

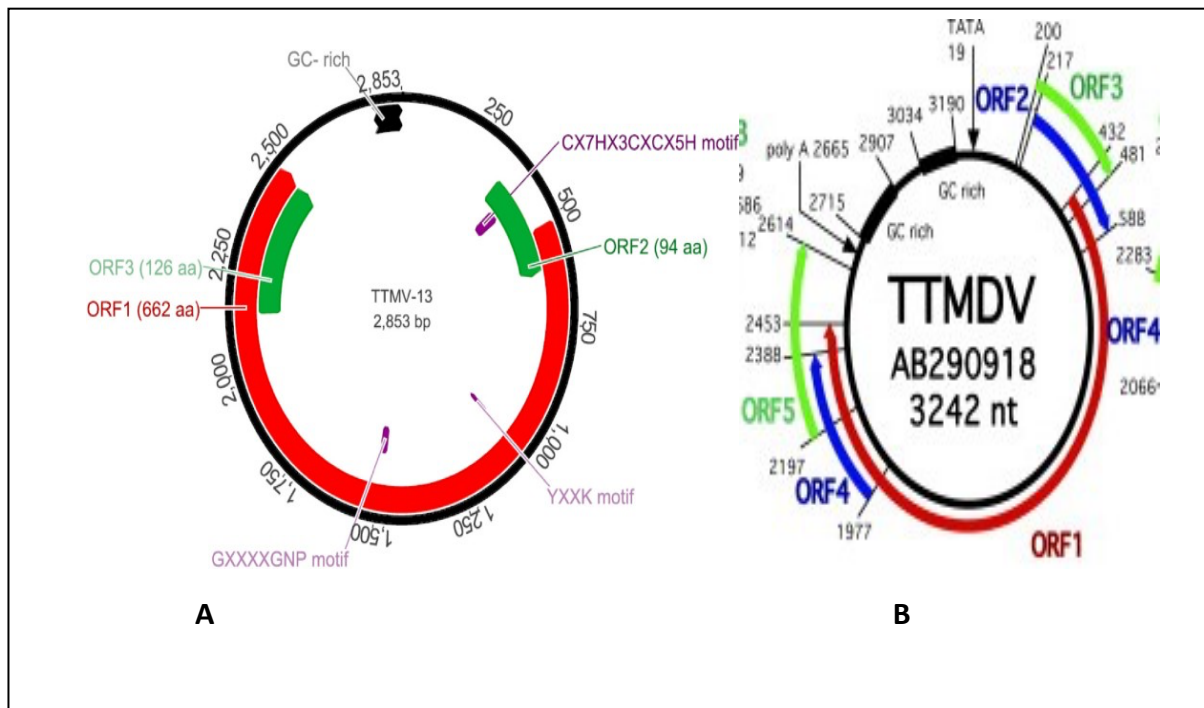


Fig 1. Genomic organization of the new TTMV and TTMDV genomes. Arrows represent ORFs, and the box indicates the GC-rich region. Putative protein sizes for each ORF are given in amino acids: Conserved motifs in both are indicated ^{1,2}.

Fig 2. TTMV Phylogenetic Tree Construction

Each of the sequences submitted to the bootstrap method phylogenetic construction for TTMV has been updated in Fig 2.

a) Phylogenetic Tree whole Genome Sequences

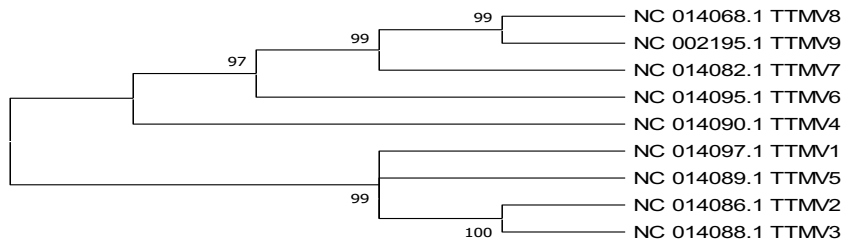


Fig 2. TTMV Phylogenetic Tree Construction

Each of the sequences submitted to the bootstrap method phylogenetic construction for TTMV has been updated in Fig 2.

b) Phylogenetic Tree ORF 1 DNA Sequences

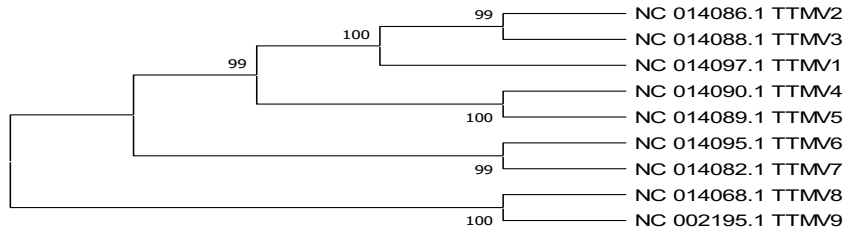


Fig 2. TTMV Phylogenetic Tree Construction

Each of the sequences submitted to the bootstrap method phylogenetic construction for TTMV has been updated in Fig 2.

c) Phylogenetic Tree ORF 2 DNA Sequences

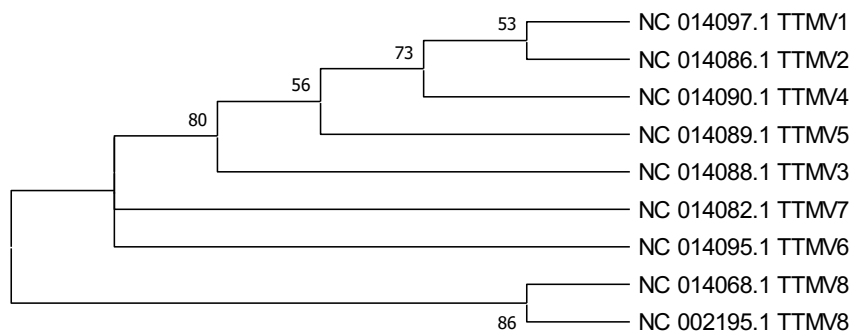


Fig 2. TTMV Phylogenetic Tree Construction

Each of the sequences submitted to the bootstrap method phylogenetic construction for TTMV has been updated in Fig 2.

