**<tentative>** Phylogenetic analysis of Sri Lankan Cassava Mosaic Virus (SLCMV) causing Cassava Mosaic Disease (CMD) in Trincomalee, Sri Lanka based on their DNA-A sequences.

**Introduction**

Cassava (*Manihot esculenta* Crantz*)* is a perennial woody shrub belonging to family Euphorbiaceae and is cultivated primarily for its tuberous roots in tropical and neotropical nations around the world. It’s native to the southern edge of Brazilian Amazon and was domesticated by indigenous South Americans roughly 15,000 years ago (Allem, 1994);(Oslen and Schaal, 1999). Sri Lanka being a tropical island, has had cassava growing on its soils for a long time and cassava has become a well-integrated element in the Sri Lankan culinary landscape. In Sri Lanka, cassava is cultivated in dry, wet as well as intermediate zones encompassing Jaffna, Trincomalee, Anuradhapura, Gampaha, Matara, Ratnapura and Kurunegala districts. Sri Lankan Department of Agriculture (DOA) lists the following eight varieties: Shani, Suranimala, Suwarne, HORDI Mu – 01, Kirikawadi, Mu 51, HORDI 06 and CAR5 555 as customarily planted cassava cultivars in Sri Lanka. Of these eight, Kirikavadi, MU51 and CARI 555 are recommended by DOA for commercial cultivation. A diverse array of products is prepared from cassava tubers e.g., curries, porridges, crispy fries, boiled tubers, ground flour from dried tubers etc. which eloquently decorate the cuisines of Sri Lankan households. Dried flour from cassava tubers even served as an alternative staple food source when Sri Lankans faced severe shortages in imported wheat flour in the1970s, which highlights the historical and dietary significance of cassava tubers in Sri Lankans’ subsistence.

However, cassava plantations in Sri Lanka have been suffering from the Cassava Mosaic Disease (CMD) caused by Cassava Mosaic Virus (CMV) for a long time. CMD in *M. esculenta* is characterized by mosaic, mottled, misshapen, twisted and puckered leaves accompanied by tuber splitting and stunting, all of which cumulatively precipitate in reduced yields. Cassava plantations in Sri Lanka are threatened by two variants of CMV, known as Indian Cassava Mosaic Virus (ICMV) and Sri Lankan Cassava Mosaic Virus (SLCMV) (Prasangika et al., 2008). Even though the disease had been known to Sri Lankan farmers since the early 1980s, the first known academic report studying CMD in Sri Lanka comes from Salim N. and Bandumala S. H (2001) (Prasangika et al., 2008). The predominant causal agent SLCMV, is a closely related variant of ICMV and infects both commercially cultivated cassava varieties and wild cassava plants (*Manihot carthaginensis* subsp. *glaziovii*, also known as Ceara rubber)(Salim, N. and Bandumala S. H 2001). Since then, this disease and its ravages have only been studied sporadically by Sri Lankan academics, albeit meriting a sustained, dedicated attention. However, SLCMV seems to have drawn intense academic interest from international researchers from countries such as Vietnam, Loas, India and Cambodia as the virus spread over to their nations, causing crop losses.

