

Complete genome sequence of a novel bipartite begomovirus associated with tomato leaf curl New Delhi disease in golden melon (*Cucumis melo*) plant leaves

Johnson Chong

UNIMAS: Universiti Malaysia Sarawak

Hung Hui Chung

UNIMAS: Universiti Malaysia Sarawak

Han Ming Gan

patriot biotech sdn bhd

WHYE KIT LEONARD LIM

limwhyekitleonard@gmail.com

UNIMAS: Universiti Malaysia Sarawak <https://orcid.org/0000-0003-4079-9630>

Research Article

Keywords: Tomato leaf curl New Delhi virus, Begomovirus, bipartite genome, golden melon, Germiniviridae family

Posted Date: July 17th, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-7079131/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

The tomato leaf curl New Delhi virus (ToLCNDV), also known as *Begomovirus solanumdelhiense* is a whitefly (*Bemisia tabaci*)-transmitted bipartite begomovirus that causes significant economic losses exceeding USD \$300 million annually. It has been reported to infect a wide range of commercially important crops such as tomato, papaya, golden melon, watermelon and luffa. The ToLCNDV genome consists of two components, DNA-A and DNA-B. In this study, we confirmed the presence of viral genome from the gDNA extracted from a symptomatic golden melon plant in Sarawak, Malaysia. The complete bipartite genome was recovered using Illumina shotgun metagenomic approach, offering a primer-free alternative to conventional Sanger-based methods, enabling detection of more divergent viral variants. Comparative genomics against publicly available ToLCNDV genomes from public database showed that the Sarawak strain clusters closely with other Southeast Asian ToLCNDV strains. Notably, the DNA-B segment demonstrated greater discriminatory power for grouping isolates by geographic origin rather than host specificity.

Full Text

The *Germiniviridae* family is one of the most diverse plant-infecting virus family known to date, encompassing viruses that utilize circular single-stranded DNA genomes from genus like *Turncurtovirus*, *Mastrevirus*, *Eragrovirus*, *Grablovirus*, *Curtovirus*, *Becurtovirus*, *Topocuvirus*, *Begomovirus* and *Capulavirus*. One of the well characterized genus of this family is the *Begomovirus* genus, which are typically transmitted by whiteflies (*Bemisia tabaci*) that inhabit tropical and sub-tropical countries across the globe. The yield loss due to begomovirus was estimated at around \$300 million per annum (Sandra & Mandal, 2024). The common practice to curb these whitefly pests is the application of chemical insecticide such as imidachloprid, abamectin, bifenthrin, pyriproxyfen and a few others which can effectively reduce the yield damage by 90%. However overuse of pesticides such as pyriproxyfen can lead to resistance, especially in greenhouse crops (Horowitz et. al., 2002). To date, *Begomovirus solanumdelhiense* (or widely known as tomato leaf curl New Delhi virus, ToLCNDV) is known to infect a plethora of economically important commercial plants such as cucumber, tomato, okra, papaya, musk melon, honeydew, cantaloupe, Chrysanthemum, capsicum, ridge gourd, bitter gourd, eggplant, luffa, squash, watermelon, zucchini, wax gourd, gourd and flowering shrub. The ToLCNDV virus is made up of bipartite genome, encompassing genome fragments DNA-A and DNA-B with sizes ranging between 2.5 to 2.8 kbp (Zhao et al., 2018).

Conventional ToLCNDV genome sequencing via PCR and Sanger's sequencing requires designing primers, manually inspecting chromatograms and overlapping read alignment. This can be an inefficient, error-prone and low-throughput process that often fails when viral sequences are divergent. In contrast, Illumina shotgun sequencing is primer-free and enables unbiased detection of novel or highly variable viral variants. However, its effectiveness depends on viral load, as host DNA can dominate in low-titer or asymptomatic samples, reducing viral coverage. Rolling circle amplification (RCA) is often used to enrich circular viral genomes, but it can introduce amplification bias and complicate assembly due to

concatemer formation. In this study, we bypassed RCA by directly converting single-stranded viral DNA into double-stranded DNA prior to Illumina library preparation. This streamlined approach reduces methodological bias, better preserves the natural genomic structure of ToLCNDV, and improves tolerance to sequence divergence. Using this strategy, we successfully recovered and characterized the complete genome of ToLCNDV from an infected golden melon plant in Sarawak, Malaysia. These data contribute to the growing body of genomic resources for begomoviruses and support ongoing efforts to monitor the genetic diversity and regional distribution of ToLCNDV in Malaysia.