d) Phylogenetic Tree ORF 1 Protein Sequences

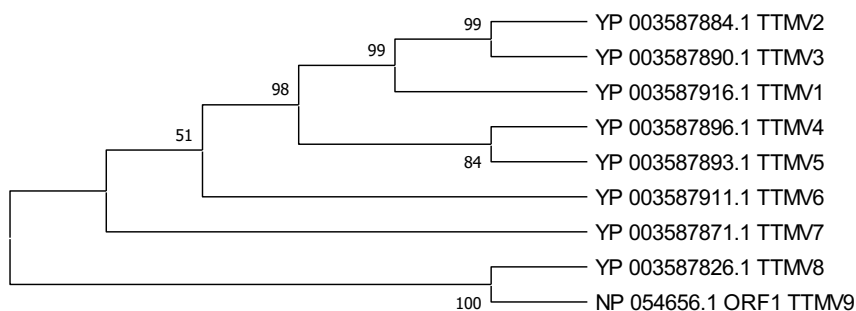
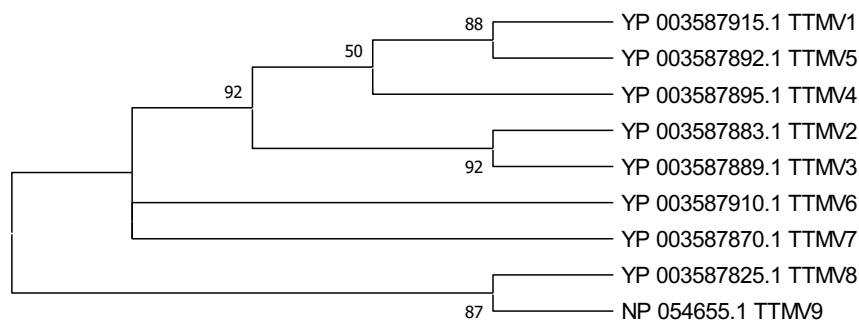


Fig 2. TTMV Phylogenetic Tree Construction

Each of the sequences submitted to the bootstrap method phylogenetic construction for TTMV has been updated in Fig 2.

e) Phylogenetic Tree ORF 2 Protein Sequences



Evolutionary analysis by Maximum Likelihood method

The evolutionary history was inferred by using the Maximum Likelihood method and JTT matrix-based model. The bootstrap consensus tree inferred from 1000 replicates represents the evolutionary history of the taxa analyzed.⁵⁷ The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. Evolutionary analyses were conducted in MEGA X in Figure 2a-e³⁻⁵.

Fig 3. TTMDV Phylogenetic Tree Construction

Each of the sequences submitted for the bootstrap method phylogenetic construction for TTMV has been updated in Fig 3

a) TTMDV Whole Genome Sequences Phylogenetic Tree

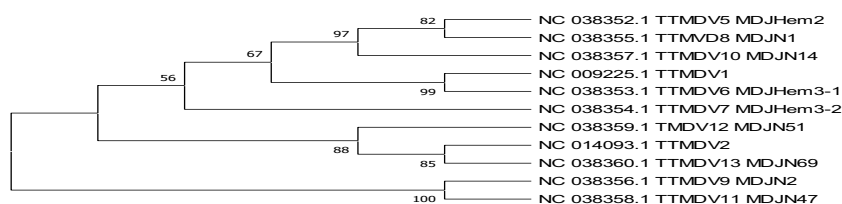


Fig 3. TTMDV Phylogenetic Tree Construction

Each of the sequences submitted for the bootstrap method phylogenetic construction for TTMV has been updated in Fig 3

b) ORF 1 DNA Sequences Phylogenetic Tree

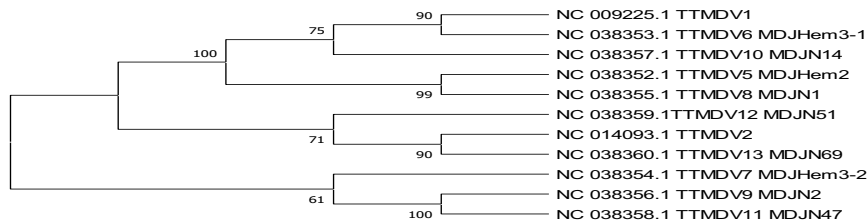


Fig 3. TTMDV Phylogenetic Tree Construction

Each of the sequences submitted for the bootstrap method phylogenetic construction for TTMV has been updated in Fig 3

c) ORF 2 DNA Sequences Phylogenetic Tree

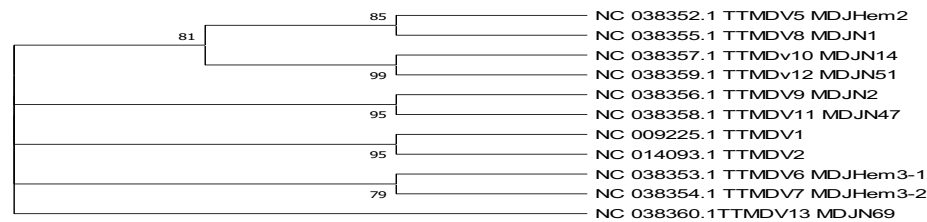


Fig 3. TTMDV Phylogenetic Tree Construction

Each of the sequences submitted for the bootstrap method phylogenetic construction for TTMV has been updated in Fig 3

d) ORF 3 DNA Sequences Phylogenetic Tree

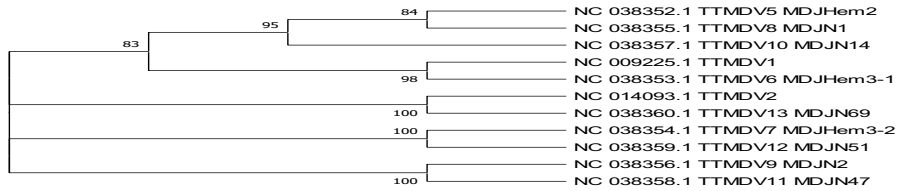


Fig 3. TTMDV Phylogenetic Tree Construction

Each of the sequences submitted for the bootstrap method phylogenetic construction for TTMV has been updated in Fig 3

e) ORF 1 Protein Sequences Phylogenetic Tree

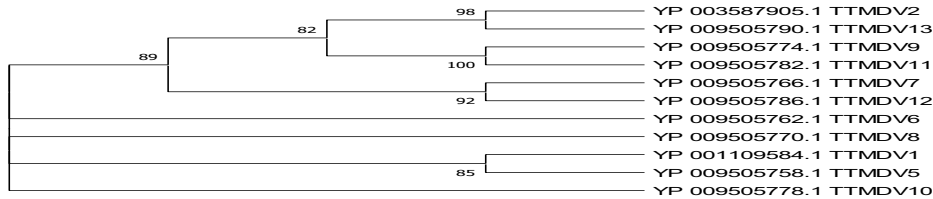


Fig 3. TTMDV Phylogenetic Tree Construction

Each of the sequences submitted for the bootstrap method phylogenetic construction for TTMV has been updated in Fig 3

