AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Complete Genome Sequence of Isolate Bari 1, a Mild Strain of Cauliflower Mosaic Virus

Kazusato Ohshima,[a,b] Rikako Ishibashi,[a] Shusuke Kawakubo[a]

[a]Laboratory of Plant Virology, Department of Biological Resource Science, Faculty of Agriculture, Saga University, Saga, Japan
[b]The United Graduate School of Agricultural Sciences, Kagoshima University, Kagoshima, Japan

ABSTRACT We present here the complete genome sequence of isolate Bari 1, a mild strain of cauliflower mosaic virus (CaMV). The isolate was collected from *Diplotaxis tenuifolia* (perennial wall-rocket) in Bari, Italy. The genome was 8,020 nucleotides long and shared ≤85.4% nucleotide identity with other CaMV isolates.

Cauliflower mosaic virus (CaMV) is the type member of the genus *Caulimovirus* in the family *Caulimoviridae* (1). A mild strain of CaMV, isolate Bari 1, was collected from *Diplotaxis tenuifolia* (perennial wall-rocket) in Bari, Italy, before 1977 (2, 3). In this study, we determined the full genome sequence of isolate Bari 1 and compared its properties with the reported CaMV genome sequences.

Bari 1-infected turnip leaves were stored in a deep freezer (Institute of Molecular Cell and Systems Biology, University of Glasgow, UK) until use. Isolate Bari 1 was sap inoculated to *Brassica rapa* cv. Hakatasuwari plants using 0.01 M potassium phosphate buffer (pH 7.0) and serially cloned through single lesions using chlorotic local lesions that appeared ~20 days after inoculation. Biologically cloned Bari 1 isolate was propagated in *B. rapa* plants. Viral DNAs were extracted from the infected leaves using the DNeasy plant minikit (Qiagen KK) and amplified using high-fidelity Platinum *Pfx* DNA polymerase (Invitrogen). We carefully sequenced three overlapping fragments of the Bari 1 genome, amplified by PCR using three sets of primer pairs (Table 1). Each PCR product was sequenced by primer walking in both directions using a BigDye Terminator v3.1 Cycle Sequencing Ready Reaction kit (Life Technologies) and an Applied Biosystems 3130 Genetic Analyzer. The sequences were assembled using BioEdit v5.0.9 (4). All tools were run with default parameters unless otherwise specified.

The full genome sequence of isolate Bari 1 was 8,020 nucleotides (nt) long. The lengths of open reading frames (ORFs) I, II, III, IV, V, VI, and VII were 984 nt, 480 nt, 384 nt, 1,476 nt, 2,019 nt, 1,569 nt, and 291 nt, respectively. The start codons in these ORFs were confirmed using the previously reported CaMV genome sequences. The GC content in the Bari 1 genome

TABLE 1 Primers used for amplifying PCR fragments

Primer[a]	Type	Sequence (5′–3′)[b]	Position (nt)[c]	Coding region[d]
C10P	Forward	5′-AGTTCCCTCACACCGGTGACC-3′	112–132	ORF VII
C22M	Reverse	5′-TGTCGATTAGGACATTCGTTGG-3′	3503–3524	ORF IV
C17P	Forward	5′-GTAGGAAATGAAGAATTAGGATC-3′	2126–2148	ORF III
C31M	Reverse	5′-GAGCCGTTTGCTCTGGAATAGC-3′	5988–6009	ORF VI
C30P	Forward	5′-AAATCCRAAGATAAGATTCCCA-3′	5693–5714	NCR
C81M	Reverse	5′-CGTCTTCTAGTTCAATTGTAGC-3′	1082–1103	ORF I

[a]The PCR fragments were amplified using the following three sets of primer pairs: C10P and C22M, C17P and C31M, and C30P and C81M.
[b]R indicates a mixture of A/G nucleotide degeneracy.
[c]The nucleotide positions correspond to those of the Xinjiang isolate genome (GenBank accession number AF140604).
[d]ORF, open reading frame; NCR, noncoding region.

Citation Ohshima K, Ishibashi R, Kawakubo S. 2021. Complete genome sequence of isolate Bari 1, a mild strain of cauliflower mosaic virus. Microbiol Resour Announc 10:e00534-21. https://doi.org/10.1128/MRA.00534-21.

Editor Simon Roux, DOE Joint Genome Institute

Address correspondence to Kazusato Ohshima, ohshimak@cc.saga-u.ac.jp.