Hybrid honeydew melon (CropPower, MY), *Cucumis melo* var. inodorus, hereafter referred to as 'golden melon' leaves suspected to be infected by tomato leaf curl New Delhi virus were harvested from Agriculture Research Centre (ARC) in Semenggoh, Sarawak, Malaysia. Genomic DNA was isolated from 100 mg of dried leaf material. The sample was first ground in a sorbitol wash buffer to eliminate polysaccharides and cellular debris. The cleaned homogenate was then mixed with a lysis buffer containing 1% SDS, 150 mM NaCl, and 50 mM Tris-HCl at pH 8.0. This mixture was incubated at 65 °C for one hour to break open the cells and release nuclear contents. Proteins and contaminants were removed by adding one-third volume of saturated NaCl, followed by a 5-minute incubation on ice. After centrifugation, the clear supernatant was combined with two volumes of ice-cold absolute ethanol and spun at 10,000 × g for 10 minutes to precipitate the DNA. The DNA pellet was rinsed twice with 70% ethanol, allowed to air dry, and then dissolved in 100 µL of TE buffer. PCR was performed using geminivirus-specific primers (Gemini-F/R) to confirm viral presence. Each reaction was carried out with the following thermal profile: initial denaturation at 95 °C for 2 minutes; 30 cycles of 95 °C for 20 seconds, 50 °C for 20 seconds, and 72 °C for 30 seconds. Amplicons were visualized by agarose gel electrophoresis to verify infection status. To convert single-stranded DNA (ssDNA) viral genomes into double-stranded DNA (dsDNA), which is essential for Illumina DNA library preparation, roughly 1000 ng of genomic DNA was mixed with random decamer primers at a final concentration of 10 µM. The mixture was first heated to 75 °C to disrupt any secondary structures, then slowly cooled to promote primer binding to the ssDNA. Double-strand synthesis was then performed using *E. coli* DNA Polymerase I (New England Biolabs, Cat. No. M0209S), in accordance with the manufacturer's protocol. The reaction was initially incubated at 25 °C for 30 minutes to start synthesis, and then at 37 °C for another 30 minutes to ensure complete dsDNA formation.

The double-stranded DNA generated from second-strand synthesis was purified using 0.6× volume of SPRI magnetic beads (Beckman Coulter, USA) to eliminate excess primers and enzymes. This purified DNA was then used as the input for library construction with the Illumina DNA Library Preparation Kit (Illumina, San Diego, CA), following standard protocol. Sequencing was carried out on a partial lane of the NovaSeq X system (Illumina, USA) using a 2×150 bp paired-end setup, producing around 5 Gb of raw sequence data. Raw paired-end reads were trimmed to remove adapters and low-quality bases using fastp v0.20.1 (Chen et al., 2018). The cleaned reads were then aligned to the *Cucumis melo* reference genome (GCF_025177605.1; NCBI Assembly) using BWA-MEM v0.7.18-r1243 (Li, 2013). Unaligned paired-end reads (those that did not map to the host genome) were isolated and used for *de novo* assembly with NOVOPlasty v4.3.5 (Dierckxsens et al., 2017), a tool designed for circular genome

reconstruction. The assembly process was seeded using *Begomovirus* sequences (Segment A: MW248649.1; Segment B: MW248650.1) to guide the iterative extension of contigs based on read overlap and k-mer structure. Two separate assembly runs were performed for each viral segment, both of which produced full, circular viral genomes. The assembled sequences were subsequently re-oriented using BLASTN to align their starting positions with those of the reference sequences. The complete genome sequences were deposited into the public GenBank database with accession number PV866663 (Segment A) and PV866664 (Segment B). The viral genome maps of both DNA-A and DNA-B fragments were generated using Proksee web server (Grant et al., 2023). The pairwise identity matrix analysis was performed utilizing Sequence Demarcation Tool v. 1.3 (Muhire et al., 2025). All publicly available complete bipartite genome sequences (containing both DNA-A and DNA-B genome fragments) of tomato leaf curl New Delhi viruses from various countries of origin and hosts were retrieved from the public GenBank database. These sequences were subjected to multiple alignment using MEGA 12 (Kumar et al., 2024) for DNA-A and DNA-B genome fragments respectively. Then, the multiple alignment is employed for Model Test via MEGA 12 (Kumar et al., 2024) to determine the best DNA model for the phylogenetic tree construction. The best DNA model determined for DNA-A genome fragment is the TN93+G+I model whereas the best DNA model calculated for DNA-B genome fragment is the T92+G+I model. The maximum likelihood phylogenetic tree was plotted based on the selected model above with MEGA 12 (Kumar et al., 2024) for DNA-A and DNA-B genome fragments across all other tomato leaf curl New Delhi viruses with tomato leaf virus from Kenya ToLCV isolate Tom14 (GenBank accession number: MN894497.1) as outgroup.