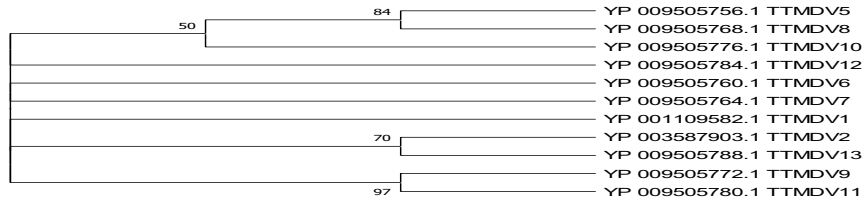
f) ORF 2 Protein Sequences Phylogenetic Tree



Fig 3. TTMDV Phylogenetic Tree Construction

Each of the sequences submitted for the bootstrap method phylogenetic construction for TTMV has been updated in Fig 3

g) ORF 3 Protein Sequences Phylogenetic Tree



Evolutionary analysis by Maximum Likelihood method

The evolutionary history was inferred by using the Maximum Likelihood method and JTT matrix-based model. The bootstrap consensus tree inferred from 1000 replicates represents the evolutionary history of the taxa analyzed.⁵⁷ The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. Evolutionary analyses were conducted in MEGA X in Figure 2a-e³⁻⁵.

Fig 4. Identified Motif for TTMV sequences.
(Each three motifs are colour coded accordingly)

