NCBI website, two conserved sequence domains: an RdRp (RdRp_1, pfam00680, E-value 7.80e-09, from aa position 479-741) and a helicase (Helicase_C, pfam00271, E-value 2.77e-09, from aa position 1229-1341) were predicted in the ORF1-encoded polyprotein, which was also demonstrated in other reported fusariviruses. The aa sequence similarities in the RdRp and Hel regions indicated that CcFV1 is a fusarivirus consisting of a (+) ssRNA genome. In addition, the protein was also predicted to contain a Phage_holin conserved domain (Phage_holin_3_6, pfam07332, E-value 9.56e-03, from aa position 17-84), which is a family of small hydrophobic proteins with two or three transmembrane domains, which is not predicted in the other three best-matched fusariviruses (Fig. 1A). Multiple alignments and comparisons of the viral RdRp domains between CcFV1 and other selected fusariviruses revealed eight conserved motifs that are characteristic of the mycoviruses (Fig. 1B).

ORF2 (4645-6144nt) encodes an Mw 56.68kDa and pI 9.06 putative protein (499aa). A BLASTp search revealed only eight similar proteins which are all fusariviruses (Table 1). The protein encoded by ORF2 has the highest sequence similarity to EnFV2 (35.42% identity, E-value 2.00e-87). A conserved domain was identified in this protein which is chromosome segregation ATPase (Smc, COG1196, E-value 1.50e-07) spanning aa positions from 89 to 246 and was also predicted in other fusariviruses (Fig. 1A).

ORF3, initiated from 6145 to 6444nt, was predicted to express a 99 aa hypothetical protein with 11.52kDa and pI 9.23. Homology search of the protein encoded by ORF3 indicated that no aa sequence similarity existed with any other proteins and conserved domains. Compared with StFV1, EnFV2, PvFV1, the ORF3 and ORF2 of CcFV1 consecutive arrangements have no overlap region (Fig. 1A).

To examine the relationship between CcFV1 and other mycoviruses, we performed phylogenetic tree analysis using alignments of the polyproteins encoded by ORF1 (Fig. 2A) and ORF2 (Fig. 2B) of CcFV1

and other selected RNA viruses. The sequences of selected *Hypoviridae* and *Totiviridae* were included to serve as outgroups. The results of the two phylogenetic trees all showed that CcFV1 clustered with the previously reported fusariviruses.

In summary, the organization and structure of the CcFV1 genome are similar to other reported fusariviruses, and the analysis of the phylogenetic trees also showed that CcFV1 is phylogenetically related to the newly proposed *Fusariviridae* family. Thus, we deduced that CcFV1 is a potential new member of the *Fusariviridae* family and is also the first report of a novel fusarivirus in *C. cassiicola*.

Declarations

Compliance with ethical standards

Funding

This work was supported by the China Agriculture Research System of MOF and MARA [grant number CARS-14], Key Project of Science and Technology of Henan Province [grant number 201300110600].

Conflict of interest

All authors declare that they have no conflicts of interest.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Acknowledgments

We are extremely grateful to Dr. Jing Wang for her valuable suggestions and help with sequence analysis.

References

1. Ghabrial SA, Castón JR, Jiang D, et al (2015) 50-plus years of fungal viruses. Virology 479–480:356–368. https://doi.org/https://doi.org/10.1016/j.virol.2015.02.034

2. Yu X, Li B, Fu Y, et al (2010) A geminivirus-related DNA mycovirus that confers hypovirulence to a plant pathogenic fungus. Proc Natl Acad Sci 107:8387 LP – 8392. https://doi.org/10.1073/pnas.0913535107

3. Liu L, Xie J, Cheng J, et al (2014) Fungal negative-stranded RNA virus that is related to bornaviruses and nyaviruses. Proc Natl Acad Sci 111:12205 LP – 12210. https://doi.org/10.1073/pnas.1401786111

4. Zhang R, Liu S, Chiba S, et al (2014) A novel single-stranded RNA virus isolated from a phytopathogenic filamentous fungus, Rosellinia necatrix, with similarity to hypo-like viruses. Front Microbiol 5:1–12. https://doi.org/10.3389/fmicb.2014.00360

5. Zhong J, Zhao SQ, Li GF, et al (2016) A novel fusarivirus isolated from the phytopathogenic fungus Nigrospora oryzae. Virus Genes 52:891–895. https://doi.org/10.1007/s11262-016-1371-5

6. Marais A, Nivault A, Faure C, et al (2018) Molecular characterization of a novel fusarivirus infecting the plant-pathogenic fungus Neofusicoccum luteum. Arch Virol 163:559–562. https://doi.org/10.1007/s00705-017-3620-x

