

https://dspace.org/

Pathogenicity determinants of severe and mild variants of tomato curly stunt virus

Brigden, Alison

2022
https://hdl.handle.net/10539/34495

Downloaded from a DSpace repository. https://dspace.org/



Pathogenicity determinants of severe and mild variants of Tomato curly stunt virus

By

Alison Brigden

(1380539)

Dissertation

BSc Masters

in

Molecular and Cell Biology

in the Faculty of Science, University of the Witwatersrand, Johannesburg, South Africa

Supervisor: Professor M.E.C Rey

April 2022

Contents

Declaration 4
Acknowledgements 5
List of Figures 6
List of Tables 13
List of abbreviations 14
ABSTRACT 18
CHAPTER 1 - LITERATURE REVIEW 20
1.1 Agricultural importance of tomato crops 20
1.2 Virus pathogenicity: a threat to tomato crops 20
1.3 The discovery and origin of *Tomato curly stunt virus* 21
1.4 The pathogenicity of *Tomato curly stunt virus* in tomato 21
1.5 Begomovirus structure and replication 23
1.6 Structure of the genomes of *Tomato curly stunt virus* 24
1.7 The significance of the intergenic region (IR) in ToCSV 26
1.8 The significance of the V2 gene in ToCSV 26
1.9 Mixed infections 27
1.9.1 Mixed infections of Geminiviruses 29
1.10 The geminiviral C5 ORF and its role in pathogenicity 29
1.11. Rationale for study 30
1.12. Aims and specific objectives 31
1.12.1 Aims 31
1.12.2 Specific objectives 31
CHAPTER 2 - METHODOLOGY 32
2.1 Infectious clone constructs 32
2.2 Investigative bioinformatics for predicting tomato and *N. benthamiana* targeting siRNA from the IR 35
2.3 Site directed mutagenesis 35
2.3.1 V22V2 single amino acid mutants 35
2.3.2 Extension of the 5’ end of the putative C5 ORF via site directed mutagenesis 40
2.4 Structural and disorder analysis on V22 mutant proteins 43
2.5 Growing *N. benthamiana* for inoculations with infectious clones 44
2.6 Agroinoculation of *N. benthamiana* with infectious clones 44
2.7 Collection of symptom progression data 45

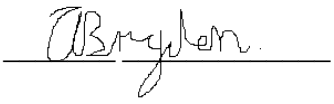
2.8 Nucleic acid extraction....46
2.9 Viral load....47
2.10 Detection of RNA transcripts....48
2.11 Data analysis and statistics....49
CHAPTER 3 - RESULTS....50
3.1 The Intergenic Region (IR)....50
3.1.1 Verification of IR swap infectious clone constructs....50
3.1.2 Height, symptom severity scores and viral load data of inoculated *N. benthamiana*....51
3.1.3 siRNA from the IR that may target *N. benthamiana* and tomato genomes....55
3.2 The V2 ORF....56
3.2.1 Bioinformatics comparison of the V30 and V22 V2 proteins....56
3.2.2 Bioinformatic analysis of V22 V2 single aa mutants....58
3.2.3 Single V22-V2 mutant amino acid clone verification....61
3.2.4 Heights, symptom scoring and viral load of *N. benthamiana* inoculated with V22ΔV2 mutants....63
3.3 Mixed infections with wildtype V30/ V22 and wildtype V30/ V30ΔIR-s clone constructs....67
3.4 Investigation of the putative C5 ORF....70
3.4.1 Bioinformatics analysis of the putative C5 protein....70
3.4.2 Infectivity trials of V22 C5 ORF mutants in *N. benthamiana*....71
CHAPTER 4 - DISCUSSION....74
4.1 The intergenic region (IR)....74
4.2 The V2 ORF....77
4.3 Mixed infections....80
4.4 Investigation of the putative C5 ORF....83
CHAPTER 5 - Concluding remarks....86
References....88
Supplementary Data....99
Section 1 – The Intergenic Region (IR)....99
Section 2 – The V2 open reading frame (ORF)....106
Section 3 – Investigation of the putative C5 open reading frame (ORF)....110

Declaration

I, Alison Brigden (student number: 1380539) am a student registered for the degree BSc (Masters by dissertation) in Microbiology and Biotechnology in the academic year of 2021.

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that ALL the work submitted for assessment for the above course is my own unaided work except where I have explicitly indicated otherwise.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is in not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.

Signature:  11 April 2022

Acknowledgements

I would like to acknowledge Wits Postgraduate Merit award (PGMA) and the National Research Fund (NRF).

I would like to thank my supervisor, Professor Rey for her supervision, advice, and support.

I would also like to thank Mr. Alexander Michael Zwolinski for his endless help and enthusiasm throughout this project. My appreciation extends to Miss Frances Ayres for her guidance and advice.

List of Figures

Figure 1 Solanum lycopersicum infected with severe and mild variants of Tomato curly stunt virus (ToCSV). Close up of the leaves show yellowing and leaf curling present. Severe symptoms induced by ToCSV-V30, and mild symptoms induced by ToCSV-V22. Images taken at 30 days post inoculation (dpi). Photos taken at the University of the Witwatersrand 2020.....22

Figure 2 Nicotiana benthamiana inoculated with ToCSV-V30 and ToCSV-V22. Symptoms induced as described in the image. N. benthamiana was inoculated with mock (pCambia vector in Agrobacterium tumefaciens) for the negative control. Images taken at 20 days post inoculation (dpi). Photos taken at the University of the Witwatersrand 2022.....23

Figure 3 Diagram showing the genome organization of Tomato curly stunt virus (ToCSV) DNA-A with a length of 2766 nucleotides, containing 6 open reading frames (ORFs), V1, V2, C1, C2, C3 and C4. The Intergenic region contains the origin of replication (ORI) as well as the short nucleotide motif hairpin (Esterhuizen, 2012). 25

Figure 4 Alignment of the 3' end of the intergenic region (IR) of Tomato curly stunt virus (ToCSV)-V30 and ToCSV-V22 (nucleotide position 61 to 139). Alignment done on BioEdit Sequence Alignment editor.....26

Figure 5 A detailed vector map of pCambia2300 with sequence length of 8743 base pairs (bp), selection genes, restriction sites, Origin of Replication (ORI), promoters and the T-DNA repeats labelled (source: Insightful Science; available at Snapgene.com). 32

Figure 6 Diagram showing the V30 and V22 genomes with the six open reading frames (ORFs), V1, V2, C1, C2, C3 and C4 indicated. In the enlarged region of the Intergenic region (IR)- swap infectious clone constructs are shown, where selected V22 IR sequences were swapped out with that of the respective V30 sequence. A conserved TATA box in the IR is shown in grey. Red lines refer to the V30 sequence, with the blue line representing V22 sequence. Nucleotide (nt) positions are relative to the Origin of Replication (ORI) which is nt 1.....33

Figure 7 Diagram of the V30 backbone *intergenic region (*IR*)*-swap infectious clone construct where the selected V30 IR sequence was swapped out with that of the respective V22 sequence. A conserved TATA box shown in grey. Nucleotide (nt) positions are relative to the Origin of Replication (ORI) which is nt 1.V22 sequence is shown in blue, with V30 represented in red.

The six open reading frames (ORFs) in V30 and V22, V1, V2, C1, C2, C3 and C4 are indicated. 35
Figure 8 Diagram of infectious clone V30 and V22 genomes with the six open reading frames (ORFs), V1, V2, C1, C2, C3 and C4 indicated. Variants with the V22 V2 amino acid (aa) substitutions in the V22ΔV2 mutant clones indicated. V30 nucleotide sequence is shown in red and V22 nucleotide sequence in blue. Valine (V), serine (S), arginine (R), leucine (L), threonine (T), lysine (K), glutamine (Q), aspartic acid (D), proline (P) are aa indicated by their respective symbols. The top boxed region in the diagram is a representation of the aa substitutions in the V22 V2 sequence. The aa position at the site of substitutions is indicated above each respective representative V2 mutant protein. The bottom boxed region is a representation of the inadvertent amino acid mutations to the overlapping V1 ORF in the V22ΔV2-T→S+ mutant clone construct. ~~QKP~~ represents the deletion of these aa. Light blue represents the V22 aa sequence and light orange represents the V30 aa sequence. The aa position at the site of substitution is indicated above each respective representative V22 V2 mutant protein. 36
Figure 9 Nucleotide (nt) and amino acid (aa) alignments (https://web.expasy.org/translate/) of the C5 protein of V22 (V22C5aa$^{134\text{-}137}$) and the full length 193 aa long putative C5 in mutant V22ΔC5aa$^{1\text{-}193}$. The amino acids have been numbered in the N-terminal to C-terminal direction using African cassava mosaic virus (ACMV, Accession number J02057) C5 as the reference (Supplementary Figure 3.1). The translated proteins are color coded: V22C5aa$^{134\text{-}193}$ (wildtype) yellow, and V22ΔC5aa$^{1\text{-}193}$ is green. Protein sequences less than 40 aa have been excluded from this diagram. The nt targeted for mutagenesis are color coded green (wildtype sequence) and purple (mutant sequence). The IUPAC-IUNMN one -letter abbreviations for the 20 standard amino acids were used. Nucleotide numbers relative to the first nucleotide of the V22 wildtype sequence are indicated below the sequences with square brackets to indicate nt numbered.....42
Figure 10 V22ΔIR-sl, V22ΔIR-s and V22ΔIR-sr were digested with BglI at 2 sites yielding 2 linear products of 1308 base pairs (bp) and 10172 bp in length as shown on a 1 % agarose gel stained with ethidium bromide. Lane 1) 1 Kb gene Ruler (ThermoScientific), lane 2) V22ΔIR-sl undigested, lane 3) V22ΔIR-sl digested, lane 4) V22ΔIR-sr undigested, lane 5) V22ΔIR-sr digested, lane 6) V22ΔIR-s undigested and lane 7) V22ΔIR-s digested. DNA samples were originally extracted from the respective transformed DH5α E. Coli cells.....50
Figure 11 Plant height relative to control (pCambia2300 empty vector in Agrobacterium tumefaciens) (%) over time measured in days post inoculation (dpi), of wildtype (WT) V22, A) V22ΔIR-s, B) V22ΔIR-sl and C) V22ΔIR-sr. D) Height (mm) of Nicotiana benthamiana

between non-inoculated and mock-inoculated (pCambia2300 vector in A. tumefaciens) treatments over time. *=p<0.05 (student T-test). Bars indicate mean ± SD.....52
Figure 12 Upward leaf roll severity score (ULR), left graphs, and Swollen vein severity score (SV), right graphs, induced in Nicotiana benthamiana for A) V22ΔIR-s, B) V22ΔIR-sr and C) V22ΔIR-sl. *=P<0.05 (student T-test). Bars indicate mean ± SD.....53
Figure 13 Nicotiana benthamiana inoculated with V22 (positive control), V22ΔIR-s, V22ΔIR-sl and V22ΔIR-sr infectious clone construct at 20 days post inoculation (dpi). Symptoms present are upward leaf roll (ULR) and swollen veins (SV). Severity of symptoms are indicated by the colour coded triangles as indicated by the key. 54
Figure 14 Nicotiana benthamiana leaves dissected from plants inoculated with pCambia2300 vector in Agrobacterium tumefaciens (Mock), V30 (positive control), V22 (positive control), V22ΔIR-s, V22ΔIR-sl and V22ΔIR-sr infectious clone construct. Leaves were viewed against the light. Symptoms present are upward leaf roll (ULR), and swollen veins (SV) Severity of symptoms are indicated by the colour coded triangles as explained by the key.....54
Figure 15 Log viral DNA fold change of wildtype (WT) ToCSV_V22 (V22) and V22ΔIR-s in Nicotiana benthamiana at 20- and 28-days post inoculation (dpi). Significance shown with an asterisk. Bars indicate mean ± SD. 55
Figure 16 PCR products, 145 bp in length, from primers targeting the IR in V22 total DNA (tDNA), V22 complementary DNA (cDNA), revertaid control (RT-), mock inoculated cDNA, Mock RT- and non-template control (NTC). A) cDNA from the forward primer (Fp), therefore the complementary sense (cs) transcript of IR. B) cDNA from the reverse primer (Rp), hence the viral sense transcript (vs). M is the DNA ladder.....55
Figure 17 CLUSTAL multiple sequence alignment by MUSCLE (3.8) of WT V30 and V22 V2 amino acid sequences with the degree of similarity of the residues by means of structure and polarity indicated by symbols below the sequences, as described by the key. The arrows indicate the amino acids that were targeted for mutagenesis in this study. 56
Figure 18 (A) PSIPRED sequence map of predicted secondary structure of wildtype (WT) V30 and V22 V2 proteins. Key indicates the predicted structures or function. (B) DISOPRED plot of predicted disorder of V30 and V22 infectious clone construct V2 proteins. QKP+ indicates the 3 additional C-terminal residues of the V22 V2 protein that is absent in V30 V2. The circled residues indicate the C-terminal residues with predicted disorder. The IUPAC-IUBMB one-letter abbreviations for the 20 standard amino acids were used. 57
Figure 19 V2 protein sequences of V22_V2 (wildtype), V22ΔV2-V→S, V22ΔV2-T→S. A) Secondary structures indicated by colour coding as indicated by the key. The amino acid

changes as a result of nucleotide mutagenesis indicated by the boxes around the residues. B) Disorder indicated by a colour coded key. The threshold for prediction of residues to be disordered was 0.5. Disorder and secondary structural predictions obtained from http://bioinf.cs.ucl.ac.uk/psipred/.....59
Figure 20 V1 protein sequences of V22 (wildtype) and V22ΔV2-T→S+. A) Secondary structures indicated by a colour coded key. The amino acid changes as a result of nucleotide mutagenesis indicated by the boxes around the residues. B) Disordered residues indicated as described in the key. The threshold for prediction of residues to be disordered was 0.5. Disorder and secondary structural predictions obtained from http://bioinf.cs.ucl.ac.uk/psipred/. 60
Figure 21 Cellular nuclear localisation (cNLS) mapper result (http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi) for (A) V22 and (B) V22ΔV-T→S+ V1 protein. Predicted NLS sequences are highlighted in red. 60
Figure 22 Putative C6 protein sequences of V22 (wildtype) and V22ΔV2-T→S+. A) Secondary structures indicated by a colour code as described in the key. The amino acid changes as a result of nucleotide mutagenesis indicated by the boxes around the residues. B) Disordered residues are indicated by a colour code as described in the key. The threshold for prediction of residues to be disordered was 0.5. Disorder and secondary structural predictions obtained from http://bioinf.cs.ucl.ac.uk/psipred/ 61
Figure 23 1 % agarose gel showing 1 kb Gene Ruler, Thermo Scientific™ (lane 1), V22 plasmid digested with SpeI (Lane 2) revealing 2 bands of 11 185 base pairs (bp) and 8 420 bp in length, 3 clones of V22ΔV2S→L+ digested with SpeI for 4 hours (lanes 3-5) and 3 clones of V22ΔV2S→L+ digested for 12 hours (lanes 6-8) revealing bands of 248 bp, 2517 bp and 8830 bp in length. The expected bands are indicated with boxes.....62
Figure 24 1 % agarose gel of 1 Kb Gene Ruler Thermo Scientific™ (lane 1), SmaI digested product of V22 (lane 2), SmaI digested products of V22ΔV2R→K+ of 6 clones (lanes 3 – 8), SalI digested of wildtype (WT) V22 clone (lane 9) SalI digested products of 5 clones of V22ΔV2T→S (lanes 10 -14). 62
Figure 25 Plant height relative to control (mock) (%) over time measured in days post inoculation (dpi), of wildtype (WT) V22, A) V22ΔV2-V→2, B) V22ΔV2-T→S+ and C) V22ΔV2-T→S. Control was Nicotiana benthamiana inoculated with mock inoculum (empty pCambia2300 vector in Agobacterium tumefaciens). 63
Figure 26 Upward leaf roll severity score (ULR) (left) and swollen vein severity score (SV) (right) induced in Nicotiana benthamiana for A) V22ΔV2-V→S, B) V22ΔV-T→S and C)

V22ΔV-T→S+. Bars indicate mean ± SD. Significance level: * = p < 0.05. The IUPAC-IUMBMB one-letter abbreviations for the 20 standard amino acids were used.....64
Figure 27 Nicotiana benthamiana inoculated with V22 (positive control) and V22ΔV2-V→S infectious clone constructs at 28 days post inoculation (dpi). Symptoms present are upward leaf roll (ULR) and swollen veins (SV). Severity of symptoms indicated by colour coded triangles as described in the key. Time point 28 dpi is depicted to show the delay in ULR recovery compared to V22.....65
Figure 28 Nicotiana benthamiana inoculated with V22 (positive control) and V22ΔV2-T→S infectious clone constructs at 20 days post inoculation (dpi). Symptoms present are upward leaf roll (ULR) and swollen veins (SV). Severity of symptoms are indicated by colour coded triangles as described in the key. Time point 20 dpi shows clear reduction in symptom severity compared to V22.....65
Figure 29 Nicotiana benthamiana inoculated with V22 (positive control) and V22ΔV2-T→S+ infectious clone constructs at 20 days post inoculation (dpi). The expanded leaves below the meristem have been shown through the light in the circle. Symptoms present are upward leaf roll (ULR) and swollen veins (SV). Severity of symptoms indicated by colour coded triangles as described in the key. The mild sunken veins observed at 20 dpi in V22ΔV2-T→S+ inoculated N. benthamiana can be seen in the close of the expanded leaf with the arrow indicating the mild sunken vein. 66
Figure 30 Nicotiana benthamiana leaves dissected from plants inoculated with pCambia2300 vector in Agrobacterium tumefaciens (Mock), V30 (positive control), V22 (positive control), V22ΔV2-V→S, V22ΔV2-T→S and V22ΔV2-T→S+ infectious clone constructs. Leaves were viewed against the light. Symptoms present are upward leaf roll (ULR) and swollen veins (SV). Severity of symptoms indicated by colour coded triangles as described in the key.....66
Figure 31 Log viral fold change of wildtype (WT) V22 and A) V22ΔV2-V→S and B) V22ΔV2-T→S in *Nicotiana benthamiana* at 20- and 36-days post inoculation (dpi). Bars indicate mean ± SD. Significance level: * = p < 0.05 (student T-test).....67
Figure 32 A) Log viral load fold change of V30, V22 and the V30+V22 (1:1 inoculum) mixed infection at 20 and 36 dpi in Nicotiana benthamiana. Bars indicate mean ± SD. Significance level: * = p < 0.05 (student T-test). B) Plant heigh relative to control (mock: empty pCambia2300 vector in Agrobacterium tumefaciens) (%) C) Swollen vein severity score (SV) observed in the leaves of N. benthamiana inoculated with V30, V22 and V30+V22. D) Upward leaf roll severity score (ULR) induced in the leaves of N. benthamiana inoculated with V30, V22 and V30+V22. Bars indicate mean ± SD. Significance level: * = p < 0.05. **=P<0.005

(student T-test). * Shows significant difference of V30+V22 to V30 or V22 as indicated above V30 or V22. 68
Figure 33 A) Swollen vein severity score (SV) observed in the leaves of *Nicotiana benthamiana* inoculated with V30, V22 and V30+V30ΔIR-s (1:1 inoculum). B) Upward leaf roll severity score (ULR) induced in the leaves of *N. benthamiana* inoculated with V30, V22 and V30+V30ΔIR-s. C) Height relative to mock (%) for *N. benthamiana* inoculated with V30, V22 and V30+V30ΔIR-s. Bars indicate mean ± SD. Significance level: * = $p < 0.05$, *$p<0.005$. * Shows significant difference of V30+V30ΔIR-s to V30 or V22 as indicated above V30 or V22.....69
Figure 34 Nicotiana benthamiana inoculated with V30, V22 (positive controls) and V30+V22 (1:1 inoculum) and V30+V30ΔIR-s (1:1 inoculum) mixed infectious clone constructs at 20 days post inoculation (dpi). Symptoms present are upward leaf roll (ULR) and swollen veins (SV). Severity of symptoms indicated by colour coded triangles as described by the key. 69
Figure 35 Nicotiana benthamiana leaves dissected from plants inoculated with pCambia2300 vector in A. tumefaciens (Mock), V30 (positive control), V22 (positive control), and a mixed infection of V30 + V22 (V30+V22) and V30 + V30ΔIR-s (V30+V30ΔIR-s) infectious clone constructs. Leaves were viewed against the light. Symptoms present are upward leaf roll (ULR) and swollen veins (SV). Severity of symptoms indicated by colour coded triangles as described by the key. 70
Figure 36 C5 protein sequences of African cassava mosaic virus (ACMV, Accession number J02057) and V22ΔC5$^{1\text{-}193.}$ A) Secondary structure prediction with confidence levels indicated using a colour key. Consensus α-helices and β-strands numbered. B) Predicted C5 protein regions of disorder with disordered residues having protein- binding potential highlighted in green. The threshold value for disordered residues was 0.5. Disorder and secondary structural predictions obtained using PSIPRED web server (http://bioinf.cs.ucl.ac.uk/psipred/). 71
Figure 37 Plant height relative to the control (mock) (%) over time (dpi) for *Nicotiana benthamiana* inoculated with V22 and V22ΔC5aa$^{1\text{-}193}$. Plant height relative to control calculated using the formula (H_i/H_m*100) where H_i is the height of inoculated plants, H_m is the height of mock plants. Formula obtained from Lapidot et al., 2006. Bars indicate mean ± SD. Significance level: * = $p < 0.05$ (Student T-test). 72
Figure 38 Upward leaf roll severity (ULR) and swollen vein severity (SV) scores over time (dpi) for *Nicotiana benthamiana* inoculated with V22ΔC5aa$^{1\text{-}193}$. Bars indicate mean ± SD. Significance level: * = $p < 0.05$ **=$P<0.005$ (Student T-test). 73

Figure 39 Nicotiana benthamiana inoculated with V22 (positive control) and V22ΔC5aa$^{1\text{-}193}$ infectious clone constructs at 20- and 32-days post inoculation (dpi). Symptoms present are upward leaf roll (ULR) and swollen veins (SV). Severity of symptoms indicated by colour coded triangles as described by the key. The time points were chosen to show the delayed onset of ULR and SV at 20 dpi induced by the V22 C5 infectious clones and the increased severity of ULR and SV at 32 dpi induced by the V22 C5 mutants when compared to V22. 73
Figure 40 *Nicotiana benthamiana* leaves dissected from plants inoculated with pCambia2300 vector in *Agrobacterium tumefaciens* (Mock), V30 (positive control), V22 (positive control) and V22ΔC5aa$^{1\text{-}193.}$Leaves at 20 days post inoculation (dpi) were viewed against the light. Symptoms present are upward leaf roll (ULR) and swollen veins (SV). Severity of symptoms indicated by colour coded triangles as described by the key. 74

List of Tables

Table 1 Summary of V22 mutant clones, amino acid changes substituted in each clone and primers used for mutagenesis.....38
Table 2 Summary of V22 C5 mutant clones names and respective amino acid changes and the primer sequences used for mutagenesis. 43
Table 3 Symptom scores for the phenotypes, upward leaf roll (ULR) and swollen veins (SV), induced in *N. benthamiana* by Tomato curly stunt virus (ToCSV). 45
Table 4 Primer names and sequences for qPCR primers used. 48
Table 5 Primer names and sequences of primers used for cDNA synthesis for the detection of a V22 IR transcript. 49
Table 6 V22 single mutant amino acid clones and amino acid positions with inadvertent mutations indicated 58

List of abbreviations

%- Percentage

°C- degree Celsius

µg- microgram

µL-microlitre

aa-amino acid

A-alanine

ACMV- African cassava mosaic virus

BLAST- Basic Local Alignment Search Tool

bp-base pair

CLCuKoV- Cotton leaf curl Kokhran virus

CLE- conserved late element

CP-Coat protein

CR- common region

cs- complementary sense

C_t- cycle threshold

CTAB- cetyltrimethylammonium bromide

D-aspartic acid

DEPC- Diethyl pyrocarbonate

DMSO-dimethyl sulfoxide

DNA- deoxyribonucleic acid

dNTP- deoxyribonucleotide triphosphate

dpi- days post inoculation

DSI- disease severity index

EACMV- East African cassava mosaic Cameroon virus

EDTA- ethylenediaminetetraacetic acid

GAPDH- Glyceraldehyde 3-phosphate dehydrogenase

h- hours

HR-hypersensitive response

Hz-Hertz

IR- intergenic region

IUPAC-IUBMB - International Union of Pure and Applied Chemistry and the International Union of Biochemistry and Molecular Biology

Kb- kilobase

K-lysine

LB- Luria-Bertani

L-Leucine

lncRNA- long non-coding RNA

mm- millimetre

mM- millimolar

M-molar

mol-mole

MP- movement protein

mRNA-messenger RNA

MUSCLE- MUltiple Sequence Comparison by Log-expectation

ng- nanogram

NLC- nuclear localisation signal

NSP- nuclear shuttle protein

nt-nucleotide

OD_{600}- optical density at 600nm

ORI-origin of replication

p- probability value

PaLCuV- Papaya leaf curl virus

PCR-polymerase chain reaction

PepVYV - Pepper vein yellows viruses

P-proline

PSIPRED- PSI-blast based secondary structure PREDiction

PTGS- post transcriptional gene silencing

PVX-potato virus X

Q-glutamine

R-arginine

RCR- rolling circle replication

Ren- replication enhancer protein

Rep- replication-associated protein

Rpm- revolutions per minute

RT-PCR- reverse transcription PCR

s- seconds

SA- South Africa

SD- standard deviation

siRNA-silencing RNA

S-serine

SV- Swollen veins

TbLCZV- Tobacco leaf curl Zimbabwe virus

TGS-transcriptional gene silencing

Tm- melting temperature

ToCSV- Tomato curly stunt virus

ToLCMLV – Tomato leaf curl Mali virus

ToRMV- Tomato rugose mosaic virus

ToYSV- Tomato yellow spot virus

TrAp- The transcriptional activator protein

T-threonine

TYLCV- Tomato yellow leaf curl virus

ULR- Upward leaf roll

USA- United States of America

vsRNA- virus derived small RNA

VS-virus sense

V-valine

WT-wildtype

α – alpha

β – Beta

βMe- beta-mercaptoethanol

ABSTRACT

Begomovirus infection causes a significant decrease in yield production in multiple economically important crops. This poses a problem to food security and livelihoods of farmers, and employers in the farming industry. Tomato curly stunt virus (ToCSV) exists as two variants, severe (ToCSV-V30) and mild (ToCSV-V22), which negatively impacts tomato production in South Africa. The non-coding intergenic region (IR) of ToCSV ssDNA is of particular interest because of its involvement in binding host factors to initiate infection and transcription of viral proteins. In addition, the IR region contains the promoter for the adjacent V2 ORF which encodes the pre-coat protein which is a known pathogenicity determinant in begomoviruses. The IR was investigated by agroinoculating *Nicotiana benthamiana*, an experimental host, with wild-type infectious clone constructs of ToCSV-V30 (V30), ToCSV-V22 (V22) and IR mutants where nucleotides of the 3'IR in V22 were replaced with the same nucleotides as found in V30. The three V22 IR mutants constructed were a full 3' IR swap mutant V22ΔIR-s (nt 61-137), and two shorter IR swaps on either side of the conserved TATA box; V22ΔIR-sl (nt 61-94) and V22ΔIR-sr (nt 108-138). The role of the V2 ORF was investigated by introducing single amino acid changes, creating V2 V27S and T58S substitution mutants, named V22ΔV2-V→S and V22ΔV2-T→S respectively. A mixed infection of wildtype (WT) V30 and V22 was also investigated as mixed infections are more likely to occur in the field as opposed to single viral infections. Another mixed infection of WT V30 and V30IR-s, the V30 backbone with the V22 IR, was done to further explore the role of the IR in a mixed infection. The putative C5 ORF in V22 was examined by extending the 5' end of the ORF using site-directed mutagenesis to generate a longer, N-terminally extended protein like that encoded by the African cassava mosaic virus (ACMV) C5 ORF. V22ΔIR-s and V22ΔIR-sl resulted in a knock-out of upward leaf roll observed in wildtype V22. V22ΔV2-V→S resulted in more severe symptoms, a delayed recovery and increased viral load at 36 dpi compared to V22. V22ΔV2-T→S resulted in a delayed onset of induced symptoms compared to V22. The wildtype mixed infection resulted in an intermediary result in symptom scores and viral load compared to the wildtypes. The V30 and V30IR-s mixed infection induced more severe symptoms than both V30 and V22. Reverse transcription and PCR was performed on V22-infected leaf material at 20 dpi which verified the presence of a V22 3' IR transcript. Furthermore, the V22 C5 mutant induced a delayed onset of symptoms compared to V22. This study concluded that a short nucleotide sequence within the 3'region of the IR plays a role in inducing upward leaf roll produced by V22 infection. Furthermore, the V22 V2 ORF was

concluded to play a role in symptom severity and recovery. The exploratory study of the putative V22 C5 ORF and the results observed open avenues to future studies to investigate the significance of putative ORFs and possible evolutionary links to these ORFs.

CHAPTER 1 - LITERATURE REVIEW

1.1 Agricultural importance of tomato crops

Food safety is of great importance, especially due to accelerated global population growth, and hence the increasing demand for food supply. Crop industries play an important role in employment and provide income to 40 % of the global population. Agriculture specifically is essential in poor rural areas with the largest source of income coming from agricultural practices. Seventy percent of Africans depend on agriculture as a means of income, with 80 % of the population in West and Central Africa depending on agriculture for a living (Leke et al., 2015).

Solanum Lycopersicum, tomato, is an important crop as the fruits produced by these plants are prevalent in the daily food intake of many people in South Africa. They belong to the dicotyledonous nightshade family *Solanaceae* and tomato fruits contain vitamins, including vitamin A and vitamin C, and minerals which are essential to the human diet (Foolad, 2007). The cultivated tomato is second to the potato in terms of worldwide consumption and is grown worldwide. Over the period from 2001 – 2011 tomato production in South Africa grew by more than 20 % and South Africa has established itself as a major tomato producer in Sub-Saharan Africa (Malherbe & Marais, 2015). As this plant is a significant contributor to the crop industry in South Africa, the concern of yield losses due to infection of these crops with viruses has increased (Pietersen et al., 2000).

1.2 Virus pathogenicity: a threat to tomato crops

Viruses are a common threat to crop production, with geminiviruses forming the second largest family of plant viruses. Of the nine genera named under the *Geminiviridae* family (Zerbini et al., 2017), the genus *Begomovirus* is particularly significant as begomoviruses infect dicotyledonous plants and have one of the most diverse viral genomes (Lima et al., 2017). Begomoviruses are of agricultural importance as they are known to infect a large range of vegetable crops as well as a few fruit crops. These viruses reduce yields of many food crops including melon, chilli, brinjal, papaya, squash, sweet potato, cassava, potato, pepper and tomato (Zerbini et al., 2017).

1.3 The discovery and origin of *Tomato curly stunt virus*

Tomato curly stunt virus (ToCSV) is a monopartite begomovirus which is of agricultural significance in South Africa. ToCSV was discovered in 1997 in cultivated tomato, in the Onderberg region of South Africa (Pietersen et al., 2008). This virus was identified based on the analysis of its core coat protein sequence. ToCSV shares the highest nucleotide sequence identity of 84% with the begomovirus Tobacco leaf curl Zimbabwe (TbLCZV) from Zimbabwe, followed by Pepper vein yellow virus (PepYVV) and Tomato leaf curl Mali virus (ToLCMLV) (Pietersen et al., 2008).

1.4 The pathogenicity of *Tomato curly stunt virus* in tomato

The infection of tomato with ToCSV results in symptoms similar to those caused by tomato yellow leaf curl virus (TYLCV). These symptoms include foliar chlorosis, leaf curling, stunting and reduced fruit size (Pietersen et al., 2000). The severity of symptoms induced in tomato depends on the infecting variant of ToCSV, which could be either ToCSV-V30 or ToCSV-V22. ToCSV-V30 causes more severe symptoms in tomato as opposed to ToCSV-V22 (Esterhuizen, 2012). Severe leaf yellowing and curling are induced in tomato by ToCSV-V30 (Figure 1). The two variants share 97.3 % nucleotide identity and are differentiated based on the presence or absence of a recombination fragment in the virion-sense gene (V2) which codes for the pre-coat protein which serves as the movement associated protein and silencing suppressor (Pietersen et al., 2008). In a previous V2 swap study, it was established that the differences in the V2 recombinant region of the two variants alone do not determine symptom differences when infecting tomato cv. Rooikhaki (Esterhuizen, 2012), suggesting other ORF mutations may play a role in symptom severity in tomato. Both mild and severe variants are found throughout the tomato-growing regions in South Africa with ToCSV-V30 being more prevalent. Mixed infection with both variants has also been found to occur on a frequent basis (Esterhuizen, 2012).

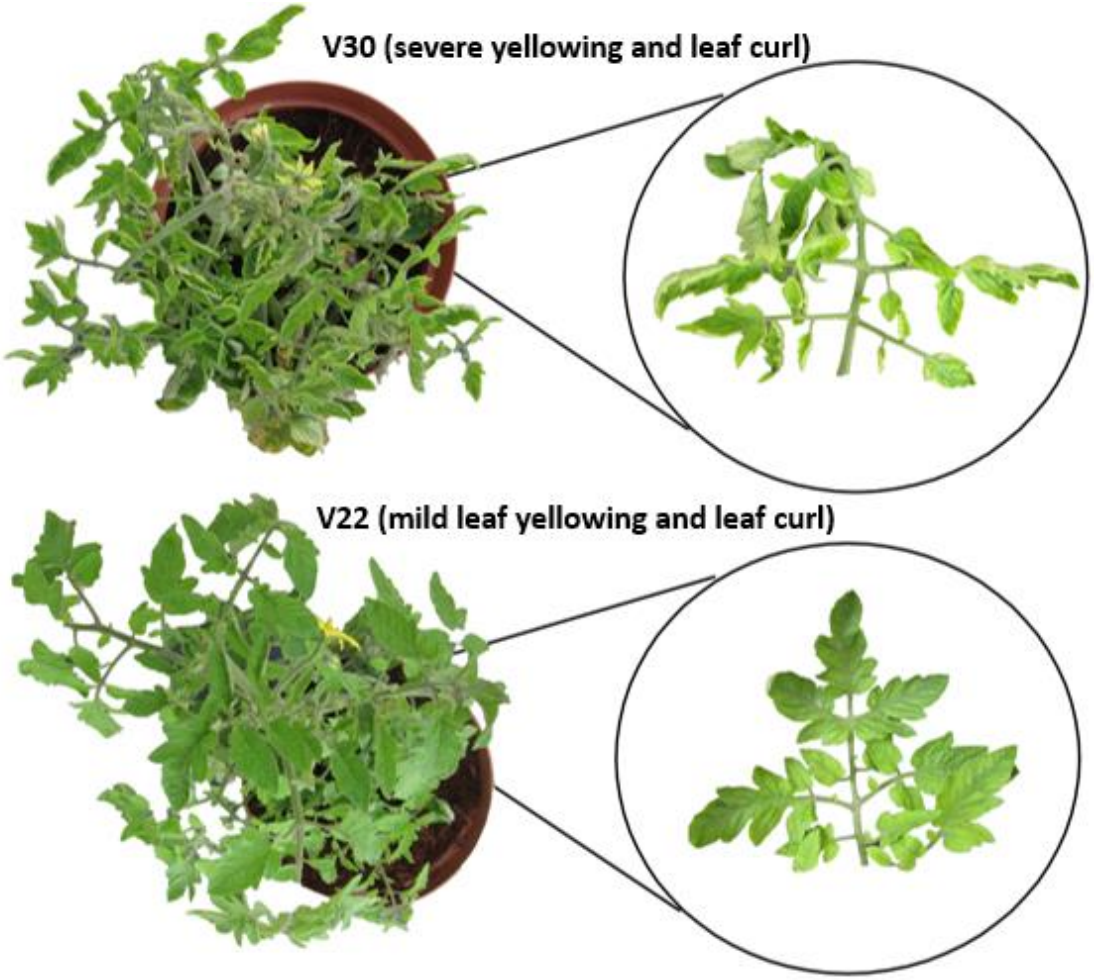


Figure 1 *Solanum lycopersicum* infected with severe and mild variants of Tomato curly stunt virus (ToCSV). Close up of the leaves show yellowing and leaf curling present. Severe symptoms induced by ToCSV-V30, and mild symptoms induced by ToCSV-V22. Images taken at 30 days post inoculation (dpi). Photos taken at the University of the Witwatersrand 2020.

Previous research [by postdoctoral fellow Dr. Ishtiaq Hassan in the Plant Biotechnology Laboratory, University of Witwatersrand 2018] found that in the experimental host *Nicotiana benthamiana,* the opposite symptom phenotypes were observed for ToCSV-V30 and ToCSV-V22 when compared to the symptoms induced in *S. lycopersicum. N. benthamiana* inoculated with ToCSV-V30 showed mild symptoms, whereas ToCSV-V22-inoculated plants resulted in more severe symptoms. Symptoms induced in *N. benthamiana* were further clarified in this study. It was established in this study that ToCSV-V30 induced downward cupping, swollen veins, rugosity and no apparent recovery whereas ToCSV-V22 induced upward leaf roll, rugosity, sunken veins and a recovery was observed in symptoms (Figure 2). *N. benthamiana,* a solanaceous plant originating in Australia, is used as a model experimental plant to study crop viruses because of its potential to be infected by a multitude of organisms including fungi,

bacteria and viruses (Ying et al., 2010). *N. benthamiana* used in laboratory studies lacks a virus inducible RNA-dependent RNA polymerase (*RdRP1*) that is responsible for limiting viral accumulation. This *N. benthamiana* is therefore hyper susceptible to infection by a wide variety of plant viruses (Ying et al., 2010).

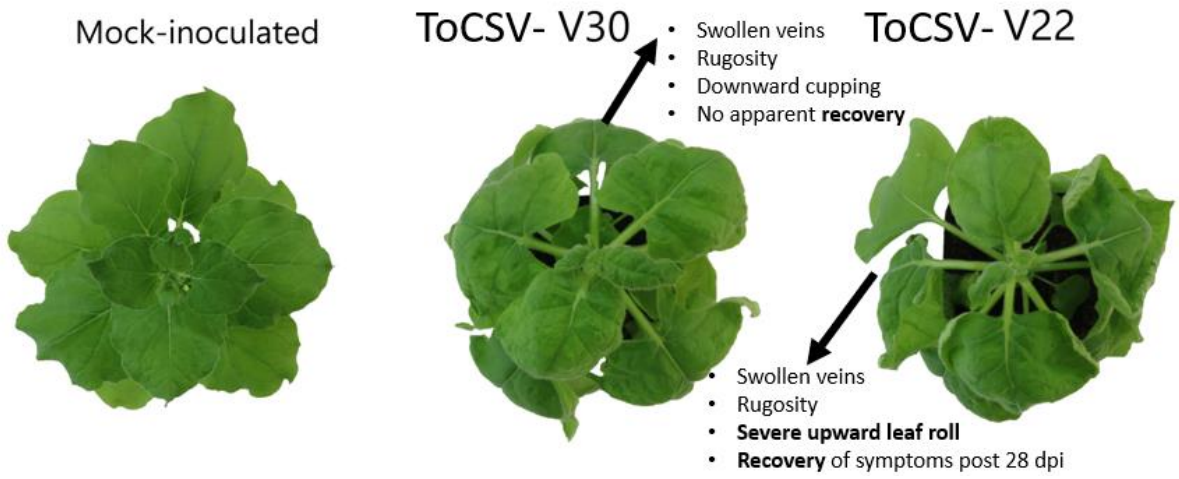


Figure 2 *Nicotiana benthamiana* inoculated with ToCSV-V30 and ToCSV-V22. Symptoms induced as described in the image. *N. benthamiana* was inoculated with mock (pCambia vector in *Agrobacterium tumefaciens*) for the negative control. Images taken at 20 days post inoculation (dpi). Photos taken at the University of the Witwatersrand 2022.

1.5 Begomovirus structure and replication

Begomoviruses can be monopartite, possessing one single stranded DNA (ssDNA) circular genome, or bipartite, possessing two ssDNA circular genomes (Collmer and Howell, 1992). The genetic material, whether monopartite or bipartite, is encased in geminate icosahedral particles (Harrison, 1985). Some begomovirus infection have been found to be associated with satellite DNAs, which are known to play a role in symptom induction (Collmer and Howell, 1992; Kumar et al., 2016). Replication of begomoviruses only occurs within their respective hosts. Symptoms and decreased crop yield induced by begomoviruses in susceptible plants are often dependent on the pathogenicity and replication ability of viruses (Jones et al., 2006). Begomoviruses infect their hosts when their whitefly vector *Bemisia tabaci* Genn. feeds on the sap of the plant (Moriones and Navas-Castillo, 2000). Once the begomovirus is inside the host cells, uncoating of the virion occurs in the cytoplasm. Following this the ssDNA is transported into the nucleus through the virus encoded coat protein (CP). Once present in the nucleus, the

viral ssDNA is converted to double stranded DNA (dsDNA) by the host DNA polymerase. The dsDNA, the transcriptionally active form of the viral DNA, becomes the template for transcription by host RNA polymerase II. Firstly, the genes encoding proteins involved in replication and transcription are transcribed from the antisense strand, followed by the expression of virion sense genes (Snehi et al., 2017). To establish systemic plant infection, the viral DNA moves out of the host nucleus into the surrounding cells or the phloem of the plant. This movement out of the nucleus is facilitated by proteins such as the movement protein (MP) and the nuclear shuttle protein (NSP) (Noueiry et al., 1994).