Fig 4(a): Whole Genome Sequences

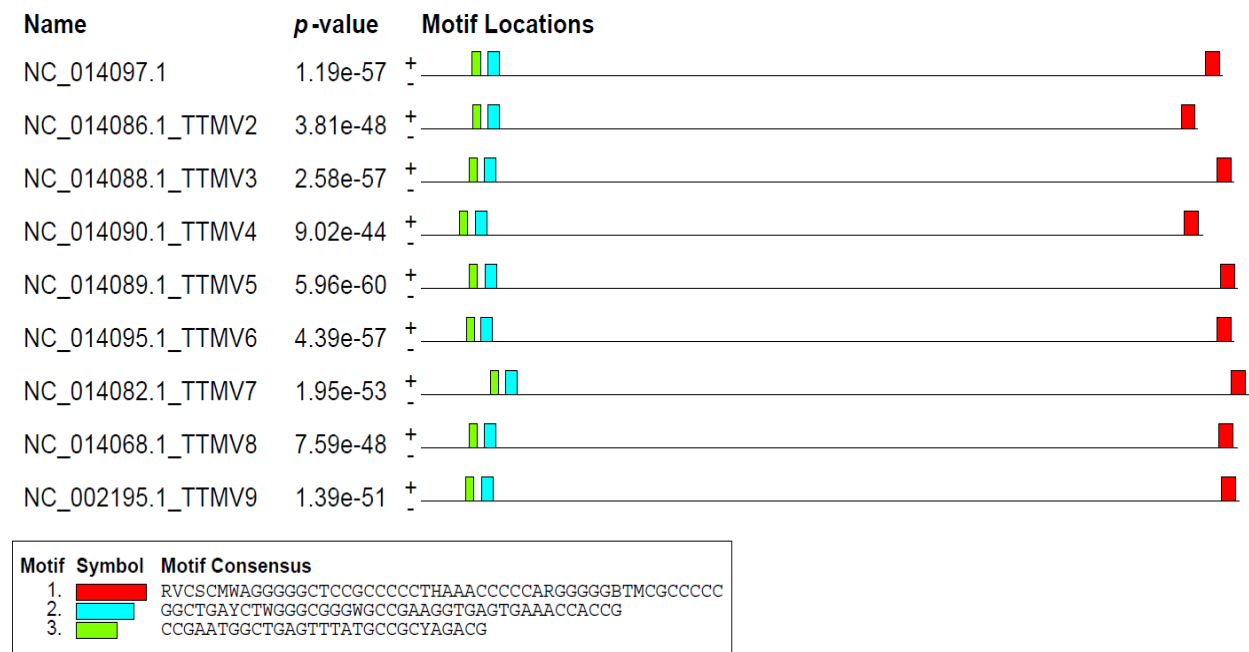


Fig 4(b): ORF 1 DNA Sequences

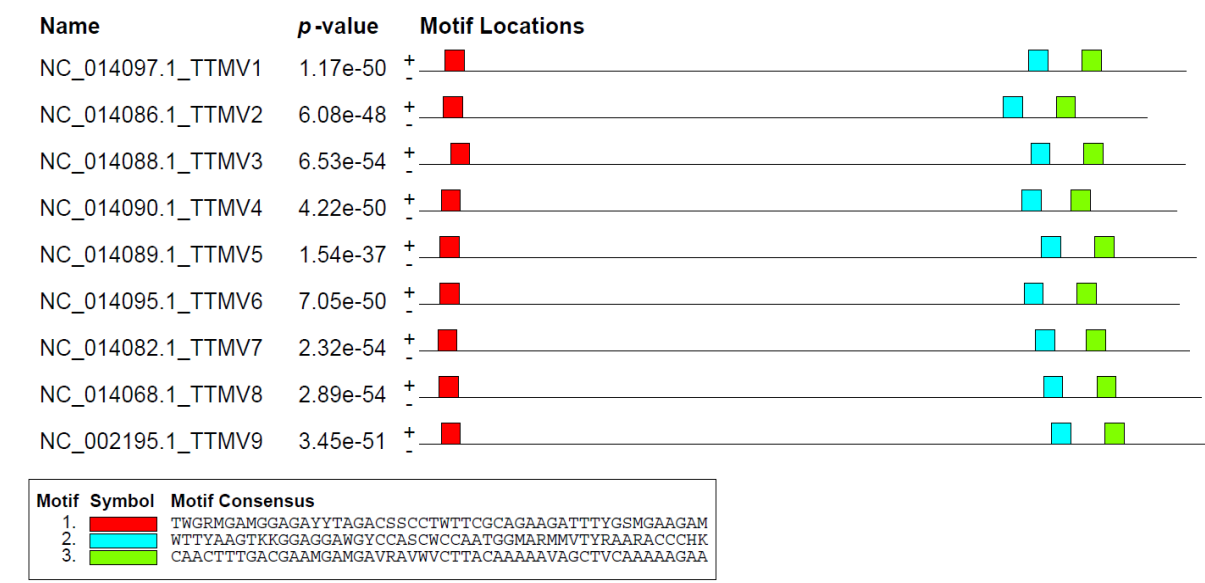


Fig 4(c): ORF 2 DNA Sequences

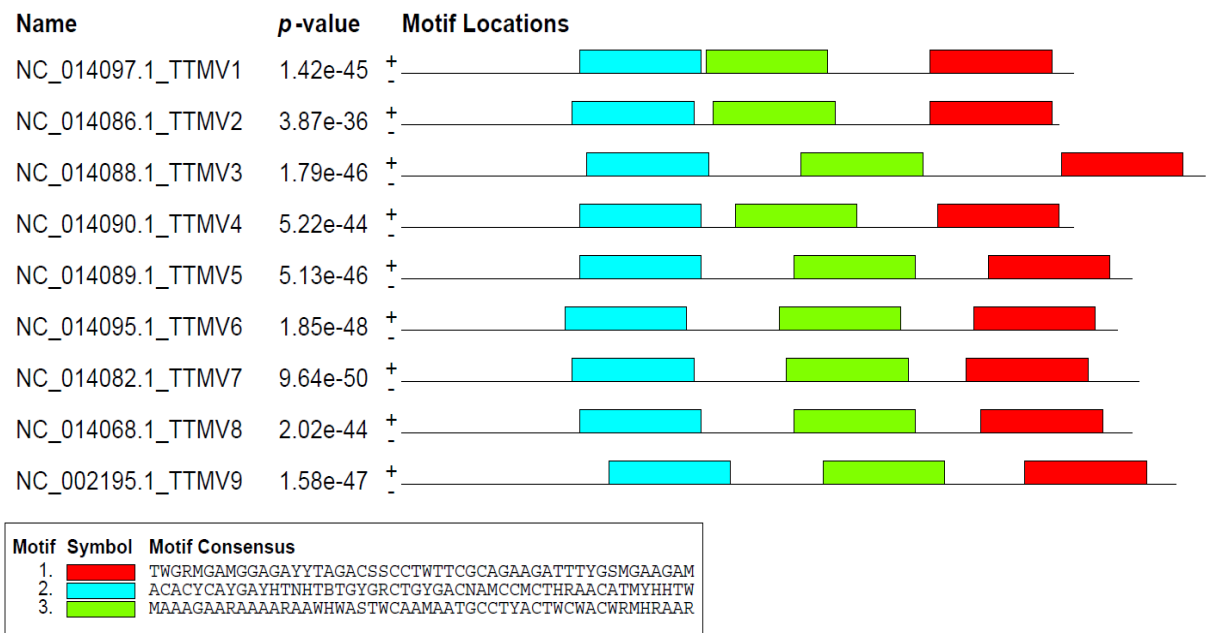


Fig 4(d): ORF 1 Protein Sequences

