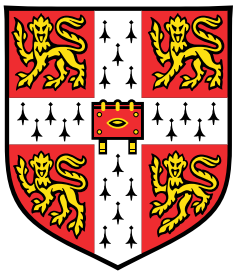
Engineering resistance to maize lethal necrosis



Luke Anthony Braidwood

Department of Plant Sciences
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Trinity College December 2017

To my family, for everything.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Luke Anthony Braidwood
December 2017

Acknowledgements

The work contained in this thesis would not have been possible, or anywhere near as enjoyable, without the contributions of my colleagues and friends. I would like to thank them all. Firstly, David Baulcombe, for welcoming an unknown DTP student into his lab and allowing me to pursue an exciting project with both guidance and freedom. To a great number of post-docs who have provided advice, demonstration, and a shoulder to whine on when times were tough – in particular Donna Bond, Adrian Valli, Betty Chung, and Sebastian Mueller. To John Welch, for guidance in performing phylogenetic analyses. For assistance, technical and administrative, I would like to thank Mel Steer, James Barlow, and Ombretta Orsini. My fellow PhD students, who have acted as teachers, friends, and students variously, especially Claire Agius, Patrick Diaz, and my comrades Alex Canto-Pastor and Catherine Griffin. Special mention goes to Sasha Blackwell for teaching me to push hard and perform reverse peristalsis whilst on bike rides.

This work took me to Kenya, where I was clueless and alone, but not for long thanks to the warm welcomes of so many. KALRO enabled me to sample and survey across Kenya, and in particular thanks goes to Anne Wangai, Jane Wamaitha, Bramwel Wanjala, and Cyrus Mugambi. The farmers who were happy to welcome me onto their land, allowed me to take samples and often talked at length about their farming practices made this work possible. Likewise at KU I would like to thank Steven Runo for his guidance and Joel Masanga for his extensive work on transforming tropical maize. Assistance in sample shipping from Musembi Mutuku at BeCA was invaluable.

Life outside the lab has been harmonious yet bassy thanks to my housemates David Barrett and Charlie Daniels, who have been perfect cohabitants thanks to their attitudes, music tastes (excluding gypsycore), and similarly creative approaches to interior design. Thanks to Luke O'Keefe, Patrick, and Rascal for the jungle gym and making life infinitely more hilarious. To Dean for being a calming influence. Finally to my family and friends for getting me this far.

Abstract

Modern agriculture is dependent on both global supply chains and crop monocultures. These features aid the evolution and spread of novel plant pathogens. Limited genetic diversity in commercial crop lines can result in widespread susceptibility to emerging pathogens. Pathogen resistance may be developed through conventional breeding approaches, or a number of transgenic strategies. This thesis focuses on the characterisation of an emerging maize disease, Maize lethal necrosis (MLN), and engineering resistant maize lines using an artificial microRNA (amiRNA) approach.

MLN is a synergistic viral disease caused by the interaction of *Maize chlorotic mottle virus* (MCMV) with any maize-infecting member of the potyviridae. I used next-generation RNA sequencing to characterise the MLN outbreak in East Africa, discovering that local and Chinese strains of the potyvirus *Sugarcane mosaic virus* (SCMV) typically coinfect with MCMV. A first global MCMV phylogeny was constructed using these samples combined with new Sanger sequencing of samples in Ecuador and Hawaii. The phylogeny supported previous hypotheses of a link between the Chinese and African outbreaks, and suggested a novel link between the Hawaiian and Ecuadorian outbreaks. The SCMV sequences generated demonstrated strong evidence of extensive recombination, in line with previous reports on SCMV and potyviruses. These data also produced first reports of a number of RNA viruses in East Africa, and five novel viral-like sequences, with their presence confirmed by RT-PCR.

RNA silencing is an important component of the plant immune response to viral infection. amiRNAs can be used to generate specific and effective viral resistance through Watson-Crick base pairing between the amiRNA and the (RNA) viral genome. Previous amiRNA approaches have targeted invariable genomic regions using consensus sequences. However, the high mutation rate of RNA viruses means single cells contain a variety of mutant genomes, collectively called a quasispecies. To deter the evolution of resistance breaking I devised a novel strategy to include intra-sample variation from NGS data in amiRNA design, and constructs, each containing five of these amiRNAs, were transformed into tropical maize lines.

RNA silencing may be hampered by the expression of viral suppressors of silencing (VSRs). Local VSR assays demonstrated that there are no local VSRs in the MCMV genome, while systemic VSR assays showed a possible systemic VSR role for the unique P32 protein, and an interesting link between photoperiod and systemic silencing more generally.

So it goes - Kurt Vonnegut

Table of contents

List of figures

List of tables

Nomenclature

Acronyms / Abbreviations

AGO Argonaute

ALPV Aphid lethal paralysis virus

amiRNA Artificial micro-RNA

APV Acyrthosiphon pisum virus

ARI Adjusted Rand index

BMV Brome mosaic virus

bp base pair

BRVF Black raspberry virus F

CarMV Carnation mottle virus

cDNA Complementary DNA

CDS Coding sequence

CIBE-ESPOL Centro de Investigaciones Biotecnológicas del Ecuador, Escuela Superior Politécnica del Litoral

CIMMYT The International Maize and Wheat Improvement Center

CITE Cap-independent translation enhancer

CMV Cucumber mosaic virus

CP Coat protein

CYVA Citrus yellow vein-associated virus

DAMP Damage associated molecular patterns

DCL Dicer-like proteins

DMSO Dimethyl sulfoxide

DNA Deoxyribonucleic acid

dNTPs Deoxynucleotide

DRB dsRNA-binding proteins

EF-Tu Elongation factor Tu

EST Expressed sequence tag

ETI Effector-triggered immunity

ETS Effector-triggered susceptibility

FAOSTAT The Food and Agriculture Organization Corporate Statistical Database

FERA Food and Environment Research Agency

gDNA Genomic DNA

GFP Green fluorescent protein

GTR Generalised time-reversible

GWAS Genome-wide association study

HC-Pro Helper-component protease

HEN1 Hua enhancer1

HR Hypersensitive cell death response

IITA The International Institute of Tropical Agriculture

IR Inverted repeat

IVT *In-vitro* transcript

JGMV Johnsongrass mosaic virus

KALRO Kenyan Agriculture and Livestock Research Organisation

kb Kilobase

KU Kenyatta University

LB Lysogeny broth

LRR Leucine-rich repeat

MATV-Ec Maize-associated totivirus-Ec

MCMC Monte carlo Markov Chain

MCMV Maize chlorotic mottle virus

MDMV Maize dwarf mosaic virus

MIR miRNA gene

miRNA Micro-RNA

MLN Maize lethal necrosis

MP Movement protein

mRNA Messenger ribonucleic acid

MS Murashige and Skoog (media)

MWLMV Maize white line mosaic virus

MYDV Maize yellow dwarf virus-RMV

NB-LRR Nucleotide-binding leucine-rich repeat

NCBI National Center for Biotechnology Information

NEB New England Biolabs

NGS Next-generation sequencing

NLHV Nilaparvata lugens honeydew virus-3

nt Nucleotide

OE-PCR Overlap-extension PCR

ORF Open reading frame

PAMP Pathogen associated molecular patterns

PCR Polymerase chain reaction

PDR Pathogen-derived resistance

PIPO Pretty interesting potyviridae ORF

PPV Plum pox virus

PRR Pattern recognition receptor

PTI Pattern-triggered immunity

PVX Potato virus X

PVY Potato virus Y

QC Quality-control

QTL Quantitative trait loci

QUASR Quality Assessment of Short Read

RACE Rapid amplification of cDNA ends

RC Replication complex (viral)

RdDM RNA-dependent DNA methylation

RDP Recombination detection programme

RdRP RNA-dependent RNA polymerase (viral)

RDR RNA-dependent RNA polymerase (plant)

RISC RNA-induced silencing complex

RNAi RNA interference

RNA Ribonucleic acid

RPV Rhopalosiphum padi virus

rRNA Ribosomal RNA

RT-PCR Reverse transcription polymerase chain reaction

RT-qPCR Quantitative reverse transcription polymerase chain reaction

RVF Ribes virus F

SABV Soybean-associated bicistronic virus

SCMV Sugarcane mosaic virus

sgRNA Subgenomic RNA

siRNA Small-interfering RNA

SNP Single-nucleotide polymorphism

sRNA Small RNA

SSC Saline sodium citrate

ssRNA Single-stranded RNA

tasiRNA Trans-acting siRNA

TAS tasiRNA gene

TBE Tris/Borate/EDTA

TBSV Tomato bushy stunt virus

T-DNA Transfer-DNA

TIR-NB-LRR Toll-interleukin-1-NB-LRR

TMV Tobacco mosaic virus

TuMV Turnip mosaic virus

TYMV Turnip yellow mosaic virus

UTR Untranslated region

VLS Viral-like sequence

vsiRNA Viral-siRNAs

VSR Viral suppressors of silencing

WSMV Wheat streak mosaic virus

Chapter 1

Introduction

1.1 Agriculture and disease

Modern monoculture-based agriculture and extensive global trade networks have increased mankind's ability to produce calories and distribute them efficiently. However, large areas of homogeneous plant species, linked by movement of crops, crop products, and machinery, are ideal for the emergence and spread of novel plant pathogens. This is because there is a large, dense population of available hosts, often with very limited genetic diversity, providing a great selective advantage to pathogens able to exploit this opportunity. Once a pathogen is prevalent in one area, movement of infected crops, vectors, or tools may transfer it huge distances across the globe, where it can threaten other growing areas. In spite of an agricultural system which selects for and spreads novel pathogens globally, future food production must increase and/or we must drastically reduce food waste. In 2010-2012, there were an estimated 850 million undernourished people, and an estimated 2 billion people suffering micronutrient deficiencies (Wheeler and von Braun, 2013). Demand for food is predicted to rise by 50% by 2030, and the limited amount of arable land available for agriculture in many areas means that sustainable intensification will be required (Godfray et al., 2010; Wheeler and von Braun, 2013). At the same time, agriculture will suffer outbreaks of both current and novel pathogens, and climate change will increase the frequency and severity of extreme weather events (Bebber et al., 2014; Wheeler and von Braun, 2013). The capability to respond rapidly to emerging pathogens will help mitigate the yield losses, and make sustainable intensification a more achievable goal.

1.2 Maize lethal necrosis

1.2.1 The spread of maize lethal necrosis

This thesis is focused on characterising and combating an emerging global disease: maize lethal necrosis (MLN). MLN was first reported in the Americas in the 1970s as a synergistic interaction between maize chlorotic mottle virus (MCMV, family *Tombusviridae*) and maize dwarf mosaic virus (MDMV, *Potyviridae*) that could cause total crop loss (Castillo, 1983; Teyssandier and Bo., 1983; Uyemoto, 1983). In the last ten years MLN has been reported for the first time in China, across East Africa, and most recently in Ecuador and Spain (fig. 1.1 and table 1.1) (Achon et al., 2017; Adams et al., 2013; Lukanda et al., 2014; Mahuku et al., 2015; Quito-Avila et al., 2016; Xie et al., 2011). MLN is caused by the interaction of MCMV with maize-infecting members of the *Potyviridae* family – MCMV has been reported with MDMV, sugarcane mosaic virus (SCMV), and wheat streak mosaic virus (WSMV). However the widespread recent reports of MLN all report SCMV as the partner virus of MCMV. The recent spread of MLN reflects the spread of MCMV; SCMV has been present in East Africa, China and South America for decades (Chen et al., 2002; Louie, 1980; Perera et al., 2009). It is clear from the rapidity and distances over which MCMV is spreading that global transport networks have been crucial to it establishing a cosmopolitan distribution, although the mechanism for this transport remains unknown.

1.2.2 Maize lethal necrosis is a threat to maize production

Maize is highly productive annual C4 monocot farmed intensively for human and animal feed. Maize plants infected with MLN present a variety of symptoms: chlorosis (yellowing) of leaf tissue, drying of leaf margins, dead heart syndrome, dwarfism and reduced grain filling (fig. 1.2). At the molecular level, as well as common viral symptoms (viral particles and inclusions in the cytoplasm), MLN infection causes mitochondrial cristae to become disorganised and the mitochondria themselves to rupture, releasing their contents into the cytoplasm (fig. 1.2) (Wang et al., 2017). MLN infection also decreased the size of starch grains visible in chloroplasts, and RT-qPCR showed that mRNA levels of the enzyme *pyruvate orthophosphate dikinase*, which is rate limiting in C4 photosynthesis, decreased. Interference with the photosynthetic process would fit with the macro-scale symptoms of chlorosis and reduced yield.

Maize is the most productive cereal crop in Sub-Saharan Africa, and is predominantly used as a dietary staple for local people (FAOSTAT, IITA). Maintaining maize yields in this region is vital to prevent famine and to protect smallholder maize farmers from increased

financial hardship. 75% of Kenyan maize is produced by smallholders, and 90% of rural households grow maize (Kang'ethe, 2011). Maize shortages also hamper economic development at the state level, as although minor shortfalls can be supplemented with duty-free imports from neighbouring East African countries, major shortages require more expensive imports from the global market. This problem is exacerbated if maize production is decreased simultaneously in multiple East-African countries. Pratt et al. (2017) found that MLN is the largest introduced pathogen threat to maize in sub-Saharan Africa, and estimated that annual East African losses owing to MLN will rise to 365-418 million dollars annually in the next 5-10 years. MLN is an aggressive viral condition, which causes 30-100% yield loss, with 80-100% yield loss reported in heavily affected areas of Kenya (De Groote et al., 2016). This resulted in MLN destroying an estimated 23% of Kenya's maize crop in 2013 (De Groote et al., 2015). Due to its rapid spread and interaction with local viruses, and the absence of resistant commercial maize lines, MCMV represents a significant threat to the most important cereal crop of the least food secure region on Earth.

Table 1.1 Maize lethal necrosis now has a global distribution. This table provides the year in which samples first tested positive for maize chlorotic mottle virus in each country it has been reported in.