SLCMV is a bipartite plant virus belonging to the family Geminiviridae, genus *Begomovirus.* The genome of SLCMV is composed of two circular ssDNAs, namely DNA-A (approximately 2,760 bases long) and DNA-B (approximately 2,737 bases long), each of which are encapsulated in separate twinned icosahedral capsids (Stanley and Gay, 1983). Genome organization of Begomovirus is characterized by the presence of a coat protein (CP) gene (ORF V1/AV1/Coat protein) in the viral sense strand and a replication associated protein (Rep) gene (C1/AC1/Rep) in the complimentary strand. Both two ORFs start off from opposite ends of a non-coding intergenic region containing the origin of replication for the viral sense strand and the promoters of CP and Rep genes (Malathi et al., 2017). DNA-A encodes six genes AC1, AC2, AC3, AC4, AV1 and AV2 that play vital roles in defense, replication, transcription and in the synthesis of viral capsids. DNA-B encodes two genes BC1 and BV1 whereby BC1 in the anti-sense strand encodes the movement protein (MP) critical for intercellular systemic movement of virus particles in host plants and BV1 in the viral sense strand encodes the nuclear shuttle protein (NSP) that facilitates intracellular transport of the virions (Hareesh et al., 2023) (Malathi et al., 2017). DNA-A and DNA-B of SLCMVs also share a common intergenic region (~200 bases long) that encodes elements involved in replication and transcription (Saunders et al., 2002). DNA-A and DNA-B of SLCMV are mutually interdependent since DNA-A depends on DNA-B for its intracellular and intercellular movement while DNA-B depends on DNA-A for its replication and encapsidation (Malathi et al., 2017). And the sizes of these two DNAs differ by a small number of nucleotides, whereby DNA-A happens to be 21 to 23 bases longer than its DNA-B counterpart. Host-to-host transmission of SLCMV is mediated by a specific biotype of whiteflies (*Bemisia tabaci*). SLCMV is also highly sap transmissible and could be successfully transmitted to 39 species in family Solanaceae via mechanical sap inoculation (Jose et al., 2008).

The genome of SLCMV is reported to be very volatile since it’s prone to high rates of mutations, which renders it more aggressive, virulent and capable of infecting a diverse array of hosts compared to ICMV (Patil et al., 2005). A study from Prasangika et al. (2008) based on *M. esculenta* germplasms concludes that most of the cassava cultivars (HORDI 28, HORDI 6, CARI 555, MU51, Kirikawadi, BW1, BW2, Wariyapola and Wagolla) planted commercially in Sri Lanka are susceptible to CMD with HORDI 6 showing the highest susceptibility and the study also identifies cultivars WA/KK/10 and 555/KK/2 to be highly resistant to CMD, based on Triple Antibody Sandwich Enzyme-Linked Immunosorbent Assay (TAS ELISA) tests performed on cassava germplasms artificially infected through wedge grafting.

This study sampled diseased *M. esculenta* leaves from select fields in Trincomalee district, Sri Lanka and compared the SLCMV viral DNA-A sequences with that of other known viral pathogens affecting *M. esculenta* plants in Sri Lanka as well as all internationally occurring variants of CMV, in a quest to discern the evolutionary relationships between them and to deduce the possible events that might have paved way for this genetic diversity to come into existence.

**Materials and Methods**

**Field Assessment of contemporary CMD situation in Trincomalee, Sri Lanka**

A brief preliminary filed survey was carried out starting from May 2021 to July 2021, in order to determine the disease incidence of CMD in Trincomalee district, Sri Lanka. Field visits were arranged to three different divisional secretariat (DS) divisions in Trincomalee district, namely Padavisripura, Muthur and Nilaveli. In each field, six plots (each 10 m2 in area) were selected and the number of diseased plants (n) showing stereotypical CMD symptoms and the number of total plants in the chosen plot (N) were documented. The percentage CMD incidence per field was calculated using the following equation (Saokham et al., 2021).

**<No other info regarding the survey>**

**Sample collection.**

Infected cassava leaves showing conspicuous and characteristic CMD symptoms were collected in June 2021 from the five different fields in Trincomalee district where cassava is being cultivated widely (Padavisripura, Muthur, Nilaveli, Kuchchaveli and Kinniya). The fields had been carefully chosen such that there’s at least a 20 - 25 km distance from one another. The harvested cassava leaves were separately packed in polythene freezer bags and stored at -20 oC in a refrigerator (Emmanuel et al., 2019).

**Extraction of total genomic DNA, PCR and sequencing.**

Leaf samples from each region were subjected to the extraction of total genomic DNA using a derivative of the conventional CTAB extraction method (Emmanuel et al., 2019). 100 mg of leaf samples were ground in CTAB extraction buffer in pre-chilled pestle and motor. Subsequently, the total genomic DNA was extracted from the lysates following Doyle and Doyle (1987).