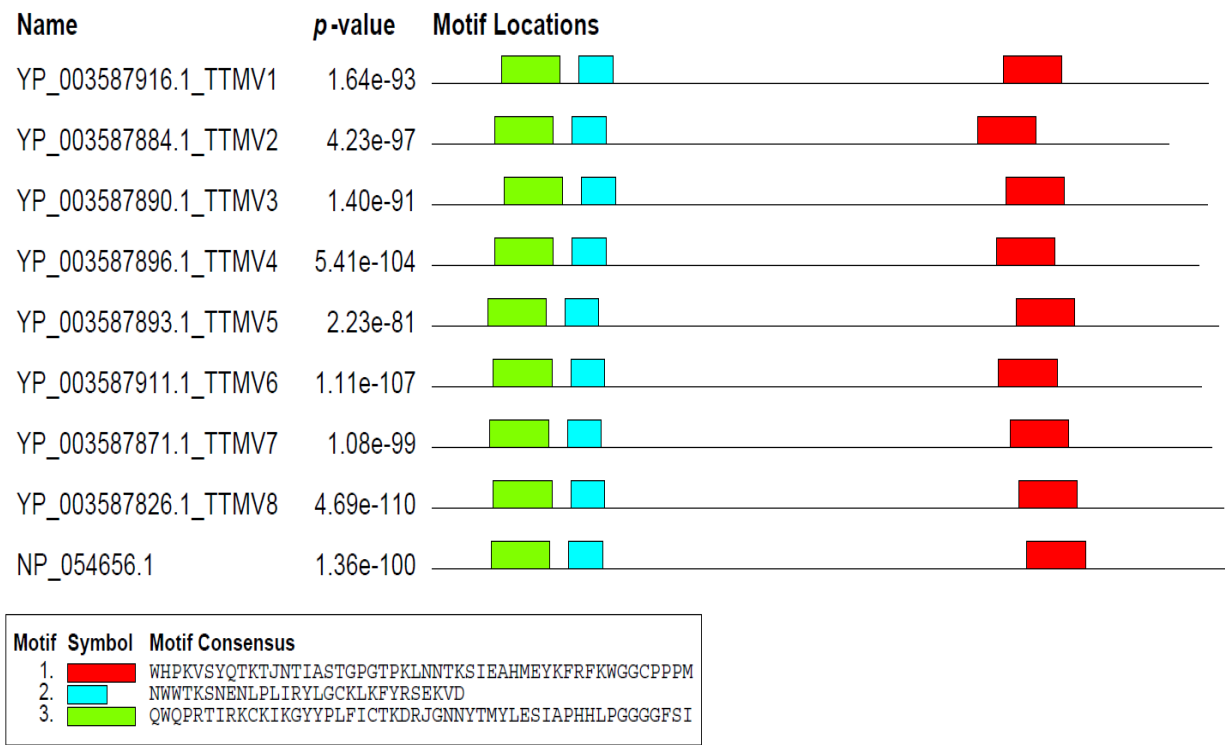


Fig 4(e): ORF 2 Protein Sequences

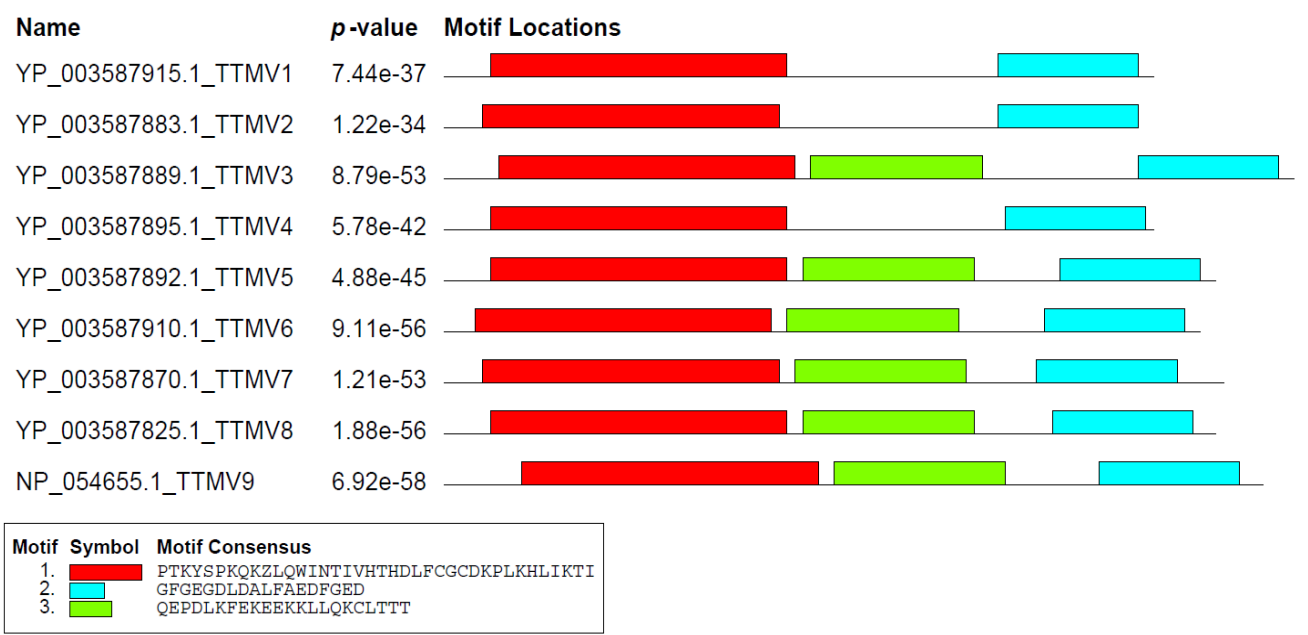
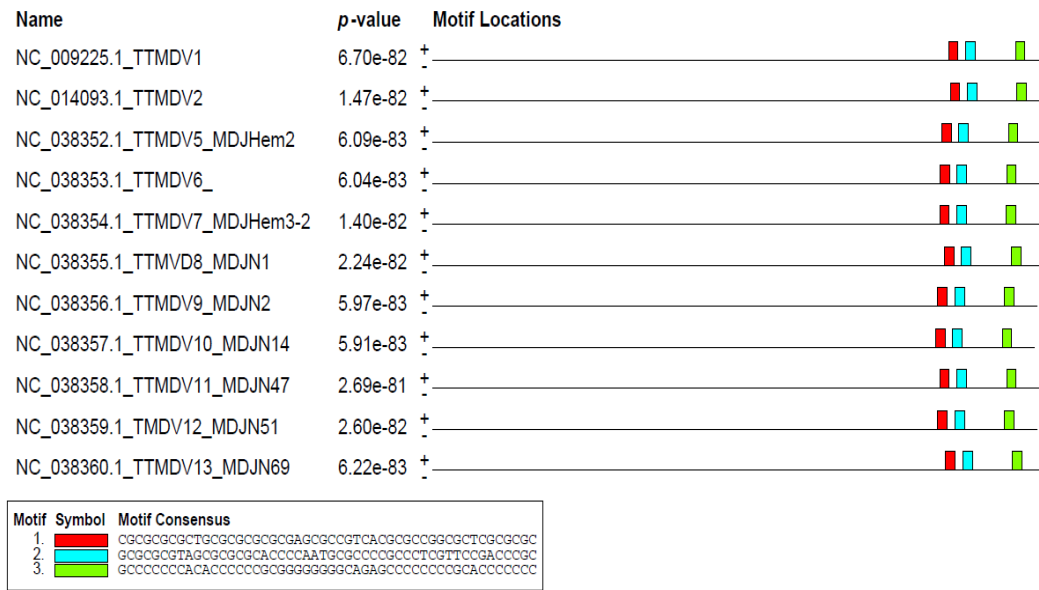
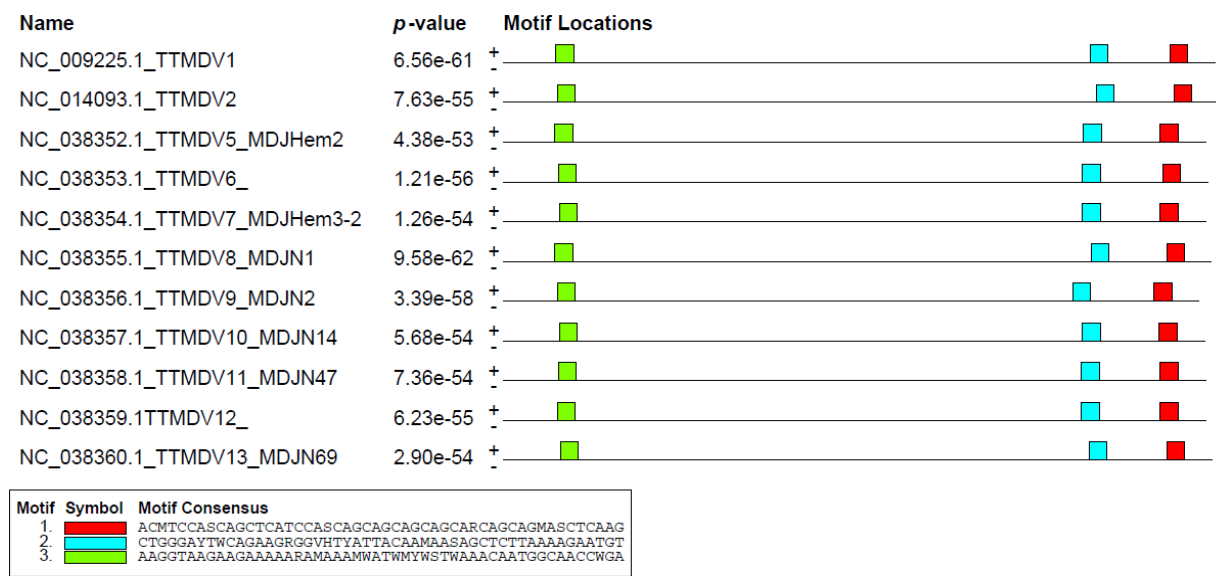


Fig 5(a). Identified Motif for TTMDV sequences.
(Each three motifs are colour coded accordingly)

