



Identification and Genome Characterization of a Dahlia Common Mosaic Virus Isolate from China

Jing Wang ^{1,2}, Jiaying Zhou ¹, Ming Chen ¹, Zhengnan Li ^{3,*} and Shuang Song ^{1,2,*}

- ¹ College of Plant Protection, Northeast Agricultural University, Harbin 150030, China
- ² College of Agriculture, Northeast Agricultural University, Harbin 150030, China
- ³ College of Horticulture and Plant Protection, Inner Mongolia Agricultural University, Hohhot 010018, China
- * Correspondence: lizhengnan@imau.edu.cn (Z.L.); songshuang.000@163.com (S.S.)

Abstract: Dahlia (*Dahlia variabilis*) is a widely cultivated ornamental and medicinal plant in China. Recently, dahlia plants with symptoms of leaf mottling and distortion were collected in Hohhot, Inner Mongolia, China. The presence of dahlia common mosaic virus (DCMV), an unassigned species in the genus *Caulimovirus*, was confirmed by high-throughput sequencing. Three fragments of DCMV Inner Mongolia isolate (DCMV-IN) were PCR-amplified with specific primers, sequenced and assembled into the complete genome sequence with a GenBank accession number of OR494328. The double-stranded circular DNA genome of DCMV-IN consists of 7949 bp and contains six open reading frames (ORFs). Sequence analysis showed that DCMV-IN shared high sequence identities with other DCMV isolates available in the GenBank database. Phylogenetic analysis of DCMV isolates and other representative caulimoviruses based on genome sequence clustered four DCMV isolates to a single branch which was closest to dahlia mosaic virus (DMV). No recombination event was detected among the four DCMV isolates.

Keywords: dahlia; high-throughput sequencing; genome; phylogenetic analysis



Citation: Wang, J.; Zhou, J.; Chen, M.; Li, Z.; Song, S. Identification and Genome Characterization of a Dahlia Common Mosaic Virus Isolate from China. *Genes* **2023**, *14*, 1833. https://doi.org/10.3390/ genes14101833

Academic Editor: Zhen He

Received: 2 September 2023 Revised: 18 September 2023 Accepted: 19 September 2023 Published: 22 September 2023



Copyright: © 2023 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

Dahlia (*Dahlia variabilis*, genus *Dahlia*, family Asteraceae), also known as Dongyangju and daliju in Chinese, is a perennial herbaceous plant. It possesses not only ornamental but also medicinal value [1]. It has been reported to contain high levels of inulin, which can be used to reduce blood glucose levels for diabetic patients [2]. Dahlia is native to Mexico and is widely cultivated in Yunnan, Guizhou, Inner Mongolia, and other regions of China. Due to the fact that dahlia is usually propagated asexually by vegetative tuberous roots in cultivation, viral infections become a significant threat to dahlias [3]. Dahlia mosaic disease is a widespread and detrimental disease in dahlia with a variety of symptoms such as distortion, mosaic and chlorosis of leaves and overall stunting of plant [4,5]. Three caulimoviruses including dahlia mosaic virus (DMV), dahlia common mosaic virus (DCMV) and DMV-D10 have been reported to be associated with dahlia mosaic disease [6–8].

DCMV is an unassigned species of the genus *Caulimovirus* in the family *Caulimoviridae*, which was first reported in the Netherlands in 2008 [9]. It has a wide range of hosts including the members of the family Asteraceae, Brassicaceae, Amaranthaceae and Chenopodiaceae, and it can be transmitted by mechanical inoculation and aphids in a non-persistent manner [8]. DCMV has an approximately 8 kb circular double-stranded genomic DNA [10]. The genome organization of DCMV is consistent with that of typical members of the genus *Caulimovirus*, consisting of at least six open reading frames (ORFs) (ORF I–VI) that successively encode movement protein (MP), aphid transmission factor (ATF), putative DNA-binding protein (DNAb), capsid protein (CP), polymerase polyprotein (PP)-containing protease, reverse transcriptase and RNAse H and an inclusion body protein (IB) [9,11].

Recently, dahlia plants displaying virus-like symptoms of leaf mottling and distortion (Figure 1) were discovered in the campus of Inner Mongolia Agricultural University. The

presence of DCMV infection in dahlia plants was confirmed through high-throughput sequencing (HTS) and PCR confirmation. Furthermore, the complete genome sequence of the virus was sequenced and analyzed.