Received 24 May 2021
Accepted 8 June 2021
Published 8 July 2021

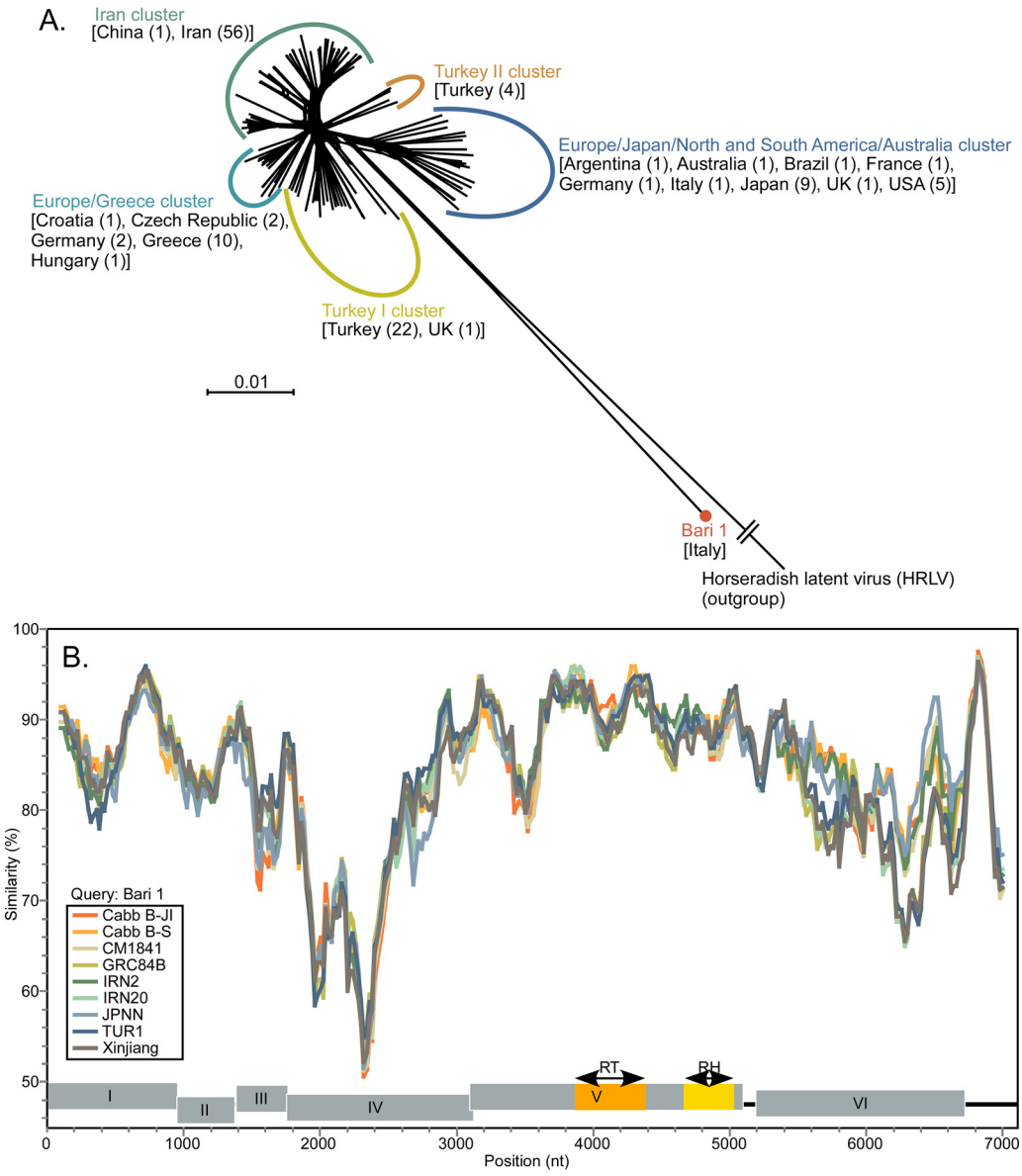

FIG 1 (A) Phylogenetic networks of cauliflower mosaic virus. Neighbor-net analysis inferred from concatenated CaMV ORF regions I to V was performed using SplitsTree v4.15.1 (8). Horseradish latent virus isolate ID1 (GenBank accession number NC_018858) was used as the outgroup taxon. The genomic sequences of 121 CaMV isolates obtained from GenBank (January 2019) and isolate Bari 1 were used. The values given in parentheses are the numbers of isolates from each country. (B) Similarity plot of the nucleotide sequences of CaMV genomes. Isolate Bari 1 (LC632935) was used as the query isolate. The following CaMV isolates were included: Cabb B-JI (KJ716236), Cabb B-S (NC_001497), CM1841 (V00140), GRC84B (AB863194), IRN2 (AB863137), IRN20 (AB863155), JPNN (AB863160), TUR1 (AB863166), and Xinjiang (AF140604). All protein coding regions of CaMV and HRLV were aligned via corresponding amino acid sequences using CLUSTAL X2 (10) with TransAlign (11). After the gaps were removed in each ORF and noncoding region, those sequences were reassembled to form concatenated genome sequences. Overlapping sequences between ORF I and ORF II (3 nt), ORF III and ORF IV (24 nt), and ORF IV and ORF V (36 to 45 nt) were removed. Finally, the concatenated genome sequences of 7,103 nt were used for analysis with SimPlot v3.5.1 (9). Nucleotide similarities were estimated using the Kimura (two-parameter) model, a window size of 200 nt, and a step size of 20 nt.