Begomoviruses replicate inside the host nucleus after the conversion of ssDNA to dsDNA, via host cell machinery, by a combination of the rolling circle replication (RCR) and recombinant-dependent replication (Jeske et al., 2001). Rolling circle replication is initiated by the viral replication-associated protein (Rep). Rep has a DNA-binding domain, a DNA nicking domain as well as an oligomerization domain. It also interacts with host factors to carry out viral replication (Orozco et al., 1997). RCR occurs when one strand of the dsDNA circular genome is nicked at a site, called the double-strand origin, creating a ssDNA circular template for multiple rounds of replication to occur. Rep binds to the 5' end of the nicked DNA. Replication around the ssDNA template displaces the nicked ssDNA strand. From this displaced ssDNA, multiple linear copies are produced which are later converted to dsDNA circular genomes, a process which is facilitated by host-encoded DNA ligase (Jeske et al., 2001). Recombination-dependent replication occurs through homologous DNA sequence recombination resulting in the formation of single-stranded as well as double-stranded DNA copies (Jeske et al., 2001).

1.6 Structure of the genomes of *Tomato curly stunt virus*

It has been found that monopartite DNA A genomes, like TYLCV, are responsible for infectivity and disease development in tomato plants (Navot et al., 1991), despite lacking DNA B that is characteristic of bipartite begomoviruses. While no satellite molecules have yet been identified in association with ToCSV infection in tomato to date (Esterhuizen, 2012), many monopartite begomoviruses require a satellite molecule for complete pathogenicity (Nawaz-ul-Rehman and Fauquet, 2009).

The circular ss-DNA ToCSV genome is 2 766 nucleotides (nt) in length. (Pietersen et al., 2008). DNA-A contains six open reading frames (ORFs), two virion-sense genes (V1 & V2)

and four complementary sense genes (cs) (C1, C2, C3 and C4) (Figure 3). V1 encodes the CP, V2 encodes the movement-associated protein / virus suppressor of gene silencing, C1 encodes for Rep and C2 encodes a protein that increases the rate of gene expression of virion-sense genes (Murayama et al., 1991; Laufs et al., 1995; Rojas et al., 2001; Saunders and Stanley, 1995). C3 encodes a replication enhancer protein (Ren) and C4 was found to determine symptom severity (Murayama et al., 1991; Laufs et al., 1995; Rojas et al., 2001; Saunders and Stanley,1995). The C4 gene plays a role in movement as well as supressing defence responses of the host thereby making C4 an important pathogenicity determinant (Pooma and Petty, 1996). The intergenic region (IR) lies between C1 and V2 and is a conserved begomovirus sequence containing an origin of replication (ORI). The IR is 259 nucleotides in length and contains a hairpin structure of a short nucleotide motif TAATATTAC, referred to as a nonanucleotide sequence, encompassing the ORI. The ORI is present in all geminiviruses and is essential to viral replication (Ikegami et al., 1988). The functions described for the abovementioned genes and regions of ToCSV is inferred from the functions established in multiple begomoviruses other than ToCSV.

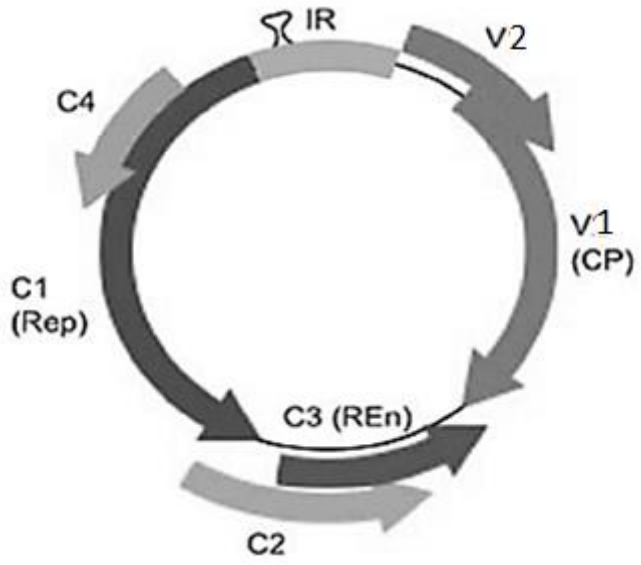


Figure 3 Diagram showing the genome organization of Tomato curly stunt virus (ToCSV) DNA-A with a length of 2766 nucleotides, containing 6 open reading frames (ORFs), V1, V2, C1, C2, C3 and C4. The Intergenic region contains the origin of replication (ORI) as well as the short nucleotide motif hairpin (Esterhuizen, 2012).

1.7 The significance of the intergenic region (IR) in ToCSV

In all geminiviruses, the IR facilitates DNA viral replication. It consists of a conserved stem loop motif that contains the replication initiation site on the 3' end of the loop which is required for DNA replication. The IR also contains bidirectional transcription promoters that are essential for transcription of ORFs on the viral and complementary strands (Paula et al., 1995). The ToCSV IR contains two direct repeat sequences of GGTAC within the nucleotide regions of 2624-2627 and 2663-2668, with these predicted to be iteron sites for the binding of Rep (Pietersen et al., 2008). The 3' IR varies greatly between ToCSV-V30 and ToCSV-V22 with multiple nucleotide differences and deletions. This can be seen in the alignment of these two variants (Figure 4). Therefore, investigating the IR region in the ToCSV may prove to be significant in determining ToCSV infectivity.

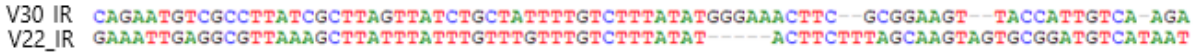


Figure 4 Alignment of the 3' end of the intergenic region (IR) of Tomato curly stunt virus (ToCSV)-V30 and ToCSV-V22 (nucleotide position 61 to 139). Alignment done on BioEdit Sequence Alignment editor.

Further research highlighting the importance of the intergenic region can be seen in a study carried out using TYLCV. TYLCV was found to modulate symptom development through IR-derived viral small-interfering RNAs (vsiRNAs) that target a long noncoding (lnc) RNA, called *SILNR1*, present in tomato. In resistant tomato cultivars, a 14 nt deletion in *SILNR1* was found, suggesting that the *SILNR1* region in conjunction with vsiRNAs from the IR is required for symptom development. *SILNR1* homologues were not found in other screened plants to date (Yang et al., 2019).

1.8 The significance of the V2 gene in ToCSV

The V2 gene, which is 431 nucleotides (nt) in length, is located on the viral strand and encodes a pre-coat protein. (Navot et al., 1991). In TYLCV, a monopartite begomovirus, it has been shown that the V2 protein, more specifically amino acid (aa) 71 of the V2 protein, is responsible for self-interaction, viral accumulation, pathogenicity and infectivity of TYLCV. This was shown in the model experimental plant *N. benthamiana* and its natural host plant, *S. lycopersicum* (Zhao et al., 2018). Furthermore, in TYLCV, V2 was found to interact with the

host protein suppressor of gene silencing 3 (SGS3), a component involved in the host RNA-silencing defence mechanism against viruses. This interaction suppresses RNA-silencing as binding of V2 to SGS3 impairs functionality of the host protein during RNA silencing (Glick et al., 2007). Additionally, TYLCV V2 was shown to suppress transcriptional gene silencing (TGS) of a green fluorescence protein (GFP) transgene in the *N. benthamiana* 16-TGS line. Suppression of the GFP was accompanied by a reduction of methylation in its promoter from which it can be concluded that V2 of TYLCV can supress methylation-mediated transcriptional gene silencing (Wang et al., 2014).

Previous research on many monopartite begomoviruses show the V2 gene to be vital for infectivity and pathogenicity, and therefore the ToCSV V2 protein may play a significant role in the pathogenicity of this plant virus. In ToCSV, the severe and mild ToCSV variants are differentiated based on the recombinant fragment present in the V2 region (Esterhuizen, 2012). As mentioned previously, a previous investigation with a V2 swap between the two ToCSV variants determined that the phenotypic differences between the ToCSV variants, upon infection of tomato cv. Rooikhaki, could not be solely attributed to V2, but further studies in other tomato cultivars, or other experimental hosts, were not performed (Esterhuizen, 2012). Furthermore, viral load was not determined in the tomato cv. Rooikhaki experiments. With the use of model experimental plants *N. benthamiana* and *N. tabacum*, studies have shown that the V2 protein in monopartite begomoviruses, Papaya leaf curl virus (PaLCuV) and Cotton leaf curl Kokhran virus (CLCuKoV), are significant in pathogenicity (Mubin et al., 2010). Additional investigation into the role of ToCSV V2 swaps between the mild and severe variants in the model experimental plant, *N. benthamiana*, may reveal significant differences in pathogenicity such as symptom severity and viral load.

1.9 Mixed infections

Mixed viral infections, although not as frequently studied as single infections, appear to be common both in the field and in wild plants or weeds (natural reservoir for viruses). Viruses in mixed infections, depending on the viruses and strains involved, play a vital role in the symptoms induced in plant diseases and it is therefore vital to look at multiple viruses when studying a plant disease (Moreno and López-Moya, 2020). Mixed infections can be attributed to multiple factors, the first being viruses which frequently mutate resulting in new variants with different host-infecting capabilities. Furthermore, the vast host range of these viruses and

virus vectors which are polyphagous and can carry a range of viruses allow for mixed infections to take place (Singhal et al., 2020). Three interactions can occur between viruses within a mixed infection, these being neutralism, synergism and antagonism. Neutralism is evident when symptoms remain unchanged compared to single infections, whereas synergism and antagonism are observed when a phenotypic or viral load difference between the single and mixed infections is present (Mascai and Gallitelli, 2016).

Synergism amongst viruses can occur via different mechanisms of action, one of these being the suppression of host defense mechanisms by one or both viruses allowing for elevated disease symptoms. Another feature of synergism is when one virus acquires the ability to invade different host tissues through viral suppressor proteins from one virus influencing movement patterns of other viruses (García-Marcos et al., 2009). Contrary to synergism, antagonism can occur when a subsequent infection is prevented due to the host silencing RNAs (siRNA), generated by the plant RNA silencing machinery, deployed against one virus consequently targets other viruses thereby preventing infection or symptom development by competing viruses (Ratcliff et al., 1997; Chávez-Calvillo et al. 2016). Therefore, in an antagonistic mixed infection, the presence of one virus is detrimental to the other virus. This would allow for the concept of cross-protection to be exploited in order to control the populations of plant viruses. Cross-protection occurs when tolerance of a plant to one virus is induced by the infection of that plant by another virus. This phenomenon can therefore be used in the field as a form of biocontrol. Antagonism resulting in induced resistance was shown in *N. benthamiana* after being infected with two geminiviruses; African cassava mosaic virus (ACMV) and East African cassava mosaic virus (EACMV) where ACMV DNA-A induced resistance to ACMV and EACMV (Reddy et al., 2012). It has previously been shown that closely related viruses or variants of a virus are more likely to result in cross-protection (Nakazono-Nagaoka et al., 2009 and Wen et al., 1991). The different interactions, those of coinfection or superinfection, between plant viruses is what establishes the outcome of a mixed infection which may also influence its host. Coinfection occurs when multiple viruses simultaneously infect a single host whereas in superinfection, different viruses infect the host at different times (Joan et al., 2003). The order of infection of viruses within a plant plays an important role in the interactions between the viruses in a mixed infection and also influences the phenomenon of cross-protection.

1.9.1 Mixed infections of Geminiviruses

Frequently mixed infections in multiple plant hosts have been identified as being geminiviruses. An important crop disease induced by a mixed viral infection of geminiviruses is cassava mosaic disease (CMD). Coinfection of multiple viruses or strains is a common occurrence in CMD often leading to increased severity in symptoms induced in cassava (Rey and Vanderschuren, 2017). An example of a mixed infection leading to CMD is the infection of two viruses belonging to the genus *Begomovirus*, these being African cassava mosaic virus (ACMV) and East African cassava mosaic virus (EACMV). These two viruses engage in a synergistic relationship with each other thereby inducing more severe symptoms in cassava (Harrison et al., 1997, Zinga et al., 2013).

Mixed infections can also promote recombination, reassortment and complementation between different viruses resulting in new variants of a virus being generated. Therefore, mixed infections play a pivotal role in the evolution of viruses and hence how their plant hosts are affected (Singhal et al., 2020). An example of this can be seen in the outbreak of a serious case of cassava mosaic disease (CMD) caused by a recombinant virulent virus strain where East African cassava mosaic virus (EACMV) gained the capsid protein sequence of African cassava mosaic virus (ACMV). This virulent strain, referred to as East African cassava mosaic virus Uganda-2 (EACMV-UG2), resulted in a devastating disease outbreak in the cassava crop in Uganda. This variant was also found to be associated with ACMV in a mixed infection where the CMD symptoms induced were significantly heightened (Pita et al., 2001).

1.10 The geminiviral C5 ORF and its role in pathogenicity

The C5 open reading frame (ORF) has only recently been identified in begomoviruses. This ORF is located downstream of the C3 ORF and overlaps with the 3’ end of the V1 ORF. The few studies that have characterised C5 in monopartite and bipartite begomoviruses have found this ORF to play an important role in determining the pathogenicity of the virus (Li et al., 2015, Li et al., 2021 and Vaghi Medina et al., 2018). The AC5 protein of mungbean yellow mosaic India virus (MYMIV) is 183 residues in length and was confirmed to play a role in RNA silencing suppression activity. The MYMIV AC5 was also found to be vital for infectivity of the virus (Li et al., 2015). The C5 protein in Ageratum leaf curl Sichuan virus (ALCScV) was shown to be transcribed and act as a virulence factor which induced symptoms in *N.*

benthamiana when expressed in the PVX vector. The ALCScV C5 was also found to supress post transcriptional gene silencing (PTGS) and thereby interfere with plants recovery as PTGS is involved in the plant antiviral defence system (Li et al., 2021 and Sijen & Kooter, 2000).

As knowledge of the C5 protein in begomoviruses grows, the mechanisms of pathogenicity of these viruses and the interaction with their respective hosts is made clearer. Therefore, exploring this novel ORF in begomoviruses for which it has not been identified or characterised yet proves to be promising. Recently, the presence of this ORF was identified in ToCSV using sequence comparison with published sequences of other begomovirus C5 ORFs (Alex Zwolinski, 2022). Furthermore, putative C6 and C7 were also identified in this manner but further studies to validate these still need to be done (Zwolinski, 2022). Investigating the putative uncharacterised C5 ORF in a begomovirus such as ToCSV may contribute to the characterisation of this virus as a putative C5 ORF was recently identified in in ToCSV.

1.11. Rationale for study

Begomovirus infection of crops is known to cause a significant decrease in yield production in a multitude of economically important crops worldwide (Saeed and Samad, 2017) thereby posing a threat to food security as well as to the livelihoods of farmers and people employed by the farming-associated industries. The importance of studying these viruses is further highlighted, as resistant crop variety development relies on the knowledge gained from this research such as this. ToCSV is a virus which has negatively impacted the yields of tomato, an important crop, in southern Africa (Pietersen et al., 2000). By comparing the pathogenicity (symptoms) and infectivity (viral load) between the mild and severe ToCSV variants as well as determining the role of the IR region and the V2 gene in ToCSV infectivity by quantitative and qualitative means, the pathogenicity determinants of ToCSV can be further elucidated. Studying a mixed infection of the two ToCSV variants proves to be more applicable for field application purposes. Furthermore, by looking into putative, uncharacterised ORFs in ToCSV, the pathogenicity of ToCSV can be further illuminated. This study will also aim to contribute towards understanding begomovirus-plant host interactions by determining the relationship between symptom severity and viral load within a mixed infection of the ToCSV variants.

1.12. Aims and specific objectives

1.12.1 Aims

1. To investigate the role of the IR and V2 ORF in infectivity (viral load) and pathogenicity (symptoms) of ToCSV-V22 in *N. benthamiana*
2. To determine the pathogenicity of both ToCSV-V30 and ToCSV-V22 variants in a mixed infection
3. To investigate the novel C5 ORF in ToCSV-V22

1.12.2 Specific objectives

- To design IR swap mutants between V30 and V22
- To introduce targeted single amino acid changes via site directed mutagenesis of the cognate V2 nucleotide region in the V22 infectious clone
- To establish pathogenicity of a mixed infection of wildtype variants, V30+V22 and a mixed infection of V30+V30ΔIR-s in *N. benthamiana* through the co-inoculation of the V30 and V22 infectious clone constructs
- To introduce potential expression of a putative C5 ORF in V22 through site-directed mutagenesis
- To determine the effects of the mutant infectious clone constructs and mixed infection, in comparison to the wild type variants and mock inoculated treatments, on symptom phenotype, height and disease severity at 4-day intervals post inoculation of *N. benthamiana*
- To determine the viral load of the IR and V2 mutant clones, and of the mixed infection using degenerate primers and real time PCR
- To determine if the IR is transcribed via cDNA synthesis and PCR

CHAPTER 2 - METHODOLOGY

2.1 Infectious clone constructs

The severe V30 and mild V22 ToCSV genetic variants are named ToCSV- [ZA: Mks30:08] and ToCSV- [ZA: Mks22:07], respectively. The infectious geminivirus clone constructs of the severe and mild variants of ToCSV used in this study were constructed using the pCAMBIA2300 vector (Esterhuizen, 2012) as this vector is suitable for infection of plants with ToCSV using *Agrobacterium tumefaciens*. The pCambia2300 vector has the kanamycin-resistant selection marker gene and is an Agrobacterium binary vector specifically used for plant transformation. This vector also contains Multiple Cloning Sites (MCS), has a high copy number in *E. coli* allowing for high DNA yields, and is stable in *Agrobacterium* therefore this plasmid backbone serves as a good vector for ToCSV in this study. The SnapGene plasmid map for this vector can be seen in Figure 5 below (from Insightful Science; available at snapgene.com).

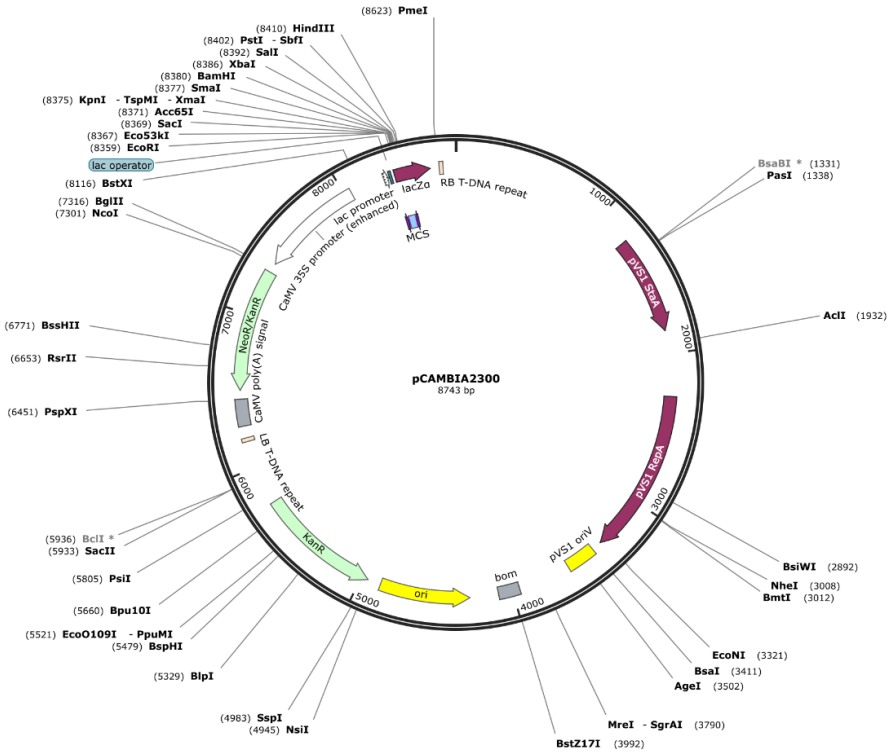


Figure 5 A detailed vector map of pCambia2300 with sequence length of 8743 base pairs (bp), selection genes, restriction sites, Origin of Replication (ORI), promoters and the T-DNA repeats labelled (source: Insightful Science; available at Snapgene.com).

Three V22 IR-swap mutants between ToCSV-V30 (V30) and ToCSV-V22 (V22) were constructed where nucleotides of the 3' region of the IR (downstream of the origin of replication loop) from V22 were replaced with that of V30 (Figure 6). The first mutant clone, V22ΔIR-s mutant, was constructed by the removal of nt 61 – 137 of the 3' IR region of V22 and replaced with the corresponding sequence from the V30 infectious clone. Due to the V30 IR having an extra nucleotide downstream of the TATA box compared to the V22 IR, nt 61-137 of V22 was swapped out for nt 61-138 of V30. This region was selected as the 3' end of the IR is least conserved between the two variants. Two shorter V22 3' IR-swaps were also constructed. V22ΔIR-sl was constructed by the removal of nt 61 – 94, a region upstream of the TATA box within the 3' IR of V22 and replaced with the corresponding region from the V30 sequence. V22ΔIR-sr was constructed by the removal of nt 108 – 137, a region downstream of the TATA box within the 3' IR region, of V22 and replaced with the corresponding region from the V30 sequence. The numbering of the nucleotides is relative to the origin or replication (ORI) (Figure 6).

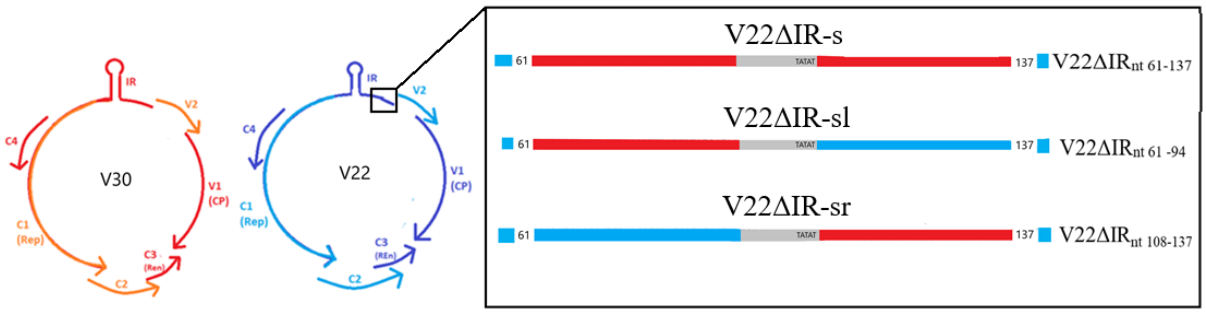


Figure 6 Diagram showing the V30 and V22 genomes with the six open reading frames (ORFs), V1, V2, C1, C2, C3 and C4 indicated. In the enlarged region of the Intergenic region (IR)- swap infectious clone constructs are shown, where selected V22 IR sequences were swapped out with that of the respective V30 sequence. A conserved TATA box in the IR is shown in grey. Red lines refer to the V30 sequence, with the blue line representing V22 sequence. Nucleotide (nt) positions are relative to the Origin of Replication (ORI) which is nt 1.

These three IR-swap constructs were synthesised by GenScript Biotech (Piscataway, NJ, US), certificate of analysis for the IR-swap clones are in supplementary page 113-115 with a virtual digest and a vector map of the V22 infectious clone construct on page 116. Once received, the constructs were transformed into competent *E. coli* DH5α cells (New England Biolabs protocol, Massachusetts, USA) to increase DNA availability of the infectious clone plasmids. To do this, aliquots of 100 µL of *E. coli* were thawed on ice and 100 ng of plasmid DNA was added. Cell-plasmid mixtures were gently tapped and exposed to 42 °C for 90 s using a water

bath. Samples were placed on ice for 5 min, followed by the addition of 900 µL of LB media to each respective sample and these were subjected to shaking at 200 rpm for 1 hour at 37 °C. After this incubation step, the cells were centrifuged, and the pellet was resuspended in 400 uL of LB media. These resuspensions were spread on kanamycin containing LB selection plates. The plates were incubated at 37 °C for 12 hours. Glycerol stocks were made using 50 % sterilized glycerol which were stored at – 80 °C.

These IR-swap infectious clone constructs were transformed, using the freeze-thaw method (Weigel and Glazebrook, 2006), into competent *Agrobacterium tumefaciens* C58C1 cells. This was done by transforming competent Agrobacterium cells on ice followed by gently mixing 100 ng of plasmid DNA into the cells. These cells were placed on ice for 10 min followed by submerging the cells in liquid nitrogen for 5 min. Cells were thawed at 37 °C using a preheated heating block and placed on ice for 2 min. A quantity of 500 µL of LB broth was added to the cells and incubated at 28 °C for 2 hours at 200 rpm. After incubation, cells were centrifuged at 14 000 xg for 1 min and pellets were resuspended in 100 µL LB. These resuspended cells were spread on LB plates containing 0.1 mg/mL kanamycin and 0.1 mg/mL rifampicin. Plates were incubated at 28 °C for 2 days.

Plasmid extraction was completed using the GeneJet plasmid miniprep kit and protocol (Thermo Scientific™, Massachusetts,US). Restriction digest reactions of the extracted plasmid was carried out with FastDigest BglI, (Thermo Scientific™) following the manufacturers protocol to verify that the correct constructs were obtained. All infectious clone constructs were digested with BglI and then subjected to Sanger sequencing (Inqaba Biotechnology Industries (Pty) Ltd, Pretoria, South Africa) for further verification.

A fourth infectious clone construct, a V30 backbone construct, named V30ΔIR-s, was previously made (Zwolinski, 2022) by removing nt 61 – 137 of the 3’ IR region of V30 and replacing it with the corresponding sequence from the V22 infectious clone (Figure 7). This V30ΔIR-s infectious clone construct was used in conjunction with WT V30 to inoculate *N. benthamiana* to establish a mixed infection of V30+V30IR-s in this study.

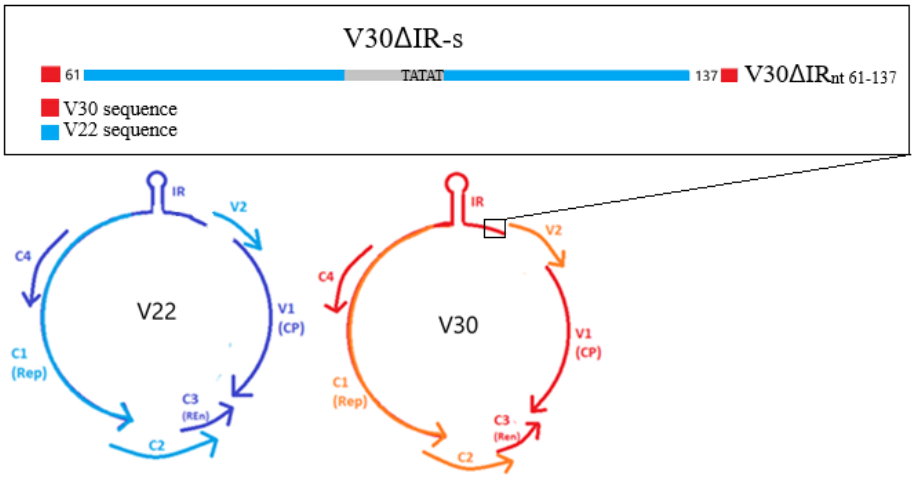


Figure 7 Diagram of the V30 backbone *intergenic region (*IR*)*-swap infectious clone construct where the selected V30 IR sequence was swapped out with that of the respective V22 sequence. A conserved TATA box shown in grey. Nucleotide (nt) positions are relative to the Origin of Replication (ORI) which is nt 1.V22 sequence is shown in blue, with V30 represented in red. The six open reading frames (ORFs) in V30 and V22, V1, V2, C1, C2, C3 and C4 are indicated.

2.2 Investigative bioinformatics for predicting tomato and *N. benthamiana* targeting siRNA from the IR

The siRNA predictor tool (www.plantgrn.noble.org) was used to predict siRNAs from the 3' IR in V30 and V22 whereby the cDNA transcript libraries of both *S. lycopersicum* and *N. benthamiana* were selected. The siRNAs that were generated were manually aligned with the IR region for both V30 and V22 (Supplementary Figure 1.1-1.4). The siRNAs were then subjected to BLAST against tomato and *N. benthamiana* genomes via both NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and SOL Genomics databases (https://solgenomics.net/).

2.3 Site directed mutagenesis

2.3.1 V22V2 single amino acid mutants

A CLUSTAL multiple sequence alignment of the V2 protein aa sequences of V30 and V22 was performed using MUSCLE (3.8) (https://www.ebi.ac.uk/Tools/msa/muscle/) to investigate the amino acid similarity and conservation between the two V2 aa sequences. Following this nucleotide sequences encoding the non-conserved amino acids between V30 and V22 V2 proteins were selected for site directed mutagenesis. The amino acids of interest

in this study are valine (V), serine (S), leucine (L), threonine (T), arginine (R), glutamine (Q), lysing (K) and proline (P). Single amino acids of the V22 V2 protein were mutated to encode for the same amino acids found in the V30 V2 protein at that same position (Table 1). The single aa V22 V2 mutants are in Figure 8 with an inadvertent mutation to V1 in the V22ΔV2-T→S+ mutant clone shown. This inadvertent mutation was introduced as a result of the primer sequences used to introduce the codon change during mutagenesis (Table 1). Mutagenesis methodology to follow. Additionally, one construct encoding for the V2 protein had three amino acids (aa 114-116) deleted. This C-terminal QKP deletion mutant (V22ΔV2Del+) was designed with the use of Benchling (https://benchling.com) and synthesised by GenScript Biotech (Piscataway, NJ, US). *E. coli* DH5α and competent *A. tumefaciens* C58C1 were then transformed with this infectious clone using the above-mentioned *Agrobacterium* and *E. coli* transformation protocols with identical parameters.

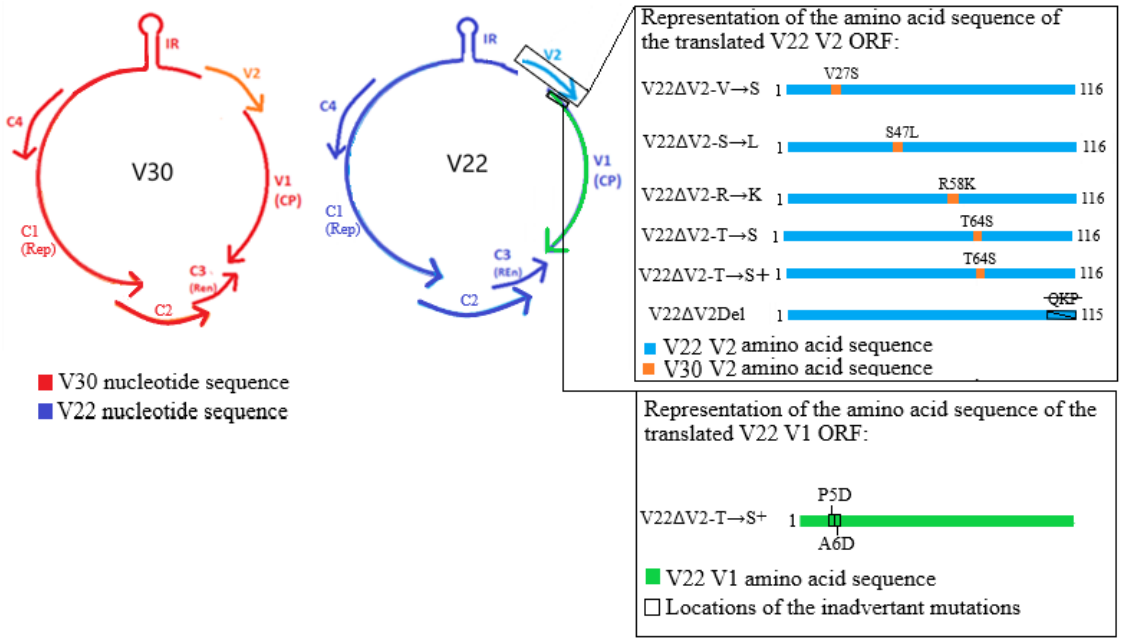


Figure 8 Diagram of infectious clone V30 and V22 genomes with the six open reading frames (ORFs), V1, V2, C1, C2, C3 and C4 indicated. Variants with the V22 V2 amino acid (aa) substitutions in the V22ΔV2 mutant clones indicated. V30 nucleotide sequence is shown in red and V22 nucleotide sequence in blue. Valine (V), serine (S), arginine (R), leucine (L), threonine (T), lysine (K), glutamine (Q), aspartic acid (D), proline (P) are aa indicated by their respective symbols. The top boxed region in the diagram is a representation of the aa substitutions in the V22 V2 sequence. The aa position at the site of substitutions is indicated above each respective representative V2 mutant protein. The bottom boxed region is a representation of the inadvertent amino acid mutations to the overlapping V1 ORF in the V22ΔV2-T→S+ mutant clone construct. ~~QKP~~ represents the deletion of these aa. Light blue represents the V22 aa sequence and light orange represents the V30 aa sequence. The aa position at the site of substitution is indicated above each respective representative V22 V2 mutant protein.

The single V22 mutant clones were constructed through site-directed mutagenesis with the use of partial overlapping primers (Table 1). Primers for V22ΔV2-T→S and V22ΔV2-T→S+ are different sets of primers as even though they introduced the same V22 V2 T→S substitution mutant at aa 64 as the V22ΔV2-T→S+ primers introduced an inadvertent change to the overlapping V1 ORF altering the overlapping V1 aa whereas the in V22ΔV2-T→S introduced no inadvertent changes to any overlapping ORFs. The primers were manually designed following the guidelines by Laible and Boonrod, 2009. Primers were synthesised and RP-cartridge purified by Inqaba Biotech (Pretoria, South Africa). RP-cartridge purification yields products which are free of organic molecules and undesired short oligonucleotides and this purity is important for applications such as mutagenesis (Semenyuk et al., 2006). Integrity of the primers, in terms of hairpin formation, homo- and heterodimer ΔG values, were evaluated using the OligoAnalyzer™ tool (https://www.idtdna.com/oligoanalyzer). Site directed mutagenesis was then carried out using a high-fidelity DNA polymerase, Phusion™ (Thermo Scientific™, Massachusetts, US).

Table 1 Summary of V22 mutant clones, amino acid changes substituted in each clone and primers used for mutagenesis.

Mutant clone name	Tm for Thermocycling conditions for mutagenesis (°C)	Forward primer sequences (Fp) and reverse primer sequences (Rp) for site directed mutagenesis (5'-3')	
V22ΔV2-V→S	65	Fp	CTTGCAatcgATTGAGTCCACTTATGAGCC
		Rp	GGACTCAATcgatTGCAAGTATTTTATAGCAAGC
V22ΔV2-S→L+	72	Fp	CGAGATCTTATactaGTTATTAGAGCCCGGGATTATGTC
		Rp	CGGGCTCTAATAACtagtATAAGATCTCGAATTAAATCGTGG
V22ΔV2-T→S+	69	Fp	GAAGCGAgtCGACGATATAATCATTTCCACGC
		Rp	GATTATATCGTCGacTCGCTTCGACATAATCCC
V22ΔV2-T→S	69	Fp	GAAGCGtCCCGCCGATATAATCATTTCCC
		Rp	GATTATATCGGCGGGaCGCTTCGACATAATC
V22ΔV2-R→K+	72	Fp	GTTATTAGAGCCaaGGATTATGTCGAAGCGACCCGCCG
		Rp	CGCTTCGACATAATCCttGGCTCTAATAACAGAAATAAGATCTCG
V22ΔV2Del+	NA	NA	

* The primer nucleotides that are not capitalized are nucleotides targeted for mutagenesis. Tm represents the melting temperature of the primer pair. The + after the mutant names indicates the presence of inadvertent mutations in established or putative ORFs. ~~QKP~~ represents deletion of QKP nucleotides in that mutant infectious clone. The IUPAC-IUBMB one-abbreviations for the 20 standard amino acids were used.

For construction of V22ΔV2-S→L, a two-step mutagenesis polymerase chain reaction (PCR) was carried out, and for the remaining four single aa V2 mutants (V22ΔV2-V→S, V22ΔV2-T→S, V22ΔV2-T→S+ and V22ΔR→K+), a three-step mutagenesis PCR was performed (Laible and Boonrod, 2009). The thermocycling conditions for the two-step mutagenesis PCR consisted of an initial denaturation step of 98 °C for 30 s, followed by 20 cycles of a 10 s 98 °C denaturation step, 72°C for 8.67 min and finally the 5 min 72 °C extension step (Thermo Scientific™, COL32688 0918). The thermocycling conditions for the three -step mutagenesis PCR was carried out as the above-mentioned conditions with the addition of an 8.69 min melting temperature (Tm) -specific annealing step before the extension step (Thermo Scientific™, COL32688 0918). Tm temperatures are shown in Table 1. The mutagenesis PCR

reaction mix consisted of 4 ng of the DNA template, the infectious clone construct of V22, 10 µL of 5X Phusion buffer (Thermo Scientific™, Massachusetts, US), 0.5 µL Phusion™ High-Fidelity Polymerase (Thermo Scientific™, Massachusetts, US), 1 µL of each respective primer, 1 µL of 25 mM deoxyribose nucleotide triphosphate (dNTPs) and nuclease free water to make up the final volume to 50 µL as per the Phusion™ High-Fidelity DNA Polymerase protocol (Thermo Scientific™, COL32688 0918). One percent Dimethyl sulfoxide (DMSO) was included in the PCR mix as DMSO relieves secondary structures allowing for improved PCR. This is an important addition to the PCR reaction mix as templates with a high GC content have more stable secondary structures (Hardjasa, et al., 2010). The target products, ~ 12000 bp in length, were run on a 0.7 % agarose gel to verify their presence and to confirm the success of the mutagenesis PCR. This was followed by DpnI digestion of the PCR products using 1 µL DpnI (Thermo Scientific™, Massachusetts,US) for 15 h at 37 °C followed by a heat deactivation step of 80 °C for 20 min. DpnI is an enzyme which requires methylated GATC recognition sites in order to cleave DNA. The purpose of digesting the mutagenesis PCR product with DpnI was to eliminate the template plasmid as this plasmid, extracted from *E. coli*, was naturally methylated whereas the newly mutated/synthesised plasmids were not methylated due to the lack of methyl transferase in the PCR mixture (Laible and Boonrod, 2009). Therefore, digesting with DpnI removed the undesired, original template DNA allowing for only the mutated, unmethylated plasmids to be transformed into the *E. coli* in the following steps.

Before transformation of these mutagenesis products into *E. coli*, Ultracompetent *E. coli* XL10-Gold cells were prepared (Sambrook and Green 2020). *E. coli* C2987 cells were streaked on LB agar for 16 h and colonies were inoculated into 10 mL LB broth and incubated at 37 °C for 16 h shaking at 120 rpm. Fifty mL of LB broth was then inoculated with 500 µL of the 10 mL cultures. This broth was incubated at 37 °C until the OD_{600} of the LB broth reached 0.5. Cultures were placed on ice for 10 min, gently swirled and transferred into 50 mL falcon tubes for centrifugation at 2000 rpm for 10 min at 4 °C. The pellet was rinsed with chilled, sterilised water and resuspended in 6 mL of $CaCl_2$ buffer (15 % glycerol and 0.08 M $CaCl_2$) and kept on ice for 30 min. The cells were centrifuged at 5000 rpm for 10 min at 4 °C, the supernatant was discarded, and the pellet was resuspended in 2 mL of ice cold $CaCl_2$ buffer. Aliquots of 100 µL of cells were dispensed, frozen using liquid nitrogen and stored at – 80 °C for further use. The purpose of the $CaCl_2$ buffer serves to encourage the bacterial cells to take up DNA from their surroundings hence improving transformation efficiency (Chang et al., 2017).