Figure 1. The observed dahlia plants with symptoms of leaf mottling and distortion.

2. Materials and Methods

2.1. Sampling of the Virus Source

In July 2022, Dahlia samples showing obvious virus-like disease symptoms of leaf mottling and distortion (Figure 1) were observed in the campus of Inner Mongolia Agricultural University in Hohhot, Inner Mongolia, China. Leaf samples were collected, immediately frozen using liquid nitrogen and transferred to -80 °C for further research.

2.2. High-Throughput Sequencing (HTS) and Data Analysis

To identify the virus infecting dahlia, total RNA was extracted from the pooled symptomatic leaf samples from three different plants, with each sample weighing approximately 200 mg. TRIzol reagent (Invitrogen, Carlsbad, CA, USA) was used for the extraction following the manufacturer's instructions. Ribosomal RNA (rRNA) was removed from the total RNA using a Ribo-Zero rRNA Removal Kit (Epicentre, Madison, WI, USA), and the rRNA-depleted RNA was used to construct a cDNA library using a TruSeq RNA Sample Prep Kit (Illumina, San Diego, CA, USA). The prepared library was sequenced using an Illumina NovaSeq 6000 platform with a read length of 2×150 bp. The raw data obtained from the original image were processed by trimming adapter sequences and getting rid of low-quality reads of length < 75 bp using the software Trimmomatic v.0.36 [12]. The taxonomy of the clean reads were then assembled de novo into contigs using MegaHit with '—min-contig-len 500' parameters [14]. The obtained contigs were subjected to BLAST analysis against the NCBI nucleotide (nt) sequences database and NCBI non-redundant protein (nr) database, with an e-value cutoff of 10^{-5} [15].

2.3. Determination of DCMV Genome

To verify the presence of DCMV in the sample identified by HTS and determine its complete genome sequence, total DNA was extracted from the pooled symptomatic leaf samples used for HTS, using EasyPure Plant Genomic DNA Kit (TransGen, Beijing, China). PCR was performed using three specific primer pairs designed based on the DCMV contigs obtained from HTS (Table 1). The PCR reaction was carried out in a total volume of 20 µL

containing 1.0 μ L of total DNA; 10.0 μ L of KOD OneTM PCR Master Mix (Toyobo, Shanghai, China); and 0.6 μ L of 10 μ M corresponding to upstream and downstream primers, with the following conditions: 35 cycles of denaturation at 98 °C for 10 s, annealing at 55 °C for 5 s and elongation at 68 °C for 200 bp/s. The PCR products were examined by 1% agarose gel electrophoresis, purified using EasyPure Quick Gel Extraction Kit (TransGen, Beijing, China), cloned into the pMD18-T simple vector (TaKaRa, Dalian, China) and sequenced by the Sanger method. For each amplicon, three independent clones were sequenced. The complete genome sequence was assembled using Vector NTI software v.10.1.1 (Invitrogen, Carlsbad, CA, USA) based on overlapping fragments, and the ORFs were predicted using ORFfinder (https://www.ncbi.nlm.nih.gov/orffinder (accessed on 1 June 2023)). The online tool MotifFinder (https://www.genome.jp/tools/motif/ (accessed on 1 June 2023)) was used to identify conserved motifs of the encoded proteins.

Table 1. Primers used for determination of DCMV genome.

Primer	Sequence (5'-3')	Position ^a
DCMV-F1	CAGTCTGGAATCGATACACC	409-428
DCMV-R1	TCTTGGTTAGCCACTGTAACCTG	3548-3570
DCMV-F2	AGGAAAGTTATCCTTTAAGGGA	3418-3439
DCMV-R2	GGACCGATTATGAGAAAGCTTC	6252-6273
DCMV-F3	ACCAGAAAACTCTCAACAGGA	6118–6138
DCMV-R3	ATAGCACAAGTTGCCTTTTGCTG	488–510

^a Binding position relative to the genome sequences of DCMV-IN (GenBank No. OR494328).