The complete DNA-A sequences of the five SLCMV samples were amplifed using three sets of degenerate PCR primers (Integrated DNA Technologies, Inc.) (OY2395F 5´-GCT CCCTGAATGTTCGGATGG- 3´, OY680R 5´-GTTCTCRTCCATCCATATCTTAC-3´; MKBEGF4 5´-ATATCTGCAGGGNAARATHTGGATGGA-3´, MKBEGR5 5´-TGGACTGCAGACNGGNAARACNATGTGGGC-3´; and GEMA 1223F 5´-TCGGAGGSTGTAAGGTCGTCCAG-3´, GE MA2454R 5´-CTCACWTAYCHCAARTGCTCTC-3´) that were designed to yield overlapping amplicons of DNA-A of old world Begomoviruses (Venkataravanappa et al., 2013), (Venkataravanappa et al., 2012). In order to eliminate the possibility of interferences from mixed infections, the primers were designed such that the amplicons were deemed to share common overlapped (approximately 200 bp) sequences at their terminals. Polymerase chain reactions were carried out in 20 µL reaction mixtures containing 10 µL ready made PCR master mix (PCR Biosystems, UK), 1 µL of forward and reverse primers (10 µM) and 1 µL DNA sample (about 50 ng). The thermocycler was programmed to iterate through 35 cycles of PCR, each consisting denaturation at 94 °C for 1 min, annealing at 58 °C for 45 s and extension at 72 °C for 90 s. This amplification process also included an initial denaturation at 94 °C for 3 min and a final extension at 72 °C for 15 min. The amplicons obtained from PCR were electrophoresed in 0.8% agarose gel incorporated with ethidium bromide (10 mg/ml) and were visually examined in a gel documentation system. **<what type of sequencing & where was it done?>** The sequences of the overlapping amplicons from each five samples were finally aligned using DNA Dragon - DNA Sequence Contig Assembler Software (SequentiX - Digital DNA Processing) leveraging the shared regions to discern the order and position of assembly to formulate the complete and contiguous DNA-A sequences.

**Genetic analysis of DNA-A sequences, pairwise alignment, BLAST and phylogenetic relationships.**

The five DNA samples from Kinniya, Kuchchaveli, Muthur, Nilaveli and Padavisripura yielded the following five DNA-A sequences Kin2021, Kuc2021, Mut2021, Nil2021 and Pd2021 respectively (GenBank accessions OK377342, OK377343, OK424595, OK377341 and OK362288). Initially, these five sequences and an SLCMV DNA-A reference sequence from Saunders et al. (2002) (GenBank accession AJ314737) were subjected to multiple sequence alignment using MUSCLE (muscle v5.1.0) (Edgar, 2004) and the sequence identity scores were computed using Sequence Demarcation Tool (SDT v1.2) (Muhire et al., 2014). Following, these five sequences were BLASTed using NCBI’s Nucleotide BLAST, to figure out the most similar sequences in GenBank database and the results were documented.

Then, these five sequences in conjunction with reference accessions for all internationally known variants of CMVs scraped from ICTV were subjected to multiple sequence alignment and sequence identity scoring using the same tools. Subsequently the sequences were compared with DNA-A sequences of a set of known Begomovirus pathogens causing diseases in Sri Lanka, using the same methods and tools.

**<needs formatting>** References:

Saunders, K., Salim, N., Mali, V.R., Malathi, V.G., Briddon, R., Markham, P.G., Stanley, J., 2002. Characterisation of Sri Lankan Cassava Mosaic Virus and Indian Cassava Mosaic Virus: Evidence for Acquisition of a DNA B Component by a Monopartite Begomovirus. Virology 293, 63–74. <https://doi.org/10.1006/viro.2001.1251>

Chi, Y., Pan, L.-L., Bouvaine, S., Fan, Y.-Y., Liu, Y.-Q., Liu, S.-S., Seal, S., Wang, X.-W., 2020. Differential transmission of Sri Lankan cassava mosaic virus by three cryptic species of the whitefly Bemisia tabaci complex. Virology 540, 141–149. <https://doi.org/10.1016/j.virol.2019.11.013>

Salim, N. and Bandumala S. H, 2001. Characterisation of a geminivirus infecting cassava in Sri Lanka, Department of Botany, University of Sri Jayewardenepura Forest Research Centre Boyagane, Kurunagala. VidyodayaJ. Of Sci. Vol. 10.pp 151-16