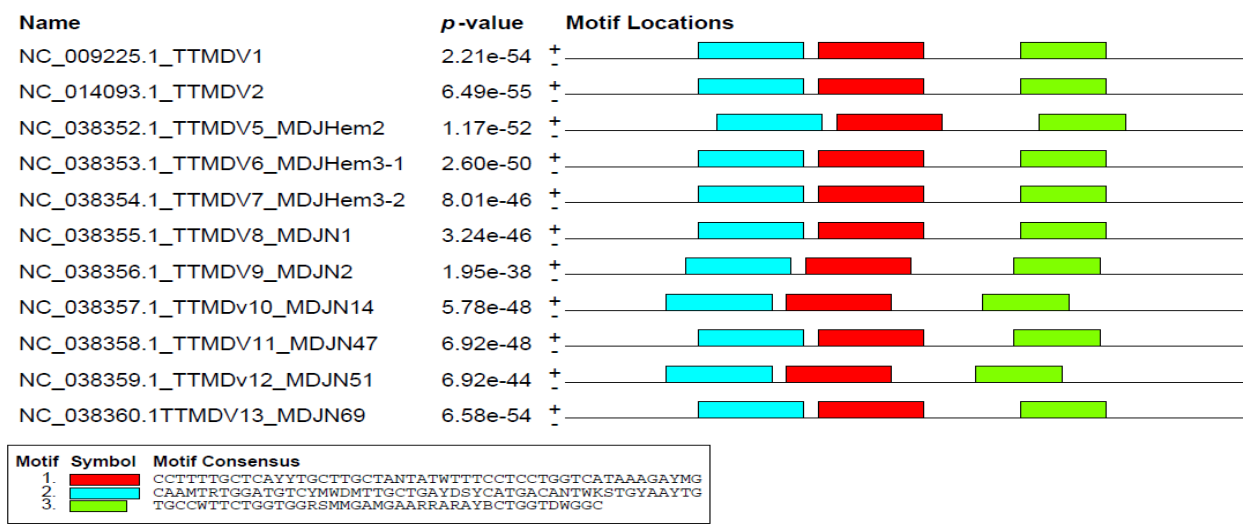
i. Fig 5(a): Whole Genome Sequences



ii. Fig 5(a): ORF 1 DNA



iii. Fig 5(a): ORF 2 DNA



iv. Fig 5(a): ORF 3 DNA Sequences

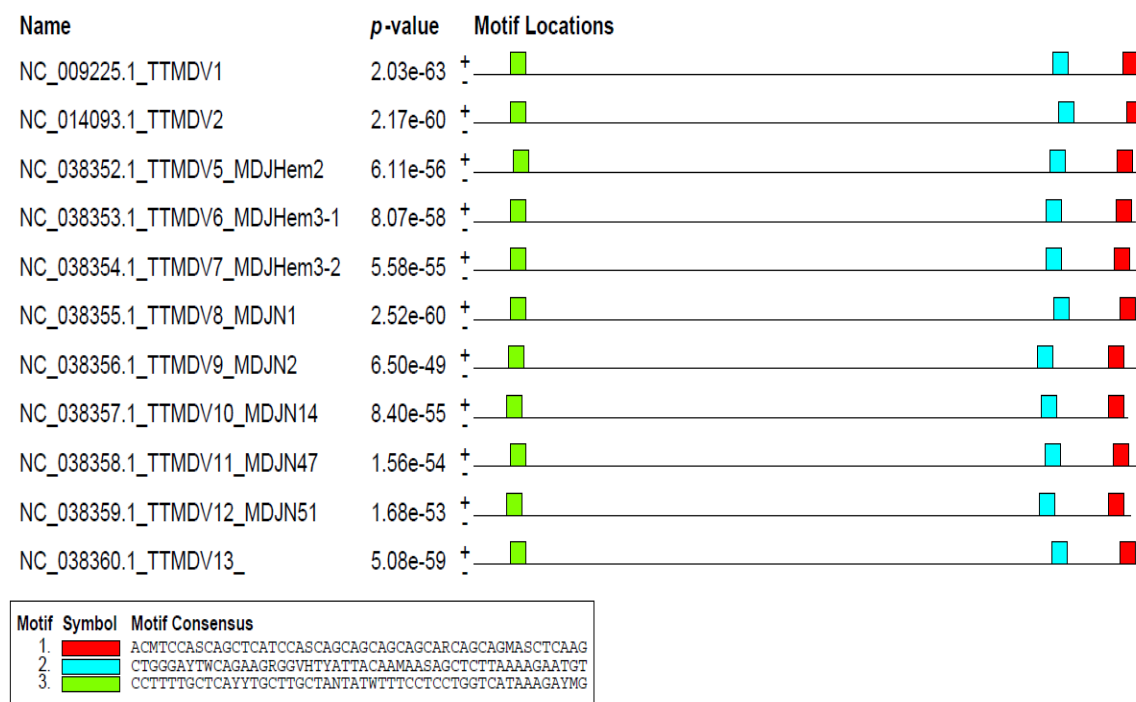


Fig 5(b). Identified Motif for TTMDV sequences

(Each three motifs are colour coded accordingly)

i. Fig 5(b): ORF 1 Protein

Name	p-value	Motif Locations
YP_001109584.1_TTMDV1	9.67e-124	
YP_003587905.1_TTMDV2	2.40e-138	
YP_009505758.1_TTMDV5	2.66e-134	
YP_009505762.1_TTMDV6	4.49e-122	
YP_009505766.1_TTMDV7	2.36e-115	
YP_009505770.1_TTMDV8	1.29e-121	
YP_009505774.1_TTMDV9	2.09e-117	
YP_009505778.1_TTMDV10	4.99e-127	
YP_009505782.1_TTMDV11	6.20e-119	
YP_009505786.1_TTMDV12	2.33e-119	
YP_009505790.1_TTMDV13	2.57e-122	

Motif	Symbol	Motif Consensus
1.		QWQPESIRKCKIKGLGLVGLGABGKQYVCYTNVKTEYTPPKAPGGGGFGV
2.		YLVEEYKARNNIWTKSNIYKDLVRYTGCKITFYRHPYTDFI IKYDRQPPF
3.		LWYPTVEHQLEJTJNSLVEAGPYVPKLDNQKNSTWELNYFYTFYFKWGGPY

ii. Fig 5(b): ORF 2 Protein

Name	p-value	Motif Locations
YP_001109583.1_TTMDV1	6.57e-68	
YP_003587904.1_TTMDV2	2.11e-69	
YP_009505757.1_TTMDV5	5.43e-67	
YP_009505761.1_TTMDV6	1.48e-63	
YP_009505765.1_TTMDV7	3.02e-61	
YP_009505769.1_TTMDV8	6.32e-67	
YP_009505773.1_TTMDV9	5.49e-47	
YP_009505777.1_TTMDV10	2.90e-60	
YP_009505781.1_TTMDV11	2.62e-65	
YP_009505785.1_TTMDV12	7.72e-60	
YP_009505789.1	1.81e-63	