The golden melon leaves infected by begomovirus were identified and discovered via the observation of the differences between healthy and diseased leaves via naked eyes (Figure 1). The healthy leaves of golden melon should look flat and not wrinkled (Figure 1A). The golden melon leaves in Figure 1B, 1C and 1D depicted a curled and wrinkled morphology with dark green and yellowish coloration. Yellow mosaic patterns were also observed on these leaves. As expected, PCR amplification using the broad-spectrum genomovirus primer produced the correct band size for DNA extracted from the symptomatic leaf (data not shown).

The viral genome of tomato leaf curl New Delhi virus from the golden melon host in Malaysia in this study was sequenced and characterized with the viral genome map depicted in Figure 1E. A total of 30 million reads totaling to 4.7 Gbp of nucleotide bases reported were generated. *De novo* assembly using the reference sequences as seed successfully recover both segments as complete and circular. Alignment of reads to both segment indicates a genome coverage of 100x. The total assembled genome size of this virus isolated in this study is 5431 bp, with DNA-A genome fragment spanning 2739 bp and DNA-B genome fragment spanning 2692 bp. BLASTn against the NCBI dataset (Date assessed: 8/7/2025) revealed that the DNA-A and DNA-B exhibited the highest identity of 96.54% and 97.62% to ToLCNDV isolated from Malaysia (host: ridge gourd) (Segment A: MW248649.1; Segment B: MW248650.1).

The DNA-A genome fragment encompasses six major open reading frames (ORFs) whereby each of them represents different protein complex, namely AV1, AV2, AC1, AC2, AC3 and AC4. The AV1 is the viral coat protein whereas AV2 is known as the pre-coat protein. AC1 is known as the replication-associates protein, AC2 is characterized as the transcription activator protein, AC3 is the replication enhancer protein and lastly AC4 is the symptom determinant protein (Akram et al., 2025). On the other hand, the DNA-B genome fragment circular DNA contains only two functional ORFs, namely the BV1 and BC1. The BV1 is the nuclear shuttle protein and the BC1 is known as the movement protein (Akram et al., 2025). The GC content of DNA-A and DNA-B genome fragments of the viral genome sequenced in this study are 44.1% and 40.79% respectively. GC-rich regions were observed across AC3 and BC1 in their respective genome fragments (Figure 1E). These regions are pivotal for their indispensable functions as orchestrators of gene expression, replication efficiency as well as transcriptional control. GC-rich zones are often associated with secondary structures that could impact the binding of host or viral replication machinery, potentially playing a role in viral pathogenicity and host adaptation (Akram et al., 2025a).

The pairwise identity matrix heatmap graphs for both DNA-A and DNA-B genome fragments of the ToLCNDV virus sequenced in this study was illustrated in Figure 2A and 3B respectively. The ToLCNDV virus isolated from the Western countries like Italy and Spain formed a cluster with high nucleotide similarities (100%) among all other DNA-A genome fragments, regardless of the host they are isolated from. Similarly, the China ToLCNDV virus are 100% identical and they formed a distinctive cluster for DNA-A genome fragment. All Malaysia ToLCNDV virus are also found to be grouped closely to one another with nucleotide similarities above 90% (Figure 2A). Notably, the pairwise identity heatmap for the DNA-B genome fragments is much more distinctive than that of the DNA-A genome fragments. Distinctive clusters (100% nucleotide identity) were observed across all DNA-B genome fragments of virus from Western countries such as Italy and Spain. Besides, China ToLCNDV viruses also formed a 100% nucleotide identity close cluster among themselves. Interestingly, the Malaysia ToLCNDV virus formed a relatively high sequence identity cluster (above 90%) with ToLCNDV from other Southeast Asia countries like Thailand, Cambodia and Indonesia (Figure 2B).