Olsen, K.M., Schaal, B.A., 1999. Evidence on the origin of cassava: Phylogeography of *Manihot esculenta*. Proc. Natl. Acad. Sci. U.S.A. 96, 5586–5591. <https://doi.org/10.1073/pnas.96.10.5586>

Allem, A.C., 1994. The origin of Manihot esculenta Crantz (Euphorbiaceae). Genet Resour Crop Evol 41, 133–150. <https://doi.org/10.1007/BF00051630>

Sri Lanka, Short of Food, Faces an Economic Crisis The New York Times May 13, 1974, Page 16 <https://www.nytimes.com/1974/05/13/archives/sri-lanka-short-of-food-faces-an-economic-crisis-people-are-well.html?smid=url-share>

Hareesh, P.S., Resmi, T.R., Sheela, M.N., Makeshkumar, T., 2023. Cassava mosaic disease in South and Southeast Asia: current status and prospects. Front. Sustain. Food Syst. 7, 1086660. <https://doi.org/10.3389/fsufs.2023.1086660>

Stanley, J., Gay, M.R., 1983. Nucleotide sequence of cassava latent virus DNA. Nature 301, 260–262. <https://doi.org/10.1038/301260a0>

Emmanuel, C.J., Inthuja, A., Keshiga, A., 2019. Status of Cassava cultivation in Jaffna peninsula and detection of Cassava mosaic disease causing agent. Vingnanam J Sci 14, 1. <https://doi.org/10.4038/vingnanam.v14i2.4150>

Malathi, V.G., Renukadevi, P., Chakraborty, S., Biswas, K.K., Roy, A., Sivalingam, P.N., Venkataravanappa, V., Mandal, B., 2017. Begomoviruses and Their Satellites Occurring in India: Distribution, Diversity and Pathogenesis, in: Mandal, B., Rao, G.P., Baranwal, V.K., Jain, R.K. (Eds.), A Century of Plant Virology in India. Springer Singapore, Singapore, pp. 75–177. <https://doi.org/10.1007/978-981-10-5672-7_5>

Saokham, K., Hemniam, N., Roekwan, S., Hunsawattanakul, S., Thawinampan, J., Siriwan, W., 2021. Survey and molecular detection of Sri Lankan cassava mosaic virus in Thailand. PLoS ONE 16, e0252846. <https://doi.org/10.1371/journal.pone.0252846>

Venkataravanappa, V., Lakshminarayana Reddy, C.N., Devaraju, A., Jalali, S., Krishna Reddy, M., 2013. Association of a recombinant Cotton leaf curl Bangalore virus with yellow vein and leaf curl disease of okra in India. Indian J. Virol. 24, 188–198. <https://doi.org/10.1007/s13337-013-0141-4>

Prasangika, H., Salim, N., Razak, M., 2008. Evaluation of susceptibility of cassava germplasm to cassava mosaic disease. J. Natn. Sci. Foundation Sri Lanka 36, 99. <https://doi.org/10.4038/jnsfsr.v36i1.137>

Patil, B.L., Rajasubramaniam, S., Bagchi, C., Dasgupta, I., 2005. Both Indian cassava mosaic virus and Sri Lankan cassava mosaic virus are found in India and exhibit high variability as assessed by PCR-RFLP. Arch Virol 150, 389–397. <https://doi.org/10.1007/s00705-004-0399-3>

Jose, A., Makeshkumar, T., Edison, S., 2008. Host Range of Sri Lankan Cassava Mosaic Virus.

Journal of Root Crops, Vol. 34 No. 1, pp. 21-25 Indian Society for Root Crops

Venkataravanappa, V., Lakshminarayana Reddy, C.N., Jalali, S., Krishna Reddy, M., 2012. Molecular characterization of distinct bipartite begomovirus infecting bhendi (Abelmoschus esculentus L.) in India. Virus Genes 44, 522–535. <https://doi.org/10.1007/s11262-012-0732-y>

Muhire, B.M., Varsani, A., Martin, D.P., 2014. SDT: A Virus Classification Tool Based on Pairwise Sequence Alignment and Identity Calculation. PLoS ONE 9, e108277. <https://doi.org/10.1371/journal.pone.0108277>