7. Liu W, Hai D, Mu F, et al (2020) Molecular characterization of a novel fusarivirus infecting the plant-pathogenic fungus Botryosphaeria dothidea. Arch Virol 165:1033–1037. https://doi.org/10.1007/s00705-020-04554-1

8. Gao Z, Cai L, Liu M, et al (2021) A novel previously undescribed fusarivirus from the phytopathogenic fungus Setosphaeria turcica. Arch Virol 166:665–669. https://doi.org/10.1007/s00705-021-04954-x

9. Gong W, Liu H, Zhu X, et al (2021) Molecular characterization of a novel fusarivirus infecting the plant-pathogenic fungus Alternaria solani. Arch Virol 166:2063–2067. https://doi.org/10.1007/s00705-021-05105-y

10. Silva WPK, Wijesundera RLC, Karunanayake EH, et al (2000) New hosts of corynespora cassiicola in Sri Lanka. Plant Dis 84:202. https://doi.org/10.1094/pdis.2000.84.2.202d

11. Madriz-Ordeana K, Jrgensen HJL, Nielsen KL, Thordal-Christensen H (2016) First Report of Kalanchoe Leaf and Stem Spot Caused by Corynespora cassiicola in Denmark. Plant Dis 101:

12. Carris LM, Glawe DA, Gray LE (1986) Isolation of the Soybean Pathogens Corynespora Cassiicola and Phialophora Gregata from Cysts of Heterodera Glycines in Illinois. Mycologia 78:503–506

13. Sato R, Kitazawa K (1980) Occurrence of Soybean Root Rot Caused by Corynespora cassiicola (Berk. & Curt.) Wei in Hokkaido. Japanese J Phytopathol 46:193–199

14. Dixon LJ, Schlub RL, Pernezny K, Datnoff LE (2009) Host specialization and phylogenetic diversity of corynespora cassiicola. Phytopathology 99:1015–1027. https://doi.org/10.1094/PHYTO-99-9-1015

15. Kusakari S, Okada K, Nakasone W, Tanaka Y (2009) Corynespora Leaf Spot of Perilla(Perilla frutescens Britt.)Caused by Corynespora cassiicola(Berk. & Curt.)Wei. Japanese J Phytopathol 57:737–740

16. Yan XX, Yu CP, Fu XA, et al (2016) CARD9 mutation linked to Corynespora cassiicola infection in a Chinese patient. Br J Dermatol 174:176–179. https://doi.org/https://doi.org/10.1111/bjd.14082

17. Smith LJ (2008) Host range, phylogenetic, and pathogenic diversity of Corynespora cassiicola (Berk. & Curt.) Wei. Diss Theses - Gradworks

18. Tamura K, Peterson D, Peterson N, et al (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol Biol Evol 28:2731–2739. https://doi.org/10.1093/molbev/msr121

Tables

Table 1. Comparison information of ORF1 and ORF2 coding proteins of CcFV1 by Blastp

	Scientific Name	Coverage	E value	Identity(%)	Length	Accession
ORF1	Setosphaeria turcica fusarivirus 1	99%	0	50.45	1546	QRI93681.1
	Erysiphe necator associated fusarivirus 2	99%	0	51.26	1545	QKN22682.1
	Erysiphe necator associated fusarivirus 1	99%	0	49.55	1549	QHD64833.1
	Plasmopara viticola lesion associated fusarivirus 1	99%	0	49.55	1549	QHD64725.1
	Pleospora typhicola fusarivirus 1	99%	0	48.94	1548	YP_009182158.1
	Plasmopara viticola lesion associated fusarivirus 3	96%	0	50.13	1502	QHD64735.1
	Sodiomyces alkalinus fusarivirus 1	99%	0	48.11	1527	YP_009551681.1
	Sodiomyces alkalinus fusarivirus 1	95%	0	48.47	1470	ATP75829.1
	Sodiomyces alkalinus fusarivirus 1	94%	0	48.87	1450	ATP75827.1
	Fusariviridae sp.	98%	0	46.9	1520	QDH87553.1
	Neurospora discreta fusarivirus 1	99%	0	46.88	1526	AZT88657.1
	Fusarium poae fusarivirus 1	98%	0	46.62	1501	YP_009272906.1
ORF2	Erysiphe necator associated fusarivirus 2	98%	2.00E-87	35.42	529	QKN22683.1
	Plasmopara viticola lesion associated fusarivirus 1	98%	9.00E-85	35.56	526	QHD64726.1
	Erysiphe necator associated fusarivirus 1	98%	5.00E-84	35.56	526	QHD64834.1
	Pleospora typhicola fusarivirus 1	98%	2.00E-80	34.03	527	YP_009182159.1
	Plasmopara viticola lesion associated fusarivirus 3	98%	1.00E-79	33.78	526	QHD64736.1
	Setosphaeria turcica fusarivirus 1	99%	5.00E-77	34.09	522	QRI93682.1
	Fusariviridae sp.	97%	4.00E-53	30.02	512	QDH87552.1
	Grapevine-associated fusarivirus 1	87%	8.00E-50	30.20	497	QXN75345.1