One hundred µL of Ultracompetent *E. coli* XL10-Gold cells, prepared as per Sambrook and Green (2020) were used for transformation. The Ultracompetent *E. coli* XL10-Gold cells were initially thawed on ice and 2 µL of the DpnI-digested PCR products with the addition of β-mercaptoethanol (β-ME) to final concentration of 24 mM were added to 50 µL aliquots of the thawed competent cells. β-ME was added as it has been shown to increase transformation efficiency (Janjua, Younis, Deeba and Naqvi, 2014). The samples were gently swirled and placed on ice for 30 min. These samples were then transferred to a 42 °C preheated water bath for 90 s and placed on ice for 2 min. Super optimum broth (SOB) media was warmed to 37 °C and 800 µL of this media was added to each tube and incubated at 37 °C at 120 rpm for 45 min. One hundred µL of cells were spread onto SOB plates containing 100 µg/mL kanamycin and incubated for a period of 24 h at 37 °C. Five colonies were selected from each plate, sub-cultured and plasmid DNA was extracted, using the GeneJET Plasmid Miniprep Kit and protocol (Thermo Scientific™, Massachusetts,US). DNA products were quantified on the NanoDropTM One Spectrophotometer (Thermo Scientific™, Massachusetts, US) and mutant clones were confirmed with Sanger Sequencing (Inqaba Biotechnology Industries (Pty) Ltd, Pretoria, South Africa).

2.3.2 Extension of the 5' end of the putative C5 ORF via site directed mutagenesis

The presence of putative C5 ORF has previously not been reported in ToCSV WT variants, however, a putative C5 ORF was recently identified in both V30 and V22 using ORF Finder (https://www.ncbi.nlm.nih.gov/orffinder/) (Zwolinksi, unpublished). The identity of this ORF was determined using sequence comparison with the published sequences of other begomovirus C5 ORFs (Zwolinksi, unpublished). The putative C5 ORF of both V30 and V22 was predicted to encode a protein 60 aa in length (https://web.expasy.org/translate/).

Alignment of the C5 coding region of African cassava mosaic virus (ACMV, Accession number J02057) with the corresponding genomic region of V22 was carried out (https://web.expasy.org/translate/) (Supplementary Figure 3.1). The alignment showed that the predicted V22 C5 protein sequence (aa 1 to 60) aligned with the predicted ACMV C5 protein sequence at aa positions 134 to 193. The C5 protein of V22 was termed C5aa$^{134\text{-}193}$ with the superscript indicating the relative ACMV C5 aa position. The start codon of ACMV C5 ORF aligned with a start codon present in the V22 genomic sequence. However, the V22 nt sequence that aligned with the ACMV C5 sequence encoding aa 1 to 134 was predicted to be a C5

untranslated region due to the presence of two in-frame stop codons. No proteins ≥ 50 aa in length were predicted to be encoded by this region in V22 (https://web.expasy.org/translate/). The first stop codon present was shown to align with ACMV C5 codon for glutamine at aa position 31. The second stop codon aligned with ACMV C5 codon for glutamine at aa position 98.

To expand the putative coding region of the V22 C5 ORF, two mutations were sequentially introduced into the untranslated V22 C5 region. The first mutation (nt 918 a→g) was introduced to change the first stop codon to one coding for glutamine (present at this position in ACMV). This mutant was predicted to encode two C5 protein segments. This mutant was used as a template for the second mutagenesis step. Two nucleotides were mutated (nt 715 t→c, nt 717 a→t) to change the second stop codon to that coding for lysine (present at this position in other begomovirus C5 proteins). Nucleotides mentioned are relative to the first nucleotide of the V22 wildtype sequence. This mutant, V22ΔC5aa$^{1\text{-}193}$, was predicted to encode a single, full length C5 protein which aligned with ACMV C5 aa 1 to 193. A V22 C5 deletion mutant (V22XC5) was also made by mutating the guanine at nt position 603 to adenine. This resulted in a change of the third codon of C5 aa$^{134\text{-}193}$ from a glutamine to a stop codon. This deletion mutant was done to prevent complete expression of C5. The mutagenic steps carried out in the construction of these C5 mutants were designed to ensure that the mutations did not introduce inadvertent mutations to any of the proteins encoded for by overlapping ORFs. Nucleotide and predicted C5 protein sequence alignments for V22 and the V22 C5 ORF mutants are shown in Figure 9. Mutagenesis primers are shown in Table 2. All DNA constructs were sent for Sanger sequencing (Inqaba Biotechnology Industries (Pty) Ltd, Pretoria, South Africa) to confirm the presence of the target mutations (Supplementary Figure 3.1).

To compare the aa similarity between the C5aa$^{1\text{-}193}$ protein encoded by the V22ΔC5aa$^{1\text{-}193}$ mutant and the C5 protein of ACMV, a CLUSTAL multiple sequence alignment (https://www.ebi.ac.uk/Tools/msa/clustalo/) of the two putative C5 protein sequences was done. DISOPRED analysis was carried out to compare the predicted disorder of the proteins C5aa$^{134\text{-}193}$ (wildtype C5) and C5aa$^{1\text{-}97}$.

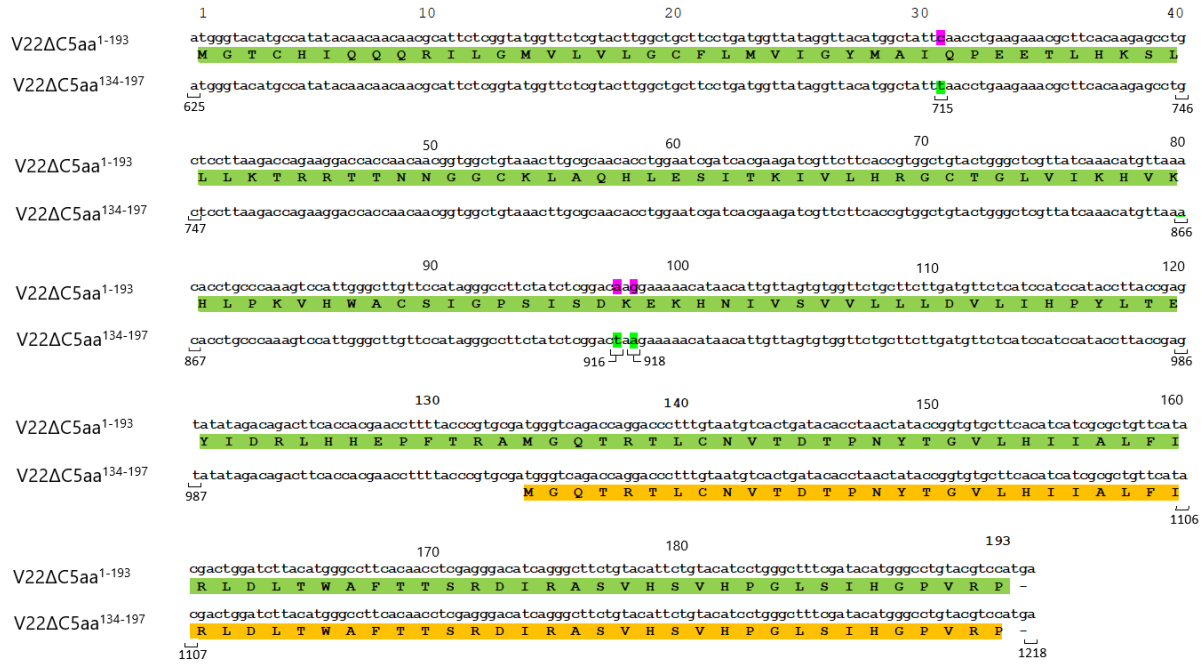


Figure 9 Nucleotide (nt) and amino acid (aa) alignments (https://web.expasy.org/translate/) of the C5 protein of V22 (V22C5aa$^{134\text{-}137}$) and the full length 193 aa long putative C5 in mutant V22ΔC5aa$^{1\text{-}193}$. The amino acids have been numbered in the N-terminal to C-terminal direction using African cassava mosaic virus (ACMV, Accession number J02057) C5 as the reference (Supplementary Figure 3.1). The translated proteins are color coded: V22C5aa$^{134\text{-}193}$ (wildtype) yellow, and V22ΔC5aa$^{1\text{-}193}$ is green. Protein sequences less than 40 aa have been excluded from this diagram. The nt targeted for mutagenesis are color coded green (wildtype sequence) and purple (mutant sequence). The IUPAC-IUNMN one -letter abbreviations for the 20 standard amino acids were used. Nucleotide numbers relative to the first nucleotide of the V22 wildtype sequence are indicated below the sequences with square brackets to indicate nt numbered.

Table 2 Summary of V22 C5 mutant clones names and respective amino acid changes and the primer sequences used for mutagenesis.

Mutant clone name	Mutation (amino acid change)	Forward primer sequences (Fp) and reverse primer sequences (Rp) for site directed mutagenesis (5'-3')	
V22ΔC5aa$^{1-97;134-193}$	Stop → Q	Fp	CTTCAGGTTGAATAGCCATGTAACCTATAACCATCAG
		Rp	CATGGCTATTCAACCTGAAGAAACGCTTCACAAG
V22ΔC5aa$^{1-193}$	Stop →K	Fp	GTTATGTTTTTCCTTGTCCGAGATAGAAGGCCCTATG
		Rp	CTATCTCGGACAAGGAAAAACATAACATTGTTAGTGTGG
V22XC5	Q →Stop	Fp	CTGGTCTAACCCATCGCACGGGTAAAAG
		Rp	CGATGGGTTAGACCAGGACCCTTTGTAATG

* The IUPAC-IUBMB one-letter abbreviations for the 20 standard amino acids were used.

2.4 Structural and disorder analysis on V22 mutant proteins

V22ΔV2-T→S and V22ΔV2-V→S mutant infectious clones introduced no inadvertent changes to other ORFs, therefore only the V2 was investigated for structural or disorder changes. On the contrary, the mutations in V22ΔV2-TS+, V22ΔV2-SL+, V22ΔV2-R→K+ and V22ΔV2Del+ clones introduced inadvertent changes to other ORFs including the V1, putative C5, putative C6 and putative C7. V22 mutant V2, V1 and putative ORFs, C5, C6 and C7, were aligned against the respective WT V22 sequences, using CLUSTAL multiple sequence alignment by MUSCLE (3.8) (https://www.ebi.ac.uk/Tools/msa/muscle/). This was to identify any changes to the inadvertent ORFs. Furthermore, PSIPRED sequence plots and DISOPRED plots (http://bioinf.cs.ucl.ac.uk/psipred/) were generated to evaluate any significant structural or protein binding differences between the WT and mutant V22 ORFs. cNLS mapper (http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi) was used to identify a possible change in cellular nuclear localisation signal (cNLS) sequences in the V1 of WT V22 and V2ΔV2-T→S+.

2.5 Growing *N. benthamiana* for inoculations with infectious clones

Nicotiana benthamiana plants were germinated and grown in a controlled growth facility. The plants were exposed to 16 h light daily and were maintained at a temperature of 26 °C with a relative humidity of 50 %. *N. benthamiana* seeds were germinated in Jiffy® 7C cocopeat pellets under clingwrap and these seeds were acclimatized by slowly piercing holes in the clingwrap over a 5-day period prior to inoculation at which the pants were at a 4-8 leaf stage. At 12 days post inoculation (dpi), the plants in the Jiffy® pellets were transferred into 7 cm x 7 cm 10 % bleach sterilized pots and filled with moist cocopeat medium. *N. benthamiana* plants were supplemented with 20 mL Kelpak seaweed at 4 dpi in addition to being fertilized with 20 mL of Seagro organic fertilizer at 8, 16, 26 and 36 dpi.

2.6 Agroinoculation of *N. benthamiana* with infectious clones

LB agar plates, containing 50 mg/mL kanamycin and 50 mg/mL rifampicin, were streaked with *A. tumefaciens,* which were transformed using the Freeze Thaw method, as described by Weigel and Glazebrook 2006, with the V30 and V22 infectious clone constructs or mutant clones. The plates were incubated for 24 hours at 28 °C and then used to inoculate 2,5 mL LB broths containing 50 mg/mL kanamycin and 50 mg/mL rifampicin. These broths were incubated at 28 °C for 12 hours shaking at 120 rpm until the OD_{600} reached 1. Cultures were centrifuged at 3300 g for 10 min at room temperature. The supernatant was discarded, and the pellet was resuspended in infiltration buffer, 0.1 µM acetosyringone and 0.01 M MgCl2, until OD_{600} was 0.8. The resuspended cells were used to inoculate two *N. benthamiana* leaves below the meristem, through injection with a 1 mL needleless syringe. One hundred µL of the respective inoculum was agroinoculated into five plants per trial, and 3 trials were carried out. For the mixed infection plants, 50 µL of V30 and 50 µL of V22 were used to inoculate the *N. benthamiana* plants at the 6-8 leaf stage. Three trials, using five plants per trial, were carried out using the above-described infectious clone constructs.

For the mixed infection trials where plants are infected with more than one infectious clone, inoculation buffers were prepared as mentioned above and then the respective inoculation buffers were mixed in a 1:1 ratio and used for inoculation of *N. benthamiana*. Three trials, using five plants per trial, were carried out using V30, V22 and V30ΔIR-s to inoculate the plants with a mixed infection of V30+V22 or V30+V30ΔIR-s.

2.7 Collection of symptom progression data

Plant symptoms were photographically recorded, height measurements were taken, and symptoms scored following a disease severity index specific to ToCSV-induced symptoms observed in *N. benthamiana*. Scoring, as per Table 3 below, was done every 4 days post inoculation on 5 plants per trial per treatment. Symptom progression scoring was repeated for all three trials. The leaves that were selected for scoring were 2 to 3 expanded leaves below the meristem of the plant.

Table 3 Symptom scores for the phenotypes, upward leaf roll (ULR) and swollen veins (SV), induced in *N. benthamiana* by Tomato curly stunt virus (ToCSV).

Score for upward leaf roll	**Description**
0	Leaf edges normal
1	Leaf edges turned up 1-7 zones, or mild ULR 1-4 zones
2	Leaf edges turned up 2-7 zones and mild/narrow ULR ≤1-4 zones, or narrow/mild ULR 5-7 zones, or broad/severe ULR 1-4 zones
3	Severe ULR ≥ 5 zones, or entire leaf edge rolled upward and onto itself
Score for swollen veins	**Description**
0	Leaf veins normal
1	Slight SV to mild SV over entire leaf, or moderate SV over less than half of leaf
2	Moderate SV more than half of leaf only, or moderate SV more than half of leaf with some severe SV, or severe SV less than half of leaf
3	Severe SV more than half of leaf, or entire leaf deep and dense SV

* For ULR scoring purposes, the perimeter of each leaf that was scored was visually divided into 7 equal sections, referred to as zones.

2.8 Nucleic acid extraction

Total nucleic acid was extracted from two 12.7 mm diameter leaf discs which were immediately flash frozen in liquid nitrogen upon collection. The leaves collected were taken from two expanded leaves below the meristem of the plant at time points 20 and 28 dpi for the V22ΔIR-s treatment, 20 and 36 dpi for the V22 mutants and the mixed infection. Twenty dpi was chosen as this is the viral load peak for both V30 and V22 as previously established by the WT infectivity curves of V30 and V22 (Zwolinski, 2022). The 28-dpi time point was chosen as it is the next interval after the viral load peak. The time point of 36 dpi was selected to see if recovery of symptom and viral load was observed in *N. benthamiana*. DNA was extracted from the leaves of *N. benthamiana* inoculated with the WT V30 and V22 variants, mixed V30+V22 infectious clones, and control plants. These control plants were mock inoculated with *A. tumefacien*s alone. DNA was extracted using a modified CTAB method originally described by Doyle (1991). The frozen leaf material samples were finely ground with the use of the Qiagen TissueLyser II system (Qiagen) and Qiagen 3 mm diameter tungsten carbide balls at 20 hertz for 2 min at 1 min intervals (Qiagen, 1064945). Five hundred µL of 2 % CTAB buffer, heated to 65 °C, and 1 µL of undiluted βMe were added to each sample and then vortexed. βMe was added as this strong reducing agent served to remove phenolic compounds in crude plant material and therefore yield a higher DNA quality (Horne et a., 2004). The samples were incubated in a heated water bath set to 65 °C for 1 h. Five hundred µL of chloroform: isoamylalcohol, at a ratio of 24: 1, was added to the samples and these were mixed by inverting the Eppendorf tubes in order to extract the nucleic acids from the samples. Samples were centrifuged at 15 000 xg for 10 min at 4 °C. Five hundred µL of the supernatant was carefully aspirated into a new tube where a second chloroform: isoamylalcohol extraction step as previously described was repeated on this supernatant. Nucleic acids were then precipitated from this supernatant by the addition of 200 µL of ice-cold isopropanol. The tubes were inverted 10 times and stored at – 80 °C for 30 min. Samples were then centrifuged for at 15 000 xg for 10 min at 4 °C. The pellet was washed twice with molecular grade ethanol. The first wash consisted of 95 % ethanol followed by 75 % ethanol. The extra wash steps allowed for purer DNA samples to be obtained. The ethanol was aspirated out and pellets were air dried in the laminar flow until visually dry. Samples were resuspended in 20 – 50 µL of Tris-EDTA (TE) buffer.

Total RNA was extracted from six homogenized 12.7 mm diameter leaf discs using TRIzol™ reagent, (Thermo Scientific™, Massachusetts, US) as per the manufacturer's protocol. The frozen leaf material was finely ground with the use of the Qiagen TissueLyser II system (Qiagen) and DEPC treated 3 mm diameter tungsten balls as per manufacturer's instructions (Qiagen, 1064945). To each homogenized sample, 1 mL of Trizol reagent (Thermo Scientific™, Massachusetts, US) was added and the tubes were inverted while thawing to thoroughly mix the contents. The samples were incubated at room temperature for 5 min followed by the addition of 300 µL of chloroform, mixed by inversion and incubated at room temperature for 5 min. The tubes were centrifuged at 12 500 rpm for a total of 15 min at 4 °C. The supernatant was transferred into new sterile 1.5 mL tubes and an equal volume of chloroform was added to the supernatant. Samples were mixed by inversion and then subjected to another round of centrifugation at 12 500 rpm for 15 min at 4 °C. The supernatant was transferred to a new sterile tube and the nucleic acids were precipitated by the addition of ice-cold isopropanol. Samples were incubated for 30 min at room temperature and then centrifuged at 12 500 rpm for 15 min. The pellets were washed with 75 % molecular grade ethanol and then centrifuged for 5 min at 4 °C. This ethanol was step was then repeated. The ethanol was aspirated out the pellets were air dried in the laminar flow. The dried pellets were resuspended in 50 µL DEPC water pre-warmed to 65 °C. Twenty µL aliquots of extracted RNA were stored at – 80 °C until further use to maintain the integrity of the RNA. All nucleic acid extractions were quantified using the spectrophotometer (NanoDropTM One, Thermo Scientific™). Extracted DNA was stored at -20 °C and RNA was stored at -80 °C in preparation for downstream analysis.

2.9 Viral load

Viral load was determined via real-time quantitative PCR, using the CFX96 Touch Real-Time PCR Detection system (BioRad, California, US), with the Maxima SYBR Green/ROX qPCR Master mix (Thermo Scientific™, Massachusetts, US), as per the manufacturer's protocol (Thermo Scientific™, #K0221). A master mix of Maxima SYBR Green (Thermo Scientific™, Massachusetts, US), forward primer, reverse primer and nuclease free water was set up according to the protocol and number of samples to be run. This master mix was then aliquoted into white shelled 96-well plates. To every 9 µL aliquot of master mix, 1 µL of 10 ng DNA was added. The negative controls consisted of the non-template control (NTC) and DNA

extracted from the *Agrobacterium* (mock)- inoculated plants. Plastic films were placed over the wells to seal the wells and plates were centrifuged at 6000 rpm for 5 min. The plates were then loaded into the BioRad CFX Connect Real-Time PCR System thermocycler (BioRad, California, US) and the two-step cycling protocol programme as in the Thermo Scientific Maxima SYBR Green manual was initiated (Thermo Scientific™, #K0221). Tm was set at 60 °C. Viral load was determined in relation to the Glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) internal control for which these optimized *N. benthamiana* specific primer sequences were sourced from Farooq et al., (2019). The target primers, previously designed and optimized to target V2 of V30 and V22 were used (Zwolinski, 2022). Primer sequences are shown in Table 4. Viral load was calculated using the delta Ct method formula: $\mathrm{Log}\Delta\mathrm{Ct} = 2^{-\Delta\mathrm{Ct}}$

Table 4 Primer names and sequences for qPCR primers used.

Primer Name	Forward primer (Fp) or Reverse primer (Rp)	Primer sequence (5'-3')
GAPDH	Fp	CTGTTATTGGAGGAGGGAACAA
	Rp	AGTCTTTCCTACCATGCCAAC
V2	Fp	CGACAGCCCGTTCACCAG
	Rp	GCCTGTACGTCCATGATCGTC

* GAPDH represents the primers used to target the *glyceraldehyde 3-phosphate dehydrogenase* (GAPDH) gene in *N. benthamaina* and the V2 represents the primers targeting the V2 ORF of the target viruses.

2.10 Detection of RNA transcripts

Since IR-derived siRNAs were reported in another begomovirus (Yang et al., 2019), investigation of a possible RNA transcripts encompassing the IR and ORI in V22 was undertaken through cDNA synthesis and PCR. Gene specific primers for cDNA synthesis were designed using Primer3 (https://bioinfo.ut.ee/primer3-0.4.0/). Quality of oligonucleotides, GC content, meting temperature and secondary structure potential, were evaluated using the Oligoanalyzer™ Tool (https://www.idtdna.com/oligoanalyzer). Primer sequences used for cDNA synthesis, as well to screen for detection of the transcripts, are listed in Table 5. The primers were designed to bind RNA transcribed from the 3' end of the IR of V22. This region of the IR was chosen as this is where the majority of the IR nucleotide differences between V30 and V22 lie. The primers were designed to target an RNA transcript of 145 bp in length in the 3' IR of V22.

Table 5 Primer names and sequences of primers used for cDNA synthesis for the detection of a V22 IR transcript.

Primer sequence name	Target nt position (5'-3')	Sequence (5' -3')
V22IR_FP	2699-79	GTAATTTGTGGCAAAGTAATTGGAATTTC
V22IR_RP	79-2699	GCTTTAACGCCTCAATTTCATTGG

cDNA from 1 ng of extracted RNA was synthesised using the RevertAid™ First Strand cDNA Synthesis Kit (Thermo Scientific™, #K1622) following the manufacturer's protocol. Genomic DNA was first removed from the RNA samples through DNase I digestion whereby 1 µL (1 U) of enzyme, in addition to 1 µL 10X Reaction Buffer with $MgCI_2$ and 1 µg of RNA, made up to a total volume of 10 µL with nuclease free water was incubated at 37 °C for 30 min. Genomic DNA was removed to ensure that there was no DNA in the sample to act as a template for RT-PCR. Enzyme deactivation was carried out with the addition of 1 µL 50 mM EDTA followed by a 10 min 65 °C incubation step. To one DNA digested RNA sample, 1 µL of the gene-specific primers (Table 5), 4 µL 5X Reaction Buffer, 1 µL Riboblock, 2 µL 10 mM dNTP mix, 1 µL Revertaid M-MuLV RT (200 U/µL) were added. The samples were gently vortexed, briefly centrifuged and incubated at 42 °C for 60 min. The reaction was then terminated by heating at 70 °C for 5 min. Negative Revertaid controls as well as *Agrobacterium*-inoculated (mock) samples were included. cDNA was stored at – 80 °C for subsequent PCR reactions. Screening PCR was carried out using DreamTaq™ PCR Master Mix (Thermo Scientific™, Massachusetts, US) as per the manufacturers protocol with a Tm of 58 °C for the V22 IR primers (Table 5). RT-PCR products were run on a 2 % agarose gel in a 1xTAE buffer at 80 V for 45 min in a Mini-Sub Cell GT Cell (BioRad) electrophoresis tank. The gels were then visualised on the Molecular Imager® Gel Doc™ XR System (BioRad, Hercules, CA, USA).

2.11 Data analysis and statistics

Standard deviations and averages were obtained from the data of 3 trials performed for each treatment. This data included plant height, disease severity scores and viral load. Statistical analysis was done using IBM SPSS Statistics Version 28.0 to determine significant differences between treatments with p value set to 0.05. Graphs were constructed on Microsoft Excel Version 2110. Significance between treatments was determined using Student T-tests.

Plant height was calculated to give a value representative of the control plant height that the inoculated plant achieved. The mock-inoculated plants were treated with the pCambia2300

empty vector in *A. tumefaciens.* This formula was obtained from Lapidot et al., 2006. Heights were calculated as percentages to give the plant height relative to the control as a percentage (%) using the formula:

$$\frac{\text{height of inoculated plants}}{\text{height of mock} - \text{inoculated plants}} \text{ x } 100$$

CHAPTER 3 - RESULTS

3.1 The Intergenic Region (IR)

3.1.1 Verification of IR swap infectious clone constructs

In order to verify the IR-swap clone constructs; V22ΔIR-sl, V22ΔIR-sr and V22ΔIR-s, the mutant clones were digested with BglII and the expected banding patterns of 1308 bp and 10172 bp were observed (Figure 10).

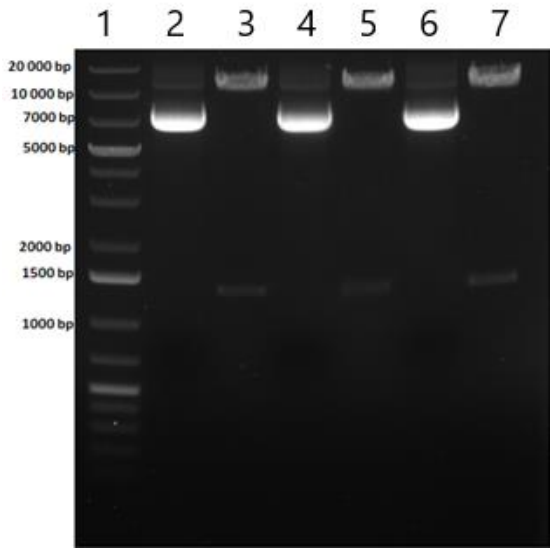


Figure 10 V22ΔIR-sl, V22ΔIR-s and V22ΔIR-sr were digested with BglI at 2 sites yielding 2 linear products of 1308 base pairs (bp) and 10172 bp in length as shown on a 1 % agarose gel stained with ethidium bromide. Lane 1) 1 Kb gene Ruler (ThermoScientific), lane 2) V22ΔIR-sl undigested, lane 3) V22ΔIR-sl digested, lane 4) V22ΔIR-sr undigested, lane 5) V22ΔIR-sr digested, lane 6) V22ΔIR-s undigested and lane 7) V22ΔIR-s digested. DNA samples were originally extracted from the respective transformed DH5α E. Coli cells.

3.1.2 Height, symptom severity scores and viral load data of inoculated *N. benthamiana*

A nucleotide alignment of the 3' IR (nucleotide 61 to 139) of V30 and V22 revealed the multiple nucleotide differences between these sequences (Figure 4 in Chapter 1).

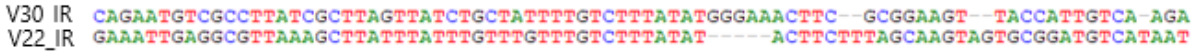


Figure 4 Alignment of V30 and V22 of the 3' end of the IR (nucleotide 61 to 139). Alignment done on BioEdit Sequence Alignment editor (https://bioedit.software.informer.com/7.2/).

Heights of the *N. benthamiana* for the IR-swap infectious clone construct trials were measured in mm at 4-day intervals post inoculation. The heights were then processed using the formula: $\frac{\text{height of inoculated plants}}{\text{height of mock-inoculated plants}}$ x 100 (Lapidot et al., 2006) to get the height relative to the mock-inoculated plants. Data was derived from mean values from three trials with five plants per trial. No significant trends between wildtype (WT) V22 mean height relative to mock height (%) and that of the IR-swap mutant clone mean mock plant heights (%) were observed. The only significant ($p<0.05$) height difference was seen at 32 dpi between WT V22 and V22ΔIR-s (Figure 11).

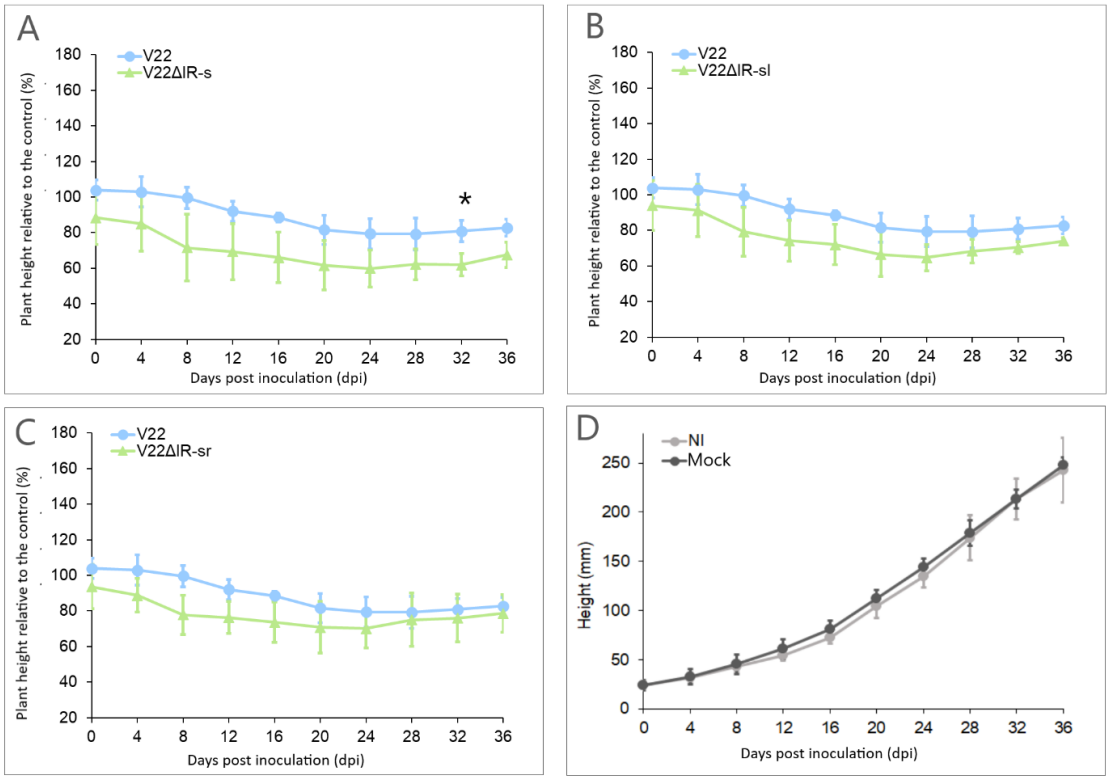


Figure 11 Plant height relative to control (pCambia2300 empty vector in *Agrobacterium tumefaciens*) (%) over time measured in days post inoculation (dpi), of wildtype (WT) V22, A) V22ΔIR-s, B) V22ΔIR-sl and C) V22ΔIR-sr. D) Height (mm) of *Nicotiana benthamiana* between non-inoculated and mock-inoculated (pCambia2300 vector in *A. tumefaciens*) treatments over time. *=p<0.05 (student T-test). Bars indicate mean ± SD.

Two expanded leaves below the meristem of the plants were scored yielding symptom severity scores. Symptom severity score data revealed that the most significant differences (p<0.05) for ULR and SV severity scores when compared to WT V22 clones was present in the V22ΔIR-s (Figure 12 A). The severity scores induced by this mutant were significantly lower than the severity scores for V22, a trend also observed with the shorter IR-swap, V22ΔIR-sr (Figure 12 B and 13**).** ULR and SV are shown at 20 dpi in Figure 13 and 14. The progression of the symptoms from 12 dpi to 36 dpi induced by the respective infectious clone constructs was visually observed (Supplementary Figure 1.8 to Figure 1.10) Progression of symptoms induced by WT V30, V22 and the mock controls were also observed (Supplementary Figure 1.5 to Figure 1.6).

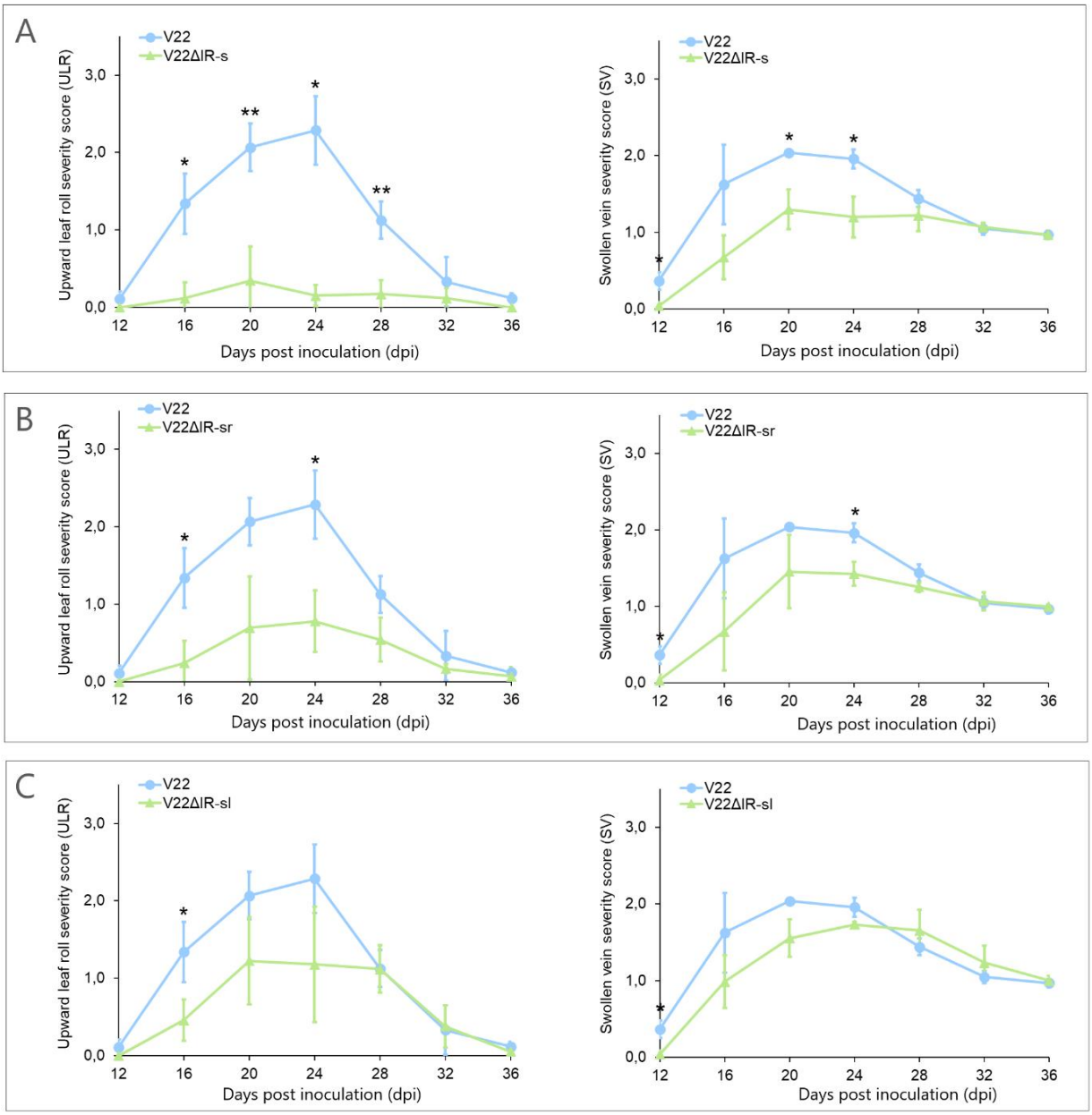


Figure 12 Upward leaf roll severity score (ULR), left graphs, and Swollen vein severity score (SV), right graphs, induced in *Nicotiana benthamiana* for A) V22ΔIR-s, B) V22ΔIR-sr and C) V22ΔIR-sl. *=P<0.05 (student T-test). Bars indicate mean ± SD.

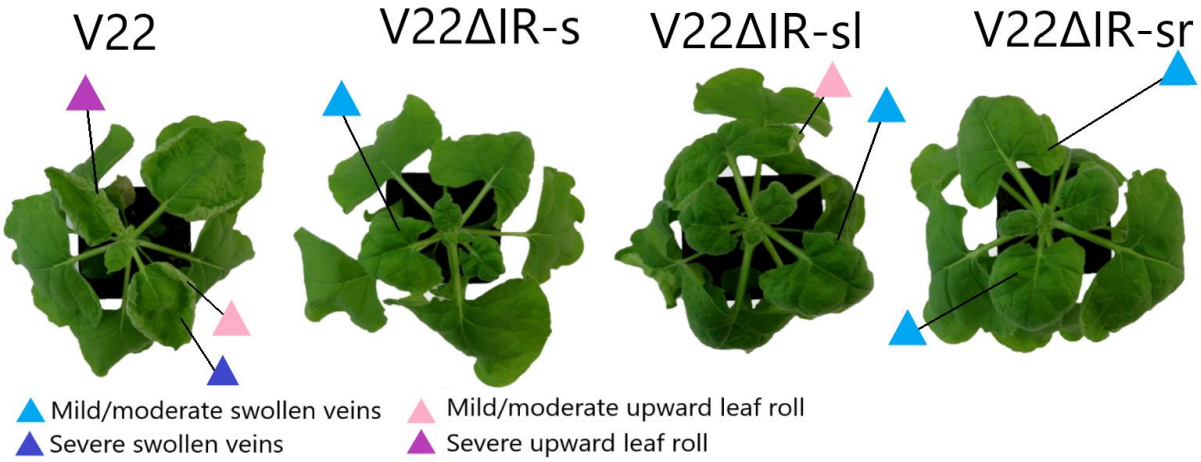


Figure 13 *Nicotiana benthamiana* inoculated with V22 (positive control), V22ΔIR-s, V22ΔIR-sl and V22ΔIR-sr infectious clone construct at 20 days post inoculation (dpi). Symptoms present are upward leaf roll (ULR) and swollen veins (SV). Severity of symptoms are indicated by the colour coded triangles as indicated by the key.

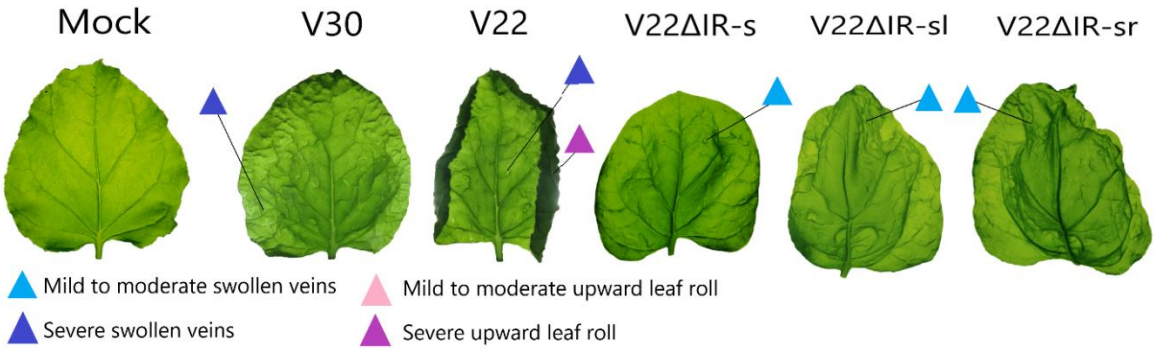


Figure 14 *Nicotiana benthamiana* leaves dissected from plants inoculated with pCambia2300 vector in *Agrobacterium tumefaciens* (Mock), V30 (positive control), V22 (positive control), V22ΔIR-s, V22ΔIR-sl and V22ΔIR-sr infectious clone construct. Leaves were viewed against the light. Symptoms present are upward leaf roll (ULR), and swollen veins (SV) Severity of symptoms are indicated by the colour coded triangles as explained by the key.

DNA was extracted and qPCR performed at the V22 viral load peak time point (20 dpi), and the time point where viral load recovery is observed (28 dpi) in *N. benthamiana*. These time points were determined from V22 viral load infectivity curve (Zwolinski, 2022). The controls in these three trials were V22 and mock-inoculated (pCambia 2300 vector in *A. tumefaciens*) plants. Viral load induced by V22 and V22ΔIR-s in *N. benthamiana* was not significantly different at both 20 and 28 dpi (Figure 15).

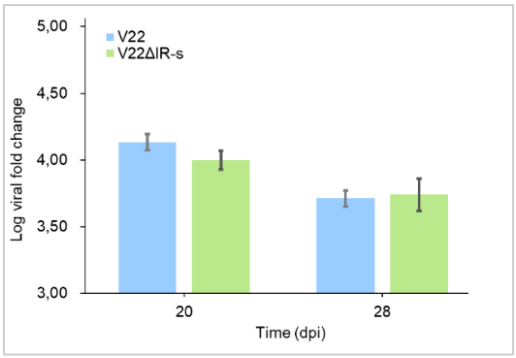


Figure 15 Log viral DNA fold change of wildtype (WT) ToCSV_V22 (V22) and V22ΔIR-s in *Nicotiana benthamiana* at 20- and 28-days post inoculation (dpi). Significance shown with an asterisk. Bars indicate mean ± SD.