2.4. Genetic Variation, Phylogenetic and Recombination Analysis

The assembled genome sequence was subjected to BLAST analysis in GenBank database (https://blast.ncbi.nlm.nih.gov/Blast.cgi; accessed on 20 June 2023). All DCMV genome sequences available in GenBank and representative *Caulimovirus* species were retrieved and analyzed together with those of the DCMV isolate determined in this work. Pairwise nucleotide and amino acid sequence identities were determined using SDT 1.0 [16]. All of the circular genomes were linearized starting at the coding region of MP [11]. Multiple sequence alignments were performed using the ClustalW algorithm in MEGA11 [17]. Using MEGA11, maximum-likelihood (ML) phylogenetic trees were constructed based on the complete genome sequences with 1000 bootstrap replicates. Genetic distances were calculated by the General Time Reversible (GTR) model which was selected as the best-fit substitution model through the 'Find Best DNA/Protein Models' program, and a discrete γ distribution was used to model evolutionary rate differences among sites [17]. Seven of the recombination detection methods, including RDP, Geneconv, Chimera, BootScan, MaxChi, SiScan and 3Seq, implemented in RDP4 software package v.3.44 [18], were used to identify potential recombination events. Only the events supported by at least five of the seven methods with p < 0.01 were accepted.

3. Results and Discussion

3.1. Virus Identification by HTS

A cDNA library was constructed from pooled leaf tissues of three symptomatic dahlia plants and sequenced using Illumina NovaSeq 6000 platform. A total of 36,031,741 clean reads with Q20 of 96.91% and Q30 of 91.59% were obtained after trimming adapter sequences and discarding low-quality reads. The virus-associated clean reads were assembled de novo into 2213 contigs ranging from 141 to 6630 bp (Supplementary Table S1). Further BLAST analysis showed that, among these contigs, the longest one with a length of 6630 bp covered 83.4% (6630/7949) of the genome of DCMV New Zealand isolate DCMV-NZ (Accession No. JN032736) [11] and shared a 99.46% pairwise identity. Therefore, the DCMV isolate associated with dahlia mosaic disease that was collected in Inner Mongolia of China in this study was named DCMV-IN.

3.2. Determination and Characterization of the DCMV-IN Genome

Three fragments with lengths of 3162 bp, 2856 bp and 2343 bp were PCR-amplified from total DNA with specific primers DCMV-F1/DCMV-R1, DCMV-F2/DCMV-R2 and DCMV-F3/DCMV-R3, respectively. The complete genome sequence of DCMV-IN was assembled based on overlapping regions and deposited in the NCBI GenBank database under accession number OR494328. DCMV-IN genome consisted of 7949 bp containing six ORFs. ORF I (1-969 nt) encoded MP of 322 aa. The transport domain GNLCYGKFMFTVY associated with cell-to-cell movement for caulimoviruses was found from 166 to 178 aa, which was consistent with other DCMV isolates although the cysteine was replaced by alanine for other caulimoviruses [9]. ORF II (962–1453 nt) encoded ATF of 163 aa, and the essential motif for aphid transmission through interaction between ATF and virus particle (IXG) was observed at the C-termianl [19]. DNAb of 120 aa and CP of 505 aa were encoded by ORF III (1450-1812 nt) and ORF IV (1797-3314 nt), respectively. An RNAbinding domain (CX2CX4HX4C) CWICAEEGHY ANEC was identified at 441-454 aa of CP [20]. ORF V (3311–5332 nt) encoded a polyprotein of 673 aa. Three conserved motifs in caulimovirus replicases including aspartyl protease, reverse transcriptase and ribonuclease H at aa positions 76–132, 290–447 and 511–645, respectively, were identified by MotifFinder in the polyprotein. ORF VI (5459–6973 nt) encoded IB of 504 aa, and a conserved motif of caulimovirus viroplasm was identified at 144–181 aa.

3.3. Comparison of DCMV Genome Sequences

BLAST analysis retrieved the complete genome sequences of three DCMV isolates, including one isolate from New Zealand (DCMV-NZ, JN032736) [16], one from Japan (DCMV-JP, LC625373) [21] and one from Beijing, China (DCMV-BJ, KX098538), and one near-full-length genome which covered all ORFs from Taiwan, China (DCMV-TW, AB740270-AB740275) [22]. DCMV-IN shared high genome sequence identities of 99.4%, 99.3% and 99.4% with DCMV-NZ, DCMV-JP and DCMV-BJ, respectively (Table 2). The pairwise nucleotide sequence identities between DCMV-IN and other DCMV isolates ranged from 97.9 to 99.1%, 97.2 to 99.6%, 98.4 to 99.7%, 98.8 to 99.7%, 99.1 to 99.5% and 99.3 to 99.6% for ORF I–VI, respectively. The amino acid sequence identities were 97.5–98.8%, 96.3–99.4%, 95.8–99.2%, 98.2–99.8%, 99.1–99.7% and 98.8–99.8% for ORF I–VI, respectively.