Country	Earliest report	Reference
Peru	1973	Castillo and Hebert (1974)
USA	1976	Niblett and Claflin (1978)
Argentina	1982	Teyssandier and Bo. (1983)
Thailand	1982	Klinkong and Sutabutra (1983)
Mexico	1984	Gordon et al. (1983)
Hawaii	1990	Jiang et al. (1992)
China	2011	Xie et al. (2011)
Kenya	2011	Wangai et al. (2012)
Tanzania	2012	Snipes (2014)
Uganda	2013	Snipes (2014)
Rwanda	2013	Adams et al. (2014)
Democratic Republic of the Congo	2013	Lukanda et al. (2014)
Ethiopia	2014	Snipes (2014)
Taiwan	2014	Deng et al. (2014)
Ecuador	2015	Quito-Avila et al. (2016)
Spain	2015	Achon et al. (2017)

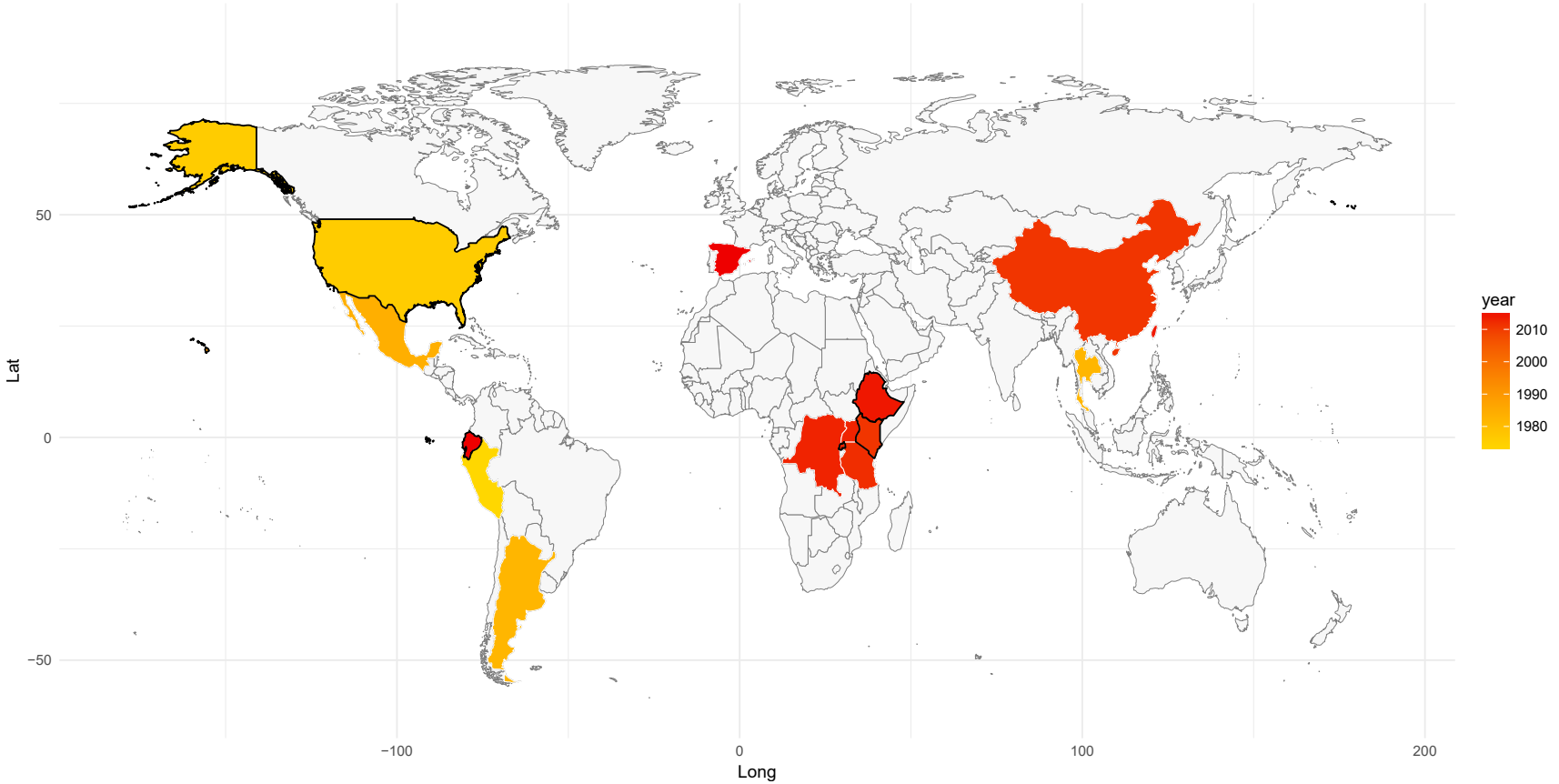


Fig. 1.1 Maize lethal necrosis (MLN) has recently spread to Europe, Africa, and Asia, meaning MLN is now present on five continents. Countries are coloured by the year in which the disease was first reported, and countries with a black outline were surveyed in this study.

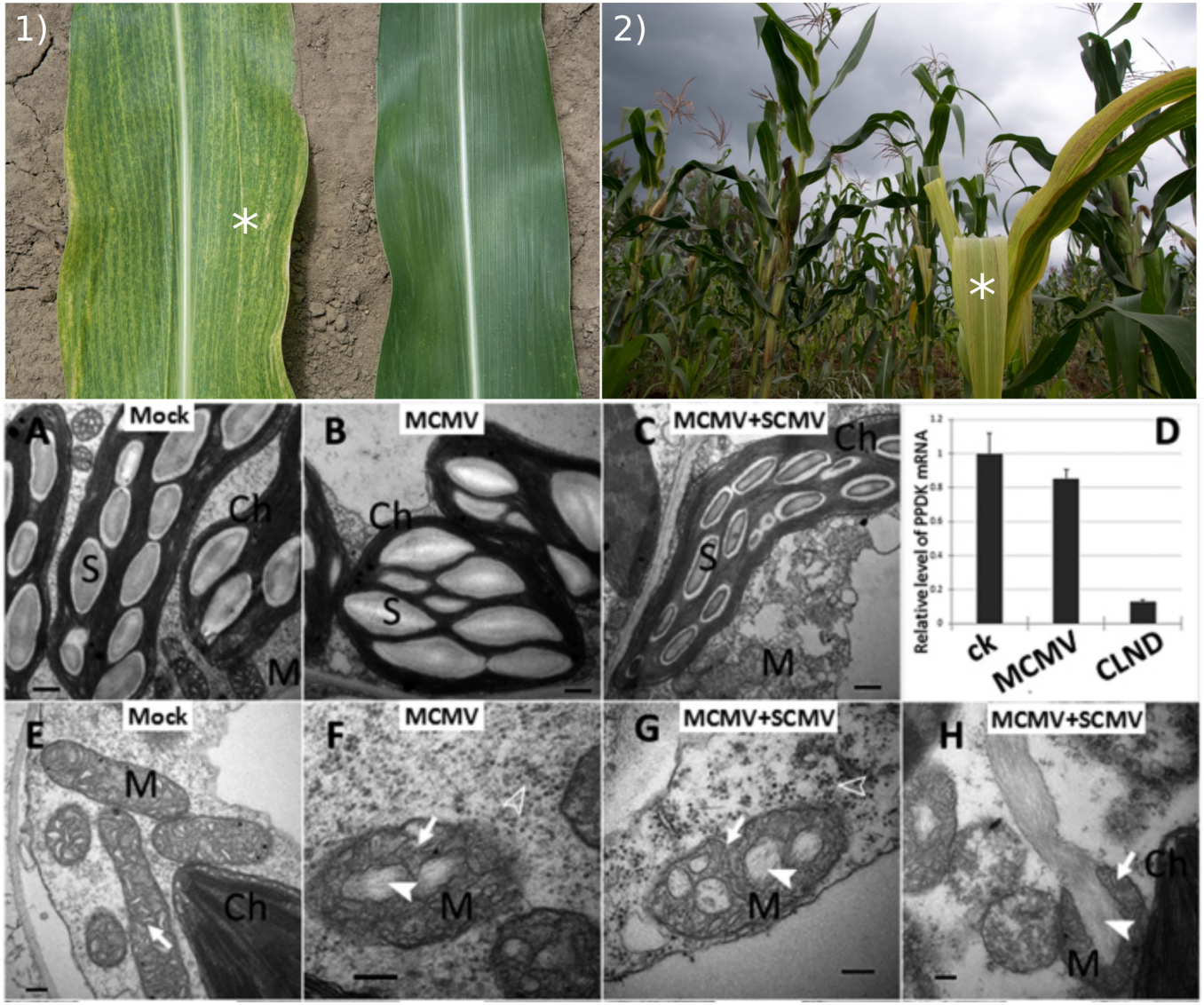


Fig. 1.2 Symptoms of maize lethal necrosis (MLN). Infected plants (asterisks) display 1) chlorosis and 2) dwarfing. A-C) Chloroplast starch grains decrease in MLN compared to uninfected or singly infected plants, and D) *pyruvate orthophosphate dikinase* expression drops. Compared to E) uninfected mitochondria, F-G) maize chlorotic mottle virus (MCMV) and MLN infected mitochondria have disorganised cristae, and H) MLN infected mitochondria rupture. Figures A-H) adapted from Wang et al. (2017).

1.2.3 Maize lethal necrosis control, and future spread

Currently there are no commercial lines in East Africa with strong resistance or tolerance to MLN, although CIMMYT's breeding programme now has a number of promising lines (CIMMYT, personal communication). Control strategies currently in use in East Africa include roguing (removal of symptomatic plants), variation in planting time, pesticide use to control insect viral vectors, and crop rotation. There is no published experimental evidence of the benefits of roguing or variation in planting time. However the benefits of crop rotation to inhibit MLN incidence have been known since the 1980s – susceptible maize fields planted with Sorghum the previous year had mean MLN incidence of 0.6%, compared with 12.2% incidence in fields planted with maize the previous year (Uyemoto, 1983). Viruses in the family *Tombusviridae* (i.e. MCMV) have extremely high titres in host roots, and infectious virions from four *Tombusviridae* genera have been isolated from environmental water sources (Mehle and Ravnikar, 2012; Sit and Lommel, 2010). Transmission in soil water or crop residues has been suggested for MCMV, and there are a number of reports of increased disease pressure after heavy rainfall and in soils with higher water capacity (Jensen, 1991; Uyemoto, 1983). Mahuku et al. (2015) found that planting clean seed in soil from MLN-affected areas resulted in 69% MCMV infection, compared to 4% in the control group planted in sterile soil. Although the mechanism for storage and transmission of MCMV in the soil is unclear, given the above evidence it is clearly an important route for the spread of MCMV.

Maize-growing regions in Kenya which have been heavily affected by MLN, such as Kisumu, have year-round growth of maize, over two main planting seasons with staggered planting times between different farmers (or single farmers hedging against uncertain rainfall). The high disease pressure may be due in part to the continuous availability of maize plants and lack of crop rotation, resulting in a build up of MCMV virions within the soil. In the absence of additional control measures or resistant varieties it is likely that MLN will spread further. Ecological niche modeling suggests that large areas of sub-Saharan Africa are at high risk of MLN outbreaks (Isabirye and Rwomushana, 2016). Epidemiological modeling suggests that clean seed and pesticides to control vectors will attain good MLN control results, which is suitable for larger farms with higher resources, while smaller farms (the majority in Kenya) would need to rely on roguing and crop rotation for lower protection (Hilker et al., 2017). The modeling also predicted that crop rotation strategies will not be effective if there are exogenous sources of infection, such as neighbouring farms, present. Although research is required to parameterise these models more accurately, this result underlines the importance of an integrated, large scale management approach for the long term control of MLN.

1.2.4 Viral synergy

Testing for the presence of viruses has been revolutionised by the advent of next-generation sequencing (NGS). Previously, testing relied predominantly on antibody and PCR based assays, and therefore their ability to detect viruses depended on the availability of suitable antibodies or primers for a specific virus. In other words, you could only find what you were looking for. NGS, in contrast, is large scale sequence-independent sequencing platform, meaning that it is possible to detect any virus present, known or unknown. The main input is the choice of whether to sequence DNA or RNA. Therefore, it is now possible to take a plant, sequence its DNA/RNA, and identify all the DNA/RNA viruses present in the tissue sampled. NGS data from wild populations in a wide variety of plant species has revealed that infection with multiple viruses is common in nature, and in some systems it may be the rule, rather than the exception (Mascia and Gallitelli, 2016; Tollenaere et al., 2015).

Different viruses inhabiting the same host can interact synergistically or antagonistically, with a wide spectrum of strengths of interaction, via a number of mechanisms (table 1.2). Synergism, in this context, describes a situation in which at least one of the coinfecting viruses increases in titre, and the other viral titre increases or remains the same. It is important to note that mixed infections that appear synergistic in terms of the enhanced host symptoms visible may not be synergistic in terms of viral titre, and vice versa. Mixed infections, rather than single, can result in different symptoms and different rates of viral transmission due to altered viral titres (and helper dependence) (Wintermantel et al., 2008).

There are a number of agriculturally damaging synergistic viral diseases in addition to MLN – for example sweet potato virus disease, caused by the interaction of sweet potato feathery mottle virus and sweet potato chlorotic stunt virus, and rice tungro disease, caused by the interaction of rice tungro bacilliform virus and rice tungro spherical virus. Understanding the incidence and impact of mixed infections in crop plants will improve our understanding of plant diseases, and increase our ability to model them in order to inform management practices.

Table 1.2 Different types of viral interactions observed in mixed infections. Note that helper-dependence is a form of trans-complementation, but trans-complementation covers a wider set of interactions.

Name	Type	Notes	Reference
Trans-complementation	Synergistic	Gene products from one virus act in trans to promote the replication or spread of the partner virus.	Moreno et al. (1997)
Helper dependence	Synergistic	One virus is dependent on a partner virus to enhance or enable its spread. Exemplified by umbraviruses, which rely on luteovirus coat proteins for encapsidation.	Robinson et al. (1999)
Cross-protection	Antagonistic	Also known as superinfection exclusion. Previous infection with a similar virus prevents or inhibits superinfection with a second virus. There is evidence for coat-protein and RNAi mediated cross-protection.	Gonzalez-Jara et al. (2009)
Mutual-exclusion	Antagonistic	Simultaneous infection of the same cell/tissue does not occur between closely related species/strains of virus. Unknown mechanism.	Dietrich and Maiss (2003)

The interactions underlying viral synergies can be direct, via interaction between protein or genetic components of the partner viruses, or indirect, via their respective impact on the host. In most plant viral synergies, the mechanism is unknown (Latham and Wilson, 2008). In synergies for which the determinant is known, movement proteins and viral suppressors of silencing (VSRs) are the most common basis, and coat proteins are also prominent (Latham and Wilson, 2008). Given the highly multifunctional nature of many viral proteins, it is possible these patterns will change as the proteins involved are better characterised and specific biochemical properties are correlated with synergy. For an example we can take the well studied synergy of potato virus X (PVX, *Alphaflexiviridae*) with potato virus Y (PVY, family *Potyviridae*) in tobacco, in which PVX titres increase around tenfold compared to single infection, and symptoms increase. This synergy was found to be dependent on PVY HC-Pro (Shi et al., 1997). However, HC-Pro has a large number of functions, including: vector transmission, RNA replication, systemic movement, suppression of RNA silencing, and protease (Hasiów-Jaroszewska et al., 2014). Mutations of the central region of tobacco etch virus (a potyvirus with which PVX is also synergistic) HC-Pro abolished synergy, but this central region still has multiple roles: RNA silencing suppression, genome amplification, and cell-to-cell movement (Hasiów-Jaroszewska et al., 2014; Shi et al., 1997). Later experiments using plum pox virus (PPV, another potyvirus with which PVX is also synergistic) HC-Pro encoded by a chimaeric PVX demonstrated that a point mutation which abolished HC-Pro VSR activity also abolished synergy (González-Jara et al., 2005). However, given the rapid evolution of viruses, this shows the basis of PVX-PPV synergy, but is not conclusive evidence of the basis of the PVX-PVY synergy. There are a large number of viral synergies in plants, with VSRs and movement proteins being the most common genetic requirements, but the difficulties inherent in studying viral proteins makes the overall picture less clear.