It was proposed that the symptoms induced by the IR-swap mutant clones could be due to transcription of a lnRNA and possible derived siRNAs derived thereof as shown in another monopartite begomovirus in a previous study (Yang *et al.*, 2019). In order to screen for an IR transcript in V22, RNA was extracted from V22 inoculated plant material at 20 dpi, cDNA synthesis and screening PCR was carried out to reveal a V22 IR transcript. An IR transcript, 146 base pairs (bp) in length, covering the 3' end of V22 IR was detected in both the cs and vs of V22 which spans the ORI in the IR (Figure 16).

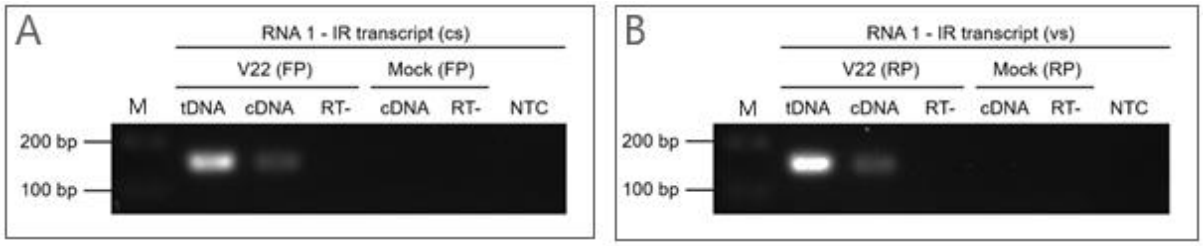


Figure 16 PCR products, 145 bp in length, from primers targeting the IR in V22 total DNA (tDNA), V22 complementary DNA (cDNA), revertaid control (RT-), mock inoculated cDNA, Mock RT- and non-template control (NTC). A) cDNA from the forward primer (Fp), therefore the complementary sense (cs) transcript of IR. B) cDNA from the reverse primer (Rp), hence the viral sense transcript (vs). M is the DNA ladder.

3.1.3 siRNA from the IR that may target *N. benthamiana* and tomato genomes

In order to determine if there could be siRNA originating from the 3' end of the IR of ToCSV V30 and V22, the siRNA predictor tool (www.plantgrn.noble.org) was used. The predicted siRNAs and their alignments of these predicted siRNAs to the ToCSV V30 and V22 genomes (Supplementary Figure 1.1 to Figure 1.4) were subjected to BLAST via SOL Genomics and

NCBI. No siRNA matches to the tomato and *N. benthamiana* genomes were identified. When the generated siRNAs were subjected to NCBI BLAST, no matches between the length of 21-24 nt were identified.

3.2 The V2 ORF

3.2.1 Bioinformatics comparison of the V30 and V22 V2 proteins

The CLUSTAL multiple sequence alignment of the V2 protein sequence of V30 and V22 showed seven aa differences as well as the C- terminal QKP residues present in V22 V2 but absent in V30 (Figure 17). Based on the amino acid differences between the two variants, V22 V2 was mutated to reflect the nucleotide sequence of V30. The five aa regions targeted for mutagenesis in this study are indicated by the arrows in Figure 17.

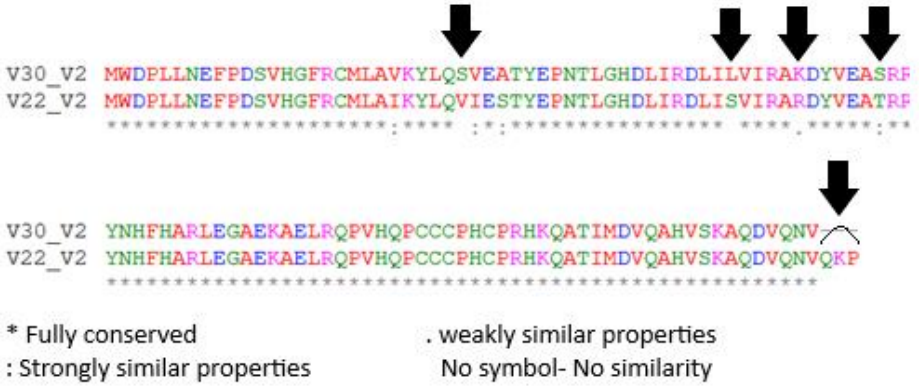


Figure 17 CLUSTAL multiple sequence alignment by MUSCLE (3.8) of WT V30 and V22 V2 amino acid sequences with the degree of similarity of the residues by means of structure and polarity indicated by symbols below the sequences, as described by the key. The arrows indicate the amino acids that were targeted for mutagenesis in this study.

Amino acid conservation and disorder between V30 and V22 V2 proteins were investigated through PSIPRED and DISOPRED plots. The PSIPRED sequence map of predicted secondary structure of V30 and V22 V2 proteins indicates that the QKP residues present in V22 are predicted to be disordered as well as involved in protein binding (Figure 18A). This indicates that this region in the V2 protein is disordered and may have the capacity to bind other proteins or macromolecules (Simon, 2020). This V22 V2 protein is therefore considered partially

disordered (Lieutaud et al., 2016). DISOPRED plots give a probability estimate of each residue within the submitted protein sequence of being disordered with these disordered regions predicted to be dynamically flexible (Ward et al., 2004). The threshold cut-off for the prediction of disorder was set at 0.5 and any amino acids exceeding this threshold were predicted to be disordered. DISOPRED plots of the V30 V2 and V22 V2 protein show that there is a predicted disorder for the C- terminal QKP residues present in the V22 V2 protein sequence (Figure 18 B). The last two amino acids at V30 C-terminus, NV, were also predicted to be disordered as well as being able to bind proteins.

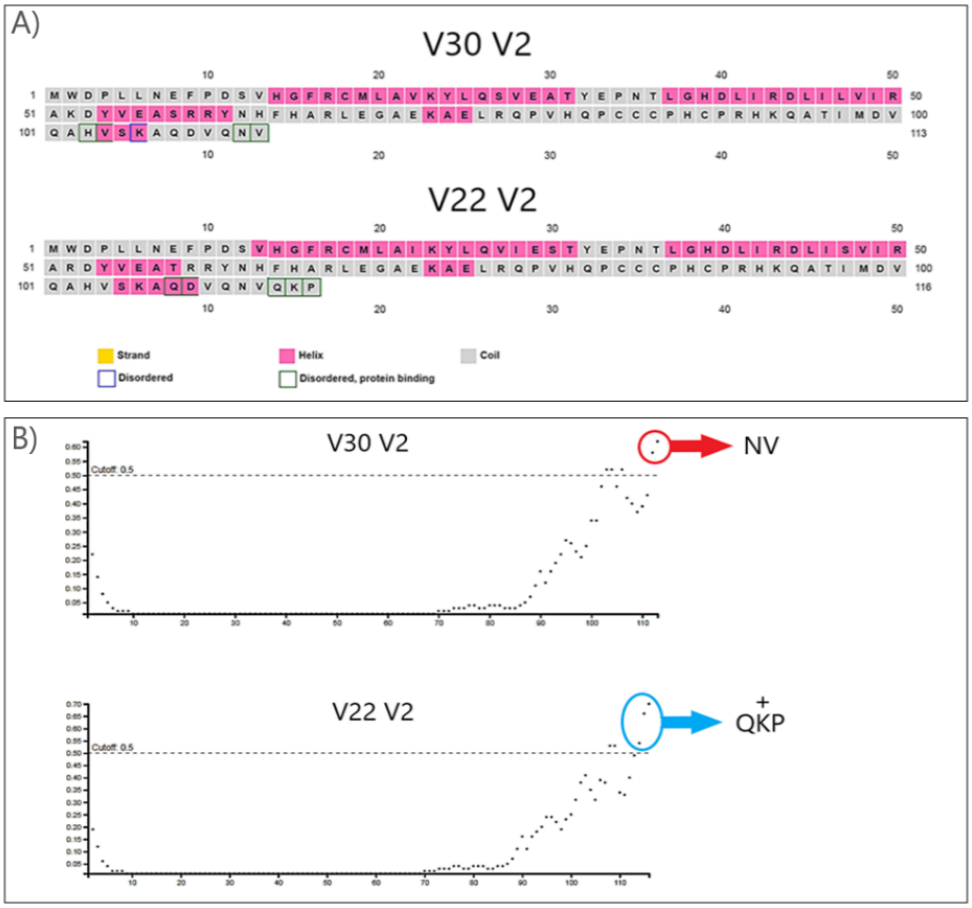


Figure 18 (A) PSIPRED sequence map of predicted secondary structure of wildtype (WT) V30 and V22 V2 proteins. Key indicates the predicted structures or function. (B) DISOPRED plot of predicted disorder of V30 and V22 infectious clone construct V2 proteins. QKP+ indicates the 3 additional C-terminal residues of the V22 V2 protein that is absent in V30 V2. The circled residues indicate the C-terminal residues with predicted disorder. The IUPAC-IUBMB one-letter abbreviations for the 20 standard amino acids were used.

3.2.2 Bioinformatic analysis of V22 V2 single aa mutants

Some of the V22 V2 single aa mutant infectious clone constructs in this study were found to have inadvertent changes to one or two of the overlapping ORFs. These ORFs were V1 as well as the putative ORFs. The putative ORFs mentioned in this study, which have not been confirmed in ToCSV, are C5, C6 and C7. A summary of these V22 V2 single aa mutants and their inadvertent mutations are summarized in Table 6.

Table 6 V22 single mutant amino acid clones and amino acid positions with inadvertent mutations indicated

Mutant clone name	Amino acid mutation in V22 V2	Inadvertent amino acid mutations to protein encoded by overlapping ORF
V22ΔV2-VS	V27S	**None**
V22ΔV2-TS	T58S	**None**
V22ΔV2-TS+	T58S	V1 (P5D and A6D) and putative C6 (A90S and G91T)
V22ΔV2-SL+	S47L	Putative C6 (102~~RNKISN~~108)
V22ΔV2-RK+	R58K	Putative C6 (P102L)
V22ΔV2Del+	114-116 (deleted)	Putative C5 (V45T) and putative C7 (C13Y)

+ indicates the presence of off targets. The mutant names in bold are the mutant clones which have no inadvertent mutations. The cells highlighted in grey are the infectious clone constructs for which 3 official trials were carried out. The non-highlighted cells are the clones which only preliminary trials were carried out. The IUPAC-IUBMB one- letter abbreviations for the 20 standard amino acids were used. ~~RNKISN~~ represents the deletion of the respective aa.

The changes in structure and disorder to the V2 protein in these V22 V2 single aa mutants, no changes to the secondary structure or disorder to the V2 protein of V22ΔV2-V→S, V22ΔV2-T→S and V22ΔV2-T→S+ relative to V22 were predicted (Figure 19). The only mutant which showed a change in the secondary structure and disorder was V22ΔV2Del+ with a shortened helix in the C-terminus accompanied by a change of disordered protein binding residues in the C-terminus of the V2 protein when compared to the V22 V2 protein structure and disorder (Supplementary Figure 3.4).

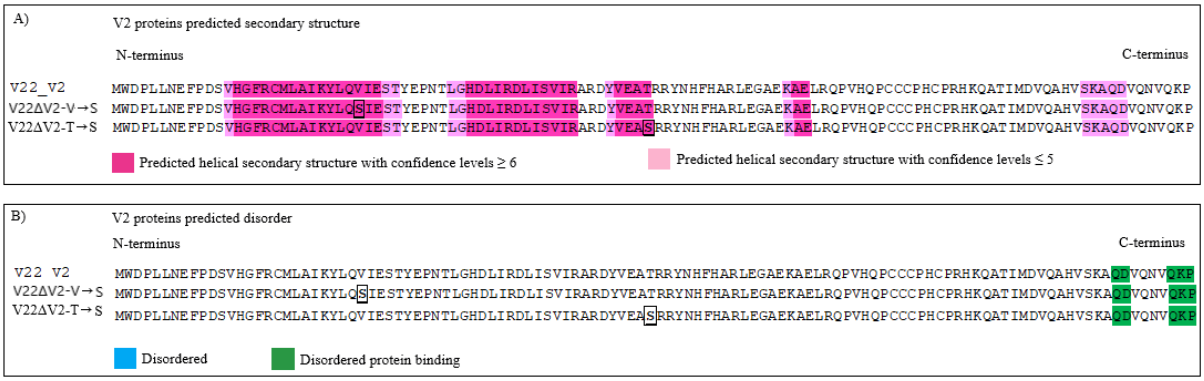


Figure 19 V2 protein sequences of V22_V2 (wildtype), V22ΔV2-V→S, V22ΔV2-T→S. A) Secondary structures indicated by colour coding as indicated by the key. The amino acid changes as a result of nucleotide mutagenesis indicated by the boxes around the residues. B) Disorder indicated by a colour coded key. The threshold for prediction of residues to be disordered was 0.5. Disorder and secondary structural predictions obtained from http://bioinf.cs.ucl.ac.uk/psipred/.

Of the V2 single mutants, V22ΔV2-T→S+ had an inadvertent change to the overlapping V1 ORF due to the nucleotide sequence change. This occurred at aa 5 and 6 where aa proline and alanine were mutated to valine and aspartic acid respectively (Figure 20). The length and confidence of the first five beta-strands predicted in the N-terminus of the V1 protein of V22ΔV2-T→S+ changed compared to the non-mutated V1 of V22 (Figure 20). This change in the structure of the strands in the N-terminus of the V1 protein of V22ΔV2-T→S+ was accompanied by a change in disordered protein binding residues in the N-terminus (Figure 20). The confidence of the prediction related to how likely the secondary structure is to occur within that protein is based on the submitted amino acid sequence and its similarity to other amino acid sequences and their secondary structure in protein databases (McGuffin *et al.*, 2000). Two cellular nuclear localisation signals (cNLs) in the N-terminus of the V1 protein in V22 were found to be altered resulting in only one cNLS present in the N-terminus of the V1 protein of V22ΔV2-T→S+ (Figure 21).

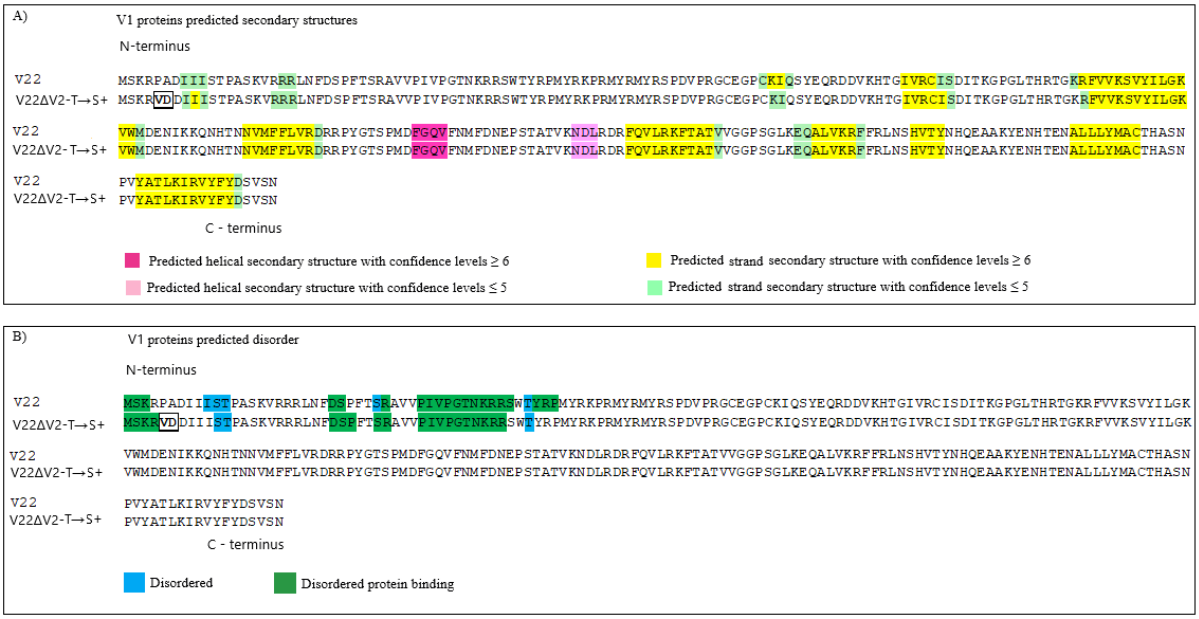


Figure 20 V1 protein sequences of V22 (wildtype) and V22ΔV2-T→S+. A) Secondary structures indicated by a colour coded key. The amino acid changes as a result of nucleotide mutagenesis indicated by the boxes around the residues. B) Disordered residues indicated as described in the key. The threshold for prediction of residues to be disordered was 0.5. Disorder and secondary structural predictions obtained from http://bioinf.cs.ucl.ac.uk/psipred/.

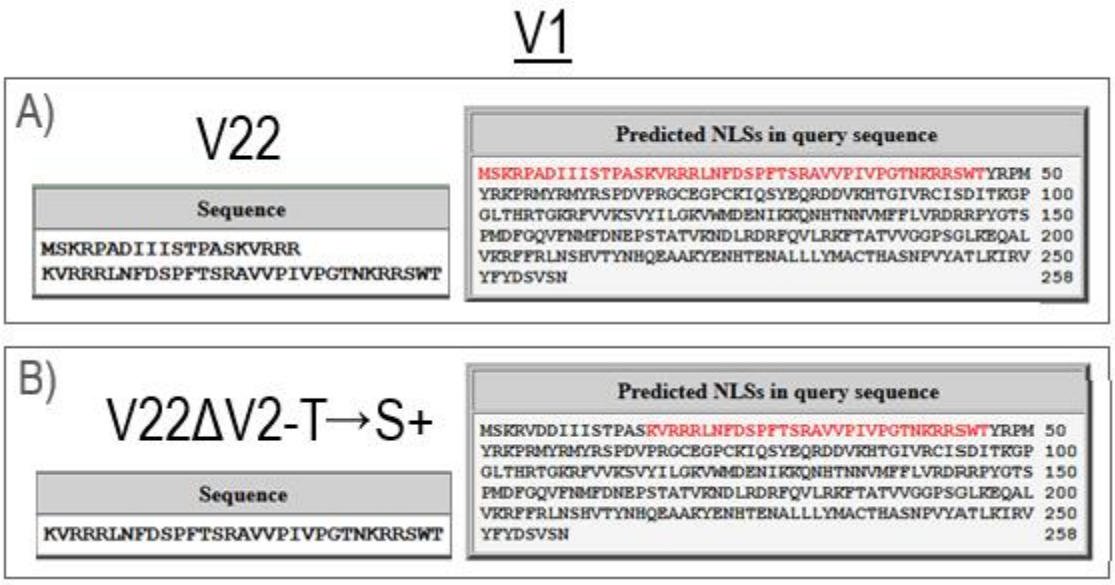


Figure 21 Cellular nuclear localisation (cNLS) mapper result (http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi) for (A) V22 and (B) V22ΔV-T→S+ V1 protein. Predicted NLS sequences are highlighted in red.

The C6 protein has not been confirmed in ToCSV and therefore it is considered an unconfirmed putative C6 protein. The changes in the putative C6 protein of V22ΔV2-T→S+ predicted due to the aa substitution was the extension of the fourth strand by only a residue in the C-terminus

of the protein. Although this small secondary structural change was observed, no change to the disorder was predicted when compared to the WT V22 putative C6 ORF (Figure 22).

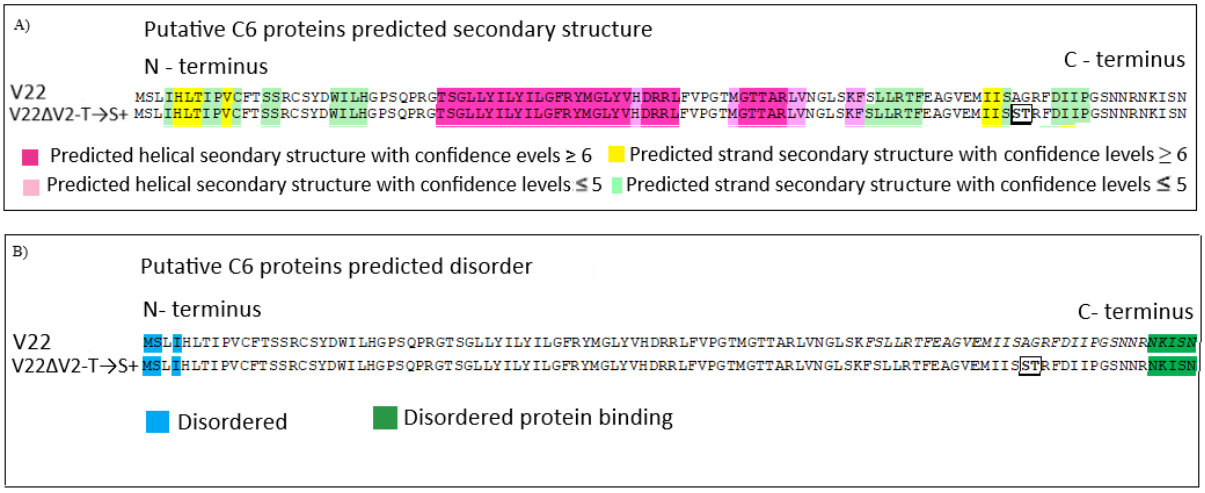


Figure 22 Putative C6 protein sequences of V22 (wildtype) and V22ΔV2-T→S+. A) Secondary structures indicated by a colour code as described in the key. The amino acid changes as a result of nucleotide mutagenesis indicated by the boxes around the residues. B) Disordered residues are indicated by a colour code as described in the key. The threshold for prediction of residues to be disordered was 0.5. Disorder and secondary structural predictions obtained from http://bioinf.cs.ucl.ac.uk/psipred/ .

The other mutants, V22ΔV2-S→L+, V22ΔV2-R→K+ and V22ΔV2Del+, for which only preliminary trials were carried out, revealed predicted changes to the overlapping putative C5, C6 and C7 ORFs. V22ΔV2-S→L+ revealed changes to the lengths of two beta-strands in the C-terminus of the putative C6 protein (Supplementary Figure 2.5 and 2.6). In addition, V22ΔV2Del+ had an extended helix in the C- terminus of the putative C7 protein where two helices were joined into one by a single residue but the disorder of the residues in C7 remained unchanged (Supplementary Figure 2.7). V22ΔV2Del+ had a predicted change where two helices were extended in the C-terminus of the putative C5 protein. Disorder remained unchanged when compared to the WT V22 C5 putative protein (Supplementary Figure 2.6).

3.2.3 Single V22-V2 mutant amino acid clone verification

To confirm the presence of mutations in the single aa V22 V2 mutant clones, V22ΔV2T→S, V22ΔV2R→K+ and V22ΔV2S→L+ were digested with their respective enzymes and all mutant clones revealed the correct banding patterns (Figure 23 and Figure 24). The bands obtained for the SpeI digested V22 infectious clone construct were 11 185 bp and 8420 bp in

length. The bands obtained for the SpeI digested V22ΔV2S→L+ were 248 bp, 2518 bp and 8830 bp in length. These mutant clones were sequenced (Sanger *et al.,* 1977) with the other V2 mutant clones; V22ΔV2-V→S, V22ΔV2-T→S+, and were verified.

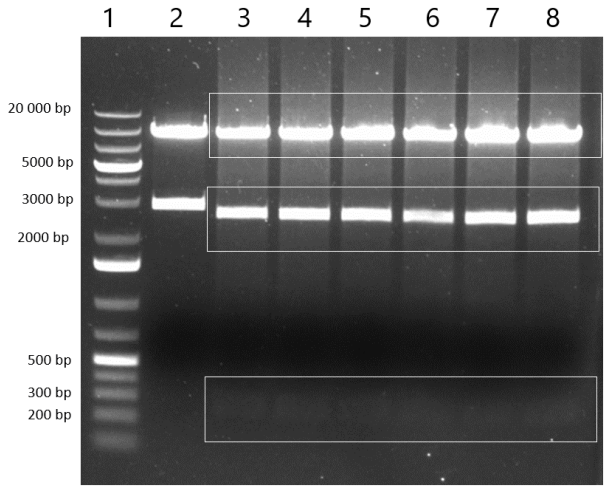


Figure 23 1 % agarose gel showing 1 kb Gene Ruler, Thermo Scientific™ (lane 1), V22 plasmid digested with SpeI (Lane 2) revealing 2 bands of 11 185 base pairs (bp) and 8 420 bp in length, 3 clones of V22ΔV2S→L+ digested with SpeI for 4 hours (lanes 3-5) and 3 clones of V22ΔV2S→L+ digested for 12 hours (lanes 6-8) revealing bands of 248 bp, 2517 bp and 8830 bp in length. The expected bands are indicated with boxes.

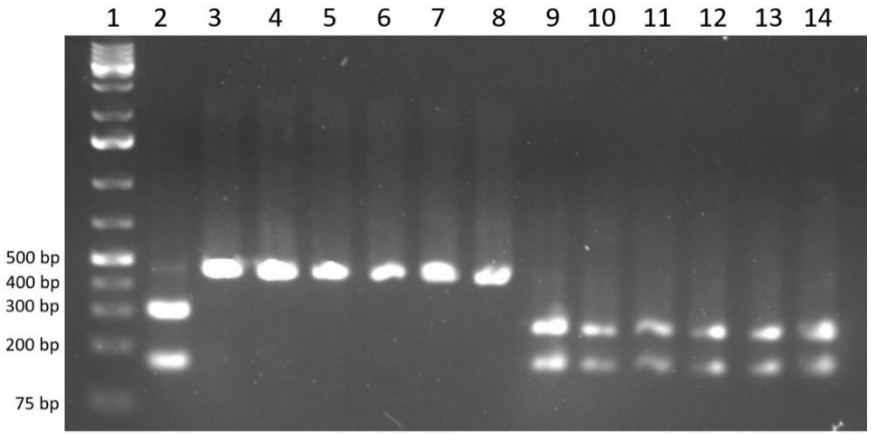


Figure 24 1 % agarose gel of 1 Kb Gene Ruler Thermo Scientific™ (lane 1), SmaI digested product of V22 (lane 2), SmaI digested products of V22ΔV2R→K+ of 6 clones (lanes 3 – 8), SalI digested of wildtype (WT) V22 clone (lane 9) SalI digested products of 5 clones of V22ΔV2T→S (lanes 10 -14).

3.2.4 Heights, symptom scoring and viral load of *N. benthamiana* inoculated with V22ΔV2 mutants

V22ΔV2-V→S, V22ΔV2-T→S and V22ΔV2-T→S+ resulted in no significant changes in the plant height relative to the control (Mock-inoculated plants) (%) in N. *benthamiana* when compared to *N. benthamiana* inoculated with WT V22 (Figure 25). The mock inoculum consisted of the empty pCambia2300 vector in *A. tumefaciens.*

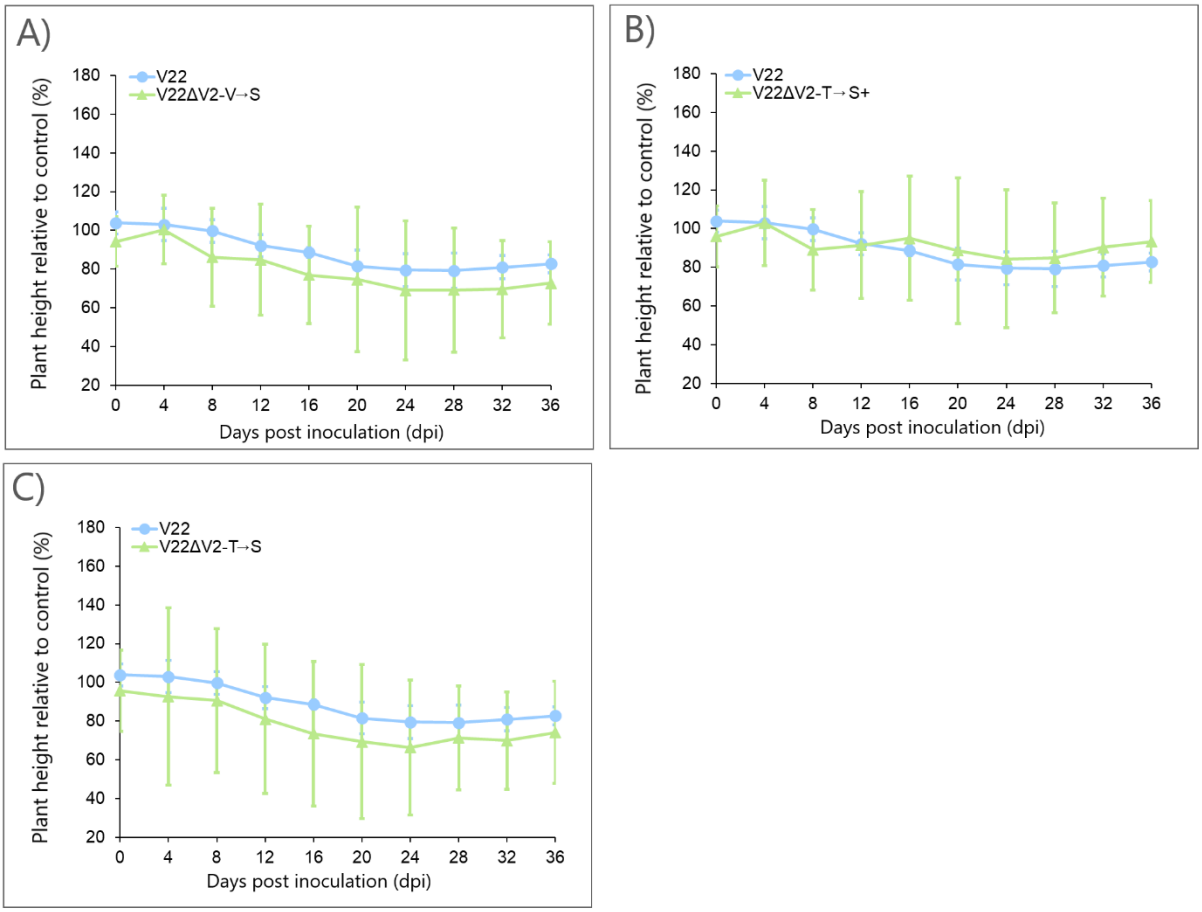


Figure 25 Plant height relative to control (mock) (%) over time measured in days post inoculation (dpi), of wildtype (WT) V22, A) V22ΔV2-V→2, B) V22ΔV2-T→S+ and C) V22ΔV2-T→S. Control was *Nicotiana benthamiana* inoculated with mock inoculum (empty pCambia2300 vector in *Agobacterium tumefaciens*).

With regards to the symptom severity scores obtained, V22ΔV2-V→S induced significantly greater ($p<0.05$) ULR and SV at 28 dpi onwards compared to V22 (Figure 26A and Figure 27). Symptom recovery in ULR was greatly reduced ($p<0.05$) compared to V22. V22ΔV2-T→S initially induced lower severity scores for SV and ULR in *N. benthamiana* at 16 to 24 dpi followed by the significant increase of SV and ULR severity scores at 28 dpi and onwards when compared to V22. This reveals that a delay in the onset of the symptoms was induced by

V22ΔV2-T→S (Figure 26 B and Figure 28). V22ΔV2-T→S+ inoculated plants resulted in the complete lack of ULR, and significantly decreased SV compared to the V22 (Figure 26 C and Figure 29). Figure 29 shows this decrease in ULR compared to WT V22 at 20 dpi. Figure 30 shows these symptom differences in the single leaf dissections of each respective abovementioned treatment. Progression of symptoms induced by V22ΔV2-V→S, V22ΔV2-T→S and V22ΔV2-T→S+ at 12 dpi to 32 dpi can be seen in Supplementary Figure 2.1, 2.2 and 2.3.

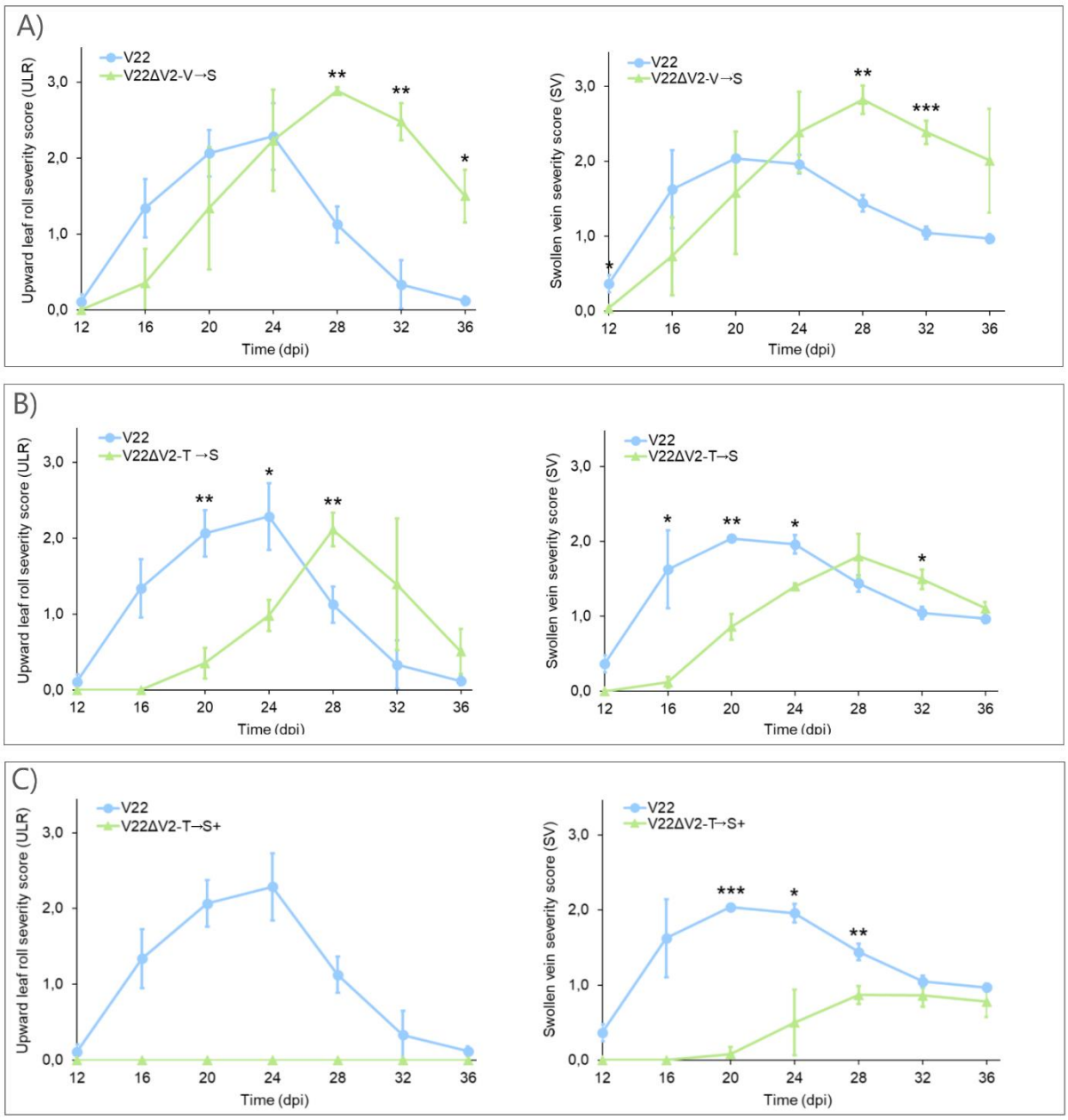


Figure 26 Upward leaf roll severity score (ULR) (left) and swollen vein severity score (SV) (right) induced in *Nicotiana benthamiana* for A) V22ΔV2-V→S, B) V22ΔV-T→S and C) V22ΔV-T→S+. Bars indicate mean ±

SD. Significance level: * = $p < 0.05$. The IUPAC-IUMBMB one-letter abbreviations for the 20 standard amino acids were used.

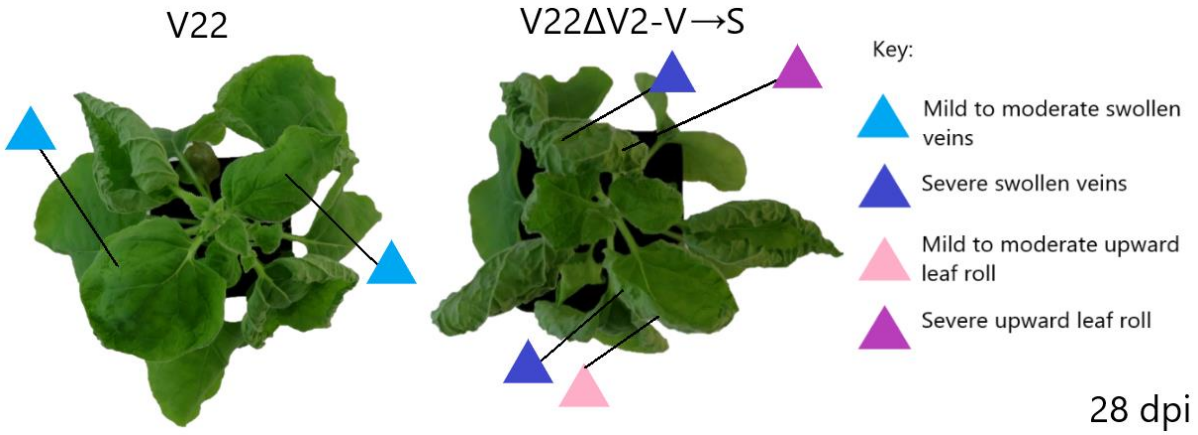


Figure 27 *Nicotiana benthamiana* inoculated with V22 (positive control) and V22ΔV2-V→S infectious clone constructs at 28 days post inoculation (dpi). Symptoms present are upward leaf roll (ULR) and swollen veins (SV). Severity of symptoms indicated by colour coded triangles as described in the key. Time point 28 dpi is depicted to show the delay in ULR recovery compared to V22.

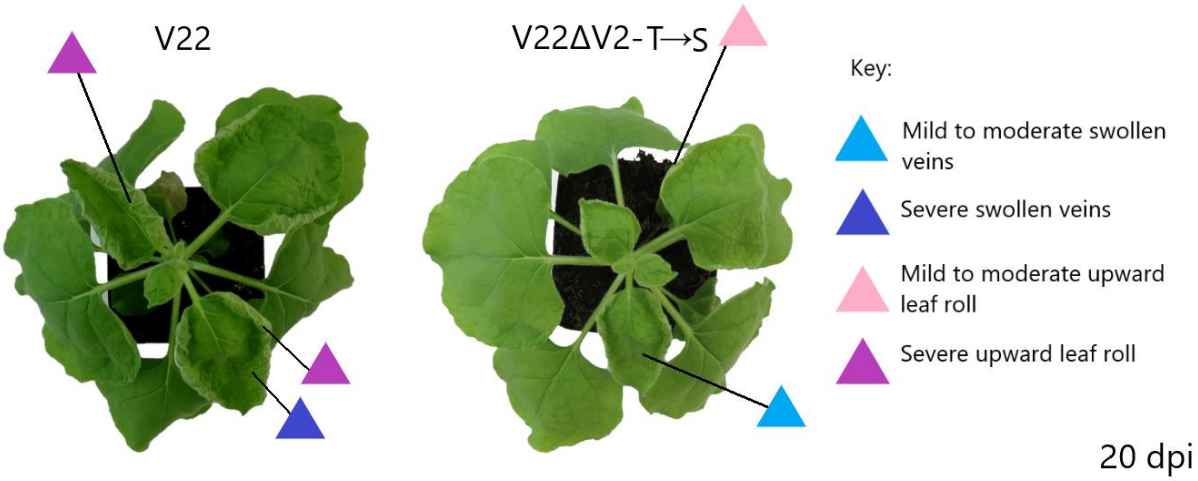


Figure 28 *Nicotiana benthamiana* inoculated with V22 (positive control) and V22ΔV2-T→S infectious clone constructs at 20 days post inoculation (dpi). Symptoms present are upward leaf roll (ULR) and swollen veins (SV). Severity of symptoms are indicated by colour coded triangles as described in the key. Time point 20 dpi shows clear reduction in symptom severity compared to V22.