Motif	Symbol	Motif Consensus
1.		DFYKPTPYNQETKNQLWMSQIADSHDSWCNCDTPFAHLLASIFPPGHKDR
2.		LTINQILERDYKZ
3.		CLSGGEEEEESG

iii. Fig 5(b): ORF 3 Protein

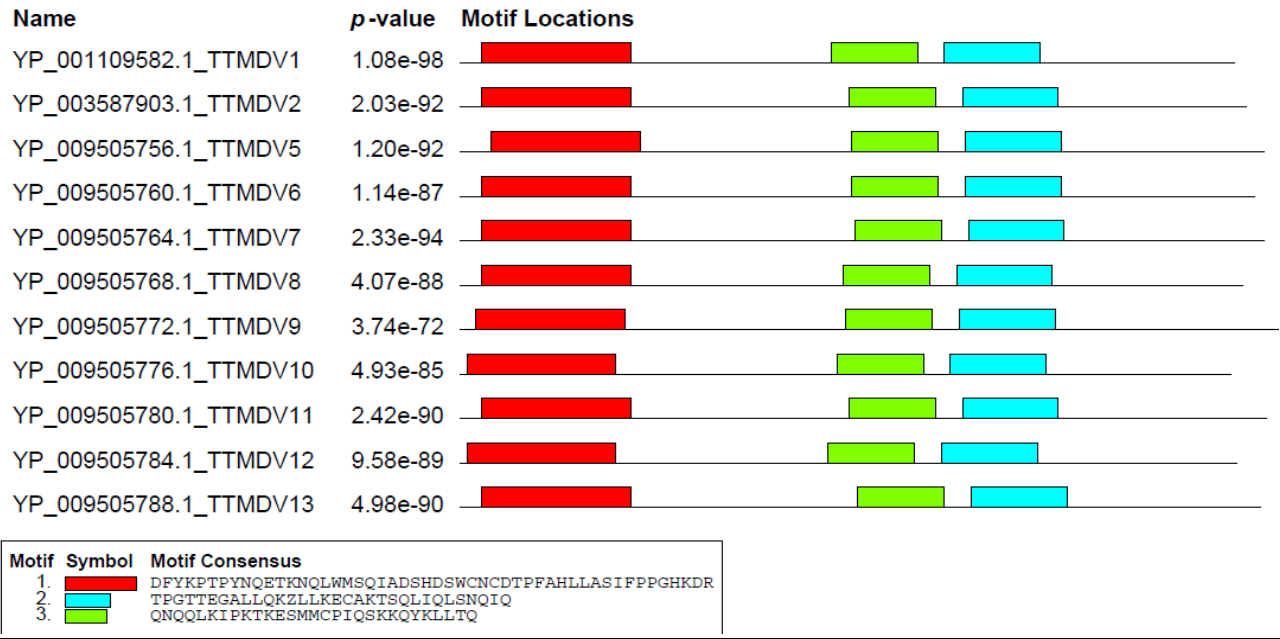


Table 1. Information of the TTMV sequences used in this study

Table 2. Information of the TTMDV sequences used in this study

Table 1. Information of the TTMV sequences used in this study

1 (a). TTMV DNA Sequences Reference Numbers

Sequence Name	Accession Number	Size (bp)	ORF DNA Regions	
			ORF 1	ORF 2
TTMV 1	NC_014097	2856	508-2496	356-631
TTMV 2	NC_014086	2765	510-2396	355-624
TTMV 3	NC_014088	2897	532-2517	341-670
TTMV 4	NC_014090	2785	532-2517	306-581
TTMV 5	NC_014089	2908	528-2543	340-639
TTMV 6	NC_014095	2897	516-2486	334-627
TTMV 7	NC_014082	2952	601-2598	416-718
TTMV 8	NC_014068	2910	529-2556	341-640
TTMV 9	NC_002195	2915	514-2565	314-631

Table 1 (b). TTMV Protein Sequence references Numbers

Sequences	Accession Number ORF1	Name of the Protein ORF 1	Accession Number ORF2	Name of the Protein ORF2
TTMV 1	YP_003587916	hypothetical protein TTMV1_gp2	YP_003587915	hypothetical protein TTMV1_gp1
TTMV 2	YP_003587884	hypothetical protein TTMV2_gp2	YP_003587883	hypothetical protein TTMV2_gp1
TTMV 3	YP_003587890	hypothetical protein TTMV3_gp2	YP_003587889	hypothetical protein TTMV3_gp1
TTMV 4	YP_003587896	hypothetical protein TTMV4_gp2	YP_003587895	hypothetical protein TTMV4_gp1
TTMV 5	YP_003587893	hypothetical protein TTMV5_gp2	YP_003587892	hypothetical protein TTMV5_gp1
TTMV 6	YP_003587911	hypothetical protein TTMV6_gp2	YP_003587910	hypothetical protein TTMV6_gp1
TTMV 7	YP_003587871	hypothetical protein TTMV7_gp2	YP_003587870	hypothetical protein TTMV7_gp1
TTMV 8	YP_003587826	ORF1	YP_003587825	ORF2
TTMV 9	NP_054656	ORF1	NP_054655	ORF2

Table 2. Information of the TTMDV sequences used in this study

2 (a). TTMDV DNA Sequences Reference Numbers

Sequence Name	Accession Number	Size (bp)	ORF Regions Range DNA		
			ORF1	ORF2	ORF3
TTMDV 1	NC_009225	3245	432-2453	217-588	217-2388
TTMDV 2	NC_014093	3253	434-2467	219-593	219-2402
TTMDV 5_MDJHem2	NC_038352	3208	441-2435	217-594	217-2388
TTMDV 6_MDJHem3-1	NC_038353	3198	432-2432	217-594	217-2367
TTMDV 7_MDJHem3-2	NC_038354	3197	435-2429	220-600	220-2376
TTMDV 8_MDJN1	NC_038355	3223	437-2446	222-590	222-2393
TTMDV 9_MDJN2	NC_038356	3187	430-2406	218-589	218-2380
TTMDV 10_MDJN14	NC_038357	3175	416-2407	219-581	219-2342
TTMDV 11_MDJN47	NC_038358	3198	431-2425	219-593	219-2378
TTMDV 12_MDJN51	NC_038359	3187	413-2407	219-572	219-2351
TTMDV 13_MDJN69	NC_038360	3230	434-2446	219-602	219-2387