MLN is a synergistic condition, with different behaviours depending on the genus of the partner (i.e. non-MCMV) virus. When MCMV was coinfecting with WSMV (tritimovirus), MCMV titre increased 3.3-11.2 fold, while WSMV titre increased 2.1-3.1 fold compared with single infections (Scheets, 1998). However, when MCMV was in mixed infection with SCMV (potyvirus), MCMV increased 1.7-5.4 fold, while SCMV concentrations remained the same (Goldberg and Brakke, 1987). The latter relationship appears to be a classical potyvirus synergy, like PVX-PVY, in which *only* the non-potyvirus titre increases in mixed infection. However the MCMV-WSMV synergy behaves differently, and WSMV HC-Pro is dispensable for the synergistic interaction (Stenger et al., 2007). It is interesting to note that WSMV HC-Pro lacks VSR activity in the context of viral synergy, but the genetic bases of the MCMV-SCMV and MCMV-WSMV synergies are unknown, so more work is needed before the mechanistic basis of MLN can be inferred. This evidence fits a hypothetical

scenario in which MCMV-SCMV synergy is effected by the VSR activity of SCMV HC-Pro, while MCMV-WSMV synergy is mediated by trans-complementation by both MCMV and WSMV gene products.

Observations of plants growing with mixed infection of MCMV and WSMV showed that lower temperatures (23ºC) favoured WSMV infection establishment, while higher temperatures (31ºC) favoured more extreme symptoms in double infections (Scheets, 1998). The impact of environmental conditions on MCMV-SCMV synergy has not been studied, but variation in a similar temperature interval to MCMV-WSMV would have important consequences for predictions of MLN spread and impact across regions with different climactic conditions. Although the basis for MLN synergy is unknown, there is a growing body of evidence about the viruses involved in the condition.

1.3 *Tombusviridae*

1.3.1 Transmission and impact

MCMV is a positive sense single-stranded RNA (ssRNA) virus, in the family *Tombusviridae*. The *Tombusviridae* are a diverse family of plant viruses with a wide range of hosts, including model viruses such as tomato bushy stunt virus (TBSV) and carnation mottle virus (CarMV). Their stable 30 nm icosahedral virions allow survival outside the host for extended periods and this, combined with their high titres in host root systems, means that soil transmission is typical in the family (Sit and Lommel, 2010). This capsid stability means that hibiscus chlorotic ringspot virus capsid has been used as a protein cage for drug delivery (Sit and Lommel, 2015). In addition to soil water, tombusvirids are known to spread via seed, pollen, grafting, and vectors (thrips, beetles, and fungi). Although most tombusvirids have limited economic impact, CarMV is one of the greatest threats to the carnation industry, TBSV harms yield in many fruit trees and the tomato growing areas of California, and MCMV, as discussed above, is a threat to maize production in multiple regions (Sit and Lommel, 2015).

1.3.2 Genome structure and protein expression

Tombusvirid genome structure is variable, but the unifying feature is an RNA-dependent RNA polymerase (RdRP) expressed by either amber stop codon readthrough or -1 ribosomal frameshifting, at approximately 5-10% the rate of the initial ORF (Sit and Lommel, 2015). Around this RdRP is a variety of different genes depending on genus; across the family there are two alternate CP forms, three different forms of MP, as well as accessory proteins with VSR activity or a role in vector transmission, suggesting a very modular evolutionary history

(fig. 1.3a). This variable structure may have been enabled by recombination, which has been shown experimentally for tombuviruses and carmoviruses, and is thought to occur by a copy-choice mechanism in which the RdRP switches templates during genome replication, using the first genome replication product as a primer to bind and continue replicating on the second genome (Cheng and Nagy, 2003).

As positive sense RNA viruses, tombusvirid genomes are functionally equivalent to mRNAs – they can be translated directly into proteins. Unlike mRNAs and many viruses, tombusvirid genomes are not 5' capped and lack a poly-A tail, but instead use RNA secondary structures to promote translation by host machinery. RNA in the 3' untranslated region (UTR) folds to form 3' cap-independent translational enhancers (3' CITEs), which directly interact with the 5' of the genome and recruit ribosomes, via interaction with the eIF4F complex, to the genome to promote its translation (fig. 1.3c and fig. 1.9). Also unlike mRNAs, tombusvirid genomes encode 5-7 proteins, which may require expression at different levels for efficient viral replication. Therefore, proteins are encoded using overlapping ORFs, readthroughs of stop codons, and proximal alternative translational start sites which requires ribosome leaky scanning, which increase the information density, but also produces differential expression from the same RNA molecule (Sit and Lommel, 2015). Additionally, tombusvirid members generate 3' subgenomic RNAs (sgRNAs), through internal RdRP initiation on minus-sense genomes, or RNA-structure stimulated premature termination of minus strand synthesis, which are then used to produce sgRNAs. The concentration of sgRNAs is decoupled from genome concentration, so the CPs, MPs, and VSRs commonly encoded on tombusvirid sgRNAs can be expressed appropriately. In addition to all these roles pertaining to viral protein expression, tombusvirid genomes are also the template for genome replication.

1.3.3 Genome replication

Recent research suggests that most or all positive sense ssRNA viruses can replicate their genomes in spherical invaginations constructed out of modified host membranes, which I will call replication complexes (RCs) (Nagy, 2016; Shulla and Randall, 2016). The roles of RCs are unknown, but some reasonable explanations are the formation of microenvironments favourable for the process of viral genome replication, exclusion of translation machinery (e.g. ribosomes) from the location of genome replication, as genomes are functionally equivalent to mRNA, and protecting the viral replication machinery from host factors, such as *R* genes and the RNA silencing machinery, that would otherwise inhibit viral replication, either by inducing a host cell response or degradation of viral factors (Nagy, 2016; Paul and Bartenschlager, 2013).

Members of the *Tombusviridae* family can form RCs from the membranes of peroxisomes, mitochondria, or the endoplasmic reticulum, resulting in multi-vesicular bodies (Jonczyk et al., 2007; Miller and Krijnse-Locker, 2008). TBSV is an important model for the study of RCs, due to the development of yeast (*Saccharomyces cerevisiae*) as a model host for TBSV, allowing the use of yeast genetic resources for screening which host factors are required for RC formation. For example, TBSV has been shown to co-opt membrane remodelling proteins ESCRT-I and ESCRT-III (ESCRT - endosomal sorting complexes required for transport), and its P33 RdRP subunit promotes redistribution of the lipid phosphatidylethanolamine to RCs via recruitment of the endosomal Rab5, and both of these processes are required for RC formation and efficient TBSV genome replication (fig. 1.3b) (Kovalev et al., 2016; Xu and Nagy, 2016). Once RCs have been formed, RdRPs replicate the genome into its negative-sense form, then use this negative sense form to produce copies of the positive-sense genome. Similar to the process of translation, long-range RNA interactions between (different) motifs are vital for genome replication and RdRP action. TBSV requires an upstream linker present in the post-readthrough region of its RdRP, and a downstream linker in its 3' UTR for efficient production of minus-sense genomes (fig. 1.3c) (Wu et al., 2009). Investigation of this process was aided by another feature of tombusvirid biology – defective interfering RNAs, genome deletion mutants which lose much of their coding capacity but retain functional RNA motifs that allow them to be replicated in RCs.

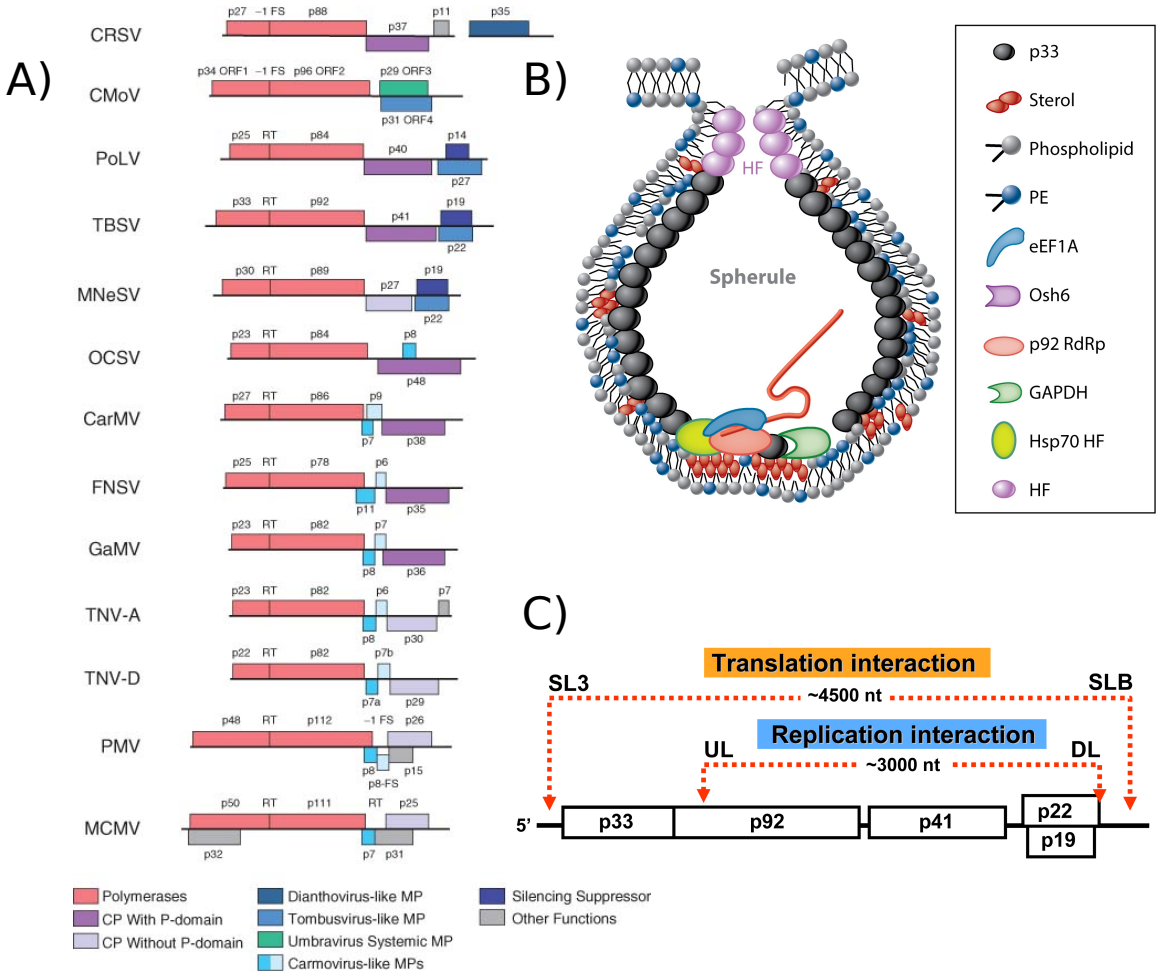


Fig. 1.3 Biological features of the tombusvirids. A) Representative genomes from each genus in the *Tombusviridae* family, with ORFs coloured according to function, highlighting the variety of genome organisation and ORFs present around the highly conserved tombusvirid RdRPs. Adapted from Sit and Lommel (2015). B) Model of the tomato bushy stunt virus (TBSV) genome replication complex, showing recruitment of host proteins and lipids. Adapted from Nagy et al. (2016). C) Long-range RNA interactions are vital for efficient TBSV translation and genome replication, adapted from Wu et al. (2009). CRSV=carnation ringspot virus, CMoV=carrot mottle virus, PoLV=pothos latent virus, MNeSV=maize necrotic streak virus, OCSV=oat chlorotic stunt virus, CarMV=carnation mottle virus, FNSV=furcraea necrotic streak virus, GaMV=galinsoga mosaic virus, TNV-A=tobacco necrosis virus-A, TNV-D=tobacco necrosis virus-D, PMV=panicum mosaic virus.

1.3.4 Maize chlorotic mottle virus biology

MCMV was discovered in the 1970s, when MLN was first characterised (Castillo and Hebert, 1974). MCMV has a unique genome structure within the *Tombusviridae* family, and accordingly is the sole member of the genus *Machlomovirus* (fig. 1.3a). The MCMV genome encodes seven proteins, of which two are unique (P32 and P31), and the small RdRP subunit has a unique amino terminal extension (fig. 1.3a) (Scheets, 2016). MCMV expresses the proteins encoded by the 3' half of its genome from a 1.47 kb sgRNA, and there is a reported shorter 3' UTR sgRNA of 0.34 kb, although this could be a degradation product rather than functional (fig. 1.4) (Scheets, 2000). A useful feature of positive sense RNA virus genomes is that because they natively encode proteins, infection can often be produced by inoculating hosts with *in-vitro* transcripts (IVTs) of the virus produced from a cDNA clone. Scheets (2016) used the MCMV cDNA plasmid pMCM41 and modified it with PCR mutagenesis to mutate each ORF, then used IVT inoculation of maize followed by Northern blotting to investigate the spread of MCMV and thereby the function of each ORF. P31 enhances systemic spread of MCMV via an unknown mechanism, while P32 is required for high titres and pronounced symptoms, again via an unknown mechanism (Scheets, 2016). The only essential proteins for genome replication are the replication complex subunits (P50 and P111), while P7a and P7b function, as cell-to-cell MPs, which is a role of the CP as well. Potentially relating to this alternative CP function (though it could be another alternative function), GFP-fusion constructs showed that MCMV CP is imported into the host nucleus, mostly in the nucleolus (Zhan et al., 2016). The CP interacts with the nuclear import factors maize importin-α1a (ZmIMPα1a) and ZmIMPα1b, and MCMV accumulation decreases if these importins are simultaneously silenced (Zhan et al., 2016). The crystal structure of the CP has recently been solved, revealing that monomers form asymmetric trimers, 60 of which make up the capsid, and identifying exposed surface loops that are important for vector transmission in other RNA viruses (fig. 1.4b-c) (Wang et al., 2015a).