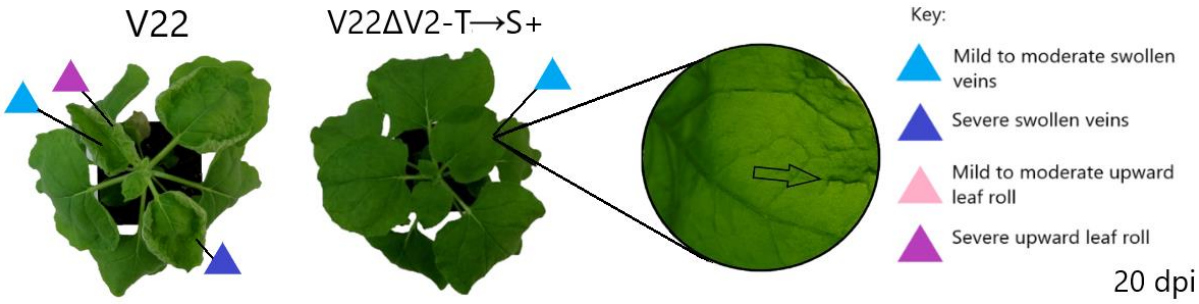


Figure 29 *Nicotiana benthamiana* inoculated with V22 (positive control) and V22ΔV2-T→S+ infectious clone constructs at 20 days post inoculation (dpi). The expanded leaves below the meristem have been shown through the light in the circle. Symptoms present are upward leaf roll (ULR) and swollen veins (SV). Severity of symptoms indicated by colour coded triangles as described in the key. The mild sunken veins observed at 20 dpi in V22ΔV2-T→S+ inoculated *N. benthamiana* can be seen in the close of the expanded leaf with the arrow indicating the mild sunken vein.

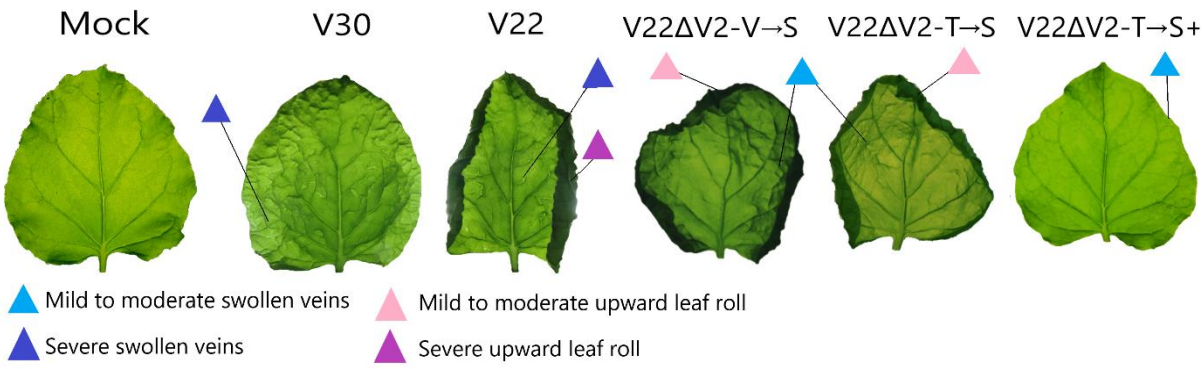


Figure 30 *Nicotiana benthamiana* leaves dissected from plants inoculated with pCambia2300 vector in *Agrobacterium tumefaciens* (Mock), V30 (positive control), V22 (positive control), V22ΔV2-V→S, V22ΔV2-T→S and V22ΔV2-T→S+ infectious clone constructs. Leaves were viewed against the light. Symptoms present are upward leaf roll (ULR) and swollen veins (SV). Severity of symptoms indicated by colour coded triangles as described in the key.

Log viral fold change of V22ΔV2-V→S was found to be significantly higher than that of V22 at 36 dpi (Figure 31). Log viral fold change at 20 dpi of V22ΔV2-V→S and that of V22ΔV2-T→S at 20 and 36 dpi revealed no significant changes when compared to V22.

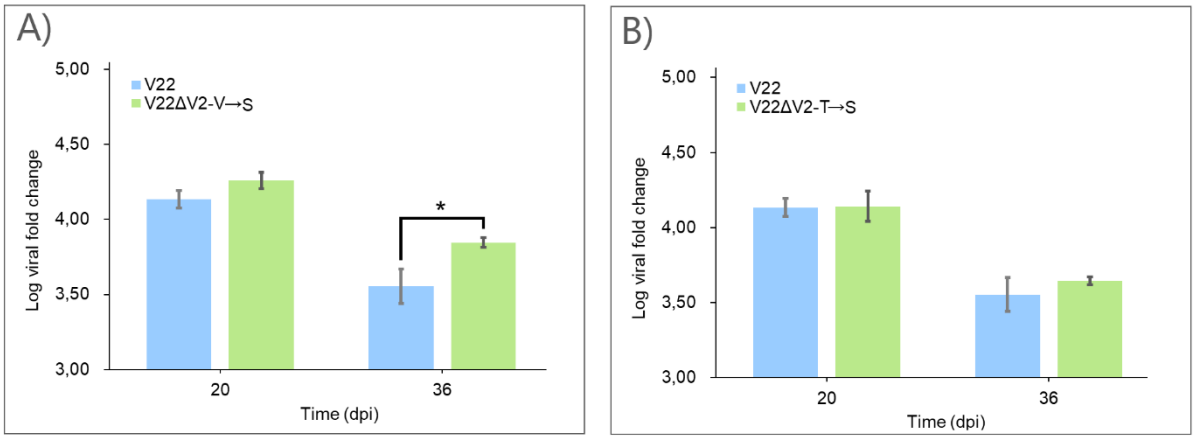


Figure 31 Log viral fold change of wildtype (WT) V22 and A) V22ΔV2-V→S and B) V22ΔV2-T→S in *Nicotiana benthamiana* at 20- and 36-days post inoculation (dpi). Bars indicate mean ± SD. Significance level: * = $p < 0.05$ (student T-test).

For V22ΔV2Del+, V22ΔV-S→L+ and V22ΔV2-R→K+, incidence scores (Supplementary Figure 2.8) were reported for the preliminary trial as full trials for these mutant clones were not carried out due to a lack of significant phenotypic changes in the preliminary trial, or these clones had too many inadvertent mutations to warrant full trials. Among these V2 mutant clones, V22ΔV2Del+ resulted in a clear decrease in ULR in *N. benthamiana* at 20, 28 and 36 dpi (Supplementary Figure 2.9).

3.3 Mixed infections with wildtype V30/ V22 and wildtype V30/ V30ΔIR-s clone constructs

Inoculation of *N. benthamiana* with the V30+V22 mixed infection and the V30+V30ΔIR-s yielded no significant changes to the plant height relative to the control (%) when compared to V30 or V22 individually infected plants (Figure 32 B). The control height was the height of *N. benthamiana* inoculated with *A. tumefaciens* containing the empty pCambia2300 vector. The log viral fold change of V30+V22 inoculated plants was found to be significantly lower than V30 but higher than that of V22 resulting in an intermediary viral load (Figure 32 A). This intermediary virus titre observed with this mixed infection is also reflected by the ULR and SV severity scores.

The SV and ULR severity scores of the V30+V22 inoculated plants resulted in intermediate scores between the scores of V30 and V22 inoculated plants whereby SV and ULR was decreased in V30+V22 compared to the WT V30 and V22 infections (Figure 32 C and Figure

32 D). V30+V22 induced downward cupping and bulging in the expanding leaves of the plant at 20 dpi which was not apparent in plants inoculated with either V30 or V22 (Figure 34). The V30+V30ΔIR-s mixed infection resulted in ULR milder than that of V22. Furthermore, SV induced were more severe than the SV induced by V30 or V22 (Figure 33). ULR differences are clearly seen in *N. benthamiana* leaf images at 20 dpi (Figure 35). Viral load was not determined for the V30+V30ΔIR-s infection.

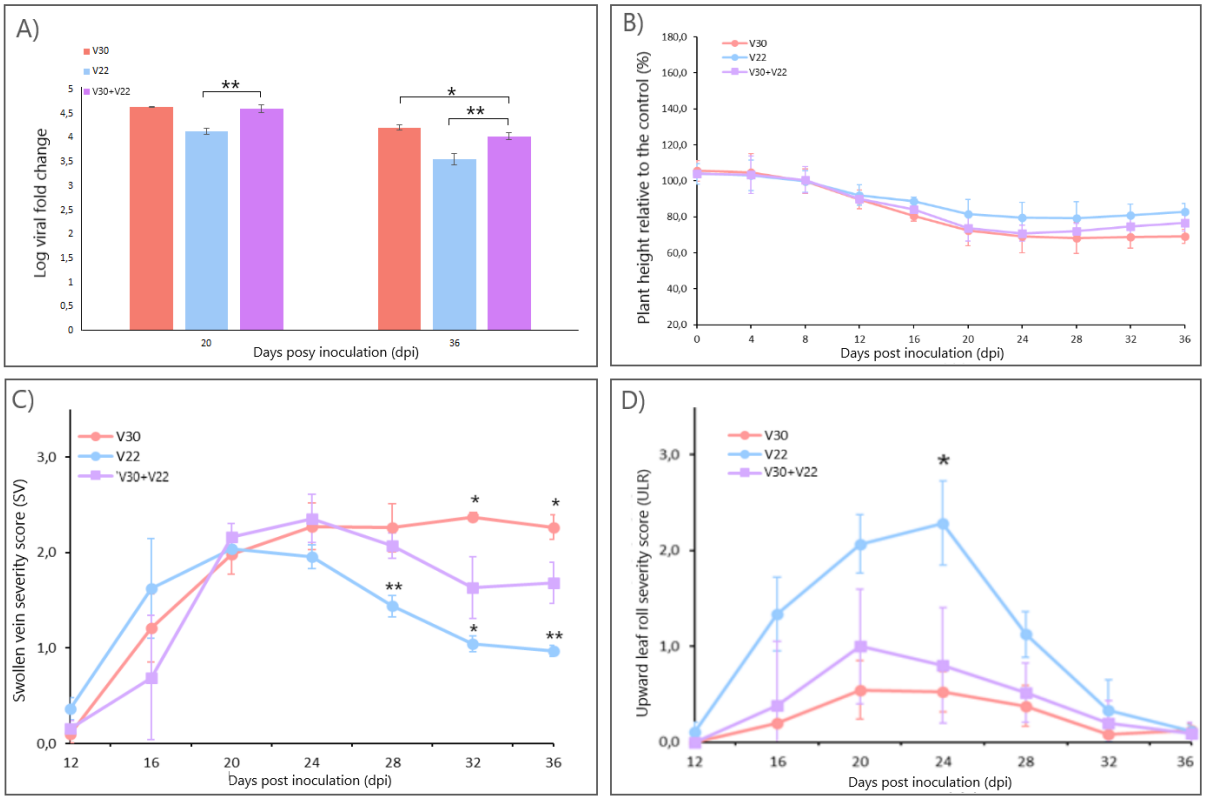


Figure 32 A) Log viral load fold change of V30, V22 and the V30+V22 (1:1 inoculum) mixed infection at 20 and 36 dpi in *Nicotiana benthamiana*. Bars indicate mean ± SD. Significance level: * = $p < 0.05$ (student T-test). B) Plant heigh relative to control (mock: empty pCambia2300 vector in *Agrobacterium tumefaciens*) (%) C) Swollen vein severity score (SV) observed in the leaves of *N. benthamiana* inoculated with V30, V22 and V30+V22. D) Upward leaf roll severity score (ULR) induced in the leaves of *N. benthamiana* inoculated with V30, V22 and V30+V22. Bars indicate mean ± SD. Significance level: * = $p < 0.05$. **=P<0.005 (student T-test). * Shows significant difference of V30+V22 to V30 or V22 as indicated above V30 or V22.

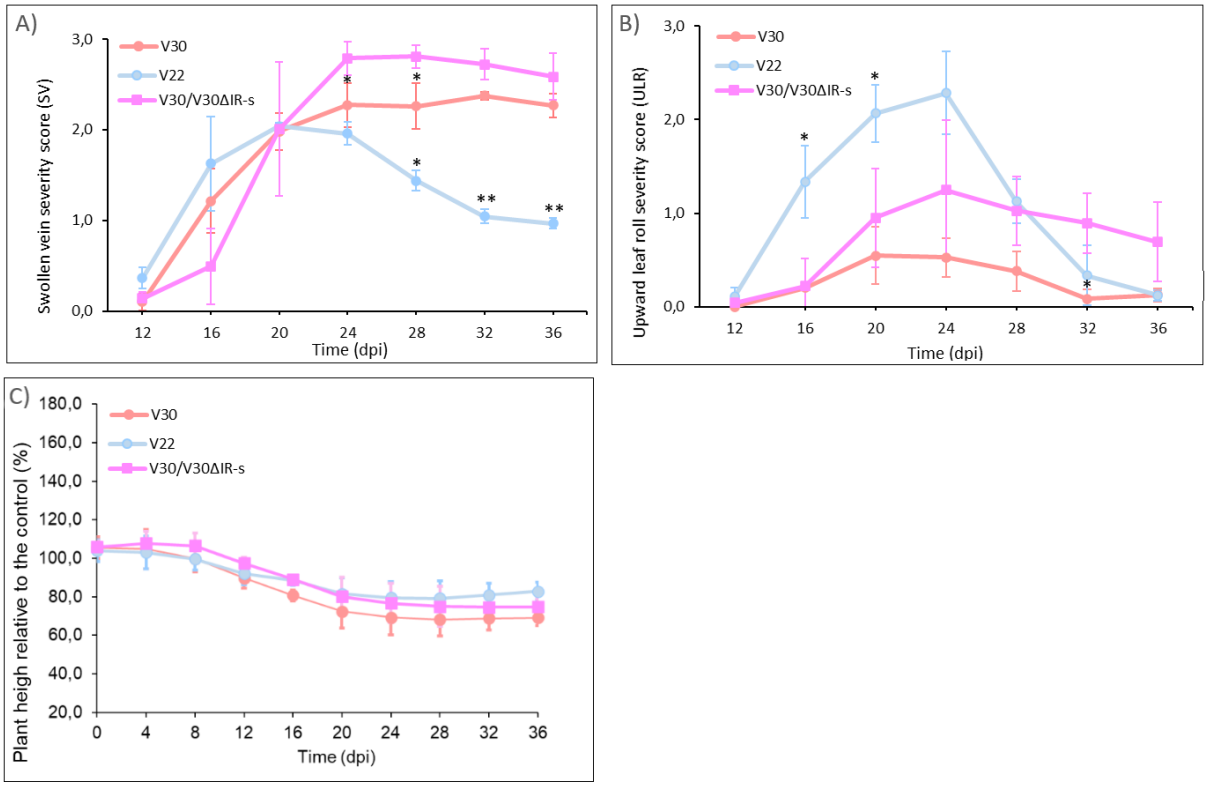


Figure 33 A) Swollen vein severity score (SV) observed in the leaves of *Nicotiana benthamiana* inoculated with V30, V22 and V30+V30ΔIR-s (1:1 inoculum). B) Upward leaf roll severity score (ULR) induced in the leaves of *N. benthamiana* inoculated with V30, V22 and V30+V30ΔIR-s. C) Height relative to mock (%) for *N. benthamiana* inoculated with V30, V22 and V30+V30ΔIR-s. Bars indicate mean ± SD. Significance level: * = $p < 0.05$, *$p<0.005$. * Shows significant difference of V30+V30ΔIR-s to V30 or V22 as indicated above V30 or V22.

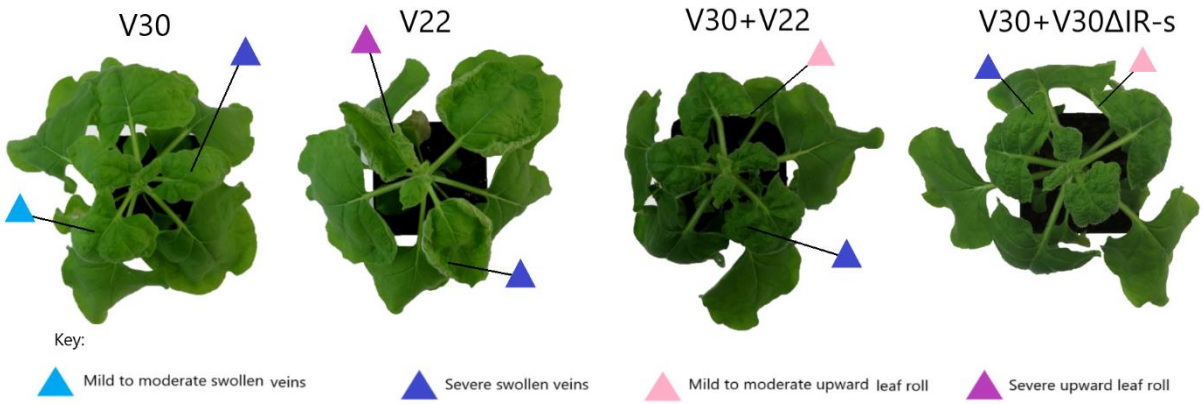


Figure 34 *Nicotiana benthamiana* inoculated with V30, V22 (positive controls) and V30+V22 (1:1 inoculum) and V30+V30ΔIR-s (1:1 inoculum) mixed infectious clone constructs at 20 days post inoculation (dpi). Symptoms present are upward leaf roll (ULR) and swollen veins (SV). Severity of symptoms indicated by colour coded triangles as described by the key.

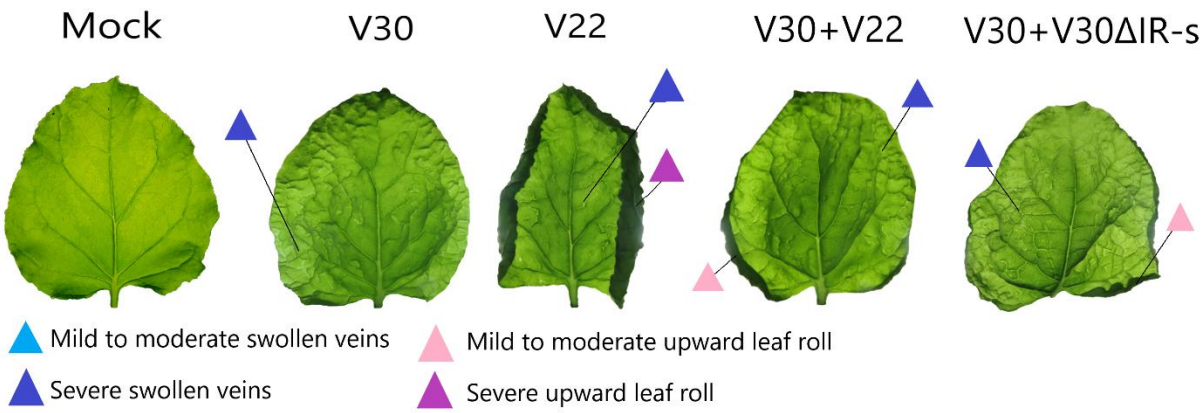


Figure 35 *Nicotiana benthamiana* leaves dissected from plants inoculated with pCambia2300 vector in *A. tumefaciens* (Mock), V30 (positive control), V22 (positive control), and a mixed infection of V30 + V22 (V30+V22) and V30 + V30ΔIR-s (V30+V30ΔIR-s) infectious clone constructs. Leaves were viewed against the light. Symptoms present are upward leaf roll (ULR) and swollen veins (SV). Severity of symptoms indicated by colour coded triangles as described by the key.

3.4 Investigation of the putative C5 ORF

3.4.1 Bioinformatics analysis of the putative C5 protein

A comparison of the predicted secondary structure (ss) of the C5 protein of African cassava mosaic virus (ACMV) and the N-terminally extended putative C5 protein of V22ΔC5$^{1\text{-}193}$ was done to see if there were similarities between the structure of begomovirus C5 and V22ΔC5$^{1\text{-}193}$. V22ΔC5$^{1\text{-}193}$ and ACMV both had seven predicted α-helices and four β-sheets. Two differences in disorder were predicted for these C5 proteins in the N-terminus and C-terminus of the protein (Figure 36).



Figure 36 C5 protein sequences of African cassava mosaic virus (ACMV, Accession number J02057) and V22ΔC5$^{1-193.}$ A) Secondary structure prediction with confidence levels indicated using a colour key. Consensus α-helices and β-strands numbered. B) Predicted C5 protein regions of disorder with disordered residues having protein- binding potential highlighted in green. The threshold value for disordered residues was 0.5. Disorder and secondary structural predictions obtained using PSIPRED web server (http://bioinf.cs.ucl.ac.uk/psipred/).

3.4.2 Infectivity trials of V22 C5 ORF mutants in *N. benthamiana*

A preliminary trial of the V22 C5 deletion mutant, V22XC5, resulted in no phenotypic or height changes compared to V22, therefore no further trials were carried out using this infectious

clone construct. V22ΔC5aa$^{1-193}$ showed a significant decrease ($p<0.05$) in *N. benthamiana* height between 4 dpi to 8 dpi when compared to V22 plant height (Figure 37).

A CLUSTAL multiple sequence alignment (https://www.ebi.ac.uk/Tools/msa/clustalo/) of the two putative C5 protein sequences, C5aa$^{1-193}$ protein encoded by the V22ΔC5aa$^{1-193}$ mutant and the C5 protein of ACMV (Supplementary Figure 3.1 A and B) revealed a high sequence conservation between the two nucleotide sequences.

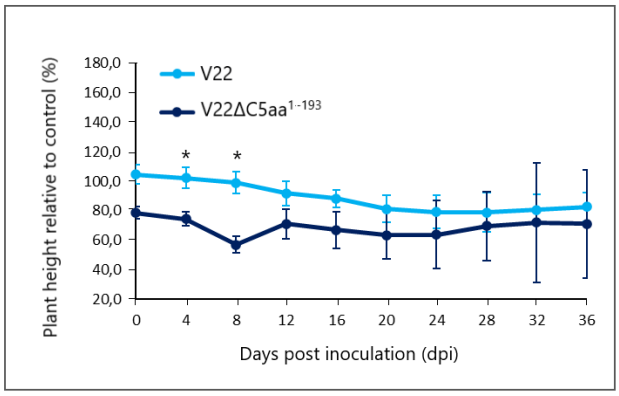


Figure 37 Plant height relative to the control (mock) (%) over time (dpi) for *Nicotiana benthamiana* inoculated with V22 and V22ΔC5aa$^{1-193}$. Plant height relative to control calculated using the formula (H_i/H_m*100) where H_i is the height of inoculated plants, H_m is the height of mock plants. Formula obtained from Lapidot et al., 2006. Bars indicate mean ± SD. Significance level: * = $p < 0.05$ (Student T-test).

The ULR and SV scores reported for V22ΔC5aa$^{1-193}$ initially induced milder SV and ULR in *N. benthamiana* compared to V22 (Figure 38 and Figure 39). The milder symptoms initially induced were seen at 20 dpi whereas more severe symptoms were induced at 32 dpi compared to V22 (Figure 39 and 40). The peak severity for both V22 and V22ΔC5aa$^{1-193}$ were similar but the time point of the peak varied. The ULR and SV symptom severity peak for V22 was 20 dpi to 24 dpi whereas these severity peaks are at time points 28 dpi to 32 dpi for V22ΔC5aa$^{1-193}$. This shows a delay in the symptom trend compared to V22 (Supplementary Figure 1.6 and 3.4).

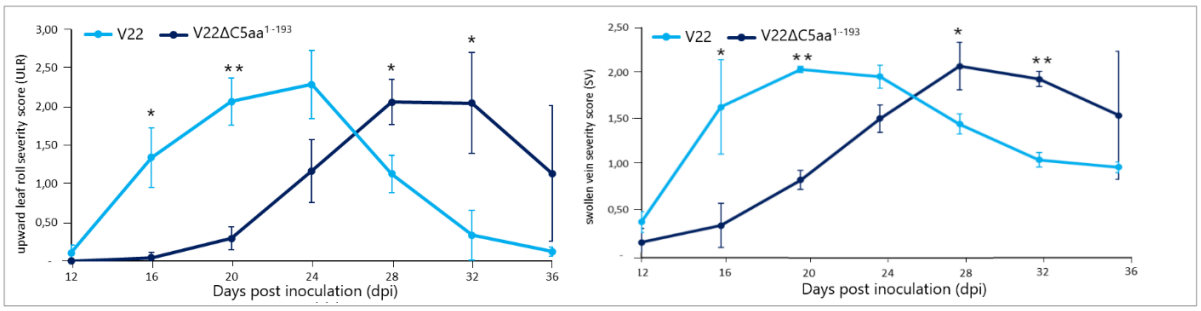


Figure 38 Upward leaf roll severity (ULR) and swollen vein severity (SV) scores over time (dpi) for *Nicotiana benthamiana* inoculated with V22ΔC5aa$^{1-193}$. Bars indicate mean ± SD. Significance level: * = $p < 0.05$ **=P<0.005 (Student T-test).

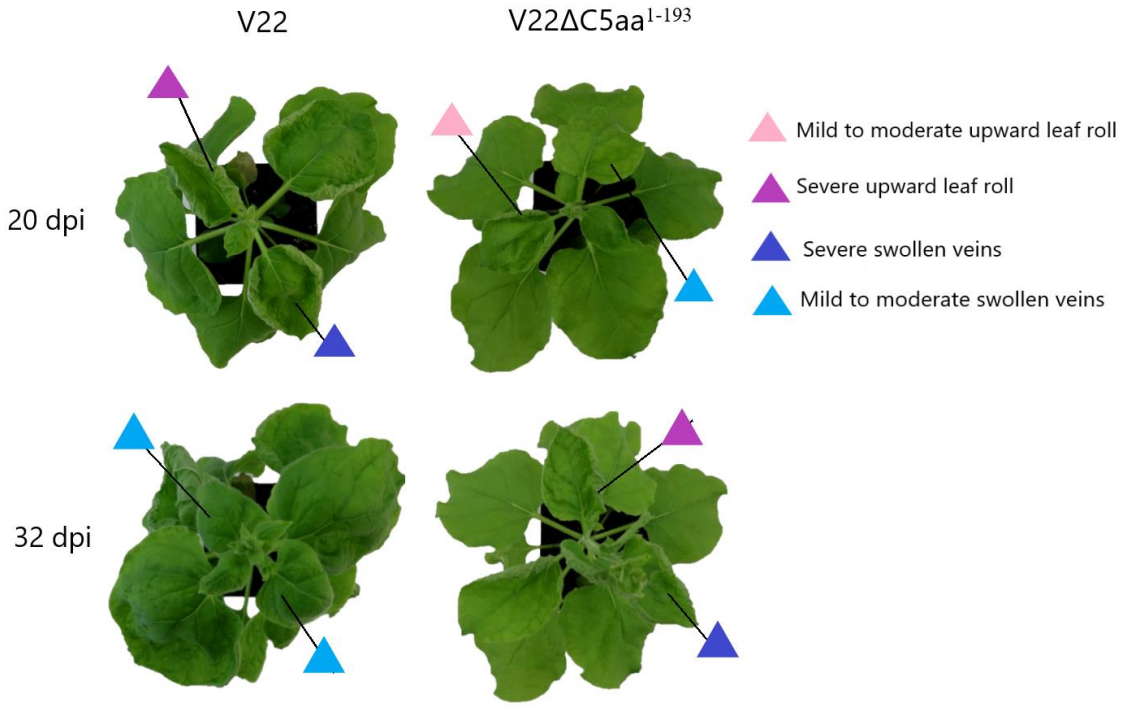


Figure 39 *Nicotiana benthamiana* inoculated with V22 (positive control) and V22ΔC5aa$^{1-193}$ infectious clone constructs at 20- and 32-days post inoculation (dpi). Symptoms present are upward leaf roll (ULR) and swollen veins (SV). Severity of symptoms indicated by colour coded triangles as described by the key. The time points were chosen to show the delayed onset of ULR and SV at 20 dpi induced by the V22 C5 infectious clones and the increased severity of ULR and SV at 32 dpi induced by the V22 C5 mutants when compared to V22.

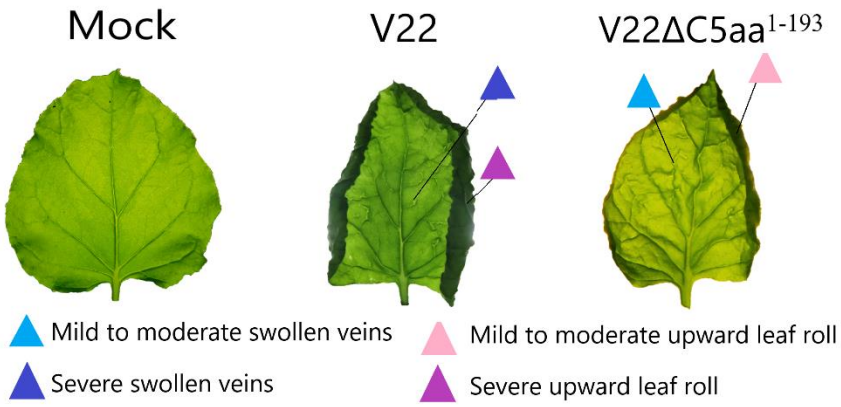


Figure 40 *Nicotiana benthamiana* leaves dissected from plants inoculated with pCambia2300 vector in *Agrobacterium tumefaciens* (Mock), V30 (positive control), V22 (positive control) and V22ΔC5aa$^{1-193}$. Leaves at 20 days post inoculation (dpi) were viewed against the light. Symptoms present are upward leaf roll (ULR) and swollen veins (SV). Severity of symptoms indicated by colour coded triangles as described by the key.

CHAPTER 4 - DISCUSSION

4.1 The intergenic region (IR)

In this study, a region in the 3' IR of V22 was identified to play a vital role in the symptom phenotype induced by V22 in *N. benthamiana*. This is evidenced by the significant decline in upward leaf roll (ULR), and swollen veins (SV) induced by V22ΔIR-s and V22ΔIR-sr when compared to wildtype (WT) V22. The results obtained for the V22ΔIR-s infectious clone construct therefore led to the conclusion that nt 61-137 in the IR of V22 is critical in symptom phenotype development, specifically ULR and SV development in *N. benthamiana*. Furthermore, V22ΔIR-sr revealed that nt 108 to 137 in the V22 IR also induced significant symptom changes such as a decrease in ULR and SV scores. Supporting the findings obtained in this study are the results obtained from trials carried out using the reverse infectious clone constructs (Zwolinski, 2022). These reverse infectious clone constructs, V30ΔIR-s and V30ΔIR-sr, were constructed using V30 as the backbone containing the respective V22 IR segments. V30ΔIR-s and V30ΔIR-sr induced ULR in *N. benthamiana* which is not observed in V30 (Zwolinkski, unpublished). These results, in addition to the results of this study, reveal that that the IR, specifically nt 108 to 137 within the IR, play a crucial role in the symptom phenotypes induced by both V30 and V22 in the experimental host plant *N. benthamiana*.

The IR has not only been known to be of significance due to it containing the Rep binding site and the ORI, but more recent studies (Yang et al., 2019) have shown it to modulate disease symptoms. Essential genetic regions such as repeated sequence elements and conserved late elements (CLE) are present in the IR which have previously been found to be associated with disease symptom development in geminiviruses (Yang et al., 2019). A study by Velten et al., (2005) found repeated sequence elements within the IR of geminiviruses to have transcriptional enhancing activity, which would make these sequences regulatory sequence elements. Furthermore, CLEs were identified within these repeats. The CLE, originally identified by comparative genome analysis of geminiviruses in 1994 (Argüello-Astorga et al., 1994), is a consensus sequence found to enhance transcription. Furthermore, these sequences have been found to be located close to the TATA box in geminiviruses hence denoting the importance of the CLE in transcription initiation (Cazzonelli et al., 2005). This study supports previous IR study results, showing that the IR plays a crucial role in geminiviral infection. This emphasizes the importance of further investigating the IR of ToCSV and its effect on viral load and symptom phenotype.

To explain the phenotypic changes induced by these IR-swap infectious clones the mechanisms at play need to be explored. Mechanisms such as RNA silencing and the inhibition of RNA silencing by viral encoded RNA silencing suppressor proteins could influence activity and symptom development (Wang et al., 2012). Regulation or interference of host gene transcription by IR-derived RNAs could influence symptom severity and phenotype. Geminiviral viral small-interfering RNAs (vsRNAs), derived from a 25-nt segment within the IR of Tomato yellow leaf curl virus (TYLCV) were identified in infected tomato (Yang et al., 2019). These vsRNAs were shown to target and silence a long noncoding RNA (lncRNA) found in the susceptible cultivar of tomato which was demonstrated to be linked to the plant height stunting and leaf curl phenotype (Yang et al., 2019).

Taking previous studies which link the IR to pathogenicity of begomoviruses into account, the V22 was then probed for a 3’ IR to explore whether the significant phenotypic changes induced by the swapping of the IR between V30 and V22 as seen in this study could be linked to the presence of an RNA transcript. An RNA transcript was found to be present in both the complementary and viral sense which extends over the origin of replication (ORI) in the IR. This shows that a continuous RNA transcript is made across this region. As the IR in this study was proven to be bidirectionally transcribed, this suggests that the transcribed IR may form double- stranded RNA (dsRNA) and hence act as a template which is processed by post-

transcriptional gene silencing (PTGS) to generate vsRNAs. Double stranded RNA (dsRNA) serves as a strong inducer for PTGS which involves guided degradation of viral and cellular RNA. Since ToCSV is a DNA virus, it triggers RNA silencing by various mechanisms including overlapping RNAs derived from transcription of overlapping ORFs which are then filled by RNA-dependent RNA polymerase 6 (RDR6). The 24 nt vsRNAs can also trigger TGS in the nucleus where geminivirus DNA replication takes place, and target host mRNA (Vaucheret, 2006). To counter the antiviral RNA mechanism induced by the plant, vsRNA generated from the IR of ToCSV may target expression of *N. benthamiana* genes, thereby altering regulation of these genes, leading to physiological disturbances and symptoms. This was indeed observed with TYLCV (Yang et al., 2019) where the study showed that bidirectional transcription of the IR of TYLCV contributed to the symptoms induced in tomato. An IR transcript was also detected in the severe ToCSV-V30 variant (Zwolinski, 2022).

The siRNAs from the IR of V30 and V22 were then predicted using the siRNA predictor tool (www.plantgrn.noble.org) (Supplementary Figure 1.1 – 1.4) but as these siRNAs had no matches to the experimental *N. benthamiana* and host *S. lycopersicum* plant genomes, no conclusions for the siRNAs generated for this region could be made in terms of siRNA interference linked to disease symptoms. siRNA predictors are not suitable for this purpose of identifying siRNAs targeting host genomes for phenotype regulation because the predictors filter out siRNAs that match host targets with high percentages. Without RNA deep sequencing data, we cannot confidently identify potential host gene targets for vsRNA-mediated silencing using the IR sequences alone. Due to these shortcomings, RNA regulation by the IR in ToCSV can be speculated but cannot be concluded without further studies. This study compliments other studies which have identified RNA transcripts in geminiviruses and further investigation for the production of siRNA or vsRNA to elucidate the possible function of this transcript in virus-induced silencing or in host gene regulation is highly encouraged (Poogin et al., 2013 and Yang et al., 2019). One way of further investigating the role of the IR in RNA regulation may be through RNA sequencing as this could identify vsRNAs targeting host RNA silencing associated with basal immunity.

The pathogenicity of V22ΔIR-s was then further explored and viral DNA load at 20 and 28 dpi was determined to have no significant difference compared to V22. Again, this points more to a pathogenicity role for IR in RNA-associated processes or mechanisms. The results obtained in this study revealed a relationship between the IR in ToCSV and the ULR and SV symptom phenotypes induced in *N. benthamiana,* but the IR region cannot be linked to ToCSV DNA

replication or height changes related to metabolic disturbances from the results obtained in this study. Further research should be carried out in the natural tomato host and compared to the results obtained for the experimental host plant used in this study.

4.2 The V2 ORF

The V2 protein of begomoviruses has been of interest in multiple studies as this protein has been linked to viral movement, gene silencing suppression, symptom phenotype induction and the hypersensitive response (HR) (Sharma and Ikegami, 2010; Zhang *et al.*, 2012 and Padidam *et al.*, 1996). The V2 protein has also been labelled a pathogenicity determinant as it has been found to induce symptoms in plants when expressed in a PVX vector (Amin *et al.*, 2011). With so many studies emphasizing the importance of the V2 with regards to the symptoms induced in their respective host plants, investigating the aa sequence differences, between the V2 protein of V30 and V22 to identify regions responsible for the different phenotypes induced by these variants proved to be interesting. In this study, the V22 variant was chosen for the V2 aa substitution mutants as V22 induces prominent ULR in *N. benthamiana* with a clear recovery evident at 32 dpi onwards. These obvious symptom phenotypes allowed for any changes to ULR or recovery in later stages to be clearly observed whereas V30 induces only mild ULR, and recovery is not as noticeable. By substituting single aa in V22 with that of V30 in the V22 V2 substitution mutants, the effect of single aa mutations to the V2 protein of V22 was investigated in terms of its influence on symptom severity and recovery observed in *N. benthamiana*.

Amongst the single V2 aa substitution mutant clones constructed, V22ΔV2-V→S and V22ΔV2-T→S had no inadvertent mutations affecting overlapping ORFs whereas V22ΔV2-T→S+ had an inadvertent mutation to the overlapping V1 ORF. Two aa changes to the V1 protein of V22ΔV2-T→S+ were incurred, a proline to aspartic acid change at aa 5 (P5D) and an alanine to aspartic acid change at aa 6 (A6D). V22ΔV2-V→S showed intriguing results with an increase in ULR, and SV severity noticed at 28 dpi onwards. The exacerbated symptom severity was accompanied by a significant increase in viral load at 36 dpi when compared to V22. In addition, the recovery of ULR at 32 and 36 dpi was greatly impaired for these mutant infected plants when compared to V22 infection. It is proposed that this increased symptom severity and viral load with reduced capabilities of recovery induced by V22ΔV2-V→S demonstrates that substituting valine in V22 to serine as found in V30 at aa 27 reveals an

association of valine in the V2 of V22 with ULR severity and recovery in *N. benthamiana*. However, the effect of this mutation remains to be shown in the natural host, tomato. V22ΔV2-V→S further suggests an association between the V22 V2 protein and symptom remission observed in *N. benthamiana*. A link between the interaction of the V2 protein with host RNA-dependent RNA polymerase and symptom remission in tomato has been previously shown in Tomato leaf curl Gujarat virus (ToLCGV) (Basu et al., 2018).

The results obtained from this study further revealed that by substituting threonine in V22 with that of serine as found in V30 at aa 58 in V22ΔV2-T→S, induced altered symptom severity. V22ΔV2-T→S induced milder symptoms initially followed by more severe symptoms at 28 dpi onwards when compared to V22. Unlike the viral load of V22ΔV2-V→S, V22ΔV2-T→S infection displayed no changes to viral load revealing that viral load is not associated with symptom development or severity in this case.

The substituted serine, a polar aa, the mutation introduced in both mutant clones is located in the predicted helices in the V22 V2 protein. Alpha-helices are the most common secondary structures (ss) to occur in proteins and they play a role in determining the function and global structure of the protein. (Haimov and Srebnik, 2016). In V22ΔV2-V→S, valine, a non-polar aa with a hydrophobic chain was substituted with serine, a non-bulky polar aa. Therefore, the change from valine to serine at aa 27 may have influenced the ss of their domains which would influence protein folding and its function ultimately resulting in the phenotypic and viral load changes seen. This aa change may have changed the protein binding properties or binding partners of the V2 protein thereby altering the symptoms induced by V22. A previous study identified a serine at position 71 in the V2 protein of geminiviruses to be involved in the pathogenicity of the virus (Zhao *et al.*, 2018), thereby placing importance on single residues in the interaction of viral proteins in host plants. The phenotypic changes observed in *N. benthmaiana* in this study cannot be attributed to any disorder changes in the mutated V2 protein compared to V22 as no changes were predicted by PSIPRED protein prediction server. Differences in symptoms and viral replication observed between the mutant and WT V2 proteins may therefore be due to the polar nature of the introduced aa.

The same T to S substitution as in V22ΔV2-T→S is displayed in V22ΔV2-T→S+ but V22ΔV2-T→S+ had an unintentional, inadvertent change to the overlapping V1 ORF resulting in two aa changes in the V1 protein of V22ΔV2-T→S+, P5D and A6D. A90S and G91T were also inadvertently introduced into the overlapping putative C6 in V22. These inadvertent

mutations were a result of the codon change introduced during mutagenesis which resulted in different aa encoded for by the overlapping ORFs. Ensuring that the codon change introduced results in the same aa to be encoded for by any overlapping ORFs is important as to avoid inadvertent aa changes in future studies. Although these inadvertent changes were unintentional, V22ΔV2-T→S+ yielded very interesting results and this mutant could be compared to the results obtained from V22ΔV2-T→S. V22ΔV2-T→S+ surprisingly resulted in a drastic reduction in symptoms induced as only mild SV were observed and ULR was abolished in *N. benthamiana*. This differs to the increased severity induced by V22ΔV2-T→S which suggests that the attenuation of symptoms observed in V22ΔV2-T→S+ infected *N. benthamiana* is a result of the aa changes to the V1 of V22. Although the putative C6 was changed, this ORF has not been confirmed and the secondary analysis of C6 of V22 and V22ΔV2-T→S+ revealed that the disorder of the putative C6 of V22ΔV2-T→S+ did not vary from that of V22. Therefore, these attenuated symptoms may be explained by the aa changes to the V1 protein and not to the aa changes in the putative C6. A change to the disorder of the N-terminal residues of the V1 protein of V22ΔV2-T→S+ relative to V22 was predicted. The change in the V1 protein structure and disorder may have altered the protein binding function and hence the interaction of V1 with other proteins. Altering the protein binding abilities of V1 could have led to lower infection ability possibly resulting in very mild symptoms as the V1 has previously been found to play a vital role in viral DNA transport to the nucleus allowing for replication and infection to take place (Liu et al., 1997). The V1 is also involved in cell-to-cell virus movement in monopartite geminiviruses which contributes to systemic infection (Unseld et al., 2001) and has previously been found to play a crucial part in viral DNA nuclear transport allowing for infection to occur (Liu et al., 1997). Further investigation of V22ΔV2-T→S+ viral load would have to be carried out to determine changes in infectivity.