Table 2. Pairwise nucleotide (upper) and amino acid (lower) sequence identities of DCMV-IN with four other DCMV isolates available in GenBank.

Virus Isolate	Genome	ORF I	ORF II	ORF III	ORF IV	ORF V	ORF VI
DCMV-NZ	99.4	99.1	99.6	98.4	99.7	99.4	99.4
	-	98.8	98.8	95.8	99.8	99.6	99.4
DCMV-JP	99.3	98.8	99.6	98.6	99.4	99.4	99.6
	-	98.8	99.4	96.7	99.0	99.7	99.8
DCMV-BJ	99.4	99.1	99.6	99.2	99.3	99.5	99.5
	-	98.8	99.4	95.8	99.4	99.6	99.4
DCMV-TW	NA	97.9	97.2	99.7	98.8	99.1	99.3
	-	97.5	96.3	99.2	98.2	99.1	98.8

3.4. Phylogenetic and Recombination Analyses

The complete genome sequences of four DCMV isolates containing DCMV-IN in this work, and other 13 representative caulimoviruses, were subjected to phylogenetic analysis. In the constructed phylogenetic ML tree (Figure 2), the four DCMV isolates (DCMV-IN, DCMV-BJ, DCMV-NZ and DCMV-JP) were clearly clustered in a single branch. The cluster of DCMV formed a clade together with DMV, mirabilis mosaic virus (MMV) and figwort mosaic virus (FMV), and DCMV was closest to DMV. The phylogenetic tree based on complete genomes in this work showed a similar topology with those based on ORF II, ORF III, ORF IV and ORF VI [4,9]. No putative recombination event was identified from the genome of the four DCMV isolates.



Figure 2. Phylogenetic analysis by maximum-likelihood (ML) method based on complete genome sequences of DCMV and other caulimoviruses. DCMV-IN determined in this study is marked in bold.

Both the results of pairwise alignment and phylogenetic analysis showed that the reported DCMV isolates around the world had low genetic diversity, suggesting that asexual propagation of dahlia plants by vegetative tuberous roots may be the main route of DCMV transmission. This work not only provides a background for an in-depth study on the epidemic and evolution of DCMV, but also highlights the necessity of virus detection in propagating materials for prevention and control of dahlia mosaic disease.

Supplementary Materials: The following supporting information can be downloaded at: https: //www.mdpi.com/article/10.3390/genes14101833/s1, Table S1: Summary of the contigs de novo assembled from virus-associated clean reads.

Author Contributions: Conceptualization, S.S. and Z.L.; methodology, J.W.; software, J.Z.; validation, J.W. and M.C.; investigation, J.W. and J.Z.; resources, Z.L.; data curation, S.S.; writing—original draft preparation, J.W. and J.Z.; writing—review and editing, S.S.; visualization, J.W.; supervision, Z.L.; funding acquisition, S.S. and Z.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by China Postdoctoral Science Foundation, grant number 2021M690584, Heilongjiang Postdoctoral Fund, grant number LBH-Z21047 and Natural Science Foundation of Inner Mongolia, China, grant number 2023LHMS03020.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The sequence data have been submitted to GenBank databases under accession number OR494328. Addresses are as follows: GenBank http://www.ncbi.nlm.nih.gov (accessed on 1 May 2023).