MCMV appears to be transmitted by soil, but can also be transmitted via seed and insect vectors. The earliest estimate of the rate of MCMV transmission via seed was 0.04% across an overall population of 42,000 seeds, with a highest rate within a single seed lot of 0.33% (Jensen, 1991). Later work suggests rates of at least 12% are possible, but more research is needed on rates of seed transmission of MCMV across different crop lines and countries (Quito-Avila et al., 2016). MCMV is transmitted semi-persistently by a number of insect vectors. The first work to identify MCMV vectors experimentally found that six species of chrysomelid beetles, including three species of corn rootworm (*Diabrotica* species) could transmit MCMV (Nault et al., 1978). However more recent outbreaks have been associated with thrip presence, and field-collected thrips (*Franliniella williamsi*) in Hawaii were capable

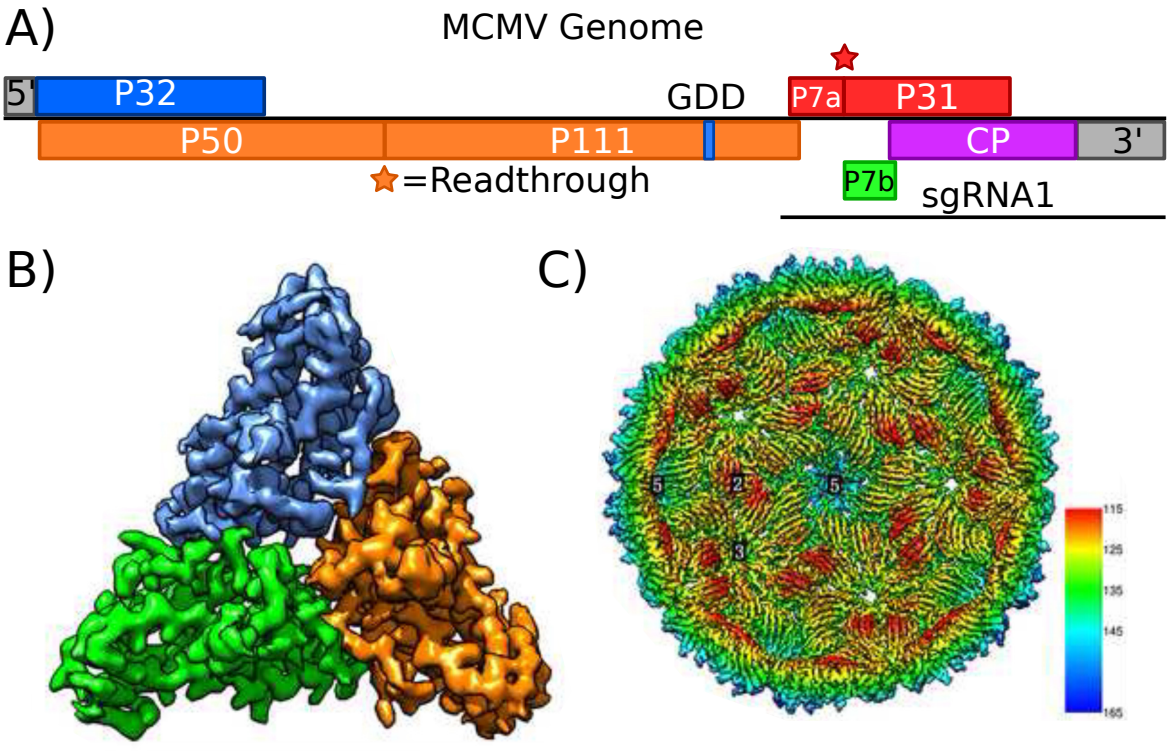


Fig. 1.4 A) Schematic of the 4.4 kb maize chlorotic mottle virus (MCMV) genome, showing the subgenomic RNA required for expression of the ORFs in the 3' half of the genome. B) A single asymmetric unit of the MCMV capsid, with monomers coloured. 60 asymmetric units make up the MCMV capsid C), coloured by radius. B) and C) adapted from Wang et al. (2015a).

of infecting maize plants (Jiang et al., 1992). Experimental work demonstrated that *F. williamsi* transmits MCMV in a semi-persistent manner for up to six days after feeding on MCMV-infected plants (Cabanas et al., 2013; Jiang et al., 1992). *F. occidentalis*, which was introduced to China in 2003 and since then has spread to a wide variety of hosts, was also found to be capable of transmitting MCMV (Zhao et al., 2014). Thrips were present in almost all maize fields surveyed in Kenya, so represent a strong candidate for the primary insect vector of MLN in East Africa (Mahuku et al., 2015). Alternative hosts of plant viruses can play an important role in epidemiology, especially in crop systems with a crop break, during which alternative hosts may act as a reservoir of inoculum for the following season. MCMV has now been reported to *naturally* infect a wide variety of alternative hosts within the Poaceae family, including weeds, fodder crops, and staple crops (table 1.3) (Bockelman, 1982; Kusia et al., 2015; Mahuku et al., 2015).

Table 1.3 Maize chlorotic mottle virus is capable of infecting a wide variety of Poaceae species, although a number of these have only been observed with experimental inoculation (in which case natural=no).

Name	Common name	Country	Natural?	Reference
Sorghum halepense	Johnson grass	Spain	Yes	Achon et al. (2017) Mahuku et al. (2015)
Eleusine coracana	Finger millet	Kenya	Yes	Kusia et al. (2015) Mahuku et al. (2015)
Sorghum bicolor	Sorghum	Kenya	Yes	Mahuku et al. (2015)
Panicum clandestinium	Kikuyu grass	Kenya	Yes	Mahuku et al. (2015)
Saccharum	Sugarcane	Kenya	Yes	Mahuku et al. (2015)
Pennisetum purpureum	Napier grass	Kenya	Yes	Mahuku et al. (2015)
Panicum miliaceum	Proso millet	NA	No	Mahuku et al. (2015)
Setaria italica	Foxtail millet	NA	No	Mahuku et al. (2015)
Andropogon scoparius	Beard grass	NA	No	Bockelman (1982)
Bromus japonicus	Cheat grass	NA	No	Bockelman (1982)
Bromus secalinus	Cheat grass	NA	No	Bockelman (1982)
Bromus tectorum	Cheat grass	NA	No	Bockelman (1982)
Bouteloua gracilis	Blue grama	NA	No	Bockelman (1982)
Calamovilfa longifolia	Prairie sandreed	NA	No	Bockelman (1982)
Digitaria sanguinalis	Purple crabgrass	NA	No	Bockelman (1982)
Eragrostis trichodes	Sand lovegrass	NA	No	Bockelman (1982)
Hordeum pusillum	Little barley	NA	No	Bockelman (1982)
Panicum dichotomiflorum	Autumn millet	NA	No	Bockelman (1982)
Panicum miliaceum	Proso millet	NA	No	Bockelman (1982)
Setaria faberi	Japanese bristlegrass	NA	No	Bockelman (1982)
Setaria viridis	Green foxtail	NA	No	Bockelman (1982)
Spartina pectinata	Prairie cordgrass	NA	No	Bockelman (1982)
Triticum aestivum	Bread wheat	NA	No	Bockelman (1982)

1.4 *Potyviridae*

1.4.1 Transmission and impact

The *Potyviridae* family is the largest and most economically damaging family of plant viruses due to the large number of virus species, and the fact most crop species are susceptible to infection by at least one *Potyviridae* species (López-Moya et al., 2009; Shukla et al., 1994). The potyvirids, like the tombusvirids, are positive-sense ssRNA viruses with an 8-10kb monopartite genome, with the exception of the bipartite bymoviruses. This genome is packaged into long (680-950 nm) and thin (11-14 nm) virions which are flexuous and constructed of around 2000 CP monomers arranged helically (fig. 1.5a-b) (López-Moya et al., 2009). *Potyviridae* members are spread by a variety of insect vectors, with MLN partner viruses spread by aphids (SCMV, MDMV) and mites (WSMV) (table 1.4) (Valli et al., 2015). Mechanical transmission is also possible, although the relevance of this in field conditions is unknown, and there are examples of seed transmission and pollen transmission (Valli et al., 2015). All of the MLN partner viruses are economically important causes of yield loss in their own right, MDMV on maize, SCMV on maize and sugarcane, and WSMV across cereals.

Table 1.4 Table summarising the insect vectors of viral genera in the *Potyviridae*. Genera containing maize chlorotic mottle virus partner viruses have the partner viruses identified.

Vector	Genus	MLN partner virus?
Aphid	*Potyvirus*	SCMV, MDMV
Aphid	*Macluravirus*	No
Whitefly	*Ipomovirus*	No
Mite	*Poacevirus*	No
Mite	*Rymovirus*	No
Mite	*Tritimovirus*	WSMV
Phytomixea	*Bymovirus*	No

1.4.2 Genome structure and protein expression

Potyvirid genomes encode a single polyprotein which spans most of the length of their genome, and is proteolytically cleaved by three viral proteases into around 10 multifunctional protein products (fig. 1.5c) (López-Moya et al., 2009). In contrast to the use of RNA motifs to generate differential expression in the *Tombusviridae*, this system implies a relatively

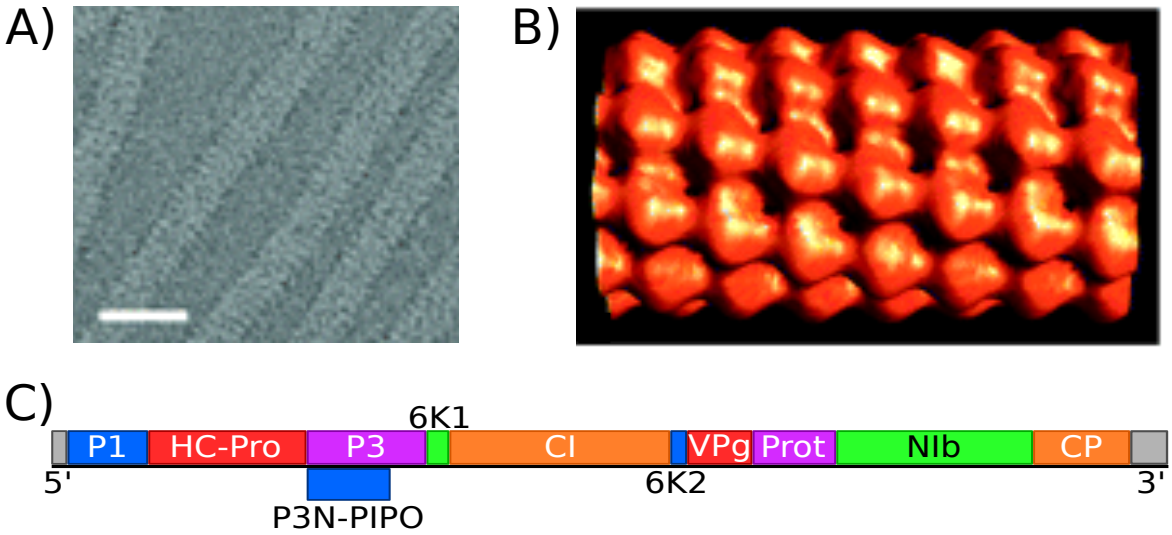


Fig. 1.5 A) Soybean mosaic virus (SMV) capsid visualised using cryo-EM, scale bar=250 Å. B) Reconstruction of SMV virion, with individual capsid subunits visible as nodules. C) Schematic of a potyviridae genome (sugarcane mosaic virus), with final proteins produced by proteolytic cleavage of the polyprotein coloured separately. A) and B) adapted from Kendall et al. (2008).

equal production of gene products from *Potyviridae* genomes (although steady-state levels may vary owing to different degradation rates). In addition to the proteins produced from the polyprotein, there is a GA_6 site in the P3 gene which causes transcriptional slippage to occur at a rate of approximately 2%, and generates an alternative fusion gene product, P3N-PIPO (Chung et al., 2008; Olspert et al., 2015). *Potyviridae* genomes are also more similar to mRNAs than *Tombusviridae*, in that they have a 5' protein cap (the 25 kDa VPg protein) and poly-adenylation of the 3' terminus on their genomes (Valli et al., 2015). Work in *Potato virus A* (potyvirus) has shown that its VPg interacts with host translation intiation factors eIF4E and its isoform eIFiso4E, which promotes viral translation and inhibits host mRNA translation (Eskelin et al., 2011).

1.4.3 Recombination in the *Potyviridae*

Intra-species recombination is important in *Potyviridae* evolution. NGS data has demonstrated the widespread occurrence of recombination amongst RNA viruses, but it appears to be particularly common in the *Potyviridae* family (Chare and Holmes, 2006; Sztuba-Solińska et al., 2011). Recombination is again thought to occur by template switching. Work on the distribution of recombination sites in the *Bromoviridae* virus *Brome mosaic virus* (BMV) showed that recombination clustered in areas of alternating GC-rich and AU-rich sequences,

which was suggested to promote dissociation of the RdRP from the initial template due to the weaker AU base pairing (Nagy and Bujarski, 1998; Sztuba-Solińska et al., 2011). *Potyviridae* recombination has not been experimentally modified, but analysing the sequences around breakpoints in the potyvirus turnip mosaic virus (TuMV) showed that recombination sites typically also had GC-rich regions upstream and AU-rich regions downstream (Ohshima et al., 2007). In addition to extensive intra-specific recombination, inter-specific recombination may have also played a role in *Potyviridae* evolution, for example in duplication and diversification of *Potyviridae* P1 proteins (Valli et al., 2007).

1.4.4 Sugarcane mosaic virus biology

There are three MCMV partner viruses in the *Potyviridae*: WSMV, MDMV, and SCMV. SCMV is the reported partner virus in the East African, Chinese, Ecuadorian, and Spanish MLN outbreaks, and because my work is on emerging MLN outbreaks I will focus on introducing SCMV. SCMV, like MDMV, is a *Potyvirus*, an extremely large viral genus containing over 150 species. SCMV can infect three major crops: sorghum, sugarcane (10-35% yield loss), and maize (20-50% yield loss) (Rybicki, 2015; Viswanathan and Balamuralikrishnan, 2005; Zhu and Ye, 2014). The first disease reports associated with SCMV date from 1916 in Puerto Rico, and shortly afterwards through southern states of the USA, and it is now known to be present in at least 25 countries across the six inhabited continents (Wu et al., 2012). This cosmopolitan distribution is likely explained by SCMV's ability to infect maize, sorghum, and sugarcane, which have all been traded extensively for hundreds of years.