Figures

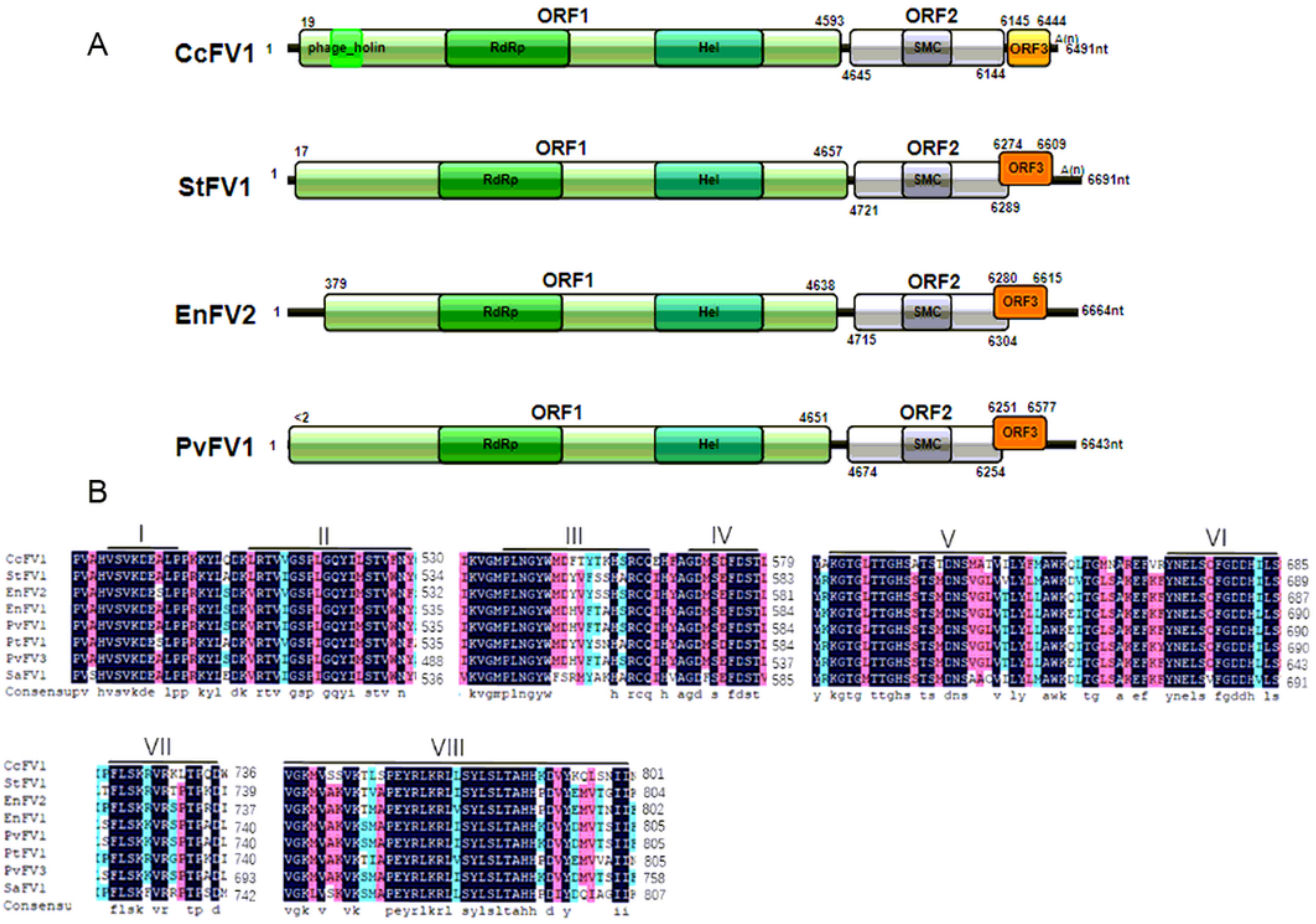

Figure 1

Schematic representations of the genomic organizations and the RdRp conserved motifs. (A) The genome organization of CcFV1 and three of its close relative fusariviruses: Setosphaeria turcica fusarivirus 1 (StFV1, MT955613.1), Erysiphe necator associated fusarivirus 2 (EnFV2, MN627464.1), and Plasmopara viticola lesion associated fusarivirus 1 (PvFV1, MN551102.1). The conserved RNA-dependent RNA polymerase (RdRp), helicase (Hel) domains, and a Phage-holin-3-6 (Phage-holin) domain of CcFV1 are indicated on the ORF1 and chromosome segregation ATPase (Smc) is indicated on the ORF2. (B) Multiple alignments of the amino acid sequence of the conserved motifs in RdRps of CcFV1 and other homologous fusariviruses: Setosphaeria turcica fusarivirus 1 (StFV1, QRI93681.1), Erysiphe necator associated fusarivirus 2 (EnFV2, QKN22682.1), Erysiphe necator associated