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

References

- 1. Costa, P.A.; Souza, D.C.; Ossani, P.C.; Mendes, M.H.; Silva, M.L.; Carvalho, E.E.; Resende, L.V. Nutritional and functional compounds in dahlia flowers and roots. *Braz. J. Food Technol.* **2022**, *25*, e2022029.
- 2. Causey, J.L.; Feirtag, J.M.; Gallaher, D.D.; Tungland, B.C.; Slavin, J.L. Effects of dietary inulin on serum lipids, blood glucose and the gastrointestinal environment in hypercholesterolemic men. *Nutr. Res.* **2000**, *20*, 191–201. [CrossRef]
- Thakur, P.; Shah, A.H.; Adhikari, Y.; Kumar, M.; Verma, S. Dahlia cultivation in India and abroad: A review. *Int. J. Plant Soil Sci.* 2022, 34, 240–251. [CrossRef]
- Geering, A.D.; McTaggart, A.R.; Teycheney, P. Untangling the taxonomy of dahlia mosaic virus. *Arch. Virol.* 2022, 167, 2325–2329. [CrossRef] [PubMed]
- 5. Eid, S.; Druffel, K.L.; Saar, D.; Pappu, H.R. Incidence of multiple and distinct species of caulimoviruses in dahlias (*D. variabilis*). *HortScience* **2009**, *44*, 1498–1500. [CrossRef]
- 6. Pappu, H.R.; Hammett, K.R.; Druffel, K.L. Dahlia mosaic virus and Tobacco streak virus in Dahlia (*D. variabilis*) in New Zealand. *Plant Dis.* **2008**, *92*, 1138. [CrossRef] [PubMed]
- 7. Pappu, H.R.; Wyatt, S.D.; Druffel, K.L. Dahlia mosaic virus: Molecular detection and distribution in dahlia in the United States. *Hortscience* **2005**, *40*, 697–699. [CrossRef]
- 8. Abduljalil, Z.A. Molecular and biological characterizations of dahlia-infecting caulimoviruses in Iraq. *Trop. Plant Pathol.* 2022, 47, 591–598. [CrossRef]
- 9. Pappu, H.R.; Druffel, K.L.; Miglino, R.; Van Schadewijk, A.R. Nucleotide sequence and genome organization of a member of a new and distinct *Caulimovirus* species from dahlia. *Arch. Virol.* **2008**, *153*, 2145–2148. [CrossRef]
- Pahalawatta, V.; Druffel, K.L.; Wyatt, S.D.; Eastwell, K.C.; Pappu, H.R. Genome structure and organization of a novel and distinct species of genus *Caulimovirus* (Family *Caulimoviridae*) associated with dahlia mosaic. *Arch. Virol.* 2008, 153, 733–738. [CrossRef]
- 11. Hadfield, J.; Linderme, D.; Shepherd, D.N.; Bezuidenhout, M.; Lefeuvre, P.; Martin, D.P.; Varsani, A. Complete genome sequence of dahlia common mosaic virus isolate from New Zealand. *Arch. Virol.* **2011**, *156*, 2297–2301. [CrossRef] [PubMed]
- 12. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 2014, 30, 2114–2120. [CrossRef] [PubMed]
- 13. Wood, D.E.; Salzberg, S.L. Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **2014**, *15*, R46. [CrossRef] [PubMed]
- 14. Li, D.; Liu, C.M.; Luo, R.; Sadakane, K.; Lam, T.W. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **2015**, *31*, 1674–1676. [CrossRef]
- 15. Wu, Q.; Luo, Y.; Lu, R.; Lan, N.; Lai, E.C.; Li, W.X.; Ding, S.W. Virus discovery by deep sequencing and assembly of virus-derived small silencing RNAs. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 1606–1611. [CrossRef] [PubMed]
- 16. 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.
- 17. Tamura, K.; Stecher, G.; Kumar, S. MEGA11: Molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 2021, *38*, 3022–3027. [CrossRef]
- 18. Martin, D.P.; Murrell, B.; Golden, M.; Khoosal, A.; Muhire, B. RDP4: Detection and analysis of recombination patters in virus genomes. *Virus Evol.* **2015**, *1*, vev003. [CrossRef]
- 19. Schmidt, I.; Blanc, S.; Esperandieu, P.; Kuhl, G.; Devauchelle, G.; Louis, C.; Cerutti, M. Interaction between the aphid transmission factor and virus particles is a part of the molecular mechanism of cauliflower mosaic virus aphid transmission. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 8885–8889. [CrossRef]
- Glasheen, B.M.; Polashock, J.J.; Lawrence, D.M.; Gillett, J.M.; Ramsdell, D.C.; Vorsa, N.; Hillman, B.I. Cloning, sequencing, and promoter identification of Blueberry red ringspot virus, a member of the family *Caulimoviridae* with similarities to the "Soybean chlorotic mottle-like" genus. *Arch. Virol.* 2002, 147, 2169–2186. [CrossRef]
- 21. Yamamoto, D.; Neriya, Y.; Suzuki, T.; Nishigawa, H.; Natsuaki, T. Construction of an infectious dahlia common mosaic virus clone. *Arch. Virol.* **2021**, *166*, 3179–3182. [PubMed]
- 22. Chao, H.Y.; Chen, Y.K. Molecular characterization of a dahlia mosaic-associated Caulimovirus. *Plant Pathol. Bull.* **2012**, *21*, 101–114.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.