SCMV has been reported to be spread in a non-persistent, non-circulative manner by aphid species, like other potyviruses (Teakle and Grylls, 1973). The mechanism in SCMV has not been investigated, but studies in other potyviruses support the bridge hypothesis, which suggests that HC-Pro binds both aphid stylets and virions, holding virions in the aphid mouthparts until its next meal, at which point the virions enter the new host cell (Flasinski and Cassidy, 1998; Peng et al., 1998). SCMV can also spread via movement of infected root cane (sugarcane), and seeds (maize) (Francisca et al., 2012). Grow-out experiments with infected maize in China showed that SCMV transmission can occur at a low rate through pollen (maximum value of 0.1%), and a much higher rate through maternal seed (maximum value of 5.6%) (Li et al., 2007). Local soil transmission of SCMV has been suggested by early work using sorghum, in which plants in the same soil had transmission between 1% and 5%, while plants identically spaced in separate pots had 0% transmission (Bond and Pirone, 1970). Further experiments demonstrated that plants only connected by a body of water allowed transmission rates of around 5% (Bond and Pirone, 1970). SCMV also has a

number of alternative hosts - studies in Australia have identified a number of Poaceae species naturally containing SCMV (table 1.5) (Karan et al., 1992; Persley and Greber, 1977; Srisink et al., 1993; Teakle and Grylls, 1973). The presence of SCMV in alternative hosts elsewhere in the world has been investigated in less detail, but it seems likely given the widespread distribution of Poaceae species. Interestingly, SCMV was found in complex with MCMV inside finger millet (*Eleusine coracana*), and appeared to induce MLN-like symptoms (Kusia et al., 2015). Finger millet is often farmed in rotation or in parallel with maize in Kenya, so this may be a relevant reservoir host for MLN in East Africa.

Table 1.5 Alternative hosts of sugarcane mosaic virus identified in studies of wild plants in Australia. Most hosts are members of the *Poaceae* family. All infections are natural.

Name	Common name	Country	Reference
Brachiara piligera	Hairy arm grass	Australia	Srisink et al. (1993)
Digitaria didactyla	Queensland blue coach	Australia	Teakle and Grylls (1973)
Dinebra retroflexa		Australia	Persley and Greber (1977)
Echinochloa colona	Jungle rice	Australia	Persley and Greber (1977)
Echinochloa frumentacea	Sawa millet	Australia	Persley and Greber (1977)
Eleusine indica	Indian goosegrass	Australia	Persley and Greber (1977)
Eragrostis cilianensis	Stinkgrass	Australia	Persley and Greber (1977)
Eriochloa procera	Cupgrass	Australia	Persley and Greber (1977)
Panicum paludosum		Australia	Persley and Greber (1977)
Paspalum conjugatum	Sour paspalum	Australia	Persley and Greber (1977)
Pennisetum glaucum	Pearl millet	Australia	Karan et al. (1992)
Setaria anceps	Bristle grass	Australia	Persley and Greber (1977)
Setaria italica	Foxtail millet	Australia	Persley and Greber (1977)
Setaria verticillata	Hooked bristlegrass	Australia	Persley and Greber (1977)
Sorghum halepense	Johnson grass	Australia	Persley and Greber (1977)
Sorghum verticilliflorum	Wild sorghum	Australia	Srisink et al. (1993)
Urochloa mosambicensis	Sabi grass	Australia	Persley and Greber (1977)

1.5 Plant immune system

1.5.1 Protein mediated immune system

Plants have a number of layers of defence against pathogen attack. Firstly, the waxy cuticle which limits water loss to the environment is a barrier to would-be pathogens, as is the cell wall (Dangl et al., 2013). Pathogens that enter plant tissues encounter two tiers of protein-mediated immune responses, the first being pattern-triggered immunity (PTI). Pathogens generate pathogen associated molecular patterns (PAMPs) and damage associated molecular patterns (DAMPs). PAMPs are typically evolutionarily conserved molecules associated with pathogen presence, such as bacterial flagellin and elongation factor Tu (EF-Tu). DAMPs are often the breakdown products of host molecules, such as cell wall components. PAMPs and DAMPs in the extracellular environment are perceived by transmembrane receptor proteins, pattern recognition receptors (PRRs), triggering PTI. PRRs are typically Leucine-rich repeat (LRR) or lysine motif kinases, which upon activation stimulates a burst of calcium ions and reactive oxygen species, and a signal transduction cascade resulting in massive transcriptional reprogramming and expression of defence genes (Monaghan and Zipfel, 2012). Xu et al. (2017) recently showed that *Arabidopsis* plants exposed to the EF-Tu epitope elf18 (a PAMP) also leads to alteration in the efficiency of translation in over 500 genes, including increased efficiency for a number of defence and defence signaling genes.

Pathogen reproduction and dispersal is inhibited by the PTI response, and most have evolved a number of proteins which interfere with PTI, often by modifying or destroying proteins involved in the host immune response. These pathogen suppressors of PTI are called effectors, they induce effector-triggered susceptibility (ETS), and accordingly plants have evolved to counter this threat (fig. 1.6). The second layer of the protein based plant immune system is effector-triggered immunity (ETI), in which intracellular receptors, most commonly nucleotide-binding LRRs (NB-LRRs), are activated by direct interaction with pathogen effectors, pathogen-triggered modification of host immune proteins (guard hypothesis), and pathogen-triggered modification of decoy host immune proteins (Jones and Dangl, 2006). The genes which produce these effector-sensors are called *R* genes, and provide monogenic dominant resistance. ETI is a more specific, enhanced and accelerated form of PTI, often resulting in the hypersensitive cell death response (HR), the localised programmed cell death of infected cells (fig. 1.6). Although the most detailed work on PTI and ETI has been performed using bacterial model PAMPs and effectors, there are over 200 known dominant antiviral R genes (Ronde et al., 2014). For example, the tobacco mosaic virus (TMV, *Virgaviridae*) P50 replication complex component is recognised by the N immune receptor, a Toll-interleukin-1-NB-LRR (TIR-NB-LRR), in complex with the chloroplastic

protein NRIP1. NRIP1 relocalises to the cytoplasm and nucleus upon TMV infection, and is required for interaction between P50 and the N receptor (Caplan et al., 2008). In addition to these two tiers of protein-targeted immune responses, there is also an intracellular antiviral RNA surveillance system.

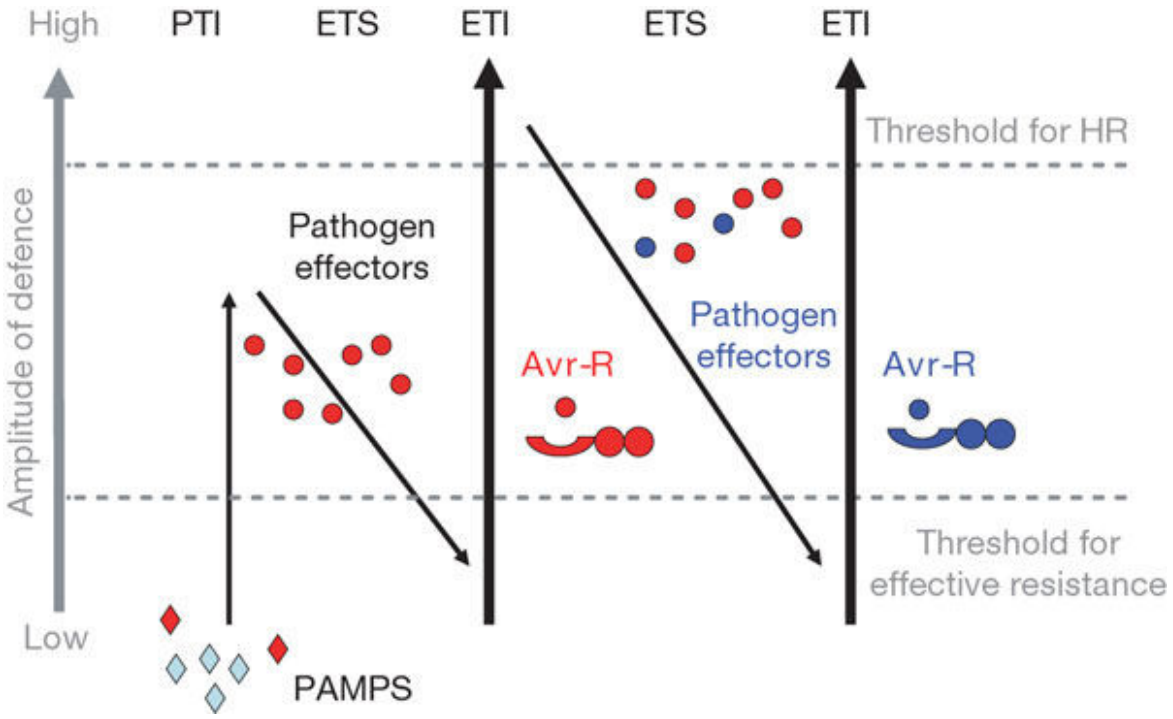


Fig. 1.6 Zigzag model of plant defence, showing pathogen associated molecular patterns (PAMPs) triggering low intensity pattern triggered immunity (PTI). Pathogens may then release effectors to subvert PTI, but these effectors can be detected by avirulence (Avr) resistance genes (or *R* genes) to trigger the higher intensity effector-triggered immunity (ETI). Over evolutionary time this results in an arms race, with pathogens evolving new effectors and hosts evolving appropriate detectors. Figure reproduced from Jones and Dangl (2006).

1.5.2 RNA silencing

The term RNA silencing encompasses a number of functionally similar, partially redundant pathways that use small RNAs (sRNAs, typically 21-24 nucleotides) to direct post-transcriptional gene regulation, cleavage and translational repression of viral RNA, and methylation of genomic DNA to alter the transcriptional state of chromatin. The general pathway is as follows: double-stranded (ds) RNA is cleaved by a member of the Dicer protein family, which contain two RNase III domains and specifically cleaves dsRNA (fig. 1.7) (Bernstein et al., 2001; Nicholson, 1999). Plants express multiple Dicer-like proteins (DCLs)

which specialise in different dsRNA substrates and produce sRNAs duplexes of different lengths which function in alternative silencing pathways. dsRNA-binding proteins (DRBs) partner specifically with DCLs to increase the efficiency and accuracy of processing, and it appears there are also inhibitory DRBs which antagonise DCL activity (Curtin et al., 2008; Tschopp et al., 2017). sRNAs duplexes (typically 21-24 nucleotides) are then protected from uridylation and exonucleolytic digestion by HUA ENHANCER1 (HEN1) which 2'-O-methylates their terminal-3' ribose (Yu et al., 2005). sRNA duplexes then associate with Argonaute family proteins (AGOs), which are the catalytic core of RNA-induced silencing complexes (RISCs) (Baulcombe, 2004). One strand of the duplex is bound by the AGO and becomes the guide sRNA, whilst the other is cleaved, with preference depending on the specific AGO protein and a general preference for the strand in the pre-sRNA duplex with a less stable 5' end (Czech and Hannon, 2011). Guide RNAs then direct RISCs to complementary RNA molecules, where the RISC either causes cleavage or translational repression of the target RNA, or to complementary genomic DNA which is methylated by the RISC, producing transcriptional silencing. The role of a RISC is dictated by the form of AGO present (Meister, 2013). Conceptually, sRNAs act as search terms to specifically direct silencing of RNA translation and DNA transcription. The initiating dsRNA may be derived from inverted repeats (IRs) in endogenous genes, miRNA genes (MIRs) which contain complementary sequences and form stem-loops, trans-acting small interfering RNA (tasiRNA) genes (TASs), transgenes, or viral infection. The sRNAs produced during silencing are classified, mostly according to their precursors, as miRNAs, small-interfering RNAs (siRNAs), heterochromatin-associated siRNAs, and tasiRNAs.

Viral infection stimulates production of viral-siRNAs (vsiRNAs), which may be primary or secondary. Primary vsiRNAs are derived from dsRNA which is present due to viral biology. dsRNA can form during genome replication of RNA viruses, secondary RNA structures present in virus genomes or transcripts, or from complementary sequences in viral transcripts. Secondary VSRs are generated by host RNA-dependent RNA polymerases (RDRs) acting on viral ssRNA to generate dsRNA, which can be primed by siRNA binding, or occur in a primer independent fashion (Devert et al., 2015). Much of the work on vsiRNAs has been performed in the model plants *Arabidopsis* and *Nicotiana benthamiana*, and gene/protein names in this section refer to *Arabidopsis*. RDR1 and RDR6 are thought to be most important for dsRNA production from viruses, as mutation of these genes results in increased susceptibility to some RNA viruses. There appears to be some specificity, for example *Arabidopsis rdr6* mutants are more susceptible to cucumber mosaic virus (CMV), but not tobacco rattle virus (Dalmay et al., 2001). In *Arabidopsis* it seems that secondary vsiRNAs are more potent than primary in antiviral silencing, though this does not exclude an initiation role for primary

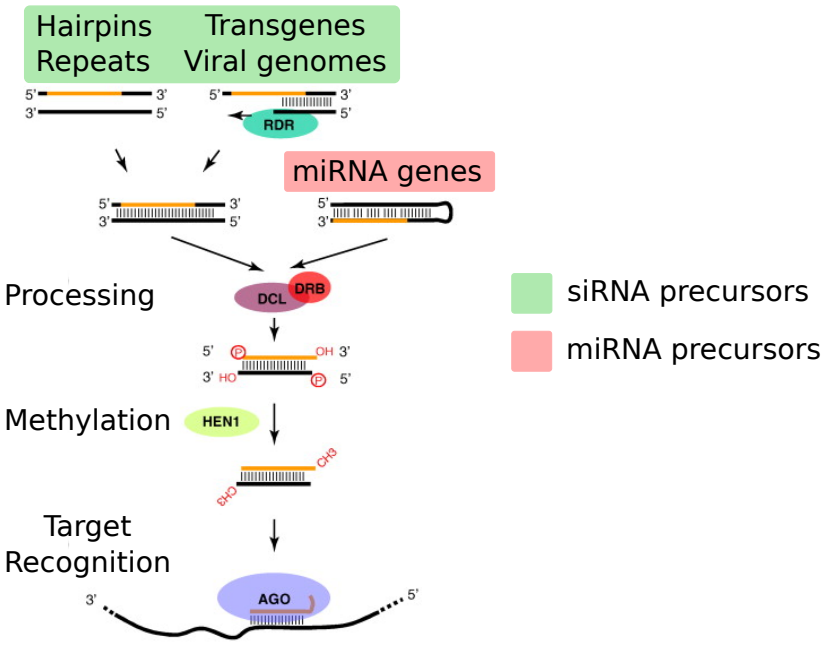


Fig. 1.7 Biogenesis of miRNAs and siRNAs in RNA silencing. Generalised RNA silencing pathways showing parallel steps for miRNA and siRNA processing. Adapted from Vazquez et al. (2010).

vsiRNAs in secondary vsiRNA generation (Wang et al., 2010b). The two most important DCLs for *Arabidopsis* vsiRNA production are DCL4 (mostly 21 nt vsiRNAs) and DCL2 (mostly 22 nt), which appear to function semi-redundantly with varying important depending on the infecting virus, although this could reflect the impact of viral suppressors of silencing (Bouché et al., 2006; Mlotshwa et al., 2008).