Previous studies have found that by altering the coat protein of maize streak virus (MSV) and TYLCV a loss of infectivity resulted in the absence of or declined symptom phenotype was observed (Liu et al., 1997 and Wartig et al., 1997), which is similar to what was observed in this study with ToCSV. Upon further investigation, it was found that in V22ΔV2-T→S +, V1 was missing a nuclear localization signal (NLS) when compared to V22 V1. This reinforces the hypothesis that the reduction in symptoms observed is correlated to the P5D and A6D mutations in V1 which could have led to the loss of the NLS function. NLS are short peptides which allows for the transportation of proteins, through the nuclear pore complex and into the nucleus. In viruses, the NLS in the capsid protein, serves to guide the virus particle through the

nuclear pore allowing for infection to take place (Macara, 2001). The importance of the NLS in other plant viruses has also been shown in previous studies where mutations targeting the NLS in potyviruses resulted in the prevention of nucleolar localization which either prevented or lowered viral replication (Rajamäki and Valkonen, 2009). The V22ΔV2-T→S mutant with the same V2 aa mutation without the overlapping V1 inadvertent change still induced symptoms in *N. benthamiana*, therefore adenine and aspartic acid at aa positions 5 and 6 in the V1 protein of V22ΔV2-T→S + are suggested to play a critical role in the infectivity of V22.

This study not only identifies two key aa residues at position 27 and 58 in the V2 protein in V22 which plays a role in pathogenicity and symptom remission, it has also identified two aa residues at position 5 and 6 in the V1 protein which are essential for infectivity of V22. This study also emphasizes the importance of taking inadvertent changes to overlapping ORFs of your target nucleotide sequence into account when making nucleotide or amino acid mutations or substitutions.

4.3 Mixed infections

Although the majority of studies look at single viral infections in plants, mixed infections of different viral species, strains, or with other microbes are more likely to take place in the field (Pita et al., 2015). Due to this, viral evolution and therefore genetic diversity occurs through recombination and horizontal gene transfer. This promotes genetic drift in a viral population allowing for a change in host range, infectivity, or pathogenicity (Pita et al., 2015). Furthermore, mixed infections have been found to directly influence virulence and symptom phenotypes induced (Alizon et al., 2013).

Inoculating *N. benthamiana* with a 1:1 ratio of WT V30 and WT V22 in this study produced an interesting outcome whereby both the symptom scores, ULR and SV, and the total viral load resulted in an intermediary result. Viral load induced by the mixed infection was greater than that of V22 but lower than that of V30 at both time points tested. Unfortunately, conclusions pertaining specifically to the viral load of either V30 or V22 in this V30+V22 mixed infection cannot be made as degenerate qPCR primers targeting both these variants were used. Due to the similarity of the variants, specific primers to differentiate between such similar variants could not be designed. However, as the total viral load was higher than V22 but lower than V30, we can hypothesize that there was either a complementary affect boosting V22 viral load

or an antagonistic affect lowering V30 viral load. Further studies on individual V30 and V22 viral load within the V30+V22 mixed infection would have to be done to determine the relationship or interaction between these two variants. Future studies could differentiate between these viral accumulations of the different variants through the addition of reporter genes to the different variants (Rupar et al., 2015). Another method which could be implemented to differentiate individual viral levels is circular DNA enrichment sequencing (CIDER-Seq) which has recently been implemented in a genetic diversity study of cotton leaf curl disease (CLCuD) (Ahmed et al., 2021). This method allows for the profiling of diverse circular DNA elements (Mehta et al., 2018).

A previous study of mixed infections of two begomoviruses, African cassava mosaic virus (ACMV) and East African cassava mosaic Cameroon virus (EACMV), reported synergism between the viruses whereby viral accumulation and symptom severity were considerably increased in cassava and *N. benthamiana* plants. This mixed infection was proposed to play a role in the suppression of PTGS therefore allowing for an increase in symptom severity (Vanitharani et al., 2004). A synergistic interaction between a mixed infection of two geminiviruses in pepper plants was also previously reported (Rentería-Canett *et al.*, 2011). Although multiple studies reveal synergism between geminiviruses, mixed infections are complex and negative interference as well as synergism have been reported to occur in the same mixed infection but at different time periods. A previous study revealed that two tomato-infecting geminiviruses, Tomato rugose mosaic virus (ToRMV) and tomato yellow spot virus (ToYSV), displayed negative interference in the beginning of infection but synergism occurred once systemic infection was reached. It was established that ToYSV releases the phloem restricted ToRMV allowing for more efficient infection and an increased viral load of ToRMV at 28 dpi. Interestingly, this study also found that interactions between geminiviruses are dependent on the host as only negative interference occurred between ToRMV and ToYSV in the tomato host. However, both negative interference and synergism were observed with the same viruses upon infection of the experimental host *N. benthamiana* (Alves-Júnior et al., 2009). This therefore shows that further studies of the V30+V22 mixed infection should be done in the natural tomato host and compared to the results obtained for *N. benthamiana* as seen in this study.

Symptom recovery was observed in *N. benthamiana* inoculated with V22 only whereas recovery was less evident in those infected with V30 only. Since recovery of symptoms in the V30+V22 co-infection was noted, this confirms that recovery seen in single V22 infections is

dominant over V30 in a V30+V22 mixed infection. In this study, V22ΔV2-V→S linked the V2 protein, specifically valine at aa position 27 to symptom recovery as the substitution of the valine to serine resulted in increased symptom severity and viral load with reduced capabilities of recovery. Therefore, since this protein was suggested to play a role in symptom remission, the V22 V2 may be outcompeting that of V30 to allow for recovery to persist in a mixed infection. An additional novel symptom phenotype was observed in the V30+V22 mixed infection in this study. This novel phenotype was prominent downward cupping and bulging of the expanding leaves. Mixed infections have previously been reported to change the symptom phenotype induced in the plant as well as having an effect on viral accumulation, virulence, and host range (Alizon, et al., 2013). A previous study revealed that a dual infection of two begomoviruses, Tomato rugose mosaic virus (ToRMV) and Tomato yellow spot virus (ToYSV), induced a unique leaf roll symptom in tomato that was not observed in the plants inoculated by each virus alone (Alves-Júnior et al., 2009). As the V30 and V22 variants are highly similar and their sequence differences lie mostly within the V2 ORF and the IR, it can be hypothesized that the V2 proteins of the two variants may complement or compete with each other in a mixed infection or compete for host factors thereby modulating disease symptoms. Direct or indirect interaction between proteins of viruses in a mixed infection has been previously linked to antagonistic interactions within a mixed infection resulting in a phenotypical change (DaPalma, et al., 2010). It can furthermore be suggested that the IR between the two variants may play a role in the novel phenotype induced by the V30+V22 mixed infection as it has been concluded in this current study that the IR is linked to the symptom phenotype induced by either V30 or V22 in *N. benthamiana*. It has also been concluded in this current study that the IR may be involved in siRNA generation as a bidirectional transcript of the IR was confirmed. Therefore, the IR from both the V30 and the V22 may play a role in terms of RNA regulation to contribute to these unique symptoms observed in this mixed infection.

Based on these observations, the role of the IR in the V30+V22 mixed infection was further explored by observing the effects of a V30+V30ΔIR-s mixed infection. This mixed infection comprised of V30 as well as the V30 mutant (V30 ΔIR-s) which encompassed the V22 IR. This mixed infection allowed for the V22 IR and its putative effect or interaction with V30 to be studied. As observed in the V30+V22 mixed infection, ULR was also milder than that of V22 in the mixed infection of V30+V30ΔIR-s. V30+V30ΔIR-s induced greater symptoms than both V30 and V22 at 24 dpi whereas in the V30+V22 mixed infection, SV induced were only

greater than that induced by V22. Therefore, as only a similar mild ULR trend between these two mixed infections was shared, the phenotypic differences observed in the V30+V22 mixed infection cannot be fully attributed to only the different IRs in the two variants. Therefore, it is speculated that both the V2 protein and the IR of the ToCSV variants may have jointly contributed to the unique symptoms observed in the V30+V22 mixed infection. This would have to be further verified by a mixed infection of V30 and a V30 mutant containing the V2 of V22. The high severity SV and mild ULR V30 phenotype is exacerbated in the V30+V30ΔIR-s mixed infection which links the IR to symptom severity as the addition of the V22 IR to V30 resulted in significantly greater SV.

As the V30+V22 mixed infection resulted in a change in viral load and symptom phenotype when compared to either V30 or V22 alone, a direct or indirect interaction between the two variants can be proposed, explaining pathogenicity of these variants. This result supports the extensive characterisation of viral variants or differing viruses in mixed infections that are likely to occur in the agricultural field as these may play a critical role in the severity of symptoms induced. Studying mixed infections to determine the interactions between viral elements is beneficial for the agricultural industry as cross-protection could be implemented with this knowledge. Cross-protection could be utilized for biological control by using a milder virus to ward off a more detrimental virus (Syller and Grupa, 2015).

4.4 Investigation of the putative C5 ORF

Four proteins encoded for by AC1/C1 to AC4/C4 from the complementary sense strand of geminiviruses is widely accepted but an additional ORF, AC5/C5, located downstream of AC3/C3 with a partial overlap with the V1 ORF has been described in the DNA-A of some monopartite and bipartite geminiviruses (Li et al., 2015, Li et al., 2021 and Vaghi Medina et al., 2018). This fifth ORF, has previously been discovered, through mutational studies, to play a role in geminiviral DNA replication in Mungbean yellow mosaic India virus (MYMIV) (Ilyas et al., 2010 and Raghavan et al., 2004). Alignments of begomovirus AC5 proteins have shown AC5 to be 83 aa to 255 aa in length, with all putative C5 ORFs having similar domain structures suggesting a conserved function. These conserved domains imply that AC5 may be involved in infection, even for those begomoviruses for which C5 has not yet been identified (Li et al., 2015).

Li et al., (2015) found that by blocking translation of the AC5 in MYMIV, a more moderate symptom phenotype was induced in *N. benthamiana* compared to the wildtype virus. On the contrary, this study revealed that a deletion of the putative C5 in V22 resulted in no difference in phenotype induced. This could imply that the putative C5 may not be important in symptom development in V22. Although this deletion mutant resulted in no phenotypic changes, the N-terminally expanded C5 in V22ΔC5aa$^{1\text{-}193}$ resulted in height and phenotypic changes when compared to V22. This shows that by extending the putative coding region of the V22 C5 ORF in this study, a change in the pathogenicity of V22 was elicited. The height of *N. benthamiana* inoculated with V22ΔC5aa$^{1\text{-}193}$ was reduced at early time points, 4 and 8 dpi, in the infection compared to that of V22. This decreased growth observed in the plants indicates that the plants may have been placed under more stress when inoculated with V22ΔC5aa1$^{\text{-}193}$ compared to V22.

Plant height reduction (stunting) at an early stage can be associated with viral infection. When plants are threatened by a pathogen, phytohormone signalling pathways are disrupted and plant growth and development signals are redirected to plant defence responses. The interactions between these hormones are critical in plant-pathogen interactions. Hormones that have been found to be involved in eliciting a response to plant pathogens are salicylic acid, ethylene and jasmonates (Robert-Seilaniantz et al 2011). Growth regulating hormones such as auxin or cytokinins play a role in the growth to defence transition in the plant resulting in stunted height (Pajerowska-Mukhtar et al., 2012). Taking this into account, it can be proposed that the extended putative coding C5 region of V22 may have elicited a change or interference in phytohormone signalling pathways in *N. benthamiana* compared to V22 at early stages of infection.

V22ΔC5aa$^{1\text{-}193}$ resulted in a delayed onset of ULR and SV development in *N. benthamiana* compared to V22 where the peak severity of ULR and SV at 20 to 24 dpi in V22 was shifted to 28 to 28 dpi in V22ΔC5aa$^{1\text{-}193}$. To see whether recovery like that of V22 was induced, the trial time of these V22ΔC5aa$^{1\text{-}193}$ trials would have to be extended. The delay in the onset of symptoms observed in this study proposes that the introduced N-terminally extended putative coding region of the C5 in V22 may act as a symptom determinant, but further demonstration of transcription and PVX vector mediated C5 expression studies need to be performed, in both *N. benthamiana* and tomato. The latent period, also referred to as the onset of symptoms, could give an indication of the viral efficiency in establishing a systemic infection (Petty et al., 1995). A shorter latent period, before the onset of symptoms, therefore, suggests that the virus may

have a better interaction between viral proteins, host factors, improved replication, or suppression of host defence responses. A longer latent period may point to a decreased efficiency of the virus to infect a plant (Petty et al., 1995). Therefore, as a longer latent period was observed with V22ΔC5aa$^{1\text{-}193}$, this mutant clone may not be as efficient in systemic infection as V22, but further studies would have to be done to conclude this. The absence of this introduced N-terminally extended putative C5 in V22, where V22 displays more severe symptoms in *N. benthamaiana* than V22ΔC5aa$^{1\text{-}193}$, may be explained by the evolutionary selection of V22 against the expression of C5 occurred. Although, it may also suggest that the C5 in V22 is still undergoing evolution to gain this C5. This study is exploratory, and no definite conclusions can be drawn from it, but future phylogeny studies may or may not confirm the suggestion of evolutionary selection against or for C5 in ToCSV. Further studies investigating viral load and other infection properties may prove to be beneficial in determining whether the extended putative C5 in V22 may elicit a change in viral accumulation. Most importantly, sensitive and accurate differentiating symptom phenotypes need to be established in tomato, if the experiments in this study are to be performed in the natural host.

In this study, the introduction in V22 of an extended putative C5 ORF has been identified to exist in other geminiviruses and was shown to have similar α-helical and β-sheet domains within the protein compared to that of the bipartite ACMV virus. This suggests that the C5 ORF may have been previously expressed in ToCSV and is no longer expressed due to evolution and selection pressure. Begomoviruses are known to have high levels of variation due to point mutations and recombination compared to other DNA viruses and their substitution rates have been found to be as high in some cases of RNA viruses (Duffy and Holmes, 2008 and Lima et al., 2013). The V22 extended putative C5 mutant resulted in a delayed onset of symptoms, the infection efficiency of the virus may have been hindered by the expression of C5 and therefore a selection against C5 in V22 occurred. It must be considered that the role of the C5 in monopartite viruses may be different to that of the characterised AC5 proteins in bipartite begomoviruses and therefore the C5 may not prove to be as beneficial to monopartite begomoviruses.

CHAPTER 5 - Concluding remarks

The data obtained from this study demonstrates that the 3' IR in ToCSV plays a vital role in the different symptom phenotypes induced by V30 and V22 in *N. benthamiana.* The presence of an IR RNA also suggests possible RNA regulation by the V22 IR to modulate symptoms induced in *N. benthamiana* but further research to confirm this would have to be done. This study then further demonstrated that individual V22 aa play a role in symptom severity and recovery observed in V22. While investigating the aa of V22 V2, two consecutive aa in V1 were found to be essential to the infectivity and symptom severity of V22. This was discovered through the unintentional ORF change in the V22ΔV2-TS+ mutant infectious clone construct which resulted in the abolishment of ULR. Therefore, one non-coding region, the IR, and two ORFs, V2 and V1, in V22 have been further characterized. This expands what is known about their contributions to the infectivity and pathogenicity of ToCSV.

The dynamics of the ToCSV variants were studied through a mixed infection which revealed that the ToCSV variants display a change in pathogenicity and viral load when in a mixed infection. This study highlights the importance of investigating viruses in mixed infections that would likely occur in the field. Furthermore, trials of the mixed infection of V30+V30ΔIR-s allowed for the role of the IR to be further understood. These results revealed that the IR may be linked to the symptom phenotype and symptom severity induced in *N. benthamiana* supporting the IR findings from the IR-swap mutant trials in this study.

By extending the putative coding region of the V22 C5 ORF in this study, a change in the pathogenicity manifested in the delayed onset of induced symptoms compared to that of V22.

Although conclusions cannot be drawn from this exploratory study without further research, this clear alteration in pathogenicity observed in the C5 mutant infection trials suggests that investigating this putative C5 further may contribute towards the understanding of infectivity mechanisms and the evolutionary history of ToCSV.

One of the challenges faced in this study was the introduction of inadvertent mutations to overlapping ORFs in some of the V22 V2 single aa mutants. This could be overcome by ensuring that the nucleotide change introduced would not result in inadvertent mutations to the overlapping ORF. A further limitation in this study was not being able to differentiate between the viral load of V30 and V22 in a mixed infection. Because the nucleotide sequences between the variants are so similar, designing efficient primers targeting the individual variants is not promising. This can be overcome through attaching green fluorescent labels (GFP) to each variant to allow for separate viral accumulations to be determined within a mixed infection (Rupar et al., 2015).

Future work would include subjecting the sequences of both variants of ToCSV to deep sequencing to allow for the IR transcript or potential vsRNA to be identified which could expand our knowledge of the role the IR plays in pathogenicity. This study could also be repeated in the natural *S. lycopersicum* host and compared to the results observed in the experimental *N. benthamiana* host to determine if the results observed with these mutant infectious clones are host specific. The putative uncharacterised C5 in V22 could be further investigated by inserting the extended C5 into a PVX vector to validate its role in pathogenicity as a symptom determinant. The presence of a C5-derived transcript could be verified through RT-PCR using gene specific primers. Furthermore, genome comparisons of ToCSV with other begomoviruses to look for signatures of recombination or mutations could be done to further investigate this protein.

References

Ahmed, N., Amin, I., Zaidi, S., Rahman, S., Farooq, M., Fauquet, C. and Mansoor, S., 2021. Circular DNA enrichment sequencing reveals the viral/satellites genetic diversity associated with the third epidemic of cotton leaf curl disease. *Biology Methods and Protocols*, 6(1).

Alizon, S., de Roode, J. and Michalakis, Y., 2013. Multiple infections and the evolution of virulence. *Ecology Letters*, 16(4), pp.556-567.

Alves-Júnior, M. et al., 2009. Synergism and negative interference during co-infection of Tomato and *Nicotiana benthamiana* with two bipartite begomoviruses. *Virology*, 387(2), pp.257–266.

Alves-Júnior, M., Alfenas-Zerbini, P., Andrade, E., Esposito, D., Silva, F., da Cruz, A., Ventrella, M., Otoni, W. and Zerbini, F., 2009. Synergism and negative interference during co-infection of tomato and *Nicotiana benthamiana* with two bipartite begomoviruses. *Virology*, 387(2), pp.257-266.

Amin, I., Patil, B., Briddon, R., Mansoor, S. and Fauquet, C., 2011. Comparison of phenotypes produced in response to transient expression of genes encoded by four distinct begomoviruses in *Nicotiana benthamiana* and their correlation with the levels of developmental miRNAs. *Virology Journal*, 8(1).

Argüello-Astorga, G., Guevara-González, R., Herrera-Estrella, L. and Rivera-Bustamante, R., 1994. Geminivirus Replication Origins Have a Group-Specific Organization of Iterative Elements: A Model for Replication. *Virology*, 203(1), pp.90-100.

Basu, S., Kumar Kushwaha, N., Kumar Singh, A., Pankaj Sahu, P., Vinoth Kumar, R. and Chakraborty, S., 2018. Dynamics of a geminivirus-encoded pre-coat protein and host RNA-dependent RNA polymerase 1 in regulating symptom recovery in tobacco. *Journal of Experimental Botany*, 69(8), pp.2085-2102.

Baulcombe D., 2005. RNA silencing. *Trends Biochem. Sci.* 30(6): 290–293. doi: 10.1016/j.tibs.2005.04.012

Benchling [Biology Software]. (2020). Retrieved from https://benchling.com.

Cazzonelli, C., Burke, J. and Velten, J., 2005. Functional characterization of the geminiviral conserved lateelement (CLE) in uninfected tobacco. Plant Molecular Biology, 58(4), pp.465-481.

Chang, A., Chau, V., Landus, J. and Pang, Y., 2017. Preparation of calcium competent Escherichia coli and heat-shock transformation. *JEMI Methods*, 1, pp.22-25.

Chen, J., Lin, B., Huang, Q., Hu, L., Zhuo, K. and Liao, J., 2017. A novel *Meloidogyne graminicola* effector, MgGPP, is secreted into host cells and undergoes glycosylation in concert with proteolysis to suppress plant defenses and promote parasitism. *PLOS Pathogens*, 13(4): p.e1006301.

Collmer, C. W. and Howell, S. H., 1992. Role of satellite RNA in the expression of symptoms caused by plant viruses. *Annual Review of Phytopathology*, 30: 419–442. doi: 10.1146/annurev.py.30.090192.002223

DaPalma, T., Doonan, B., Trager, N. and Kasman, L., 2010. A systematic approach to virus–virus interactions. *Virus Research*, 149(1), pp.1-9.

Doyle, J., Doyle, J. and Doyle, F., 1987 A rapid DNA isolation procedure for small amounts of fresh leaf tissue. *Phytochemistry Bulletin*, 19 :11-15.

Duffy, S. and Holmes, E., 2008. Phylogenetic Evidence for Rapid Rates of Molecular Evolution in the Single-Stranded DNA Begomovirus Tomato Yellow Leaf Curl Virus. *Journal of Virology*, 82(2), pp.957-965.

Edgar, Robert C. (2004), MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Research* 32(5), 1792-97.

Esterhuizen, L.L., 2012. A study of the South African Tomato curly stunt virus pathosystem: Epidemiology, molecular diversity and resistance (Doctoral dissertation, University of Johannesburg).

Etessami, P., Callis, R., Ellwood, S. and Stanley, J., 1988. Delimitation of essential genes of cassava latent virus DNA 2. *Nucleic Acids Research*, 16 :4811–4829. doi: 10.1093/nar/16.11.4811

Foolad, M., 2007. Genome Mapping and Molecular Breeding of Tomato. *International Journal of Plant Genomics*, 2007: 1-52 doi: 10.1155/2007/64358.

Ghoshal, B. and Sanfaçon, H., 2015. Symptom recovery in virus-infected plants: Revisiting the role of RNA silencing mechanisms. *Virology*, 480, 167-179. doi: 10.1016/j.virol.2015.01.008

Green, M. and Sambrook, J., 2020. The Inoue Method for Preparation and Transformation of Competent Escherichia coli: "Ultracompetent" Cells. *Cold Spring Harbor Protocols*, 2020(6).

Haimov, B. and Srebnik, S., 2016. A closer look into the α-helix basin. *Scientific Reports*, 6(1).

Hardjasa, A., Ling, M., Ma, K. and Yu, H., 2010. Investigating the Effects of DMSO on PCR Fidelity Using a Restriction Digest-Based Method. *Journal of Experimental Microbiology and Immunology (JEMI)*, 14, pp.161-164.

Harrison, B. D., 1985. Advances in Geminivirus Research. *Annual Review of Phytopathology*, 23. 55–82. doi: https://doi.org/10.1146/annurev.py.23.090185.000415

Mehta, D., Hirsch-Hoffmann, M., Were, M., Patrignani, A., Zaidi, S., Were, H., Gruissem, W. and Vanderschuren, H., 2018. A new full-length circular DNA sequencing method for viral-sized genomes reveals that RNAi transgenic plants provoke a shift in geminivirus populations in the field. *Nucleic Acids Research*, 47(2), pp.e9-e9.

Hofgen, R. and Willmitzer, L., 1988. Storage of competent cells For Agrobacterium transformation. *Nucleic Acids Research*, 16 (20), 9877-9877. doi: 10.1093/nar/16.20.9877

Horne EC, Kumpatla SP, Patterson MG, Thompson SA 2004 Improved high-throughput sunflower and cotton genomic DNA extraction and PCR fidelity. Plant Mol Biol Rep. 22:83

Ilyas, M., Qazi, J., Mansoor, S. and Briddon, R., 2010. Genetic diversity and phylogeography of begomoviruses infecting legumes in Pakistan. *Journal of General Virology*, 91(8), pp.2091-2101.

Jameson P, Clarke S., 2002. Hormone-virus interaction in plants. Crit. Rev. *Plant Science.* 21, 205–28. doi: 10.1080/0735-260291044241

Janjua, S., Younis, S., Deeba, F. and Naqvi, S., 2014. High Efficiency DNA Transformation Protocol for Escherichia coli using Combination of Physico-chemical Methods. INTERNATIONAL JOURNAL OF AGRICULTURE & BIOLOGY, pp.1814–9596.

Jeske, H., Lütgemeier, M. and Preiß, W., 2001. DNA forms indicate rolling circle and recombination-dependent replication of Abutilon mosaic virus. *The EMBO Journal*, 20, 6158–6167. doi: 10.1093/emboj/20.21.6158

Jones J.D.G., Dangl J.L., 2006. The plant immune system. *Nature* 444, 323–329. doi: https://doi.org/10.1038/nature05286

Jyothishwaran, G., Kotresha, D., Selvaraj, T., Srideshikan, S. M., Rajvanshi, P. K., & Jayabaskaran, C. (2007). A modified freeze–thaw method for efficient transformation of *Agrobacterium tumefaciens*. *Current Science*, *93*(6), 770–772.

Kalaiselvi, R., Rajasekar, M. and Gomathi, S., 2017. Cryopreservation of Plant Materials - A Review. *International Journal of Chemical Studies*, 5 (5), 560-564.

Krake, L., Rezaian, M. and Dry, I., 1998. Expression of the Tomato Leaf Curl GeminivirusC4Gene Produces Viruslike Symptoms in Transgenic Plants. *Molecular Plant-Microbe Interactions*, 11 (5), 413-417. doi: https://doi.org/10.1094/MPMI.1998.11.5.413

Kumar, S., Srivastava, A., Jaidi, M., Chauhan, P. and Raj, S., 2016. Molecular Characterization of a Begomovirus, α-Satellite, and β-Satellite Associated with Leaf Curl Disease of Parthenium hysterophorus in India. *Plant Disease*, 100(11), pp.2299-2305.

Laible, M. and Boonrod, K., 2009. Homemade Site Directed Mutagenesis of Whole Plasmids. *Journal of Visualized Experiments*, (27).

Lapidot, M., Ben-Joseph, R., Cohen, L., Machbash, Z. and Levy, D., 2006. Development of a Scale for Evaluation of Tomato yellow leaf curl virus Resistance Level in Tomato Plants. P*hytopathology*, 96 ,1404–1408. doi: 10.1094/PHYTO-96-1404

Lapidot, M., Ben-Joseph, R., Cohen, L., Machbash, Z. and Levy, D., 2006. Development of a Scale for Evaluation of Tomato yellow leaf curl virus Resistance Level in Tomato Plants. *Phytopathology®*, 96(12), pp.1404-1408.

Leke, W., Mignouna, D., Brown, J. and Kvarnheden, A., 2015. Begomovirus disease complex: emerging threat to vegetable production systems of West and Central Africa. *Agriculture & Food Security*, 4 (1), 1-14. doi: 10.1186/s40066-014-0020-2

Li, F. et al., 2015. The AC 5 protein encoded by mungbean yellow mosaic india virus is a pathogenicity determinant that suppresses RNA silencing-based antiviral defenses. *New Phytologist*, 208(2), pp.555–569.

Li, F., Xu, X., Huang, C., Gu, Z., Cao, L., Hu, T., Ding, M., Li, Z. and Zhou, X., 2015. The AC 5 protein encoded by Mungbean yellow mosaic India virus is a pathogenicity determinant that suppresses RNA silencing-based antiviral defenses. *New Phytologist*, 208(2), pp.555-569.

Li, P. et al., 2021. The C5 protein encoded by ageratum leaf curl sichuan virus is a virulence factor and contributes to the virus infection. *Molecular Plant Pathology*, 22(9), pp.1149–1158.

Lieutaud, P., Ferron, F., Uversky, A., Kurgan, L., Uversky, V. and Longhi, S., 2016. How disordered is my protein and what is its disorder for? A guide through the "dark side" of the protein universe. *Intrinsically Disordered Proteins*, 4(1), p.e1259708.

Lieutaud, P., Ferron, F., Uversky, A., Kurgan, L., Uversky, V. and Longhi, S., 2016. How disordered is my protein and what is its disorder for? A guide through the "dark side" of the protein universe. Intrinsically Disordered Proteins, 4(1), p.e1259708.

Lima, A. T. M., Silva, J. C. F., Silva, F. N., Castillo-Urquiza, G. P., Silva, F. F., Seah, Y. M., Mizubuti, E. S. G., Duffy, S. and Zerbini, F. M., 2017. The diversification of begomovirus populations is predominantly driven by mutational dynamics. *Virus Evolution* ,3 (1),1-14. doi: 10.1093/ve/vex005

Lima, A., Sobrinho, R., González-Aguilera, J., Rocha, C., Silva, S., Xavier, C., Silva, F., Duffy, S. and Zerbini, F., 2013. Synonymous site variation due to recombination explains higher genetic variability in begomovirus populations infecting non-cultivated hosts. *Journal of General Virology*, 94(2), pp.418-431.

Liu, H., Davies, J. and Boulton, M., 1997. Maize streak virus coat protein binds single- and double-stranded DNA in vitro. Journal of *General Virology*, 78(6), pp.1265-1270.

Lovdal, T. and Lillo, C., 2009. Reference gene selection for quantitative real-time PCR normalization in tomato subjected to nitrogen, cold, and light stress. *Elsevier*, 387 (2), 238-242. doi: 10.1016/j.ab.2009.01.024

Ma, L., Lukasik, E., Gawehns, F. and Takken, F., 2012. The Use of Agroinfiltration for Transient Expression of Plant Resistance and Fungal Effector Proteins in *Nicotiana benthamiana* Leaves. *Methods in molecular biology*, 835, 61-74. doi: 10.1007/978-1-61779-501-5_4.

Macara, I., 2001. Transport into and out of the Nucleus. *Microbiology and Molecular Biology Reviews*, 65(4), pp.570-594.

Malherbe, S. and Marais, D. (2015). Economics, Yield and Ecology: A Case Study from the South African Tomato Industry. *Outlook on Agriculture*, 44, 37–47. doi: 10.5367/oa.2015.0195

Mascia, T., Santovito, E., Gallitelli D. and Cillo, F., 2010. Evaluation of reference genes for quantitative reverse-transcription polymerase chain reaction normalization in infected tomato plants. *Molecular Plant Pathology*, 11(6), 805-814. doi: 10.1111/j.1364-3703.2010.00646.x.

McGuffin, L., Bryson, K. and Jones, D., 2000. The PSIPRED protein structure prediction server. *Bioinformatics*, 16(4), pp.404-405.

Moriones, E. and Navas-Castillo, J., 2000. Tomato yellow leaf curl virus, an emerging virus complex causing epidemics worldwide. *Virus Research*, 71, 123–134. doi: 10.1016/s0168-1702(00)00193-3

Mubin, M., Amin, I., Amrao, L., Briddon, R. and Mansoor, S., 2010. The hypersensitive response induced by the V2 protein of a monopartite begomovirus is countered by the C2 protein. *Molecular Plant Pathology*, 11 (2), 245-254. doi: 10.1111/j.1364-3703.2009.00601.x.

Nakazono-Nagaoka, E. et al., 2009. Cross-protection against Bean Yellow Mosaic Virus (BYMV) and Clover Yellow Vein virus by attenuated BYMV isolate M11. *Phytopathology*®, 99(3), pp.251–257.

Navot, N., Pichersky, E., Zamir, D., Zamir, D. and Czosnek, H., 1991. Tomato yellow leaf curl virus: a whitefly-transmitted geminivirus with a single genomic component. *Virology* 185 (1): 151. doi: 10.1016/0042-6822(91)90763-2

Nawaz-ul-Rehman, M. and Fauquet, C., 2009. Evolution of geminiviruses and their satellites. *FEBS Letters*, 583(12) :1825-1832. doi: 10.1016/j.febslet.2009.05.045

Nejat, N. and Mantri, N. 2017. Plant Immune System: Crosstalk Between Responses to Biotic and Abiotic Stresses the Missing Link in Understanding Plant Defence. *Current Issues in Molecular Biology*, 1-16. doi: 10.21775/cimb.023.001

Nelson RS, Citovsky V., 2005. Plant viruses. Invaders of cells and pirates of cellular pathways. *Plant Physiology.* 138 :1809–14. doi: https://doi.org/10.1104/pp.104.900167

New England Biolabs. Protocol guidelines from NEB 5-alpha competent E. coli (High Efficiency).

Noueiry, A. O., Lucas, W. J. and Gilbertson, R. L. (1994). Two proteins of a plant DNA virus coordinate nuclear and plasmodesmal transport. *Cell*, 76: 925–932. doi: 10.1016/0092-8674(94)90366-2

Orozco, B., Miller, A., Settlage, S. and Hanley-Bowdoin, L., 1997. Functional Domains of a Geminivirus Replication Protein. *Journal of Biological Chemistry*, 272 (15): 9840-9846. doi: 10.1074/jbc.272.15.9840

Padidam, M., Beachy, R. and Fauquet, C., 1996. The Role of AV2 ("Precoat") and Coat Protein in Viral Replication and Movement in Tomato Leaf Curl Geminivirus. *Virology,* 224(2), pp.390-404.

Pajerowska-Mukhtar, K. M., Wang, W., Tada, Y., Oka, N., Tucker, C. L., Fonseca, J. P., et al. (2012). The HSF-like transcription factor TBF1 is a major molecular switch for plant growth-to-defense transition. *Curr. Biol.* 22, 103–112. doi: 10.1016/j.cub.2011.12.015

Pallas. V, J.A. Garcia., 2011. How do plant viruses induce disease? Interactions and interference with host components J. *The journal of general virology.*, 92: 2691-2705. doi: 10.1099/vir.0.034603-0

Parthenium hysterophorus in India, *The American Phytopathological Society* 100 (11) ,2299-2305. doi https://doi.org/10.1094/PDIS-09-15-0982-RE

Paula Suärez, López Encarnación, Martínez Salas, Pablo Hernández, Crisanto Gutiérrez., 1995. Bent DNA in the Large Intergenic Region of Wheat Dwarf Geminivirus, *Virology* 208 (1):303-311. doi: 10.1006/viro.1995.1153

Petty, I.T.D. et al., 1995. Complementable and noncomplementable host adaptation defects in bipartite geminiviruses. *Virology*, 212(1), pp.263–267.

Pietersen, G., Idris, A. M., Krüger, K. and Brown, J. K., 2008. Characterization of Tomato curly stunt virus: a new tomato-infecting begomovirus from South Africa. *Plant Pathology*, 57: 809–818. doi: 10.1111/j.1365-3059.2008.01882.x

Pita, J., Morris, V. and Roossinck, M., 2015. Mutation and Recombination Frequencies Reveal a Biological Contrast within Strains of Cucumber Mosaic Virus. *Journal of Virology*, 89(13), pp.6817-6823.

Pooma, W. and Petty, I. T., 1996. Tomato golden mosaic virus open reading frame AL4 is genetically distinct from its C4 analogue in monopartite geminiviruses. *The Journal of General Virology*, 77 (8): 1947–1951. doi: 10.1099/0022-1317-77-8-1947

Raghavan, V., Malik, P., Choudhury, N. and Mukherjee, S., 2004. The DNA-A Component of a Plant Geminivirus (Indian Mung Bean Yellow Mosaic Virus) Replicates in Budding Yeast Cells. *Journal of Virology*, 78(5), pp.2405-2413.

Rajamäki, M. L., & Valkonen, J. P. 2009. Control of nuclear and nucleolar localization of nuclear inclusion protein a of picorna-like Potato virus A in Nicotiana species. *The Plant cell*, 21(8), 2485–2502. https://doi.org/10.1105/tpc.108.064147

Ramesh, S., Sahu, P., Prasad, M., Praveen, S. and Pappu, H. 2017. Geminiviruses and Plant Hosts: A Closer Examination of the Molecular Arms Race. *Viruses*, 9 (9), 256. doi: 10.3390/v9090256

Reddy, R.V.C. et al., 2012. Molecular interaction between two cassava geminiviruses exhibiting cross-protection. *Virus Research*, 163(1), pp.169–177.

Rentería-Canett, I., Xoconostle-Cázares, B., Ruiz-Medrano, R. and Rivera-Bustamante, R., 2011. Geminivirus mixed infection on pepper plants: Synergistic interaction between PHYVV and PepGMV. *Virology Journal*, 8(1).

Robert-Seilaniantz, A., Grant, M., and Jones, J. D. 2011. Hormone crosstalk in plant disease and defense: more than just jasmonate-salicylate antagonism. *Annu. Rev. Phytopathol.* 49, 317–343. doi: 10.1146/annurev-phyto-073009-114447

Rupar, M., Faurez, F., Tribodet, M., Gutiérrez-Aguirre, I., Delaunay, A., Glais, L., Kriznik, M., Dobnik, D., Gruden, K., Jacquot, E. and Ravnikar, M., 2015. Fluorescently Tagged Potato virus Y: A Versatile Tool for Functional Analysis of Plant-Virus Interactions. *Molecular Plant-Microbe Interactions*®, 28(7), pp.739-750.

Sanger, F., Nicklen, S. and Coulson, A., 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12), pp.5463-5467.

Saunders, K., Lucy, A. and Stanley, J., 1992. RNA-primed complementary-sense DNA synthesis of the geminivirus African cassava mosaic virus. *Nucleic Acids Research*, 20, 6311–6315. doi: 10.1093/nar/20.23.6311

Semenyuk, A., Ahnfelt, M., Estmer Nilsson, C., Yong Hao, X., Földesi, A., Kao, Y., Chen, H., Kao, W., Peck, K. and Kwiatkowski, M., 2006. Cartridge-based high-throughput purification of oligonucleotides for reliable oligonucleotide arrays. *Analytical Biochemistry*, 356(1), pp.132-141.

Sharma, P. and Ikegami, M., 2010. Tomato leaf curl Java virus V2 protein is a determinant of virulence, hypersensitive response and suppression of posttranscriptional gene silencing. *Virology*, 396 (1), 85-93. doi: https://doi.org/10.1016/j.virol.2009.10.012

Sharma, P. and Ikegami, M., 2010. Tomato leaf curl Java virus V2 protein is a determinant of virulence, hypersensitive response and suppression of posttranscriptional gene silencing. *Virology*, 396(1), pp.85-93.

Sijen, T. & Kooter, J.M., 2000. Post-transcriptional gene-silencing: RNAS on the attack or on the defense? *BioEssays*, 22(6), pp.520–531.

Simon, I., 2020. Macromolecular Interactions of Disordered Proteins. *International Journal of Molecular Sciences*, 21(2), p.504.

Snehi, S. K., Purvua, A. S., Parihar, S. S., Gupta, G., Singh, V. and Raj, S. K. ,2017. Overview of Begomovirus Genomic Organization and its Impact, *International Journal of current research*,9(11): 61368-61380. doi: http://www.journalcra.com

Syller, J. and Grupa, A., 2015. Antagonistic within-host interactions between plant viruses: molecular basis and impact on viral and host fitness. *Molecular Plant Pathology*, 17(5), pp.769-782.

Syller, J. and Grupa, A., 2015. Antagonistic within-host interactions between plant viruses: molecular basis and impact on viral and host fitness. *Molecular Plant Pathology*, 17(5), pp.769-782.

Taylor, S., Nadeau, K., Abbasi, M., Lachance, C., Nguyen, M. and Fenrich, J., 2019. The Ultimate qPCR Experiment: Producing Publication Quality, Reproducible Data the First Time. *Trends in Biotechnology*, 37 (7), 761-774. doi:https://doi.org/10.1016/j.tibtech.2018.12.002

Unseld, S., Höhnle, M., Ringel, M. and Frischmuth, T., 2001. Subcellular Targeting of the Coat Protein of African Cassava Mosaic Geminivirus. *Virology*, 286(2), pp.373-383.

Urbino, C., Thébaud, G., Granier, M., Blanc, S. and Peterschmitt, M., 2008. A novel cloning strategy for isolating, genotyping and phenotyping genetic variants of geminiviruses. *Virology Journal*, 5 (1), 135. doi: https://doi.org/10.1186/1743-422X-5-135

Vaghi Medina, C.G. et al., 2018. Tomato apical leaf curl virus: A novel, monopartite geminivirus detected in tomatoes in Argentina. *Frontiers in Microbiology*, 8.