Maximum likelihood tree based on the DNA-A fragment reveals five major clades among publicly available ToLCNDV viral genomes (Figure 2C). Clade 1 includes all Malaysian ToLCNDV along with sequences from other Southeast Asian countries such as Cambodia, Indonesia and Thailand. Clade 2 consists of most of the India isolates and all isolates from Bangladesh. Clade 3 contains all ToLCNDV genomes from Western countries, including Italy and Spain. Clade 4 consists of isolates from Iran and Seychelles, while Clade 5 includes all Chinese isolates and one from Pakistan. Notably, one Indian isolate from *Duranta erecta* did not cluster with any major clade, indicating distinct divergence. Bootstrap values for the DNA-A tree ranged from 55.62% to 72.97%.

Similarly, DNA-B maximum likelihood tree also formed five distinct clades, with generally higher bootstrap support ranging from 54.6% to 100% (Figure 2D). Clade 1 again contains all Malaysian isolates, alongside those from Southeast Asia, including Thailand, Indonesia, and Cambodia, as well as Indian isolates from *Duranta erecta* and bitter melon. Clade 2 consists of several Indian isolates and all

Chinese isolates. Clade 3 comprises the remaining Indian isolates together with all Bangladeshi sequences. Clade 4 includes ToLCNDV genomes from Iran and Seychelles, while Clade 5 encompasses all isolates from Italy and Spain. Compared to DNA-A, the DNA-B tree displays clearer and more geographically consistent clustering. Notable clades with strong bootstrap support (100%) include those from China, Spain/Italy, and Malaysia.

It is interesting to note that the clustering of DNA-A and DNA-B genome fragments within the phylogenetic tree is solely based on geographical locations rather than the host plant type and species. This trend was also reported by Zhao et al. (2018), Avendi et al. (2021), Chen et al. (2021), Batista et al. (2022) as well as Sahu and Sanan-Mishra (2020) based on the phylogenetic tree they plotted across ToLCNDV isolates for both DNA-A and DNA-B genome fragments. Furthermore, the phylogenetic tree analyses performed on other plant viral genomes such as piper yellow mottle virus, cucumber mosaic virus, melon necrotic spot virus and rice yellow mottle virus had also supported this evidence (Lim et al., 2022a-b; Lim, 2022; Lim, 2023). Taken together, the DNA-B genome fragment is more powerful in providing a higher cluster resolution in distinguishing ToLCNDV isolates by geographic locality compared to the DNA-A genome fragments.

We report the complete bipartite genome of *Tomato leaf curl New Delhi virus* (ToLCNDV) isolated from golden melon in Sarawak, Malaysia, using an RCA-free Illumina shotgun sequencing approach. This method allowed recovery of the viral genome without primer bias, preserving its natural sequence structure. Phylogenetic analysis revealed clear geographic clustering, with DNA-B showing higher resolution than DNA-A. For routine surveillance where sequencing both segments may not be feasible, DNA-B is the preferred marker for tracking ToLCNDV diversity and spread.

Declarations

Acknowledgement

This work is funded by the Ministry of Higher Education Malaysia through the Fundamental Research Grant Scheme, with grant number FRGS/1/2022/STG01/UNIMAS/02/2 awarded to H. H. Chung and Pusat Penyelidikan Pertanian, Jabatan Pertanian Sarawak.

Ethics declarations

Conflict of interest

The authors declare they have no conflict of interest.