In addition to RNA silencing inhibiting viral replication within a single cell, there is also local cell-to-cell movement of the silencing signal, presumed to be through plasmodesmata. Longer distance systemic movement of the silencing signal between tissues occurs through phloem transport of a silencing signal, most likely 21 or 24 nucleotide siRNAs, from source to sink tissues, although this phenomenon has not been investigated in the context of antiviral sRNAs (Melnyk et al., 2011; Molnar et al., 2011; Tournier et al., 2006). Plants are capable of excluding most viruses from their shoot apical meristem, their primary growing point and eventual source of germline cells, despite the fact this is a sink tissue, in a process known as meristem exclusion (Foster et al., 2002; Martin-Hernandez and Baulcombe, 2008). This appears to rely on RNA silencing, as ectopic expression of the potexvirus TGBp1, which suppresses systemic movement of RNA silencing, in *N. benthamiana* produced plants susceptible to viral meristem invasion, and knockdown of *RDR6* (necessary for reception but

not production of the systemic silencing signal) in *N. benthamiana* allowed PVX to invade the meristem (Foster et al., 2002; Schwach et al., 2005).

There are two indications that RNA silencing is an effective defence against viruses. Firstly, RNA silencing appears to restrict viral host ranges: *Arabidopsis* is not considered a PVX host, but becomes one if multiple DCL genes or the AGO2 gene is mutated (Jaubert et al., 2011). *N. benthamiana* is a host model due to its unusual susceptibility to a wide variety of viruses, and has a natural loss of function mutation in the RDR *NbRdRP1m*, which is partially responsible for its susceptibility (Yang et al., 2004). The second indication is the ubiquity of viral VSRs which suppress RNA silencing pathways. Most plant viruses express VSRs, a general term used to describe a huge variety of independently evolved proteins which target multiple stages of the RNA silencing pathway. Typically, VSRs function by binding dsRNAs to inhibit DCL activity, sequestering sRNA duplexes before they enter RISCs, or targeting of proteins involved in sRNA processing, resulting in inhibition of their activity or their degradation (Incarbone and Dunoyer, 2013). This demonstrates that across a wide variety of host plants, strong selection is placed on a wide variety of viruses to produce VSRs, which in turn implies the importance of RNA silencing in antiviral defence.

RNA silencing in maize has received less study, but transgene silencing occurs, vsiRNAs are generated upon viral infection, and at least 25 families of miRNAs are expressed (Xia et al., 2014; Zhang et al., 2009). The maize genome is thought to contain 5 *DCL* genes, 5 *RDR* genes, and 18 *AGO* genes (Qian et al., 2011). This is a similar gene number for *DCL* and *RDR* compared to *Arabidopsis*, but almost double the number of *AGO* genes. Phylogenetically, maize has RNA silencing genes falling into the same clusters as *Arabidopsis*, although this doesn't confirm they have the same function (Qian et al., 2011). Additionally, maize and rice have *AGO18* genes, which cluster separately from dicot *AGOs* (Qian et al., 2011). In rice, AGO18 sequesters miR168 (a miRNA), which in turn downregulates the antiviral AGO1 (Wu et al., 2015). *AGO18* is induced by viral infection in rice and maize, and provides broad-spectrum viral resistance in rice (Wu et al., 2015; Xia et al., 2016). Sequencing of vsiRNAs in maize shows that they are predominantly 21 and 22 nt in length, which in *Arabidopsis* would suggest action by DCL2 and DCL4 (Li et al., 2017; Xia et al., 2014, 2016). However, phylogenetic clustering does not prove functional equivalence. In SCMV infection vsiRNAs are mostly 21 and 22 nt in length, and while DCL2 (22 nt vsiRNA production in *Arabidopsis*) is induced by SCMV infection, DCL4 (21 nt vsiRNAs in *Arabidopsis*) expression decreases, but 21 nt vsiRNAs are the most common form (Xia et al., 2014). This could be explained by differences in DCL function between monocots and dicots, or differences in protein levels caused by varying degradation rates or alterations in translational efficiency. vsiRNA production has been examined in the context of MLN infection, finding

that in mixed infection there are approximately three times the amount of SCMV-vsiRNAs than MCMV-vsiRNAs, despite the fact MCMV titre is much higher than SCMV in mixed infection (Xia et al., 2016). Additionally, mixed infection appeared to promote the production of 22 nt over 21 nt vsiRNAs targeting both viruses (fig. 1.8). Single and mixed infection promoted *DCL2, AGO2a*, and *AGO18a* expression, a pattern also seen in maize plants infected with rice black streaked dwarf virus (Li et al., 2017; Xia et al., 2016). This suggests their involvement in the antiviral RNA silencing pathways in maize, although mutational analysis would be required to confirm this.

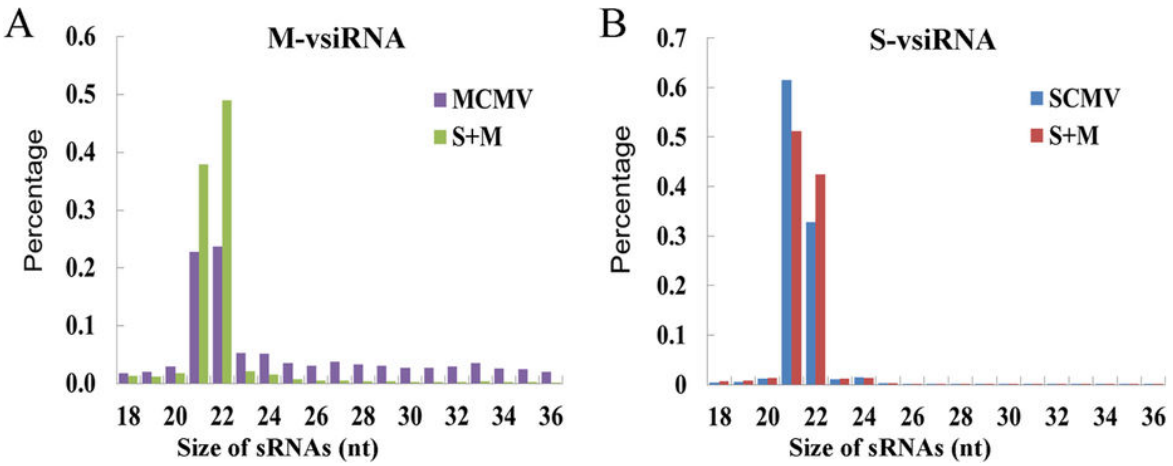


Fig. 1.8 Size distribution of viral siRNAs (vsiRNAs) generated in maize chlorotic mottle virus (MCMV), sugarcane mosaic virus (SCMV), and maize lethal necrosis (MLN) infected plants. A) MCMV targeting vsiRNAs, the proportion of 22nt vsiRNAs increases in mixed infection (S+M). B) SCMV targeting vsiRNAs, with proportional increase in 22nt vsiRNAs in mixed infection. Figure reproduced from Xia et al. (2016).

1.6 Engineering resistance to viral disease

Engineering resistance to viral diseases has been the subject of much basic research, resulting in a large number of lab-based examples and a small number of products in commercial production, such as papaya ringspot virus resistant papaya (*Carica papaya*) and potyvirus resistant squash (*Cucurbita pepo*) (Lindbo and Falk, 2017). Understanding of the mechanisms at work in genetically engineered plants, and the possibilities for engineering, has evolved with understanding generated through basic research. I will introduce natural and artificial mechanisms of resistance (as both can be used in engineering), and discuss appropriate strategies for engineering MLN resistance.

1.6.1 Natural resistance

Dominant monogenic resistance genes

The *R* genes which produce the proteins responsible for ETI are typically dominant and monogenic. This means they are attractive in conventional breeding programmes for their ability to function in the heterozygous state and avoid segregation, as would be the case for a multigenic trait. *R* genes most commonly encode NB-LRR proteins, and most commonly target bacterial or fungal pathogens (Ronde et al., 2014). However, there are around 25 cloned viral *R* genes, of which 80% are NB-LRRs (summarised in Ronde et al. (2014)). There are two major effect *R* genes against SCMV in maize, *Scmv1* and *Scmv2*. However, it appears that neither encodes a NB-LRR, as the best candidate in the *Scmv2* interval is *GRMZM2G116204*, an auxin-binding protein, and the causal gene for *Scmv1* is a *Thioredoxin h-type* gene (Leng et al., 2015; Li et al., 2016; Liu et al., 2017).

Recessive resistance genes

Viruses are obligate intracellular pathogens, so cannot replicate without a host cell. They rely on co-opting host proteins and membranes to replicate their genomes and translate genetic information into viral proteins. Mutations in co-opted host factors that prevent them from being used by the virus therefore inhibits virus replication. However, if this mutated host protein is present in the heterozygous state, there will be a functional form of the protein present for the virus to co-opt. Therefore, mutations of this kind generate resistance genes which are recessive, and less attractive to breeders as a result. The most common source of recessive genes are mutations in translation factors, which are commonly recruited by a wide variety of viruses for translation of viral proteins (fig. 1.9) (Sanfaçon, 2015). For example, mutations in translation initiation factor eIF4E and its isoform eIFiso4E, which interact with the VPg cap of potyviridae genomes, generates recessive resistance to a wide variety of potyviridae viruses in twelve different host plants (Sanfaçon, 2015).

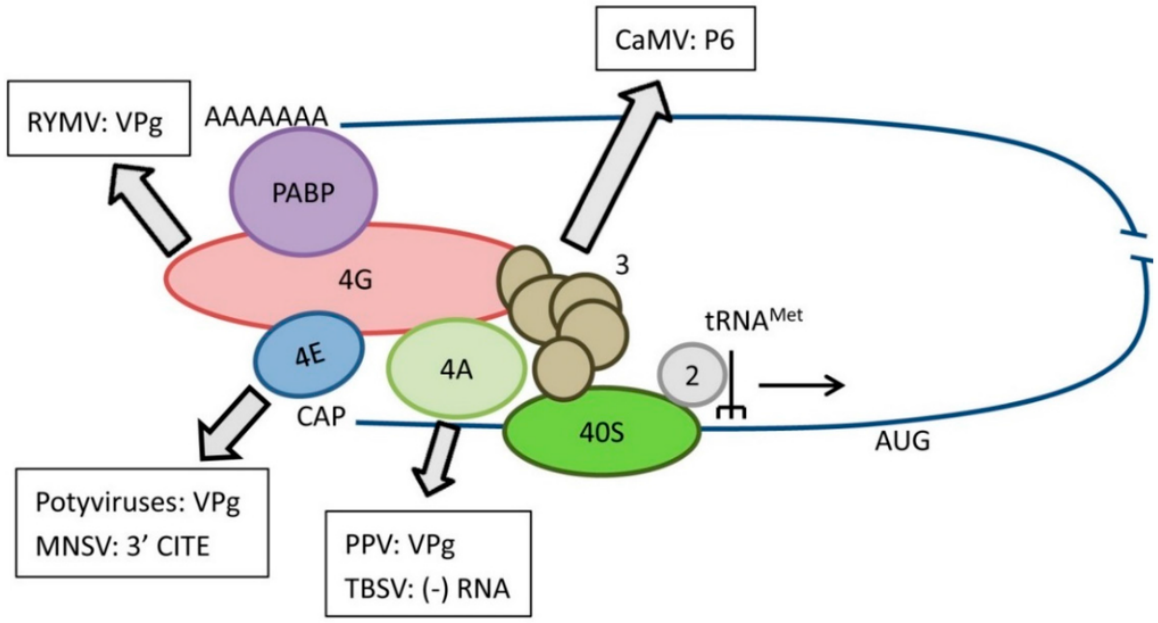


Fig. 1.9 Simplified diagram of translation initiation complex, showing which factors are associated with different viruses. Mutants of the translation initiation complex subunits are common sources of recessive resistance. CaMV=cauliflower mosaic virus, MNSV=melon necrotic spot virus, PPV=plum pox virus, RYMV=rice yellow mottle virus, TBSV=tomato bushy stunt virus. Adapted from Sanfaçon (2015).

1.6.2 Engineered resistance

Pathogen-derived resistance

The initial concept behind pathogen-derived resistance (PDR) was the idea that mis-expression of a viral protein, although harmless to the host, would interfere with the biology of that same virus upon infection of the modified host. This concept was introduced from results using the Qβ bacteriophage in *Escherichia coli*, and shortly afterwards Abel et al. (1986) showed that *Nicotiana tabacum* plants expressing tobacco mosaic virus (TMV) CP had delayed or absent symptoms after inoculation with TMV (Sanford and Johnston, 1985). Although the mechanism behind PDR was unknown, it was an attractive approach because it required little prior knowledge of viral protein function, each virus provided the means to engineer resistance against it (its own genes), and there are a small number of genes to test in viral genomes (Lindbo and Falk, 2017). There are examples of replicase and movement protein transgene expression inhibiting viral infection, but the genes most commonly used in PDR engineering were CPs.

Coat protein mediated resistance

Expressing viral CPs to generate resistance is effective against viruses from a number of different families. However the mechanism behind CP mediated resistance has not been shown for many of these examples. In the best studied case, using TMV CP expression, it seems transgenic expression of CPs interferes with the uncoating of TMV virions as they enter the host cell. TMV CP mutants which are unable to aggregate don't provide TMV resistance, and those with an enhanced ability to form (non-helical) aggregates provide enhanced resistance against TMV (Asurmendi et al., 2007; Bendahmane et al., 2007). Transgenic squash plants expressing the CPs of watermelon mosaic virus, CMV, and zucchini yellow mosaic virus were resistant to all three viruses and became the first commercially released crop with genetically engineered virus resistance (Lindbo and Falk, 2017). CP mediated resistance relies on the presence of CP protein, so it was expected that the level of host CP expression would correlate with the level of viral resistance, but this was not true in a number of cases, hinting at an alternative mechanism.

Sense and antisense RNA mediated resistance

The expression of CPs can generate viral resistance, but so can the expression of a truncated CP or a non-coding region of viral genomes (Duan et al., 2012). This suggested that instead of the CP protein generating resistance, it was the RNA molecule encoding it (Baulcombe, 1996). If inserted RNA sequences similar to a virus were silenced by the host plant, then there would be resistance against that virus (Stam et al., 1997). In addition to sense RNA molecules (i.e. protein coding) being able to induce viral resistance, it was found that antisense viral RNA molecules were more efficient at producing resistance, and plants expressing both sense and antisense viral sequences were the most efficient at producing resistance (Waterhouse et al., 1998). This result, in combination with contemporaneous papers exposing the processes of RNA silencing for the first time, opened a new avenue of research based on using dsRNA to induce antiviral resistance.

