



## Cucumber mosaic virus: Global genome comparison and beyond

Leonard Whye Kit Lim\*, Ing Mee Hung and Hung Hui Chung\*

Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia.

Email: [lwkl1993@gmail.com](mailto:lwkl1993@gmail.com); [hhchung@unimas.my](mailto:hhchung@unimas.my)

Received 17 August 2021; Received in revised form 20 September 2021; Accepted 20 December 2021

### ABSTRACT

**Aims:** The cucumber mosaic virus (CMV) is categorized under the genus *Cucumovirus* and family Bromoviridae. This virus is known to infect over 1200 plant species from 100 families, including ornamental and horticultural plants. In this study, we pioneered a global genome comparison to decipher the unknown orchestrators behind the virulence and pathogenicity of CMV via the discovery of important single nucleotide polymorphic markers.

**Methodology and results:** As a result, the genome size was found to be a potential preliminary country-specific marker for South Korea and the GC content can be utilized to preliminarily differentiate Turkey isolates from the others. The motif analysis as well as whole genome and coat protein phylogenetic trees were unable to form country-specific clusters. However, the coat protein haplotype analysis had successfully unconcealed country-specific single nucleotide polymorphic markers for Iran, Turkey and Japan isolates. Moreover, coat protein modelling and gene ontology prediction depicted high conservation across CMV isolates from different countries.

**Conclusion, significance and impact of study:** The country-specific single nucleotide polymorphic markers unearthed in this study may provide significant data towards the profiling of varying virulence and pathogenicity of CMV across the globe in time to combat the yield loss driven by this virus thru the most efficacious biological control measures in the future.

**Keywords:** Cucumber mosaic virus, genome comparison, country-specific SNPs, haplotype, coat protein

### INTRODUCTION

The cucumber mosaic virus (CMV) is a *Cucumovirus* genus member grouped under the family Bromoviridae. It has the capability to infect over 1200 plant species from plant families exceeding 100. The host range of this virus is not limited to ornamental plants only but it also infects horticultural plants (Zitter and Murphy, 2009) and most virus host plants are perennial or biennial in nature (Draegar, 2016).

The 29 nm diameter CMV virions are icosahedral in nature and they possess 180 subunits of a single capsid protein (CP) as well as 18% RNA (Jacquemond, 2012). This virus composed of a sum of three single-stranded positive-sense RNAs, with names 1 to 3, in the order of diminishing length. The monocistronic RNA1 encodes for protein 1a, in which the N-terminal encompasses the putative methyltransferase domain whereas the C-terminal houses the helicase motif (Jacquemond, 2012). The RNA2 encodes for the larger 2a protein and smaller 2b protein whereby the 2a protein contains GDD motif pivotal to RNA-dependant RNA polymerase (RdRp) functioning (Jacquemond, 2012). The 2b protein disrupts

the RNA interfering (RNAi) pathway of the host. The RNA3 is bicistronic and it encodes for the coat protein (CP) as well as the movement protein (MP) or also known as the 3a protein (Jacquemond, 2012). The total lengths of RNAs differ across different strain groupings and strains and this has resulted in varying genome sizes.

The CMV genome sizes range from 8500 to 8700 bp and it has been sequenced from various host species like radish, turnip weed, wild cabbage, tobacco, musk melon, field pumpkin, tomato, cucumber, spinach, Indian shot, black mustard seed, balsam, Adzuki bean, fungus, naranjilla and Jew's melon, as well as countries like Iran, Japan, Turkey, South Korea, Brazil, China, Colombia, France, Germany, Poland and USA (GenBank, 2021). A sum of 249 full genomes were found deposited in the public GenBank database to date but we only select the 40 genomes published in renowned high impact journal articles to ensure high fidelity and confidence on the country of origin and host species profiles reported (GenBank, 2021). The commonest host species is the radish, followed by the field pumpkin. The three major CMV genome sequence contributors are researchers from Iran, Japan and Turkey (GenBank, 2021). With the

\*Corresponding author