was 40.1%. The genomic regions of isolate Bari 1 were assessed for nucleotide identity with horseradish latent virus (HRLV) because Bari 1 formed a single lineage distinct from all the CaMV lineages and close to HRLV in the ORF VI phylogenetic tree (5). The reverse transcriptase (RT) and RNase H1 (RH) regions in ORF V of isolate Bari 1 shared 79.1% and 91.1 to 92.7% nucleotide identity, analyzed using EMBOSS Needle (https://www.ebi.ac.uk/Tools/psa/emboss_needle/) (6), with those of HRLV isolate ID1 (GenBank accession number NC_018858) and

other CaMV isolates, respectively. The Bari 1 genome shared 67.1% nucleotide identity with the HRLV ID1 genome. Although Bari 1 is biologically distinct from other CaMV isolates, the demarcation criteria for nucleotide identities of species (7) showed that this isolate is a member of CaMV.

Network analysis inferred from concatenated ORFs I to V (Fig. 1A) was performed using SplitsTree v4.15.1 (8). The network of CaMV isolates, except Bari 1, had short internal branches. The results of a SimPlot v3.5.1 (9) analysis using the Bari 1 genome sequence as the query isolate are shown in Fig. 1B. The nucleotide similarities between the 5′ half of the ORF IV region of Bari 1 and those of the other CaMV isolates were the lowest. The genome of Bari 1 shared ≤85.4% nucleotide identity with those of the CaMV isolates. This study reports the first complete genome sequence of isolate Bari 1 and confirms that Bari 1 is biologically distinct but a member of CaMV.

Data availability. The complete genome sequence of CaMV isolate Bari 1 has been deposited in DDBJ/ENA/GenBank under the accession number LC632935.

ACKNOWLEDGMENTS

We thank Joel Milner (University of Glasgow) for providing the Bari 1-infected turnip leaves. We thank Ryosuke Yasaka (Laboratory of Plant Virology, Saga University) for his careful technical assistance and Adrian Gibbs (emeritus faculty, Australian National University) for his helpful discussions.

This work was in part funded by Saga University and Kagoshima University and supported by the Japanese Society for the Promotion of Science (KAKENHI grant numbers 18K05653 and 21K05601). The genomic sequences were determined at the Analytical Research Center for Experimental Sciences, Saga University. The funders had no role in study design, data collection, interpretation, and analysis, or the decision to submit the work for publication.

REFERENCES

1. Franck A, Guilley H, Jonard G, Richards K, Hirth L. 1980. Nucleotide sequence of cauliflower mosaic virus DNA. Cell 21:285–294. https://doi.org/10.1016/0092-8674(80)90136-1.
2. Hull R. 1980. Structure of the cauliflower mosaic virus genome III. Restriction endonuclease mapping of thirty-three isolates. Virology 100:76–90. https://doi.org/10.1016/0042-6822(80)90553-x.
3. Stratford R, Plaskitt KA, Turner DS, Markham PG, Covey SN. 1988. Molecular properties of Bari 1, a mild strain of cauliflower mosaic virus. J Gen Virol 69:2375–2386. https://doi.org/10.1099/0022-1317-69-9-2375.
4. Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser (Oxf) 41:95–98.
5. Yasaka R, Nguyen HD, Ho SYW, Duchêne S, Korkmaz S, Katis N, Takahashi H, Gibbs AJ, Ohshima K. 2014. The temporal evolution and global spread of *Cauliflower mosaic virus*, a plant pararetrovirus. PLoS One 9:e85641. https://doi.org/10.1371/journal.pone.0085641.
6. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res 47:W636–W641. https://doi.org/10.1093/nar/gkz268.
7. Teycheney P-Y, Geering ADW, Dasgupta I, Hull R, Kreuze JF, Lockhart B, Muller E, Olszewski N, Pappu H, Pooggin MM, Richert-Pöggeler KR, Schoelz JE, Seal S, Stavolone L, Umber M, ICTV Report Consortium. 2020. ICTV virus taxonomy profile: *Caulimoviridae*. J Gen Virol 101:1025–1026. https://doi.org/10.1099/jgv.0.001497.
8. Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23:254–267. https://doi.org/10.1093/molbev/msj030.
9. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, Ingersoll R, Sheppard HW, Ray SC. 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J Virol 73:152–160. https://doi.org/10.1128/JVI.73.1.152-160.1999.
10. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948. https://doi.org/10.1093/bioinformatics/btm404.
11. Weiller GF. 1999. TransAlign version 1.0. Research School of Biological Sciences, Canberra, Australia.