Hairpin RNA mediated resistance

The use of a long sense and antisense RNA sequence, either as an inverted repeat or with a small loop region between them, is referred to as a hairpin, due to its secondary structure. Hairpins are more efficient than sense or antisense mediated resistance as they form dsRNA directly, rather than requiring host RDR activity or binding to a separate complementary RNA molecule. Therefore, DCL proteins can act on the dsRNA hairpin constitutively to generate high levels of siRNAs with sequence similarity to viral genomes. Hairpin technology has

been used to engineer resistance to a number of viruses in different plant species, for example a 700bp hairpin of PVY sequence produced resistance in 60% of tobacco plants, rising to 100% if the loop region was an intron (Smith et al., 2000). There are some downsides to hairpin RNA technology. Firstly, because a large number of siRNAs are produced from the hairpin, there is an increased chance that similarity between host mRNAs and hairpin siRNAs will result in knockdown of a random subset of host genes. The hairpin itself, if virus-derived, is typically over 200bp and can contain RNA structural motifs and/or coding information for proteins or protein domains. Given the prevalence of recombination in RNA viruses, these extended viral sequences are a potential biosafety concern, as they could be recombined into the target genome, or other viral genomes. Additionally, hairpin RNAs are prone to self-silencing at the transciptional level. The discovery of miRNA genes suggested another RNA-based approach to engineering resistance.

amiRNA mediated resistance

miRNA genes form short stretches of imperfectly complementary dsRNA, which are processed by DCL1 to miRNAs, which downregulate host genes. However, the miRNA sequence, and its complementary miRNA* sequence, can be replaced with viral sequences to produce artificial miRNAs (amiRNAs) that target viral genomes. Niu et al. (2006a) was the first to use amiRNAs to produce viral resistance, inserting amiRNAs targeting TuMV and turnip yellow mosaic virus (TYMV) into *Arabidopsis*, producing lines with specific and effective resistance to both viruses (fig. 1.10). Direct comparison of hairpin and amiRNA mediated resistance suggested that amiRNAs are more efficient at generating resistance, and additionally are less inhibited by cold temperatures (Niu et al., 2006a; Qu et al., 2007). Other benefits of amiRNAs are that there is a lower chance of off-targets as fewer sRNA species are produced (and screening for them is easier), that the small miRNA gene unit means that multiplexing is more feasible, and the very small fragment of viral genome sequence used minimises biosafety concerns. There are currently no commercially available crops that use amiRNAs to provide viral resistance (Lindbo and Falk, 2017).

Engineering maize lethal necrosis resistance

Genetic modification can be used to insert any of the above sources of resistance, natural or engineered, into host plants. In this thesis I set out to characterise the MLN outbreak in East Africa, and engineer maize lines with some level of resistance against MLN. At the onset of this project (2014), large scale screening of commercial maize lines in East Africa for MLN resistance had limited success, with only 30 promising lines in 25,000,

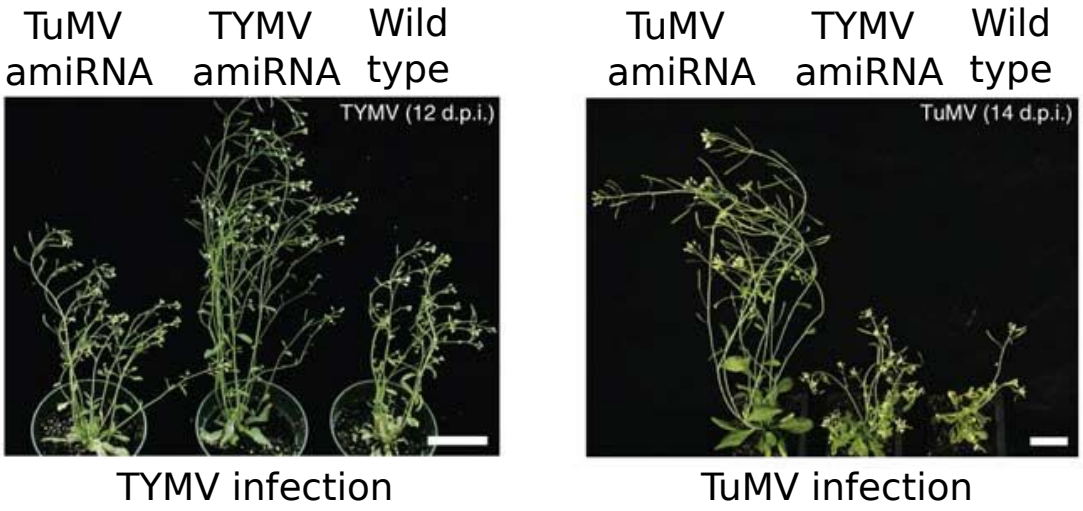


Fig. 1.10 amiRNAs produce specific and effective antiviral resistance. Two different amiRNA-expressing *Arabidopsis* lines challenged with turnip yellow mosaic virus (TYMV) and turnip mosaic virus (TuMV) show resistance to the virus targeted by their amiRNA. Figure adapted from Niu et al. (2006b).

none of which were breeding stock (CIMMYT). This meant that *R* genes against MLN were not available, although there were the two known SCMV resistance genes *Scmv1* and *Scmv2*. These genes were found to be effective in reducing (not eliminating) MLN symptoms in incubator and some field conditions, but overwhelmed in regions of high disease pressure (Bulegeya, 2016). A 2015 genome-wide association study (GWAS) searching for determinants of MLN resistance found 18 minor (explains <10% of variation in MLN resistance) and 6 medium (explains 10-20% of variation in MLN resistance) quantitative trait loci (QTL), suggesting that natural MLN resistance may be multigenic, although it's impossible to exclude monogenic resistance outside of the breeding panels used (Gowda et al., 2015). We decided to survey MLN using next-generation sequencing (NGS), which provides information on the variability of viruses both within and between samples. NGS provides the genome sequences of viruses in samples, so RNA-based resistance is a good choice as we would already have the data required for rational design. Out of the well characterised options for RNA-based resistance, amiRNAs are the most attractive for the reasons detailed above. Therefore, as an initial stage in a strategy to engineer resistance against MLN I set out to survey MLN in East Africa, determine its causal agents, and choose optimal targets for amiRNA mediated resistance.

Chapter 2

Methodological reference

2.1 Summary

Each experimental chapter contains a methodology section for the specific experiments described within that chapter. The purpose of this chapter is to act as a reference for media and growth conditions etc. used in techniques described later. Restriction digests, ligations and Gateway cloning were all performed according to manufacturers instructions.

2.2 PCR master mixes

The tables below show the volumes of reagents generally used for a single PCR reaction using the given polymerase. Volumes are scaled up as necessary to increase reaction volume and/or number of reactions.

Table 2.1 Makeup of a single Phusion (NEB) PCR reaction. Phusion PCR was generally used for cloning and related reactions, to minimise the probability of PCR errors.

Reagent	Volume
H2O	11.8
5x Phusion Buffer	4
10μM F primer	1
10μM R primer	1
10mM dNTPs	0.4
Phusion Taq	0.2
DMSO	0.6
Template	1

Table 2.2 Makeup of a single Dreamtaq (Thermo Scientific) PCR reaction. Dreamtaq PCR was generally used for genotyping and amplification from genomic DNA.

Reagent	Volume
H2O	15.7
10x dreamtaq Buffer	2
10μM F primer	0.4
10μM R primer	0.4
10mM dNTPs	0.4
DreamTaq polymerase	0.1
Template	1

Table 2.3 Makeup of a single KOD Xtreme (MilliporeSigma) PCR reaction. KOD PCR was used for amplification from maize genomic DNA.

Reagent	Volume
H2O	3.4
2x Xtreme Buffer	10
10μM F primer	0.6
10μM R primer	0.6
2mM dNTPs	4
KOD Xtreme polymerase	0.4
Template	1

2.3 Maize transformation

Table 2.4 Media compositions used for *Agrobacterium tumefaciens* transformation of maize embryos. 2,4-D = 2,4-Dichlorophenoxyacetic acid.

Media	Ingredient	Amount (per litre)
Infection medium	MS with vitamins	4.4g
	Sucrose	30g
	Glucose	20g
	2,4-D	1.5mg
	Casein hydrolysate	1mg
	Acetosyringone	100uM
	Adjust pH to 5.2	

Table 2.4 Media compositions used for *Agrobacterium tumefaciens* transformation of maize embryos. 2,4-D = 2,4-Dichlorophenoxyacetic acid.

Media	Ingredient	Amount (per litre)
Co-cultivation medium	MS with vitamins	4.4g
	Sucrose	20g
	Glucose	10g
	Proline	0.7g
	2,4-D	1.5mg
	CuSO4	100mM
	MES monohydrate	0.5g
	Agar	8g
Resting medium	MS with vitamins	4.4g
	Sucrose	20g
	Proline	0.7g
	2,4-D	1.5mg
	MES monohydrate	0.5g
	Agar	8g
	Silver nitrate	1.6mg
	Carbenicillin	250mg
First selection medium	MS with vitamins	4.4g
	Sucrose	30g
	Proline	0.7g
	2,4-D	1.5mg
	MES monohydrate	0.5g
	Agar	8g
	Carbenicillin	250mg
	Basta	1.5mg

Table 2.4 Media compositions used for *Agrobacterium tumefaciens* transformation of maize embryos. 2,4-D = 2,4-Dichlorophenoxyacetic acid.

Media	Ingredient	Amount (per litre)
Second selection medium	MS with vitamins	4.4g
	Sucrose	30g
	Proline	0.7g
	2,4-D	1.5mg
	MES monohydrate	0.5g
	Agar	8g
	Carbenicillin	250mg
	Basta	3mg
Embryo maturation medium	MS with vitamins	4.4g
	Sucrose	60g
	Proline	0.7g
	MES monohydrate	0.5g
	Agar	8g
	Carbenicillin	250mg
	Adjust pH to 5.8	
Regeneration medium	MS with vitamins	4.4g
	Sucrose	30g
	Proline	0.7g
	MES monohydrate	0.5g
	Agar	8g
	Adjust pH to 5.8	

2.4 Silencing suppressor assays

To generate inoculation cultures, single *A. tumefaciens* colonies on selective plates are transferred to selective lysogeny broth (LB) and grown at 28ºC for 48 hours. The infiltration buffer is used to resuspend *A. tumefaciens* strains $\geq$ 2 hours before inoculation.

Table 2.5 Lysogeny broth (LB) buffer, autoclaved before use.

Ingredient	Amount (per litre)
Tryptone	10g
Yeast extract	5g
NaCl	10g

Table 2.6 Ingredients required for one litre of *Agrobacterium tumefaciens* infiltration buffer. Strains are left in buffer for at least two hours before infiltration.

Ingredient	Amount (per litre)
MgSO4 (1M)	10ml
MES (0.5M, pH 5.6)	20ml
Acetosyringone (0.1M)	1.5ml
H2O	968.5

Chapter 3

Characterisation of maize chlorotic mottle virus in East Africa and globally

3.1 Summary and objectives

The first and only reported MCMV genome sequence from East Africa was generated from pooled Kenyan maize samples and was most closely related to Chinese MCMV isolates (Adams et al., 2013). Increased knowledge of MCMV genomic variation would provide insight into the epidemiology of MCMV globally, and inform the engineering of sequence-mediated resistance. Therefore, I set out to survey the MCMV outbreak in East Africa, and construct a first global phylogeny of MCMV. I show that East African MCMV isolates are highly homogeneous, and the sister clade of Chinese isolates. The first genome sequences from Hawaii and South America (Ecuador) combine to form the sister clade to previously reported North American isolates. The majority of global MCMV sequence diversity is between genetically differentiated and geographically isolated clades. I establish the use of the adjusted Rand index to extract clade-specific SNPs and document natural mutation within the systemic movement protein P31 in Hawaiian and Ecuadorian isolates.

Methodology

3.2 Maize chlorotic mottle virus genome sequence generation

3.2.1 East-African maize survey

During August 2014, I collected 25 maize leaf samples from Kenya, stored them on dry ice in RNA-later (Ambion), then extracted RNA using Trizol (Ambion) according to manufacturer's instructions. Farmers were asked to fill out a questionnaire with the help of a local translator to provide information on farm type and maize lines (table 3.1). A Kenyan Agriculture and Livestock Research Organisation (KALRO) colleague collected four leaf samples from Ethiopia. Food and Environment Research Agency (FERA, UK) provided RNA from an additional Rwandan maize leaf sample.

3.2.2 NGS library preparation

I depleted ribosomal RNA (rRNA) using the Ribo-Zero Magnetic Kit (Plant Leaf – Epicentre), and produced indexed stranded libraries using Scriptseq V2 RNA-Seq Library Preparation kits and Scriptseq Index PCR primers (Epicentre). I performed purification steps using Agencourt AMPure XP beads, then checked library quantity and quality using Qubit (Life Technologies) and a Bioanalyzer High Sensitivity DNA Chip (Agilent Technologies). Libraries were sent to Beijing Genomics Institute for 100bp paired-end sequencing on one lane of a HiSeq 2000 (Illumina).

3.2.3 Maize chlorotic mottle virus consensus sequence generation

I de-multiplexed libraries allowing one error within the index sequence using a custom python script, then trimmed adaptors using Trim galore! (parameters: *–phred64 –fastQC –illumina –length 30 –output_dir adaptors_removed_trimgalore_06_01_16 –paired -retain_unpaired input_1.fq input_2.fq*), which also removes bases at the 3' ends of reads with PHRED <20 (Krueger, 2015; Martin, 2011). Most deduplication methods use the start and end points of a read aligning to a reference, and delete all reads with identical start and end points. These methods are suitable for alignment to large genomes, where the chance of this occurring is low, but for short viral genomes with thousands-fold coverage this will occur many times. Therefore, I performed deduplication by string-matching using the script *fastq_duplicate_remover.py* from the Quality Assessment of Short Read (QUASR) pipeline

(Watson et al., 2013). I also used the QUASR script *quality_control.py* for quality trimming (parameters: *-m 30 -l 50* - removing bases from the 3' of reads until median PHRED score of the read is 30, with a minimum length of 50). Library quality was checked using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). I constructed a bowtie reference containing all 35 publicly available MCMV sequences using the *bowtie2-build* command, producing an index with each MCMV sequence as a separate "chromosome". To capture all MCMV-like reads I aligned to this index using bowtie2 (parameters: *-D 20 -R 2 -N 1 -L 20 -i S,0,2.50 –phred64 –maxins 1000 –fr*) (Langmead and Salzberg, 2012)).