Table 2 (b). TTMDV Protein Sequence references Numbers

Sequences	Accession Number ORF1	Name of the Protein ORF 1	Accession Number ORF2	Name of the Protein ORF2	Accession Number ORF3	Name of the Protein ORF 3
TTMDV 1	YP_001109584	hypothetical protein TTMidV_gp1	YP_001109583	hypothetical protein TTMidV_gp2	YP_001109582	hypothetical protein TTMidV_gp3
TTMDV 2	YP_003587905	hypothetical protein TTmidV2_gp4 [TTMDV] ORF 1	YP_003587904	hypothetical protein TTmidV2_gp3	YP_003587903	hypothetical protein TTmidV2_gp2
TTMDV 5_MDJHem2	YP_009505758	hypothetical protein	YP_009505757	hypothetical protein	YP_009505756	hypothetical protein
TTMDV 6_MDJHem3-1	YP_009505762	hypothetical protein	YP_009505761	hypothetical protein	YP_009505760	hypothetical protein
TTMDV 7_MDJHem3-2	YP_009505766	hypothetical protein	YP_009505765	hypothetical protein	YP_009505764	1 hypothetical protein
TTMDV 8_MDJN1	YP_009505770	hypothetical protein	YP_009505769	hypothetical protein	YP_009505768	hypothetical protein
TTMDV 9_MDJN2	YP_009505774	hypothetical protein	YP_009505773	hypothetical protein	YP_009505772	hypothetical protein
TTMDV 10_MDJN14	YP_009505778	hypothetical protein	YP_009505777	hypothetical protein	YP_009505776	hypothetical protein
TTMDV 11_MDJN47	YP_009505782	hypothetical protein	YP_009505781	hypothetical protein	YP_009505780	hypothetical protein
TTMDV 12_MDJN51	YP_009505786	hypothetical protein	YP_009505785	hypothetical protein	YP_009505784	hypothetical protein
TTMDV 13_MDJN69	YP_009505790	hypothetical protein	YP_009505789	hypothetical protein	YP_009505788	hypothetical protein

Table 3.(a). Information content (IC) values of each three identified TTMV motifs

Source	Multilevel Consensus Sequences Motifs*	IC Value
Whole Genome TTMV DNA Sequences	RVCSCMWAGGGGGCTCCGCCCCCTHAAACCCCCARGGGGGBTMCGCCCCC	64.8
	GGCTGAYCTWGGGCGGGWGCCGAAGGTGAGTGAAACCACCG	71.2
	CCGAATGGCTGAGTTTATGCCGCTYAGACG	55.3
ORF 1 TTMV DNA Sequences	TWGRMGAMGGAGAYYTAGACSSCCTWTCGCAGAAGATTTYGSMGAAGAM	67.8
	WTTYAAGTKKGGAGGAWGYCCASCWCCAATGGMARMVITYRAARACCCHK	63.5
	CAACTTTGACGAAMGAMGAVRAVWVCTTACAAAAAVAGCTVCAAAAAGAA	65.3
ORF 2 TTMV DNA Sequences	TWGRMGAMGGAGAYYTAGACSSCCTWTCGCAGAAGATTTYGSMGAAGAM	67.9
	ACACYCAYGAYHTNHTBTGYGRCTGYGACNAMCCMCTHRAACATMYHHTW	55.4
	MAAAGAARAAAARAASHWASTWCAAMAATGCCTYACTWCWACWRMHRAAR	50.6
ORF 1 TTMV Protein Sequences	WHPKVSQTKTJNTIASTGPGTPKLNNTKSIEAHMEYKFRFKWGGCPPPM	146.8
	NWWTKSNENLPLIRYLGCKLKFYRSEKVD	91.4
	QWQPRITRKCKIKGYYPFICTKDRJGNNYTMYLESIAPHHLPGGGGFSI	147.3
ORF 2 TTMV Protein Sequences	PTKYSPKQKZLQWINTIVHTHDLFCGCDKPLKHLIKTI	105.6
	GFGEGLDALFAEDFGED	59.2
	QEPDLKFEKEEKLLQKCLTTT	68.6

* Consensus sequence of a motif set is a pseudo sequence that has a minimum average distance to every motif in a set.

Table 3.(b) Information content (IC) values of each three identified TTMDV motifs are listed below

Source	Multilevel Consensus Sequences Motifs* (5' to 3')	IC Value
Whole Genome TTMDV DNA Sequences	CGCGCGCGCTGCGCGCGCGCGAGCGCCGTCACGCGCCGGCGCTCGCGCGC	99.1
	GCGCGCGTAGCGCGCGCACCCCAATGCGCCCCGCCCTCGTTCCGACCCGC	100
	GCCCCCCCACACCCCCCGCGGGGGGGGCAGAGCCCCCCCCGCACCCCCC	97.4
ORF 1 TTMDV DNA Sequences	ACMTCCASCAGCTCATCCASCAGCAGCAGCAGCARCAGCAGMASCTCAAG	73.9
	CTGGGAYTWCAGAAAGRGVHTYATTACAAMAASAGCTCTTAAAAGAATGT	73.9
	AAGGTAAGAAGAAAAARAMAAAMWATWMYWSTWAAACAATGGCAACCWGA	75.7
ORF 2 TTMDV DNA Sequences	CCTTTTGCTCAYYTGCTTGCTANTATWTTTCCTCCTGGTCATAAAGAYMG	72.5
	CAAMTRTGGATGTCYMWDMTGCTGAYDSYCATGACANTWKSTGYAAYTG	63.1
	TGCCWTTCTGGTGGRSMMGAMGAARRARAYBCTGGTDWGGC	51.1
ORF 3 TTMDV DNA Sequences	ACMTCCASCAGCTCATCCASCAGCAGCAGCAGCARCAGCAGMASCTCAAG	73.9
	CTGGGAYTWCAGAAAGRGVHTYATTACAAMAASAGCTCTTAAAAGAATGT	73.9
	CCTTTTGCTCAYYTGCTTGCTANTATWTTTCCTCCTGGTCATAAAGAYMG	72.5
ORF 1 TTMDV Protein Sequences	QWQPESIRKCKIKGLGLTLVGAEGKQYVCYTNVKTEYTPPKAPGGGGFGV	161.2
	YLYEEYKARNNIWTKSNIYKDLVRYTGCKITFYRHPYTDFIIKYDRQPPF	157.3
	LWYPTVEHQLETJNSLVEAGPYVPKLDNQKNSTWELNYFYTFYFKWGGPY	150.3
ORF 2 TTMDV Protein Sequences	DFYKPTPYNQETKNQLWMSQIADSHDSWCNCDTPFAHLLASIFPPGHKDR	165.6
	LTINQILERDYKZ	39.8
	CLSGGEEEEEDSG	36.3
ORF 3 TTMDV Protein Sequences	DFYKPTPYNQETKNQLWMSQIADSHDSWCNCDTPFAHLLASIFPPGHKDR	165.6
	TPGTTEGALLQKZLLKECAKTSQLIQLSNQIQ	100.8
	QNQQLKIPKTKESMMCPIQSKKQYKLLTQ	78.1

* Consensus sequence of a motif set is a pseudo sequence that has a minimum average distance to every motif in a set.