Vanitharani, R., Chellappan, P., Pita, J. and Fauquet, C., 2004. Differential Roles of AC2 and AC4 of Cassava Geminiviruses in Mediating Synergism and Suppression of Posttranscriptional Gene Silencing. *Journal of Virology*, 78(17), pp.9487-9498.

Varma, A. and Malathi, V. G., 2003. Emerging geminivirus problems: A serious threat to crop production. *Annals of Applied Biology*, 142, 145–164. doi: https://doi.org/10.1111/j.1744-7348.2003.tb00240.x

Vaucheret, H., 2006. Post-transcriptional small RNA pathways in plants: mechanisms and regulations. *Genes & Development*, 20(7), pp.759-771.

Velten, J., Morey, K. and Cazzonelli, C., 2005. Plant viral intergenic DNA sequence repeats with transcription enhancing activity. *Virology Journal*, 2(16).

Vinogradov, A. E., 2001. Mirrored Genome Size Distributions in Monocot and Dicot Plants. Acta *Biotheoretica*, 49, 43–51. doi: https://doi.org/10.1023/A:1010233608680

Wang, B., Li, F., Huang, C., Yang, X., Qian, Y., Xie, Y. and Zhou, X., 2014. V2 of tomato yellow leaf curl virus can suppress methylation-mediated transcriptional gene silencing in plants. *Journal of General Virology*, 95 (1) :225-230. doi: 10.1099/vir.0.055798-0

Wang, M., Masuta, C., Smith, N. and Shimura, H., 2012. RNA Silencing and Plant Viral Diseases. Molecular Plant-Microbe Interactions®, 25(10), pp.1275-1285.

Ward, J., McGuffin, L., Bryson, K., Buxton, B. and Jones, D., 2004. The DISOPRED server for the prediction of protein disorder. *BIOINFORMATICS APPLICATIONS NOTE*, 20(13), pp.2138-2139.

Weigel, D. and Glazebrook, J., 2006. Transformation of Agrobacterium Using the Freeze-Thaw Method. Cold Spring Harbor Protocols, 2006(7), p.pdb. *prot*4666.

Wen, F., Lister, R.M. & Fattouh, F.A., 1991. Cross-protection among strains of Barley yellow dwarf virus. *Journal of General Virology*, 72(4), pp.791–799.

Yang, Y., Liu, T., Shen, D., Wang, J., Ling, X., Hu, Z., Chen, T., Hu, J., Huang, J., Yu, W., Dou, D., Wang, M. and Zhang, B., 2019. Tomato yellow leaf curl virus intergenic siRNAs target a host long noncoding RNA to modulate disease symptoms. *PLOS Pathogens*, 15 (1): 1007-1034. doi: https://doi.org/10.1371/journal.ppat.1007534

Ying, X.-B., Dong, L., Zhu, H., Duan, C.-G., Du, Q.-S., Lv, D.-Q., Fang, Y.-Y., Garcia, J. A., Fang, R.-X. and Guo, H.-S., 2010. RNA-dependent RNA polymerase 1 from Nicotiana tabacum suppresses RNA silencing and enhances viral infection in *Nicotiana benthamiana*. *The Plant Cell*, 22: 1358–1372. doi: 10.1105/tpc.109.072058

Zerbini, F. M., Briddon, R. W., Idris, A., Martin, D. P., Moriones, E., Navas-Castillo, J., Rivera-Bustamante, R., Roumagnac, P., Varsani, A. and ICTV Report Consortium.,2017. ICTV Virus Taxonomy Profile: Geminiviridae. *Journal of General Virology*, 98: 131–133. doi: 10.1099/jgv.0.000738.

Zhang, J., Dong, J., Xu, Y. and Wu, J., 2012. V2 protein encoded by Tomato yellow leaf curl China virus is an RNA silencing suppressor. *Virus Research*, 163(1), pp.51-58.

Zhao, W., Ji, Y., Wu, S., Ma, X., Li, S., Sun, F., Cheng, Z., Zhou, Y. and Fan, Y., 2018. Single amino acid in V2 encoded by TYLCV is responsible for its self-interaction, aggregates and pathogenicity. *Scientific Reports*, **8**(1), 3561. doi: https://doi.org/10.1038/s41598-018-21446-2.

Zhao, W., Ji, Y., Wu, S., Ma, X., Li, S., Sun, F., Cheng, Z., Zhou, Y. and Fan, Y., 2018. Single amino acid in V2 encoded by TYLCV is responsible for its self-interaction, aggregates and pathogenicity. *Scientific Reports*, 8(1).

Zwolinksi, A. M, 2022 *PhD thesis*. Mutational and structural analyses of Tomato curly stunt virus pathogenicity determinants. University of the Witwatersrand (South Africa).

Supplementary Data

Section 1 – The Intergenic Region (IR)

V22 3’ IR - *N. benthamiana cDNA* library

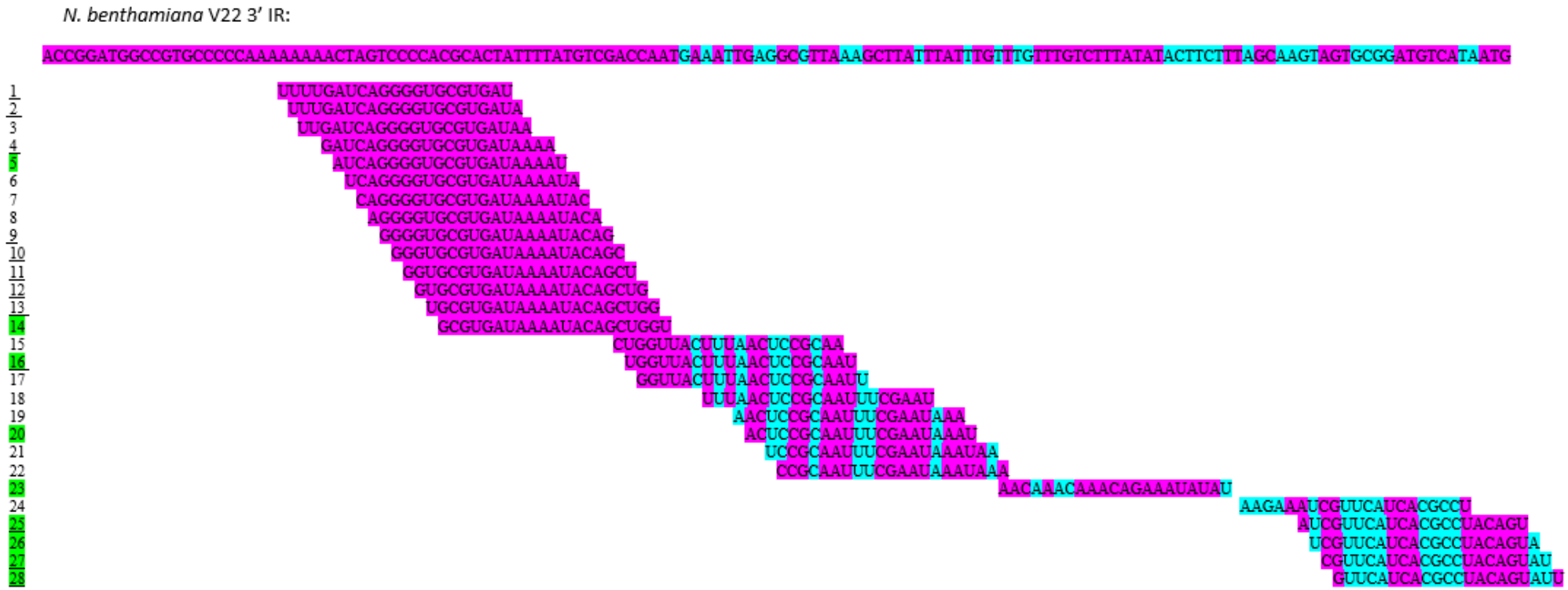


Figure 1.1 Alignment of 3’ end of Tomato curly stunt virus (ToCSV) V22 intergenic region and 29 siRNAs generated using www.plantgrn.noble.org against *Nicotiana benthamiana* cDNA library. Magenta indicates sequences common to ToCSV-V30 and ToCSV-V22, cyan indicates sequence specific to ToCSV_V22. Green indicates efficiency between 9 and 10. Underlined indicates off targets below or equal to 15.

V30 3’ IR - *N. benthamiana* cDNA library

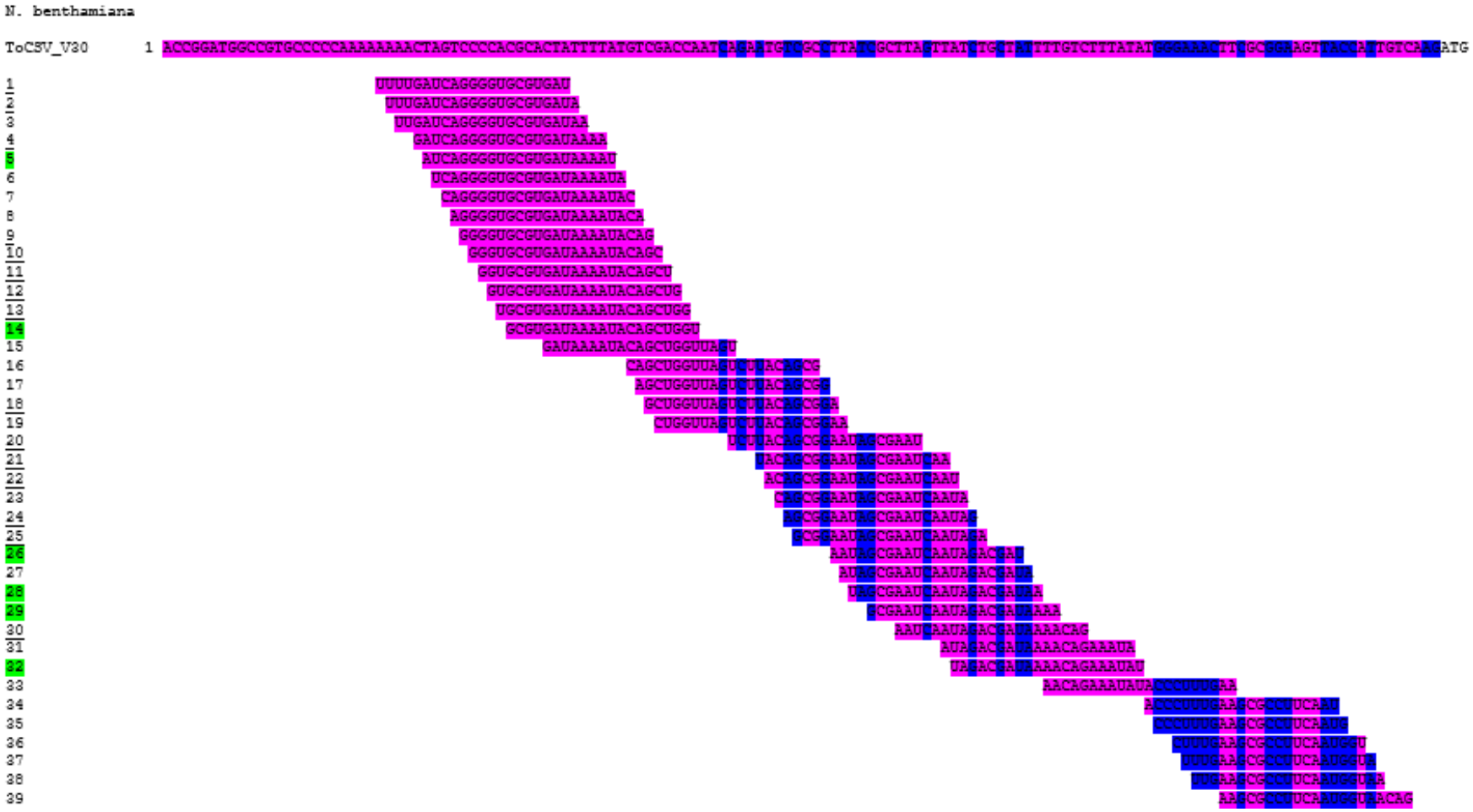


Figure 1.2 Alignment of 3’ end of Tomato curly sunt virus (ToCSV)-V22 intergenic region and 29 siRNAs generated using www.plantgrn.noble.org against *Solanum lycopersicum* cDNA library. Magenta indicates sequences common to ToCSV-V22 and ToCSV-V30, navy indicates sequence specific to ToCSV-V22. Green indicates efficiency between 9 and 10. Green indicates efficiency between 9 and 10. Efficiency denotes the effectiveness of designed siRNA to silence the transcripts. Underlined indicates off targets below or equal to 15.

V22 3' IR - *S. lycopersicum* cDNA library

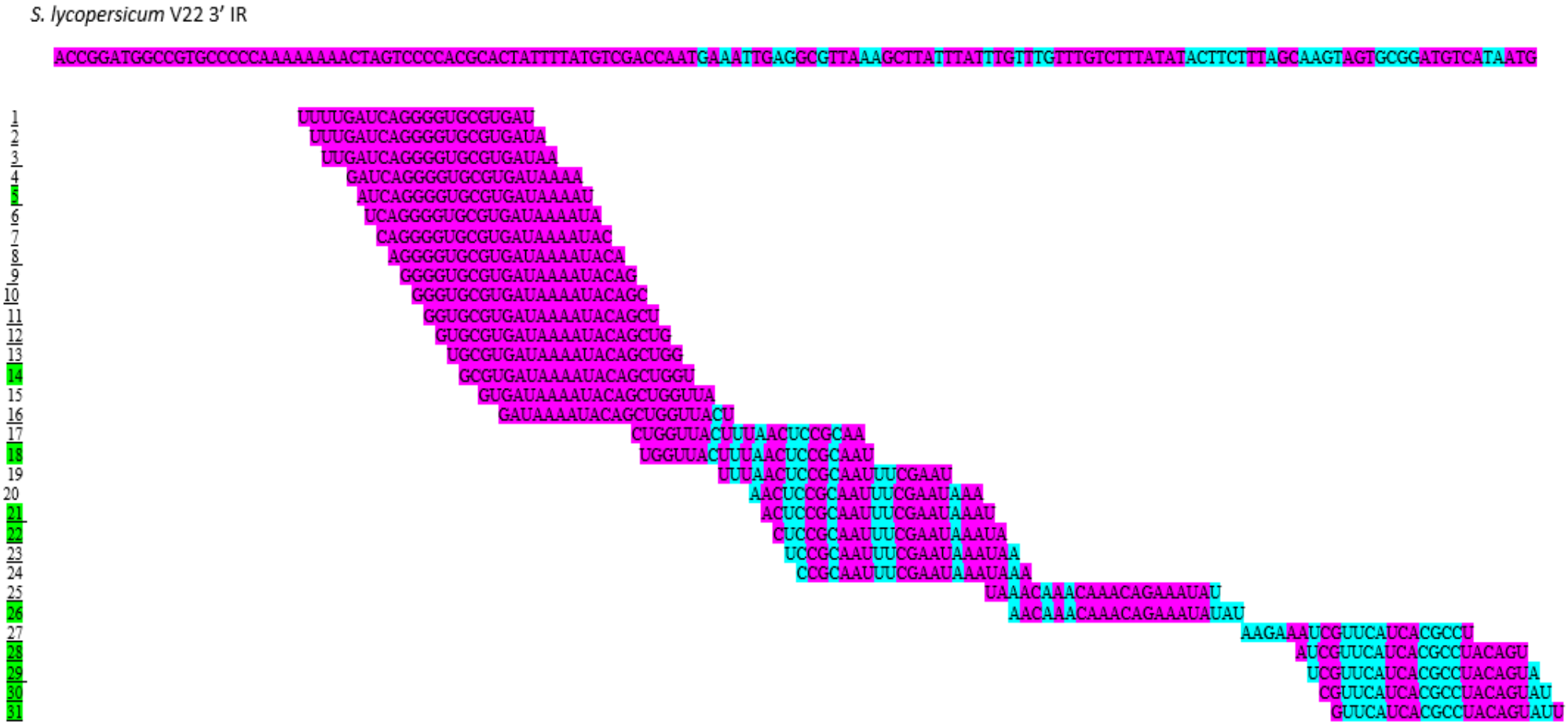


Figure 1.3 Alignment of 3' end of Tomato curly stunt virus (ToCSV)-V22 intergenic region and 29 siRNAs generated using www.plantgrn.noble.org against *Solanum lycopersicum* cDNA library. Magenta indicates sequences common to ToCSV-V22 and ToCSV-V30, cyan indicates sequence specific to ToCSV_V22. Green indicates efficiency between 9 and 10. Efficiency denotes the effectiveness of designed siRNA to silence the transcripts. Underlined numbers indicate off targets below or equal to 15.

V30 3’ IR - *S. lycopersicum* cDNA library

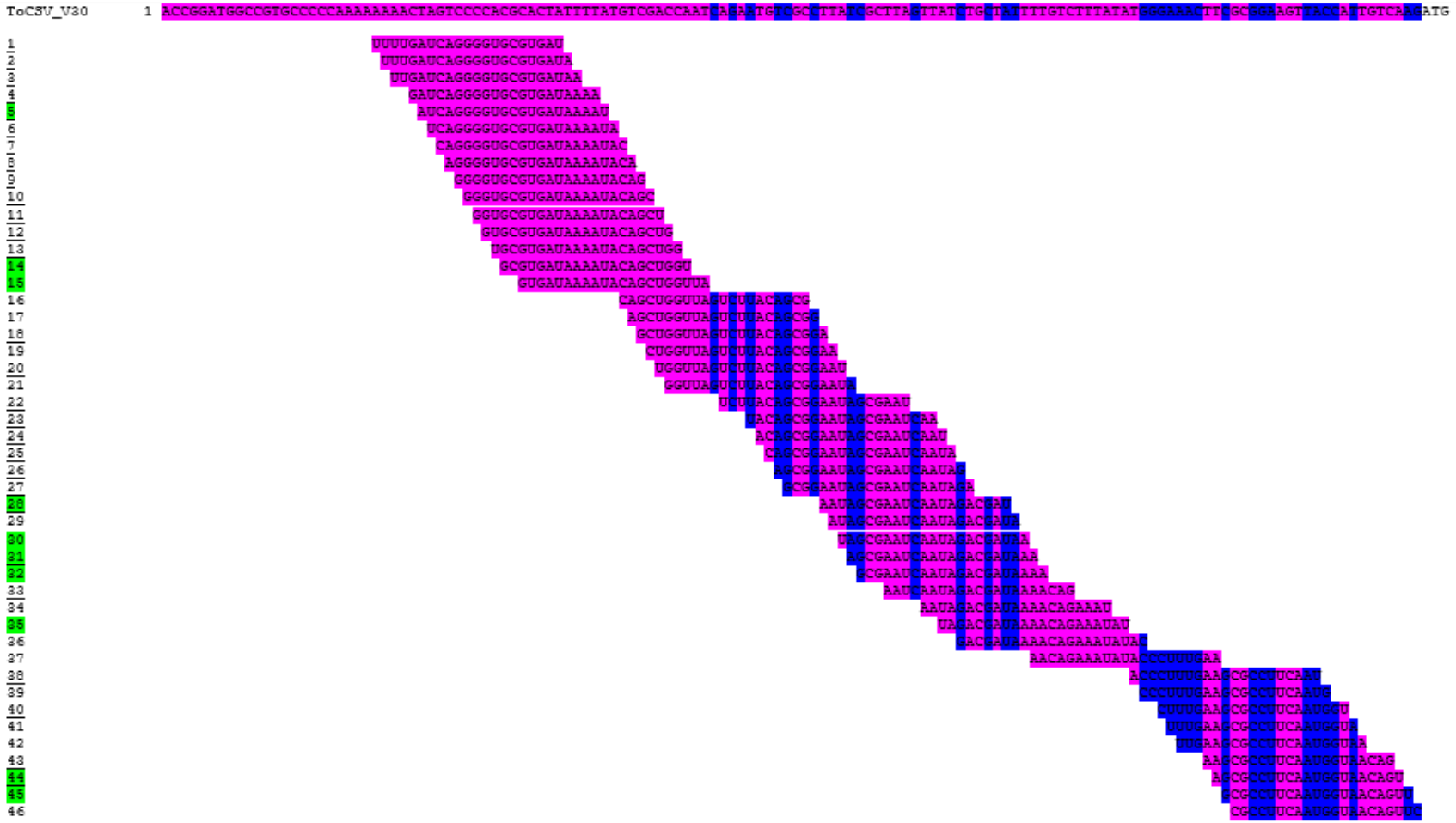


Figure 1.4 Alignment of 3’ end of Tomato curly stunt virus (ToCSV)-V30 intergenic region and 29 siRNAs generated using www.plantgrn.noble.org against *Solanum lycopersicum* cDNA library. Magenta indicates sequences common to ToCSV-V30 and ToCSV-V22, navy indicates sequence specific to ToCSV-V30. Green indicates efficiency between 9 and 10. Green indicates efficiency between 9 and 10. Efficiency denotes the effectiveness of designed siRNA to silence the transcripts. Underlined numbers indicate off targets below or equal to 1.

Section 1 – IR section

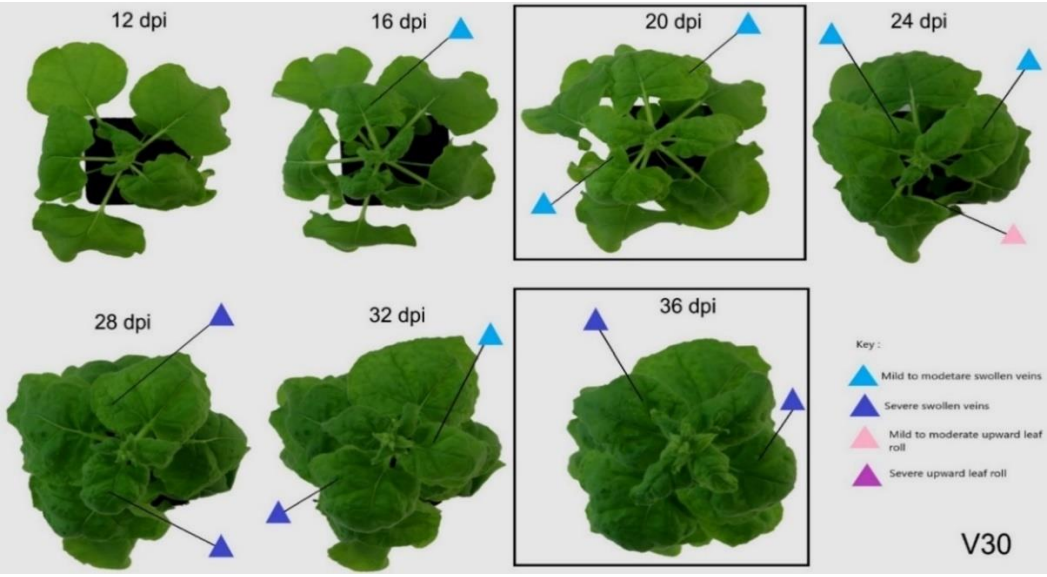


Figure 1.5 Symptom progression images of *Nicotiana benthamiana* inoculated with V30 (Positive control) at 12-, 16-, 20-, 24-, 28-, 32-, and 36 days post inoculation (dpi). Symptoms present are upward leaf roll (ULR) with pink and purple triangles indicating either mild to moderate or severe ULR. Symptoms present are upward leaf roll (ULR) and swollen veins (SV). Severity of symptoms shown by colour coded triangles as indicated in key. Twenty days post inoculation (dpi) is boxed as this is where symptom differences are clear, and 36 dpi is boxed as this is the time point post recovery of ULR and SV. SV and mild SV seen at 20 dpi and severe SV at 36 dpi with no symptom recovery.

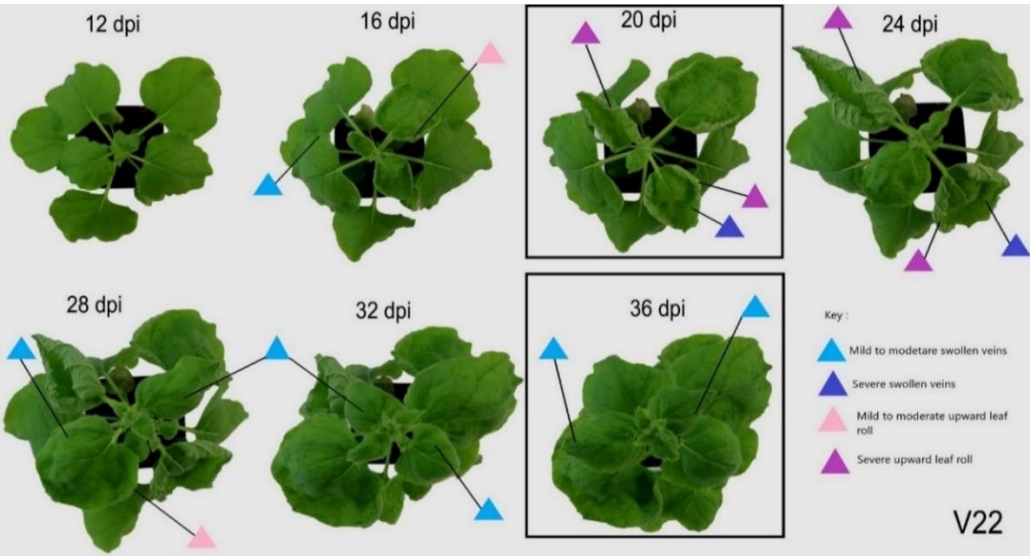


Figure 1.6 Symptom progression images of *Nicotiana benthamiana* inoculated with V22 (Positive control) at 12-, 16-, 20-, 24-, 28-, 32-, and 36 days post inoculation (dpi). Symptoms present are upward leaf roll (ULR) and swollen veins (SV). Severity of symptoms shown by colour coded triangles as indicated in key. Twenty days post inoculation (dpi) is boxed as this is where symptom differences are clear and 36 dpi is boxed as this is the time point post recovery of ULR and SV. Severe ULR and SV evident at 20 dpi. Mild and SV no ULR at 36 dpi showing a recovery in ULR and SV.

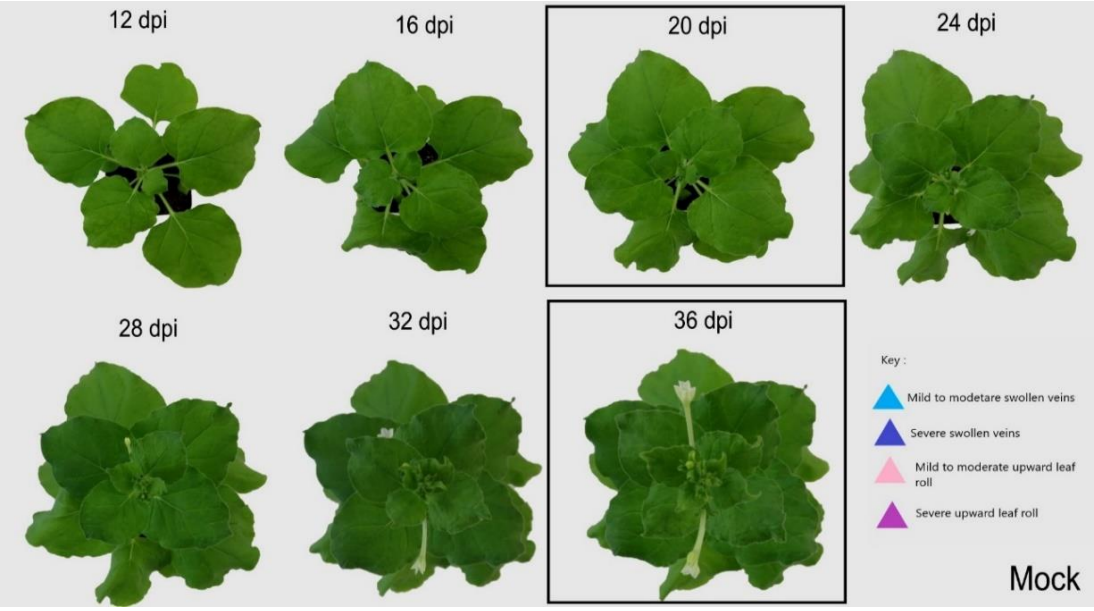


Figure 1.7 Symptom progression images of *Nicotiana benthamiana* inoculated with the empty pCambia2300 vector in *A. tumefaciens* (Mock) at 12-, 16-, 20-, 24-, 28-, 32-, and 36 days post inoculation (dpi). No symptoms induced.

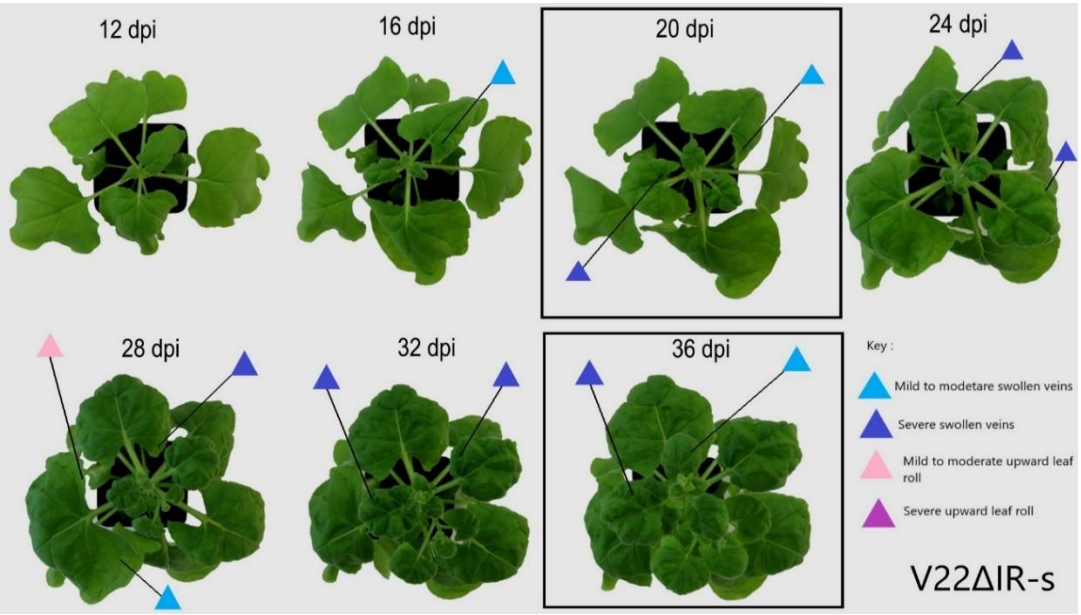


Figure 1.8 Symptom progression images of *Nicotiana benthamiana* inoculated with the V22ΔIR-s infectious clone construct at 12-, 16-, 20-, 24-, 28-, 32-, and 36 days post inoculation (dpi). Symptoms present are upward leaf roll (ULR) and swollen veins (SV). Severity of symptoms shown by colour coded triangles as indicated in key. Twenty days post inoculation (dpi) is boxed as this is where symptom differences are clear and 36 dpi is boxed as this is the time point post recovery of ULR and SV. Only SV are observed at 20 and 36 dpi.

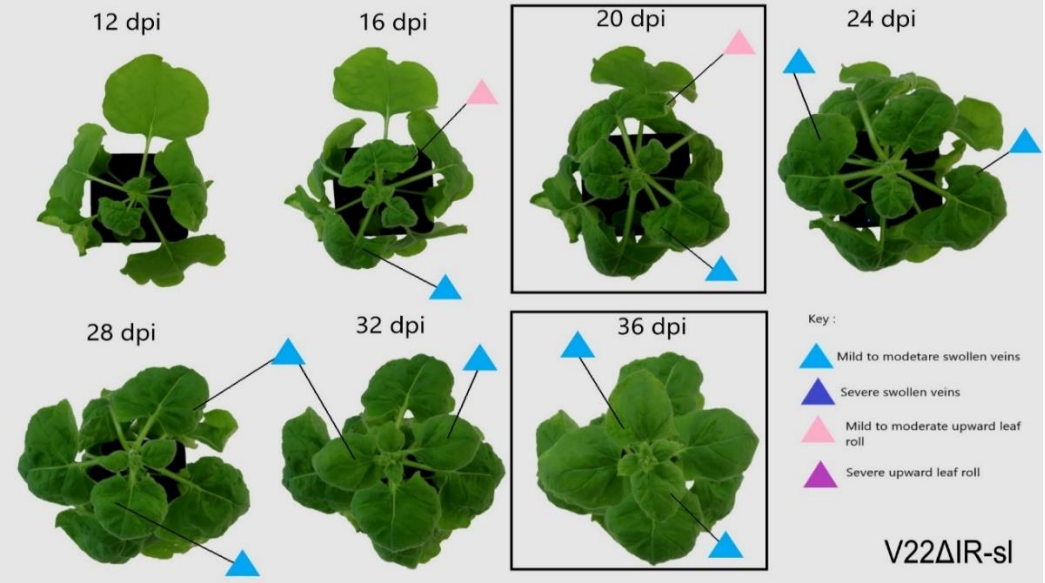


Figure 1.9 Symptom progression images of *Nicotiana benthamiana* inoculated with the V22ΔIR-sl infectious clone construct at 12-, 16-, 20-, 24-, 28-, 32-, and 36 days post inoculation (dpi). Symptoms present are upward leaf roll (ULR) and swollen veins (SV). Severity of symptoms shown by colour coded triangles as indicated in key. Twenty days post inoculation (dpi) is boxed as it is a point of interest as this is the time point where viral load is at its peak and also the time point at which we see majority of differences in ULR and SV induced by the mutant infectious clone constructs in this study. 36 dpi is boxed as this is the time point post recovery of ULR and SV. Mild ULR and SV at 20 dpi and only mild SV at 36 dpi.

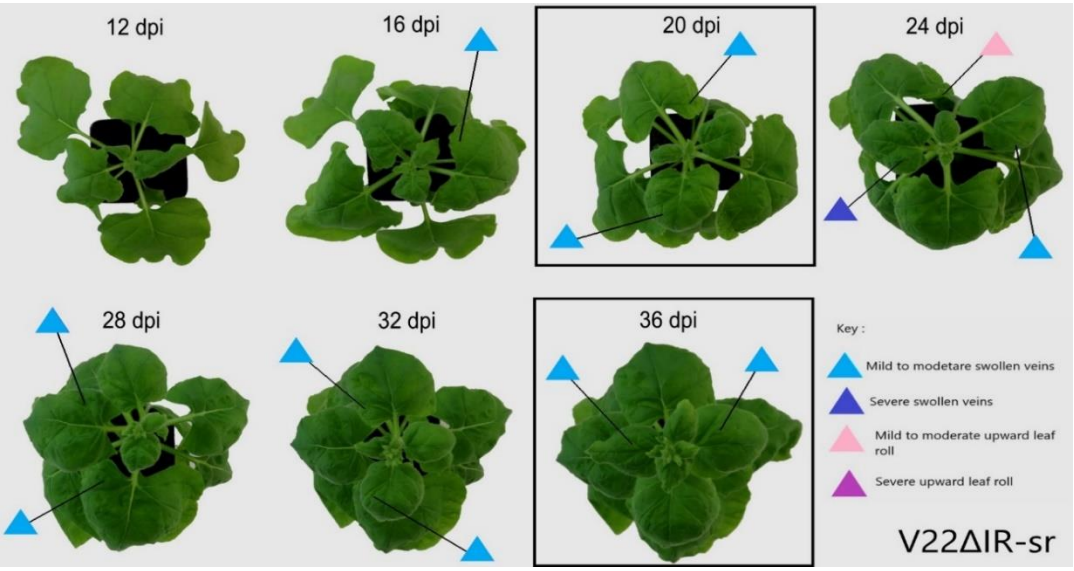


Figure 1.10 Symptom progression images of *Nicotiana benthamiana* inoculated with the V22ΔIR-sr infectious clone construct at 12-, 16-, 20-, 24-, 28-, 32-, and 36 days post inoculation (dpi). Symptoms present are upward leaf roll (ULR) and swollen veins (SV). Severity of symptoms shown by colour coded triangles as indicated in key. Twenty days post inoculation (dpi) is boxed as it is a point of interest as this is the time point where viral load is at its peak and also the time point at which we see majority of differences in ULR and SV induced by the mutant infectious clone constructs in this study. 36 dpi is boxed as this is the time point post recovery of ULR and SV. Only mild SV at 20 and 36 dpi were induced.

Section 2 – The V2 open reading frame (ORF)

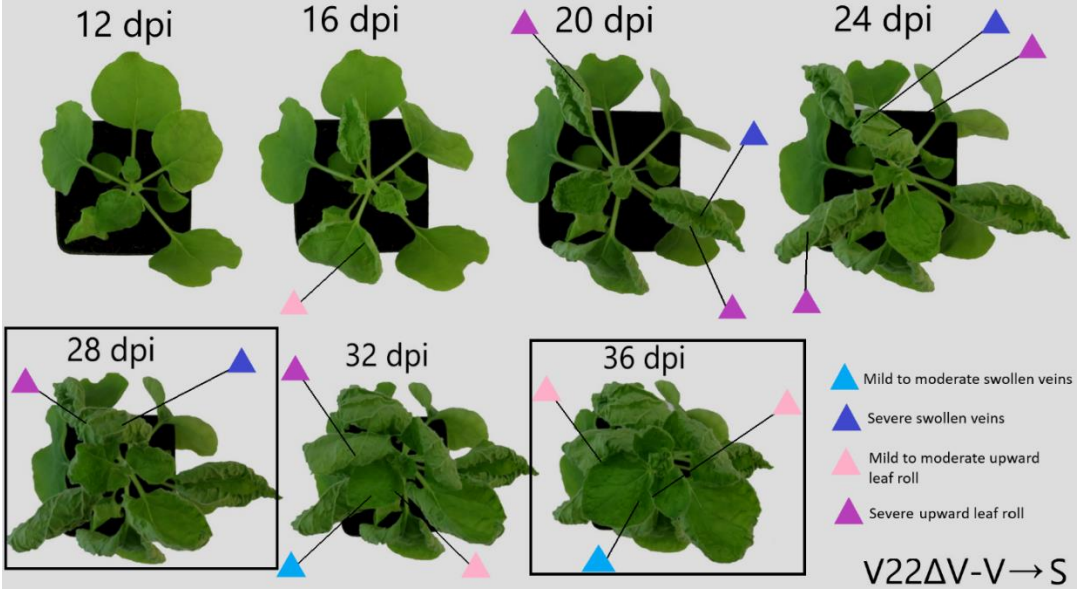


Figure 2.1 Symptom progression images of *N. benthamiana* inoculated with the V22ΔV-V→S infectious clone construct at 12-, 16-, 20-, 24-, 28-, 32-, and 36 days post inoculation (dpi). Symptoms present are upward leaf roll (ULR) and swollen veins (SV). Severity of symptoms shown by colour coded triangles as indicated in key. Twenty-eight dpi and 36 dpi are boxed to show time points at which the differences in ULR and SV when compared to V22 are the most apparent. Severe ULR and SV at 28 dpi and mild SV and ULR at 36 dpi are present.

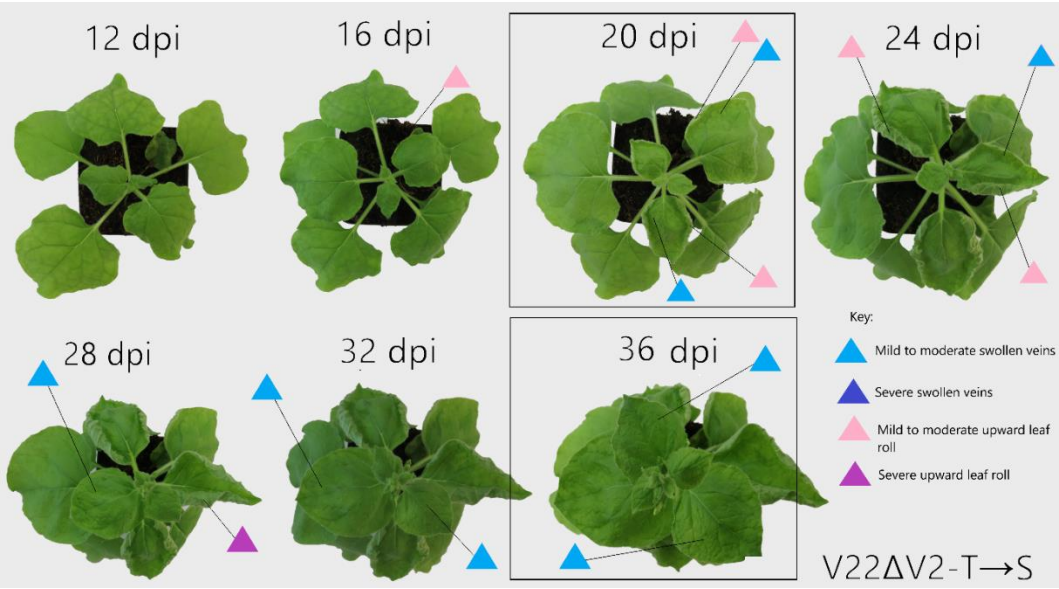


Figure 2.2 Symptom progression images of *N. benthamiana* inoculated with the V22ΔV2-T→S infectious clone construct at 12-, 16-, 20-, 24-, 28-, 32-, and 36 days post inoculation (dpi). Symptoms present are upward leaf roll (ULR) and swollen veins (SV). Severity of symptoms shown by colour coded triangles as indicated in key. Twenty days post inoculation (dpi) is boxed as it is a point of interest as this is the time point where viral load is at its peak and also the time point at which we see majority of differences in ULR and SV induced by the mutant infectious clone constructs in this study. 36 dpi is boxed as this is the time point post recovery of ULR and SV. Mild SV and ULR seen at 20 dpi and only mild SV observed at 36 dpi.