fusarivirus 1 (EnFV1, QHD64833.1), Plasmopara viticola lesion associated fusarivirus 1 (PvFV1, QHD64725.1), Pleospora typhicola fusarivirus 1 (PtFV1, YP_009182158.1), Plasmopara viticola lesion associated fusarivirus 3 (PvFV3, QHD64735.1), Sodiomyces alkalinus fusarivirus 1 (SaFV1, YP_009551681.1). The I–VIII delegate the eight conserved motifs of RdRps in these referenced fusariviruses.

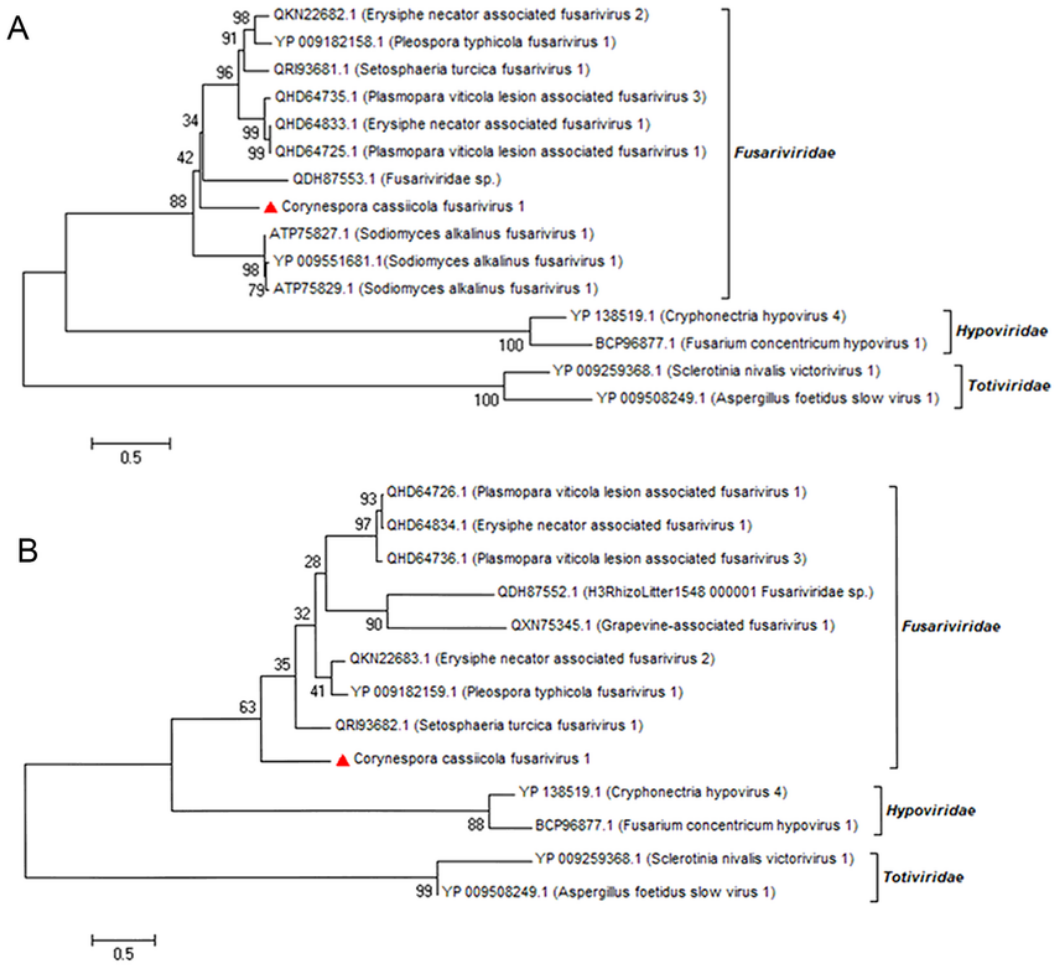

Figure 2

Phylogenetic analyses of CcFV1 and other related RNA viruses. Two hypoviruses and two totiviruses were included as outgroups. (A) The phylogenetic tree was constructed using protein alignments of the polyprotein from ORF1 of CcFV1 and other selected viruses. (B) The phylogenetic tree was constructed using alignments of the polyprotein coded by ORF2 of CcFV1 and other relative viruses. The CcFV1 was marked with a red triangle.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1.docx