Ethical approval

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

References

1. Akram, M., Kumar, D., & Kamaal, N. (2025). Complete genome sequence of a novel bipartite begomovirus infecting butterfly pea (*Clitoria ternatea* L.) in India. *Archives of Virology*, 170, 7.
2. Avendi, E.K., Adediji, A.O., Kilalo, D.C., Olubayo, F.M., Macharia, I., Ateka, E.M., Machuka, E.M., & Mutuku, J.M. (2021). Metagenomic analyses and genetic diversity of *Tomato leaf curl Arusha virus* affecting tomato plants in Kenya. *Virology Journal*, 18, 2.
3. Batista, J.G., Nery, F.M., Melo, F.F.S., Malheiros, M.F., Rezende, D.V., Boiteux, L.S., Fonseca, M.E.N., de Miranda, B.E.C., & Pereira-Carvalho, R. (2022). Complete genome sequence of a novel bipartite begomovirus infecting the legume weed *Macroptilium erythroloma*. *Archives of Virology*, 167, 1597-1602.
4. Chen, B.Z., Yang, Z.J., Wang, W.B., Hao, T.T., Yu, P.B., Dong, Y., & Yu, W.B.. (2024). Chromosome-level genome assembly and annotation of *Flueggea virosa* (Phyllanthaceae). *Sci Data*. 13;11(1), 875.
5. Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34(17), i884–i890.
6. Chen, Y.J., Lai, H.C., Lin, C.C., Neoh, Z.Y., & Tsai, W.S. (2021). Genetic diversity, pathogenicity and pseudorecombination of cucurbit-infecting begomovirus in Malaysia. *Plants*, 10, 2396.
7. Dierckxsens, N., Mardulyn, P., & Smits, G. (2017). NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research*, 45(4), e18.
8. Grant, J.R., Enns, E., Marinier, E., Mandal, A., Herman, E.K., Chen, C.Y., ... Stothard, P. (2023). Proksee: In-depth characterization and visualization of bacterial genomes. *Nucleic Acids Research*, 51(W1), W484-W492.
9. Horowitz AR, Kontsedalov S, Denholm I, Ishaaya I. (2002) Dynamics of insecticide resistance in *Bemisia tabaci*: a case study with the insect growth regulator pyriproxyfen. *Pest Manag Sci.*, 58(11),1096-100.
10. Kumar, S., Stecher, G., Suleski, M., Sanderford, M., Sharma, S., & Tamura, K. (2024). MEGA12: Molecular evolutionary genetic analysis version 12 for adaptive and green computing. *Molecular Biology and Evolution*, 41(12), msae263.
11. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.
12. Lim, LWK. (2023). Rice Yellow Mottle Virus: Genomic Dissection and Global Genome Comparison. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4372844
13. Lim, L.W.K. (2022). Comparative genomic analysis reveals the origin and global distribution of melon necrotic virus isolates. *Gene Reports*, 29, 101685.
14. Lim LWK, Hung IM, & Chung HH (2022a) Cucumber mosaic virus: global genome comparison and beyond. *Malays J Microbiol*, 18(1), 79-92.
15. Lim LWK, Liew JX, & Chung HH (2022b) Piper yellow mottle virus: A deep dive into the genome. *Gene Reports*, 29, 101680.

16. Muhire, B.M., Roumagnac, P., Varsani, A., & Martin, D.P. (2025). Sequence Demarcation Tool (SDT), a Free User-Friendly Computer Program Using Pairwise Genetic Identity Calculations to Classify Nucleotide or Amino Acid Sequences. *Methods Mol Biol.*, 2912, 71-79.
17. Sahu, A.K., & Sana-Mishra, N. (2020). Complete genome sequence of a new bipartite begomovirus associated with leaf curl disease of *Capsicum annum*. *3 Biotech*, 10, 235.
18. Sandra, N., & Mandal, B. (2024). Emerging evidence of seed transmission of begomoviruses: Implications in global circulation and disease outbreak. *Front. Plant Sci.*, 15, 1376284.
19. Zhao, L., Zhong, J., Zhang, X., Ding, M., & Zhang, Z. (2018). Complete genome sequence of a new bipartite begomovirus infecting *Boehmeria leiophylla* in China. *Archives of Virology*, 163, 1989-1992.