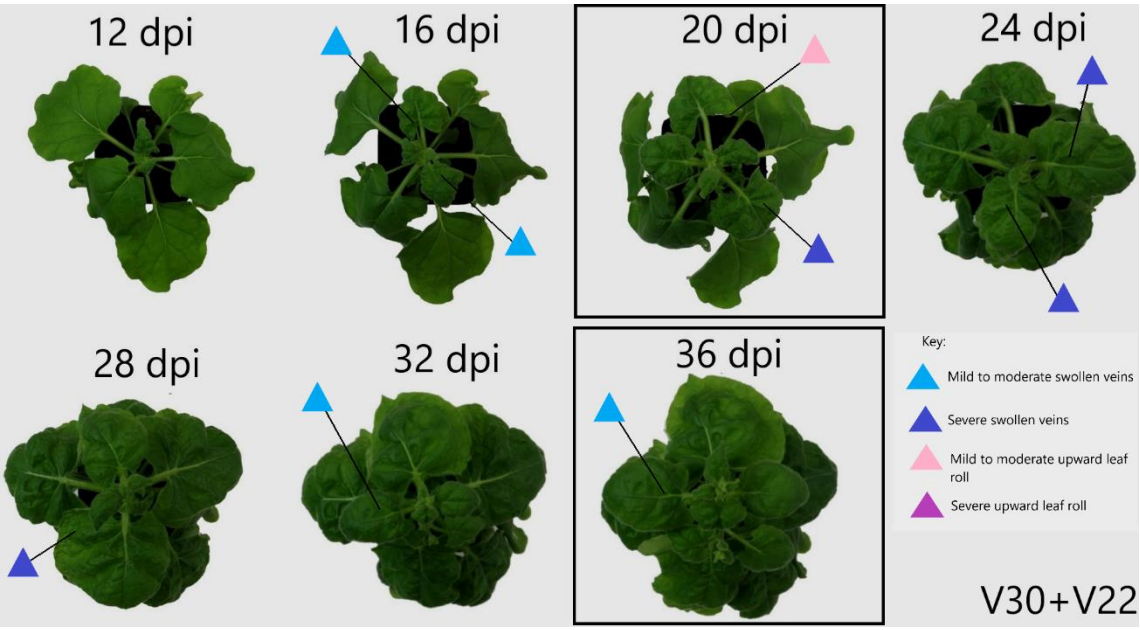


Figure 2.3 Symptom progression images of *N. benthamiana* inoculated with the V30+V22 mixed infection construct at 12-, 16-, 20-, 24-, 28-, 32-, and 36 days post inoculation (dpi). Symptoms present are upward leaf roll (ULR) and swollen veins (SV). Severity of symptoms shown by colour coded triangles as indicated in key. Twenty days post inoculation (dpi) is boxed as it is a point of interest as this is the time point where viral load is at its peak and also the time point at which we see majority of differences in ULR and SV induced by the mutant infectious clone constructs in this study. 36 dpi is boxed as this is the time point post recovery of ULR and SV. Downward cupping and bulging evident at 20 dpi with mild ULR and severe SV. Only mild SV evident at 36 dpi.

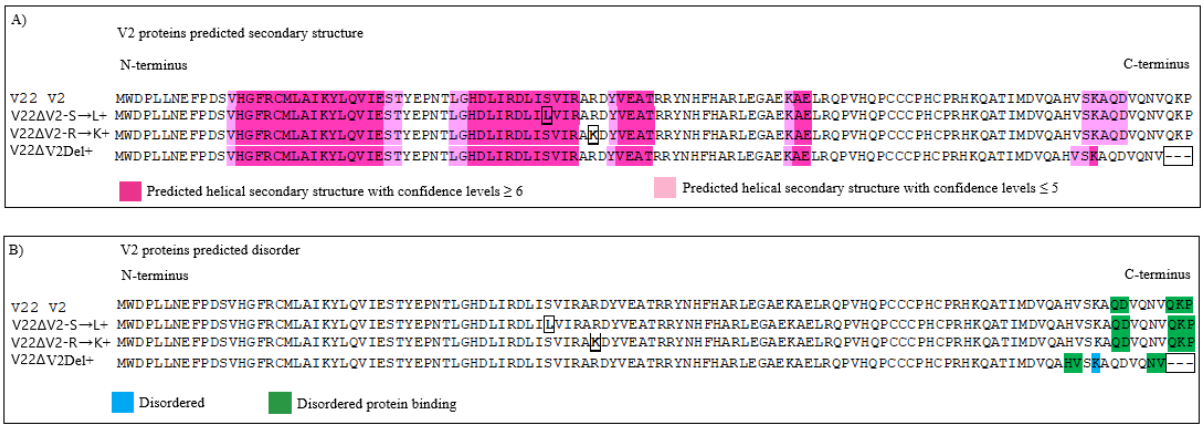


Figure 2.4 V2 protein sequences of V22 (wildtype), V22ΔV2-S→L+, V22ΔV2-R→K+ and V2Del+. A) Secondary structures indicated by dark pink (high confidence level prediction of ≥ 6) and by light pink (lower confidence level prediction of ≤ 5). The amino acid changes as a result of nucleotide mutagenesis indicated by the boxes around the residues. B) Disordered residues are highlighted in blue and disordered protein binding residues are highlighted in green. The threshold for prediction of residues to be disordered was 0.5. Disorder and secondary structural predictions obtained from http://bioinf.cs.ucl.ac.uk/psipred/ .



Figure 2.5 Putative C6 protein sequences of V22 (wildtype) and V22ΔV2-S→L+ and V22ΔV2-R→K+. A) Secondary structures indicated by dark pink (high confidence level prediction of ≥ 6) and by light pink (lower confidence level prediction of ≤ 5). The amino acid changes as a result of nucleotide mutagenesis indicated by the boxes around the residues. B) Disordered residues are highlighted in blue and disordered protein binding residues are highlighted in green. The threshold for prediction of residues to be disordered was 0.5. Disorder and secondary structural predictions obtained from http://bioinf.cs.ucl.ac.uk/psipred/ .

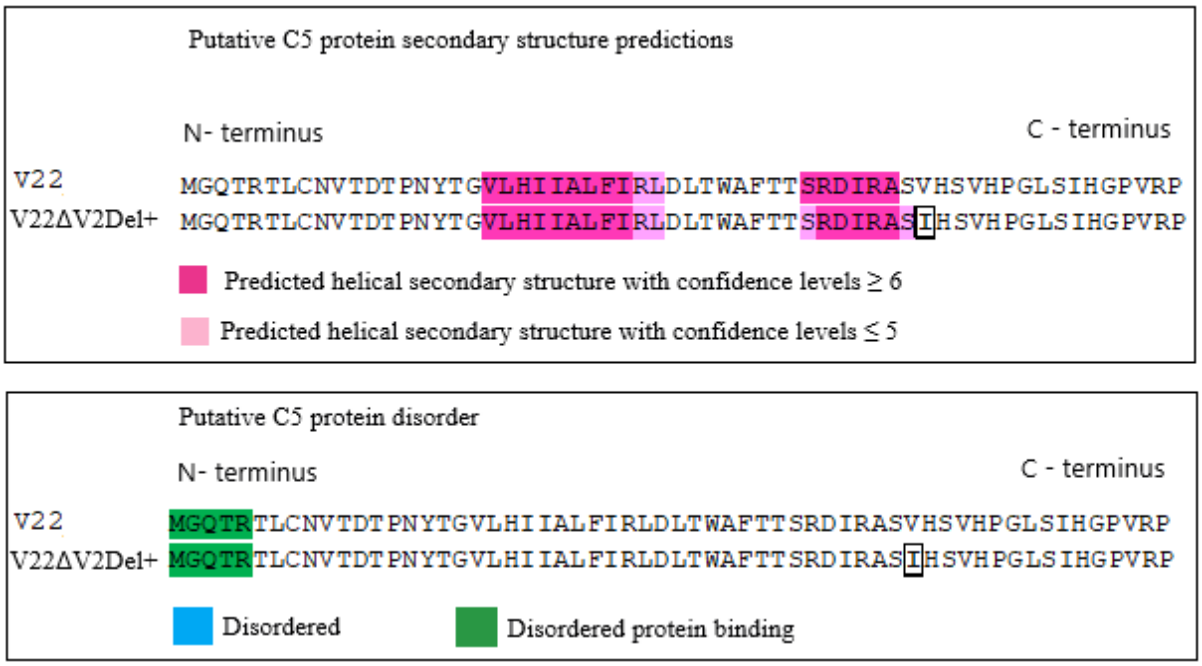


Figure 2.6 Putative C5 protein sequences of V22 (wildtype) and V22ΔV2Del+ A) Secondary structures indicated by dark pink (high confidence level prediction of ≥ 6) and by light pink (lower confidence level prediction of ≤ 5). The amino acid changes as a result of nucleotide mutagenesis indicated by the boxes around the residues. B) Disordered residues are highlighted in blue and disordered protein binding residues are highlighted in green. The threshold for prediction of residues to be disordered was 0.5. Disorder and secondary structural predictions obtained from http://bioinf.cs.ucl.ac.uk/psipred/ .

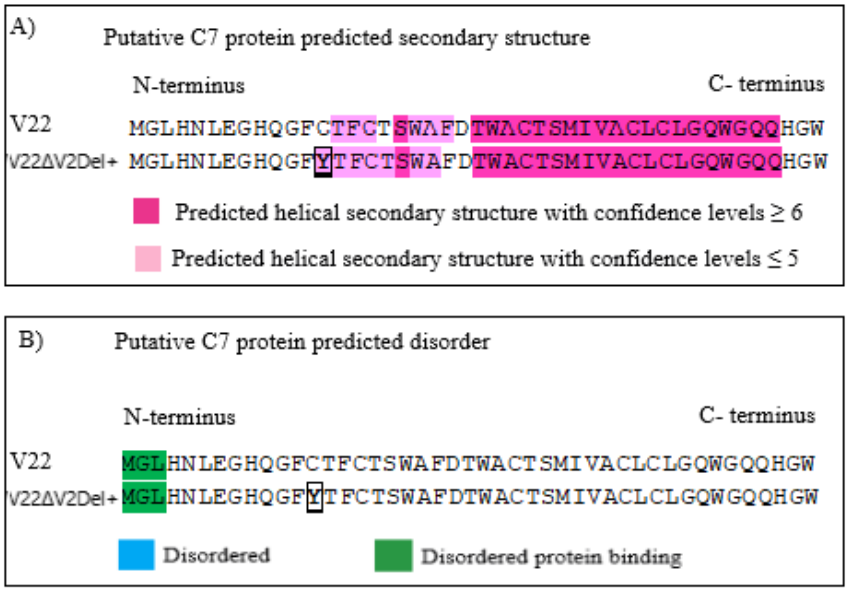


Figure 2.7 Putative C7 protein sequences of V22 (wildtype) and V22ΔV2Del+ A) Secondary structures indicated by dark pink (high confidence level prediction of ≥ 6) and by light pink (lower confidence level prediction of ≤ 5). The amino acid changes as a result of nucleotide mutagenesis indicated by the boxes around the residues. B) Disordered residues are highlighted in blue and disordered protein binding residues are highlighted in green. The threshold for prediction of residues to be disordered was 0.5. Disorder and secondary structural predictions obtained from http://bioinf.cs.ucl.ac.uk/psipred/ .

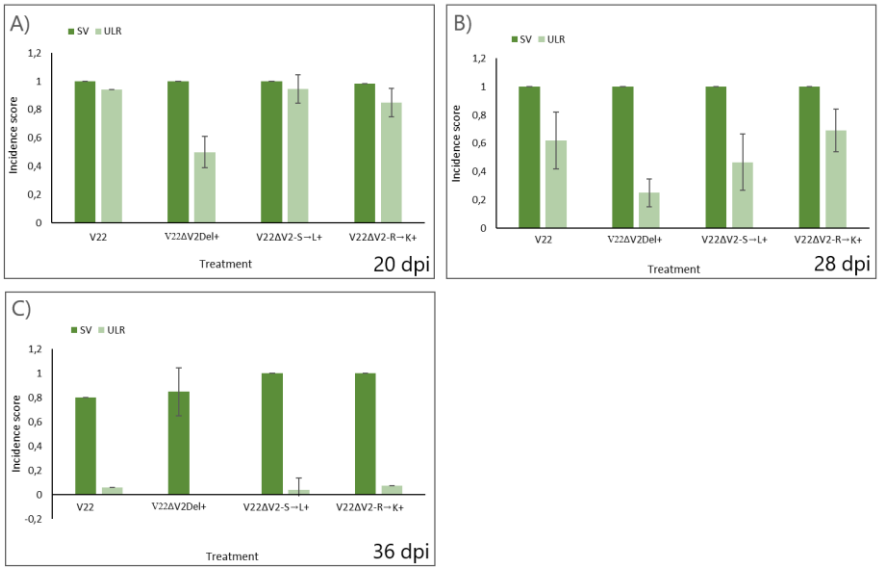


Figure 2.8 Incidence scores of V22, V22ΔV2Del+, V22ΔV-S→L+, V22ΔV2-R→K+, at A) 20 dpi, B) 28 dpi and C) 36 dpi from Prelim trials. Standard deviation was calculated between the individual plants from the preliminary trial. score of 1 was given to a plant if a symptom was present and a score of 0 was given if the symptom was not observed.

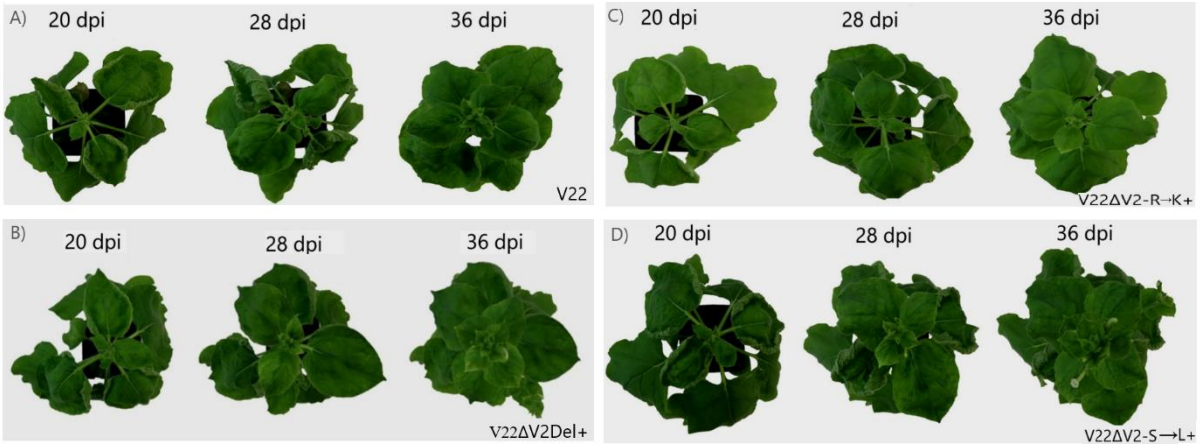


Figure 2.9 Images of *Nicotiana benthamiana* inoculated with A) V22 wildtype (WT), B) V22ΔV2Del+, C) V22ΔV2-R→K+ and D) V22ΔV2-S→L+ at 20-, 28- and 36-days post inoculation (dpi).

Section 3 – Investigation of the putative C5 open reading frame (ORF)

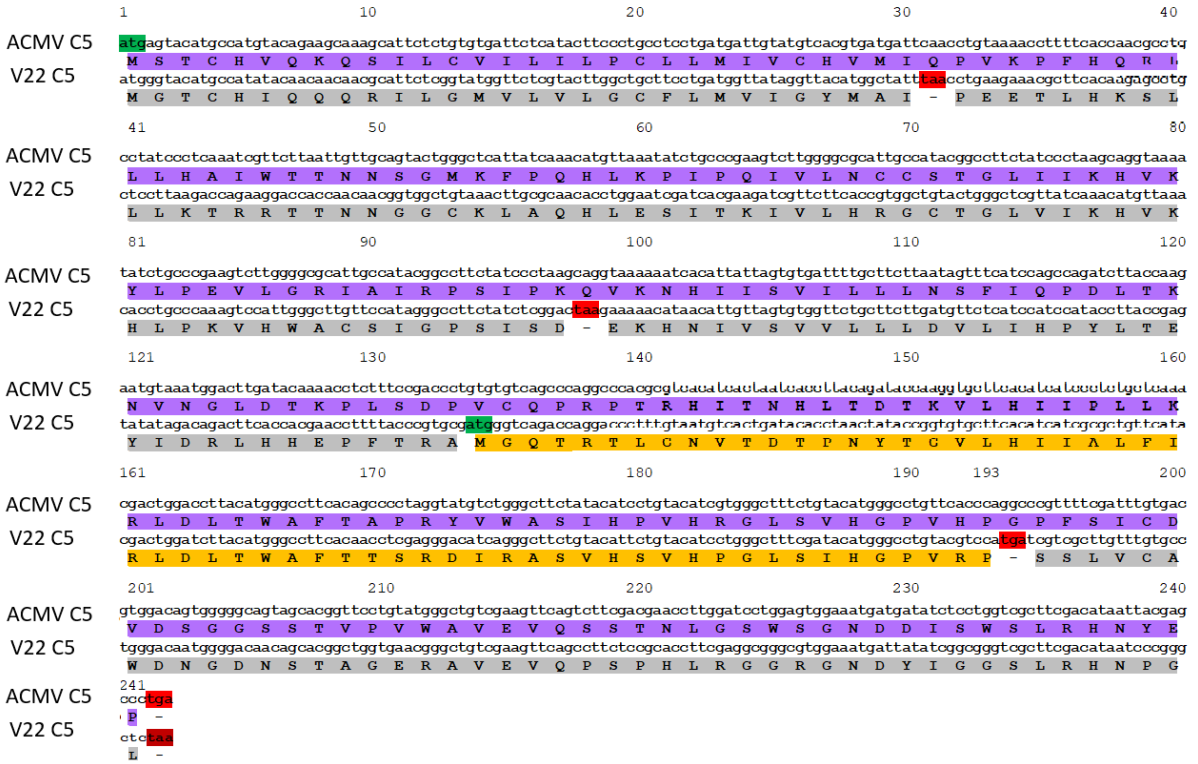


Figure 3.1 A) Nucleotide (nt) and amino acid (aa) alignments (https://web.expasy.org/translate/) of the C5 protein of African cassava mosaic virus (ACMV, Accession number J02057) and V22. The start codons are highlighted in green, and the stop codons are highlighted in red. The ACMV C5 predicted translation of 50 amino acids (aa) or greater is highlighted in violet. The V22 C5 protein sequence predicted to be translated (≤50aa) is highlighted in orange. The V22 untranslated C5 protein sequence is highlighted in grey.



Figure 3.1 B) CLUSTAL multiple sequence alignment by MUSCLE (3.8) of wild type AC5 of African Cassava Mozaic virus (ACMV) and the full length 193 aa long putative C5 in mutant V22ΔC5aa$^{1\text{-}193}$ with the degree of similarity of the residues by means of structure and polarity indicated by symbols below the sequences.

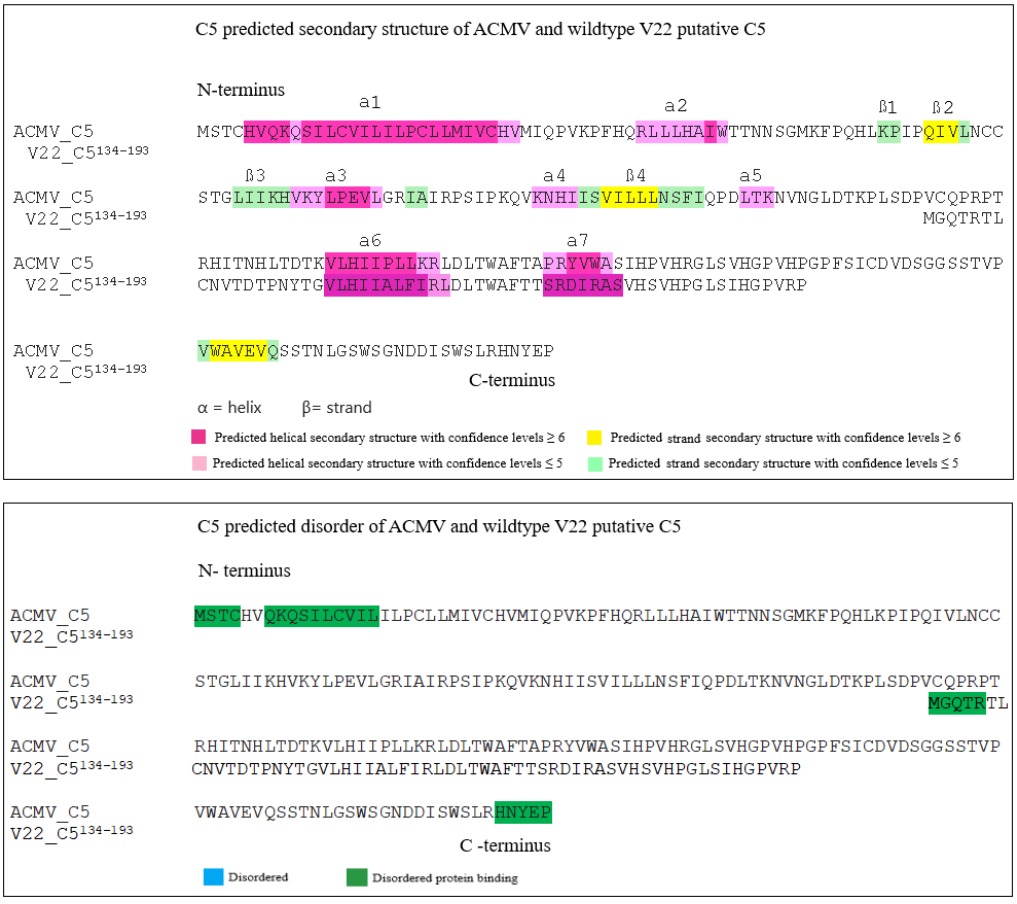


Figure 3.2 C5 protein sequences of African cassava mosaic virus (ACMV), (Accession number J02057) and C5ΔC5$^{134\text{-}193.}$ A) Predicted secondary structural features of alpha helices indicated by dark pink (high confidence level prediction of ≥ 6) and by light pink (lower confidence level prediction of ≤ 5). Beta strands predicted at a high confidence level ≥6 are shown in yellow, those predicted with a low level of confidence are shown in mint.

The amino acid changes as a result of nucleotide mutagenesis indicated by the boxes around the residues. α represents a helix and β represents a strand. B) Disordered residues are highlighted in blue and disordered protein binding residues are highlighted in green. The threshold for prediction of residues to be disordered was 0.5. Disorder and secondary structural predictions obtained from http://bioinf.cs.ucl.ac.uk/psipred/ .

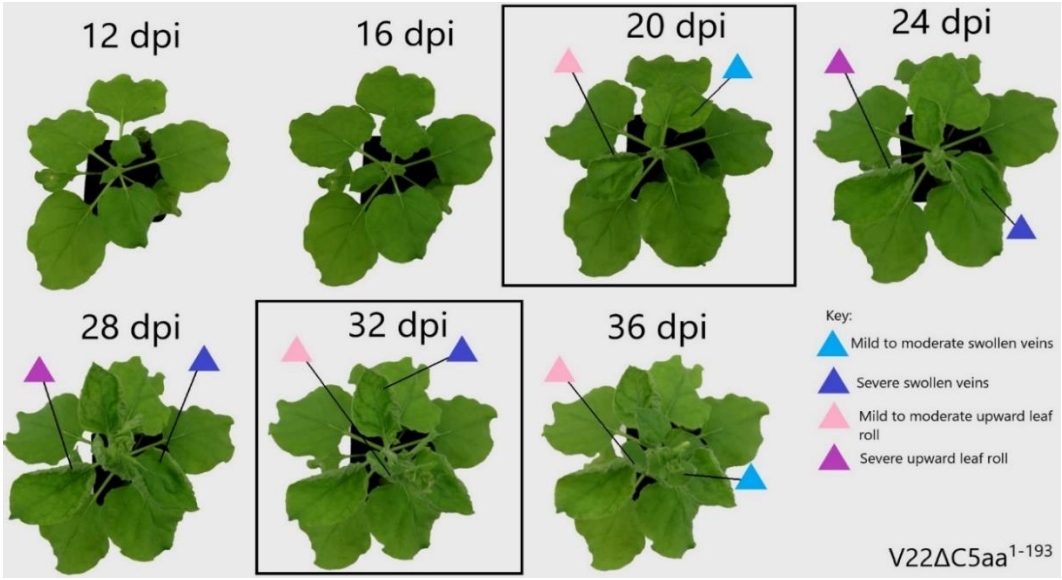


Figure 3.3 Symptom progression images of *Nicotiana benthamiana* inoculated with the V22ΔC5aa$^{1-193}$ infection construct at 12-, 16-, 20-, 24-, 28-, 32-, and 36 days post inoculation (dpi). Symptoms present are upward leaf roll (ULR) and swollen veins (SV). Severity of symptoms shown by colour coded triangles as indicated in key. Twenty- and 32-days dpi are boxed as they are the time points at which we see majority of differences in ULR and SV induced by the mutant infectious clone constructs when compared to wildtype V22 in this study.

GenScript

Make Research Easy

CONFIDENTIAL

Certificate of Analysis

Project ID: U0149FG070-8

Construct Information:

Gene Name: pCAMBIA2300_V22_MOD5(V22ΔIR-s)

Clone ID: K10791 **Gene Length:** 179 bp

Cloning Vector: pCAMBIA2300_V22 **Cloning Strategy:** SacI/BamHI

QC Items	Specifications	Results	
Sequencing Alignment	Sequencing results are consistent with the targeted insert sequence.	Pass	Consistent
Vector Sequence	The flanking sequences of the cloning site are correct.	Pass	Correct
			Shown in the SQD file
Restriction Digests	The size of inserted fragment is correct and free of unexpected bands.	Pass	Correct
			Shown in attachment 1
DNA Quality	Miniprep: 4 μg	Pass	≥ 4 μg
	OD260/280=1.8~2.0		OD260/280=1.88
	Free of contamination		Pure
Quality Grade	Research Grade	Pass	Research Grade
Appearance	Clear and free of foreign particles.	Pass	Clear
			Free of foreign particles
Additional Test	N/A	N/A	

NOTE

Shipping at	Plasmid Storing at	Bacstab Storing at	Glycerol Stock Storing at
Room Temperature	-20°C	4°C	-20°C/-80°C

Certified by: Morgan Wu **Date:** 08/29/2020

Thank you for your patronage to our Gene services! To maintain this working relationship, we shall be grateful if you can add our webpage URL into your lab website. As a token of appreciation, you will be rewarded by 1,000 EZcoupon™ points. For more information, please contact us by e-mail at webt@genscript.com.

GenScript

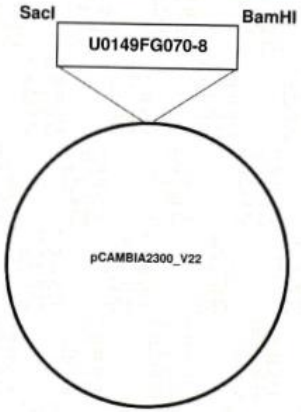
Make Research Easy

CONFIDENTIAL

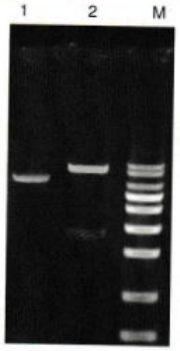
Attachment 1

Plasmid Construct Map

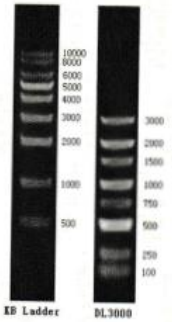
The gene was cloned in pCAMBIA2300_V22 by SacI/BamHI .



Enzyme Digestion



Lane M: KB Ladder
Lane 1:U0149FG070-8 plasmid
Lane 2:U0149FG070-8 plasmid digested by BamHI



Digestion Conditions:
About 300ng plasmid digested
Digestion in water-bath,37°C for 40 minutes
1% Agarose Gel

GenScript

Make Research Easy

CONFIDENTIAL

Certificate of Analysis

Project ID: U0149FG070-2

Construct Information:

Gene Name: pCAMBIA2300_V22_MOD3(V22ΔIR-sl)

Clone ID: K10781 **Gene Length:** 178 bp

Cloning Vector: pCAMBIA2300_V22 **Cloning Strategy:** SacI/BamHI

QC Items	Specifications	Results	
Sequencing Alignment	Sequencing results are consistent with the targeted insert sequence.	Pass	Consistent
Vector Sequence	The flanking sequences of the cloning site are correct.	Pass	Correct Shown in the SQD file
Restriction Digests	The size of inserted fragment is correct and free of unexpected bands.	Pass	Correct Shown in attachment 1
DNA Quality	Miniprep: 4 μg OD260/280=1.8~2.0 Free of contamination	Pass	≥ 4 μg OD260/280=1.88 Pure
Quality Grade	Research Grade	Pass	Research Grade
Appearance	Clear and free of foreign particles.	Pass	Clear Free of foreign particles
Additional Test	N/A	N/A	

NOTE

Shipping at	Plasmid Storing at	Bacstab Storing at	Glycerol Stock Storing at
Room Temperature	-20°C	4°C	-20°C/-80°C

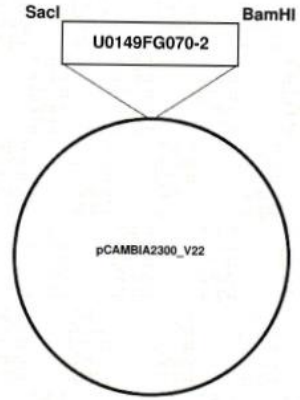
Certified by: Morgan Wu **Date:** 08/29/2020

Thank you for your patronage to our Gene services! To maintain this working relationship, we shall be grateful if you can add our webpage URL into your lab website. As a token of appreciation, you will be rewarded by 1,000 EZcoupon™ points. For more information, please contact us by e-mail at web@genscript.com.

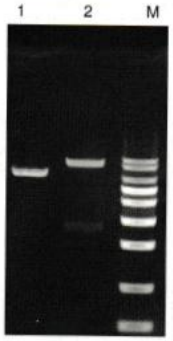
Attachment 1

Plasmid Construct Map

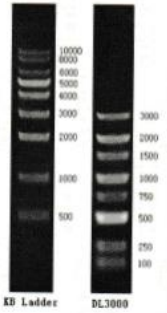
The gene was cloned in pCAMBIA2300_V22 by SacI/BamHI .



Enzyme Digestion



Lane M: KB Ladder
Lane 1:U0149FG070-2 plasmid
Lane 2:U0149FG070-2 plasmid digested by BamHI



Digestion Conditions:
About 300ng plasmid digested
Digestion in water-bath,37°C for 40 minutes
1% Agarose Gel

GenScript

Make Research Easy

CONFIDENTIAL

Certificate of Analysis

Project ID: U0149FG070-6

Construct Information:

Gene Name: pCAMBIA2300_V22_MOD4 (V22ΔIR-sr)

Clone ID: K10787 **Gene Length:** 179 bp

Cloning Vector: pCAMBIA2300_V22 **Cloning Strategy:** SacI/BamHI

QC Items	Specifications	Results	
Sequencing Alignment	Sequencing results are consistent with the targeted insert sequence.	Pass	Consistent
Vector Sequence	The flanking sequences of the cloning site are correct.	Pass	Correct
			Shown in the SQD file
Restriction Digests	The size of inserted fragment is correct and free of unexpected bands.	Pass	Correct
			Shown in attachment 1
DNA Quality	Miniprep: 4 μg	Pass	≥ 4 μg
	OD260/280=1.8~2.0		OD260/280=1.88
	Free of contamination		Pure
Quality Grade	Research Grade	Pass	Research Grade
Appearance	Clear and free of foreign particles.	Pass	Clear
			Free of foreign particles
Additional Test	N/A	N/A	

NOTE

Shipping at	Plasmid Storing at	Bacstab Storing at	Glycerol Stock Storing at
Room Temperature	-20°C	4°C	-20°C/-80°C

Certified by: Morgan Wu **Date:** 08/29/2020

Thank you for your patronage to our Gene services! To maintain this working relationship, we shall be grateful if you can add our webpage URL into your lab website. As a token of appreciation, you will be rewarded by 1,000 EZcoupon™ points. For more information, please contact us by e-mail at web@genscript.com.

GenScript

Make Research Easy

CONFIDENTIAL

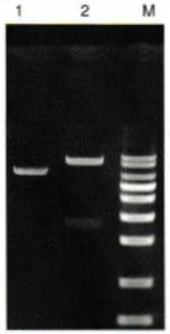
Attachment 1

Plasmid Construct Map

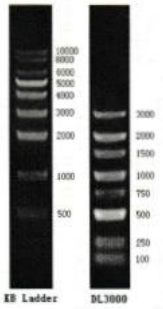
The gene was cloned in pCAMBIA2300_V22 by SacI/BamHI .



Enzyme Digestion



Lane M: KB Ladder
Lane 1:U0149FG070-6 plasmid
Lane 2:U0149FG070-6 plasmid digested by BamHI



Digestion Conditions:
About 300ng plasmid digested
Digestion in water-bath,37°C for 40 minutes
1% Agarose Gel

For research use only

860 Centennial Ave., Piscataway, NJ 08854, USA

Toll-Free: 1-877-436-7274 Tel: 1-732-885-9188 Fax: 1-732-210-0262 Email: order@genscript.com Web: www.genscript.com

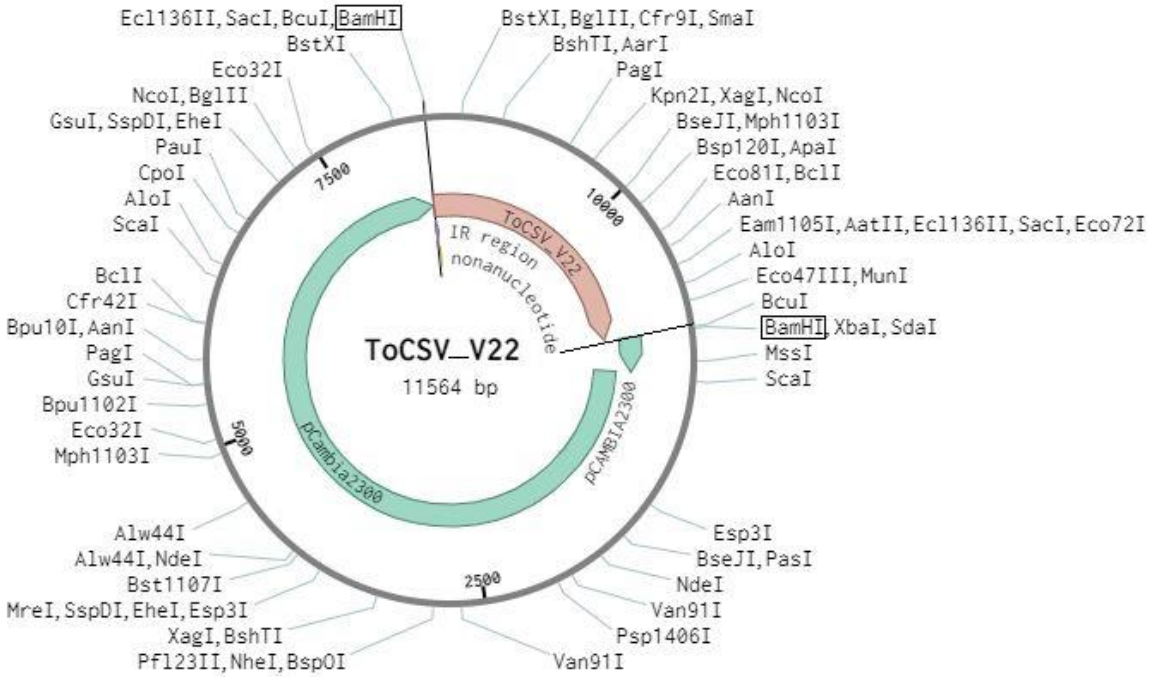


Figure 3.3 A vector map of the V22 infectious clone construct with a sequence length of 11 564 base pairs (bp) Nonanucleotide in the IR and restriction sites labelled. BamHI is boxed as these are the sites used for verification of the infectious clones by Genscript. Vector map generated on Benchling (https://benchling.com/).

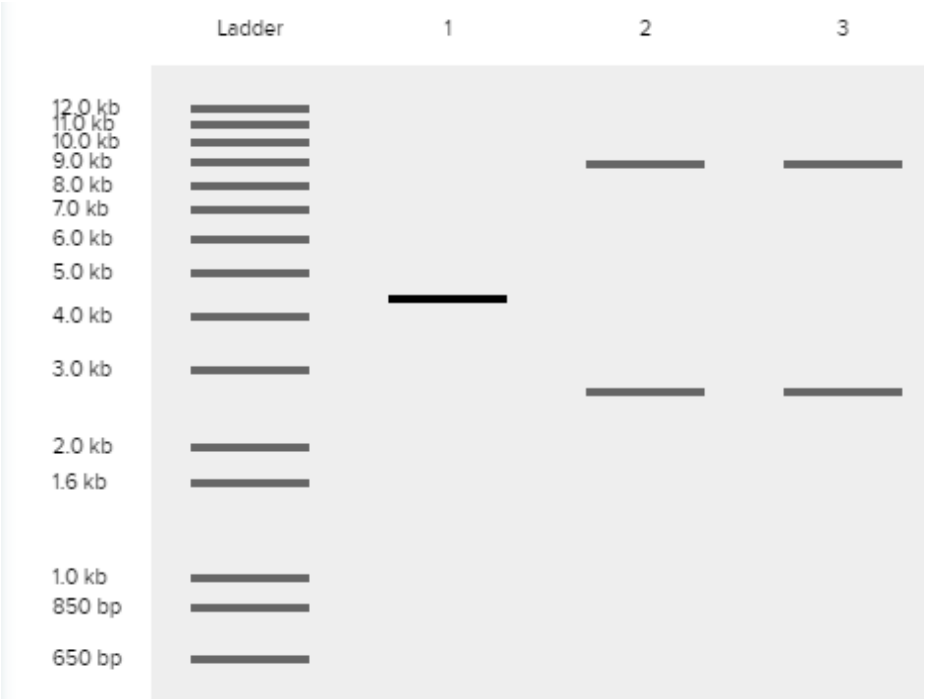


Figure 3.4 Virtual digest of ToCSV_V22 targeting BamHI shown in lane 2 and undigested ToCSV_V22 in lane 1. Product lengths depicted in lane 2 are 2665 and 8899 base pairs (bp) in length. Digest of V22ΔIR-s with BamHI in lane 3 has the same product sizes as seen in lane 2. Virtual digest generated using Benchling (https://benchling.com/).

Table 3. 1 Primer sequences for sequencing of clone constructs.

Primer	**Sequence (5'-3')**	**Templates and target sequences.**
ToCSVseq_FP	GTCGTGCTCCACCATGTTG	V22 IR-swap mutants: V22ΔIR-s, V22ΔIR-sl and V22ΔIR-sr (Primers target either side of the IR for sequencing with a product size of 634)
ToCSVseq_RP	CACCACGAACCTTTTACCCG	
V22IR_FP2	AAACTCCTAAAGCGGCCATC	V30 IR-swap mutant: V30ΔIR-s (Primers target either side of the IR for sequencing a product size of 630)
ToCSVseq_RP	CACCACGAACCTTTTACCCG	
V22V2_FP	GTGCGGATGTCATAATGTGG	Single V2 mutants: V22ΔV2-V→S, V22ΔV2-S→L, V22ΔV2-R→K, V22ΔV2-T→S, V22ΔV2T→S+ (Primers target either side of V2 for sequencing with a product size of 478 bp)
V22V2_RP	GACCAGGACCCTTTGTAATGTC	
C5seq_FP	ACGATCATGGACGTACAGGC	V22 C5 mutant: V22ΔC5aa$^{1\text{-}193}$ (Primers target either side of C5 for sequencing with a product size of 563 bp).
C5seq_RP	ACAACGCATTCTCGGTATGGT	

*bp= base pairs