Figures

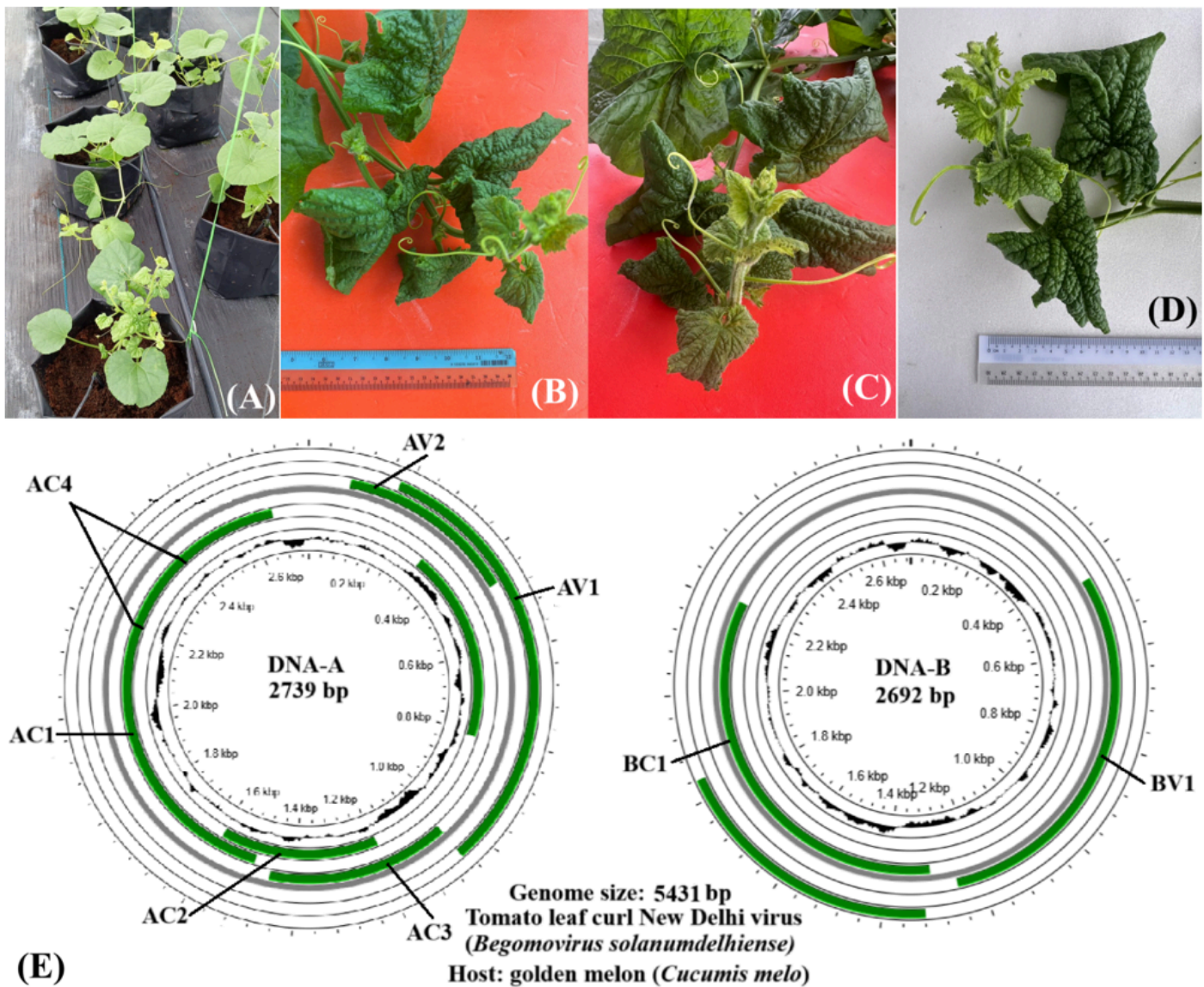


Figure 1

(A) Healthy golden melon plant. (B, C, D) Golden melon plant leaves infected by *Begomovirus solanumdelhiense* (or widely known as tomato leaf curl New Delhi virus). (E) The genome map of two circular single strand DNA genome fragments (DNA-A and DNA-B) of the tomato leaf curl New Delhi virus isolated from golden melon in this study. The black curves represent GC content across the circular genome.

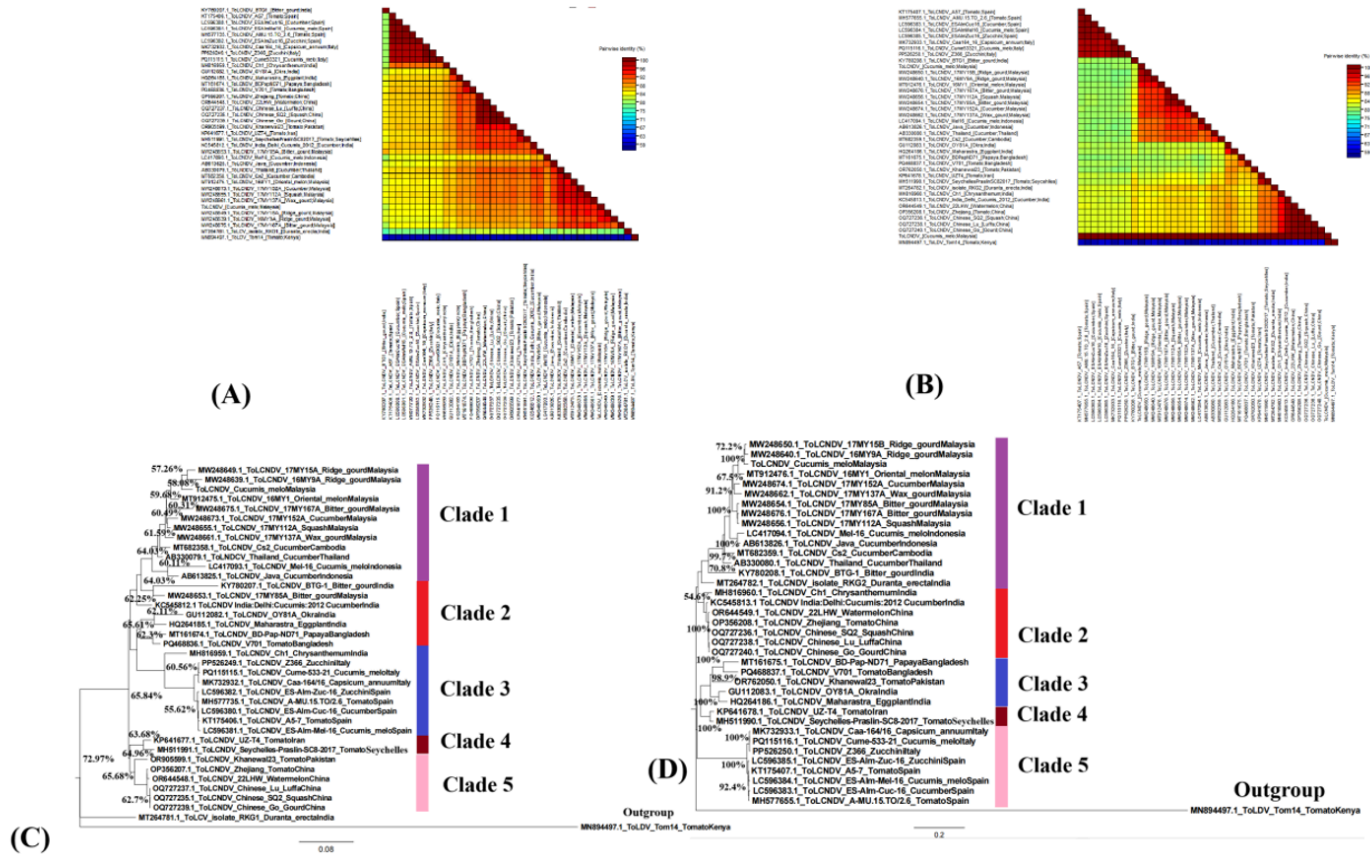


Figure 2

The pairwise identity matrix heatmap of (A) DNA-A and (B) DNA-B genome fragments of all publicly available ToLCNDV complete genomes across the globe, with tomato leaf virus from Kenya ToLCV isolate Tom14 (GenBank accession number: MN894497.1) as outgroup. (C) The maximum likelihood phylogenetic tree of DNA-A genome fragments of all publicly available ToLCNDV complete genomes across the globe, with tomato leaf virus from Kenya ToLCV isolate Tom14 (GenBank accession number: MN894497.1) as outgroup, with 1000 bootstrap replications. (D) The maximum likelihood phylogenetic tree of DNA-B genome fragments of all publicly available ToLCNDV complete genomes across the globe, with tomato leaf virus from Kenya ToLCV isolate Tom14 (GenBank accession number: MN894497.1) as outgroup, with 1000 bootstrap replications.