I extracted the reads aligning to MCMV and performed *de-novo* assembly using Trinity, extracted MCMV contigs of 1kb or more using blast, then manually checked contigs and assembled contigs from the same library if necessary (Grabherr et al., 2011). To generate consensus sequences, I aligned each library to its respective Trinity MCMV contig using bowtie2, generated pileups using samtools, and called sequences using the QUASR script *pileup_consensus.py*, with a threshold of zero or ten % of reads to call bases with ambiguity codes (parameters: *–ambiguity 0|10 –dependent –cutoff 25 –lowcoverage 20*) (Li, 2011).

3.2.4 Sanger sequencing of maize chlorotic mottle virus isolates

Colleagues at the University of Hawaii purified MCMV from infected maize leaves collected on the island of Kauai. They extracted RNA using Trizol, then performed RT-PCR. They used primers based on the Nebraska isolate (EU358605.1) to amplify six amplicons covering the genome (Stenger and French, 2008), which they cloned into pCR4-TOPO TA plasmids and Sanger sequenced using M13 and internal primers to obtain total coverage. A consensus sequence was determined by sequencing six full-length clones. A total of three Ecuadorean isolates were fully sequenced for this study by colleagues at Centro de Investigaciones Biotecnológicas del Ecuador, Escuela Superior Politécnica del Litoral (CIBE-ESPOL, Ecuador). They designed primers from conserved regions using an alignment of available MCMV sequences, including a local isolate obtained previously by degenerate-oligonucleotide-primed RT-PCR using double-stranded RNA as template (Quito-Avila et al., 2016). They Sanger sequenced overlapping PCR amplification products (three independent reactions for each primer set) in both directions (Macrogen, South Korea).

3.3 Maize chlorotic mottle virus genome analysis

3.3.1 Sequence alignment, recombination and phylogenetic analyses

I aligned MCMV genomes using MUSCLE in MEGA6, with a gap extension cost of -1000 (other settings default) (Tamura et al., 2013). I checked and refined the alignment manually in JALview (Clamp et al., 2004). To visualise evidence of possible recombination, I used SplitsTree4 to generate splits networks, using default settings - distances calculated by uncorrected P (match option for ambiguous bases), and network generated by neighbour-net (Huson and Bryant, 2006).

To generate a MCMV phylogeny I used Bayesian inference; I split Alignment sites into three partitions: A) non-coding B) codon positions one and two, and overlapping ORFs (i.e. little to no degeneracy) C) codon position three. I generated phylogenetic trees using two runs of four Monte Carlo Markov Chain (MCMC) computations run for 1,000,000 generations under a general-time-reversible (GTR) model with a gamma distribution of rate variation between sites in MrBayes 3.2 (Ronquist et al., 2012). The first 10% of generations were discarded as burn-in (default). I used Tracer (http://beast.community/tracer) to examine convergence and effective sample size to confirm that estimated sample sizes for each parameter exceeded 200, as recommended by the MrBayes manual. I pooled 1800 trees, which were sampled every 500 generations (default) and constructed consensus trees, then visualised them in FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

3.3.2 SNP and population genetics analysis

I extracted alignment sites at which SNPs clustered similarly to phylogeny using the Adjusted Rand Index (ARI) to measure the equivalence between phylogenetic groupings and SNP segregation at each alignment site using custom R scripts (more detailed explanation in results section).

To produce population genetics indices and statistics I used an alignment of MCMV sequences without ambiguity codes (i.e. 0% threshold) in DnaSP v5, as DnaSP v5 does not accept ambiguity codes (Librado and Rozas, 2009). To test statistical significance, I used a permutation test with 1000 replications.

Results

3.4 Survey of maize lethal necrosis-symptomatic maize in East Africa

3.4.1 Maize lethal necrosis survey rationale

PCR- and antibody-based techniques for viral surveys are widely used, cheap, and require less specialist equipment than NGS. However, they are also heavily biased by the information input to the tests i.e. the primers or antibodies chosen. Even if the primers and antibodies selected are targeted to the correct viruses, variation in nucleotide sequence may prevent efficient primer binding, or mutations in the CP may prevent antibody binding, resulting in false-negatives. The latter phenomenon was observed in East-Africa; the first isolate of African MCMV is serologically distinct from North American isolates, against which commercially available MCMV-testing antibodies were raised (Adams et al., 2013).

NGS requires no prior knowledge to detect viruses, beyond the nucleic acid type (Adams et al., 2013). This makes it a valuable technique for surveying emerging viruses, especially in areas with limited agricultural research activity - i.e. areas likely to contain undiscovered viruses. This feature is especially useful for surveying MLN, owing to the synergistic nature of the condition, and the phylogenetic diversity of partner viruses for MCMV. To maximise the information collected on MCMV, I collaborated with scientists who could provide Sanger sequencing data on MCMV in other regions of the world.

3.4.2 Sampling strategy in Kenya

I went to Kenya to collect MLN-infected maize samples in August 2014. Sampling sites were chosen using a number of factors: presence of planted maize in August, avoidance of areas sampled by KALRO for their own NGS survey of MLN, variety in ecological zone, and the security situation (Eastern coastal areas were inaccessible at this time). I had oral agreement from colleagues at KALRO that we could pool data, so to maximise the amount of MCMV diversity captured we sampled different areas. Individual sampling sites were separated by at least 15km, again to promote variation between samples, with the exception of three sites adjacent to the Tana river (T2F2, T2F3, T2F4). Two sampling expeditions were undertaken, with 13 sites sampled in total (fig. 3.1a). Nomenclature of samples is as follows: T2F3S1 corresponds to sampling trip two, farm three, sample one. Details of each site can be found in table 3.1.

3.4.3 Global sampling strategy

To increase the amount of data available for analysis I collaborated with colleagues from the Food and Environment Research Agency (FERA, UK), CIBE-ESPOL, and the University of Hawaii (USA). Colleagues provided NGS data from Ethiopia (KALRO, fig. 3.1b) and Rwanda (FERA), and Sanger sequencing data from Hawaii (University of Hawaii) and Ecuador (CIBE-ESPOL) (table 3.1).

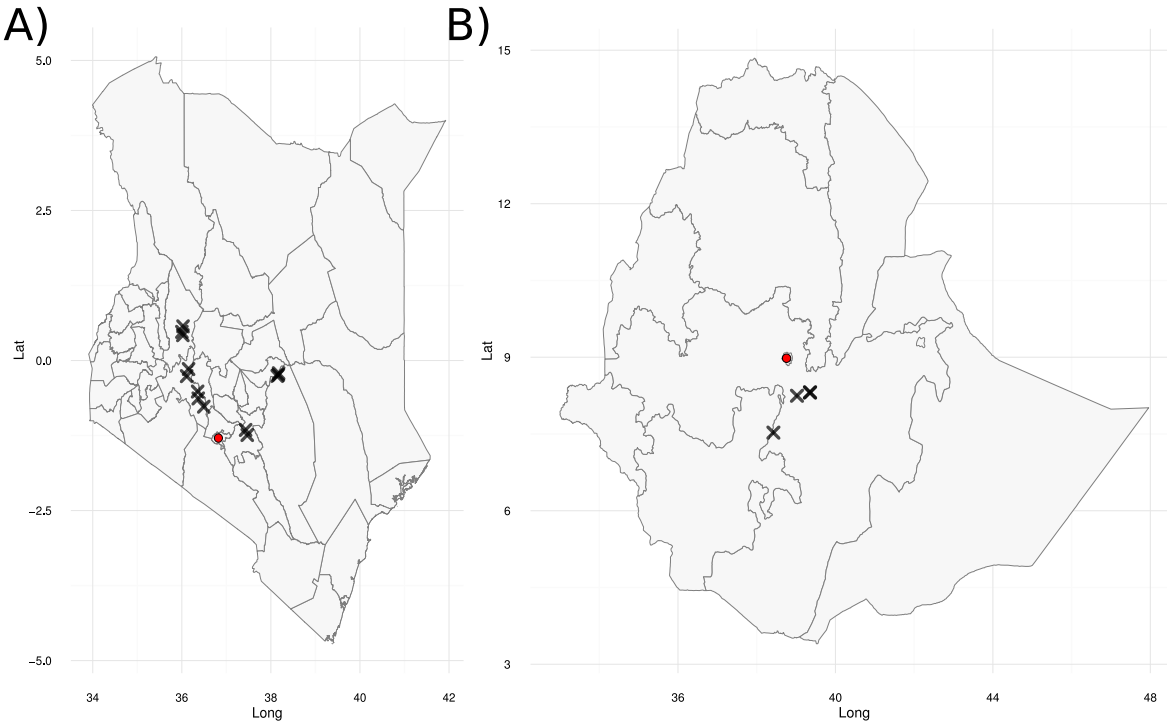


Fig. 3.1 Sampling sites in Kenya (A) and Ethiopia (B). Capital cities in red. Note that scale is different between A) and B).

Table 3.1 Maize sampling site details

Site	Country	Region	Latitude	Longitude	Altitude (m)	Date	Farm type	Maize lines	Crop stage
T1F1	Kenya	Nakuru	-0.260	36.109	1912	06/08/14	Mixed	KS H629	R5
T1F2	Kenya	Nakuru	-0.142	36.151	2086	06/08/14	Mixed	KS H614	R1
T1F3	Kenya	Baringo	0.477	36.001	1022	07/08/14	Arable	DPP-K3	V2-V8
T1F4	Kenya	Baringo	0.422	36.023	1005	07/08/14	Arable	Unknown	V1-R1
T1F5	Kenya	Baringo	0.569	36.027	1023	07/08/14	Arable	KS H4	V2-V8
T1F6	Kenya	Nakuru	-0.517	36.355	1985	08/08/14	Arable	Pioneer, Pannar	V4-V8
T1F7	Kenya	Nakuru	-0.632	36.369	1908	08/08/14	Arable	KS H614, H513	V5-R1
T1F8	Kenya	Nakuru	-0.769	36.494	2155	08/08/14	Arable	Unknown	V2-V9
T2F1	Kenya	Machakos	-1.241	37.469	1185	13/08/14	Arable	Pioneer 3253	V3-V6
T2F2	Kenya	Kitui	-0.257	38.142	442	14/08/14	Arable	Pannar 4m-21	VT
T2F3	Kenya	Kitui	-0.239	38.167	438	14/08/14	Arable	Pannar 4m-21	VT
T2F4	Kenya	Kitui	-0.212	38.174	419	14/08/14	Arable	Pannar 4m-21	V4-V6
T2F5	Kenya	Machakos	-1.156	37.433	1284	15/08/14	Arable	Pioneer	VT-R5
ETF1	Ethiopia	Oromia	8.310	39.352	1241	08/08/14	NR	Unknown	NR
ETF2	Ethiopia	Oromia	8.247	39.022	1595	09/08/14	NR	BH661	NR
ETF3	Ethiopia	Oromia	7.527	38.424	1642	09/08/14	NR	Unknown	NR
ETF4	Ethiopia	Oromia	8.314	39.356	1235	09/08/14	NR	Unknown	NR
B1	Rwanda	N. Province	NR	NR	NR	11/03/2013	NR	NR	NR
Kauai	USA	Hawaii	NR	NR	NR	12/2009	Arable	NR	NR
Portoviejo	Ecuador	Manami	NR	NR	44	12/2015	NR	Yellow hybrids	NR
Sta Elena	Ecuador	Santa Elena	NR	NR	0	09/2016	NR	Yellow hybrids	NR

3.5 A first global phylogeny of maize chlorotic mottle virus

3.5.1 Maize chlorotic mottle virus exhibits low sequence diversity

The initial aim of the MCMV survey was to investigate the level of structural and nucleotide diversity within and between regions containing MCMV, and examine the evidence for diversifying and purifying selection across the MCMV genome, predominantly using dN/dS based methods. I extracted MCMV genomes from the NGS data by aligning each library to a reference of all NCBI MCMV genomes, extracting aligned reads, and performing *de novo* assembly for each library using Trinity (v2.0.2). I then aligned each Library to its individually assembled MCMV genomes and consensus sequences called. This is a more convoluted approach than aligning to a reference then calling a consensus, but it is more sensitive to indels, especially short indels (see Samtools documentation). This produced 33 novel MCMV genomes, from Kenya (n=24), Ethiopia (n=5), and Rwanda (n=4). Sequence lengths ranged from 4399-4440bp; which is 99-100% coverage compared to the longest previously reported MCMV genome sequence (KF744393.1). Including all Genbank isolates, nucleotide identity between genomes ranged from 100% to 96.55%. Average nucleotide diversity across all sequences was 0.01, which is very low for viruses generally and similar to the lowest recorded diversity values in the tombusviridae (García-Arenal et al., 2001; Varanda et al., 2014a,b) (table 3.2). Most of the diversity estimates in table 3.2 are derived from CP sequences (typically subject to strong purifying selection), rather than whole genomes, and from a smaller number of sequences across a smaller geographic range, reinforcing that MCMV is atypically homogeneous. To have a more comparable measure, I also calculated the nucleotide diversity across the MCMV CP gene, which is close to that of the whole genome, and therefore still has low nucleotide diversity compared to most values in table 3.2 and the literature (García-Arenal et al., 2001). This may be a result of intense purifying selection across the genome, recent divergence, or a combination of both.

3.5.2 Bayesian inference of maize chlorotic mottle virus phylogeny

Phylogenetic inference relies on the partitioning of different states of characters (in molecular biology, typically sites in an alignment) between groups of individuals. The distribution of these character states are then used to infer which individuals are more or less closely related. The average number of SNPs between MCMV sequences is only 45.5 (table 3.3). Therefore, I decided to use nucleotide, rather than amino acid, alignments as the basis of my phylogenetic analyses to avoid decreasing the (small) number of characters available.

Phylogenetic analyses assume a single evolutionary history of each genome, an assumption which may be violated in the case of recombination. Recombination is a widespread phenomenon in RNA viruses and the tombusviridae (Boulila, 2011; Chare and Holmes, 2006; Wang et al., 2015b). I therefore performed a splits network analysis, which detects inconsistent phylogenetic signals across an alignment. The further a splits network is from a bifurcating tree in shape, the more inconsistent the signals, and therefore the stronger the evidence for recombination or rate variation (fig. 3.2). The splits network for MCMV shows clear separation of isolates from different regions, indicating that there has been no or little recombination between MCMV genomes in geographically isolated regions (fig. 3.3a). Conventional phylogenetic analysis is suitable, therefore, for investigating the relationships between these regions.

Mr Bayes (3.2) was used for Bayesian inference of MCMV phylogeny (fig. 3.3b). Bayesian inference was selected over maximum-likelihood as it can handle ambiguous bases and generate confidence estimates without data subsetting. Clearly resolved clades contain North American isolates, Hawaiian with South American isolates, and Chinese with African isolates.

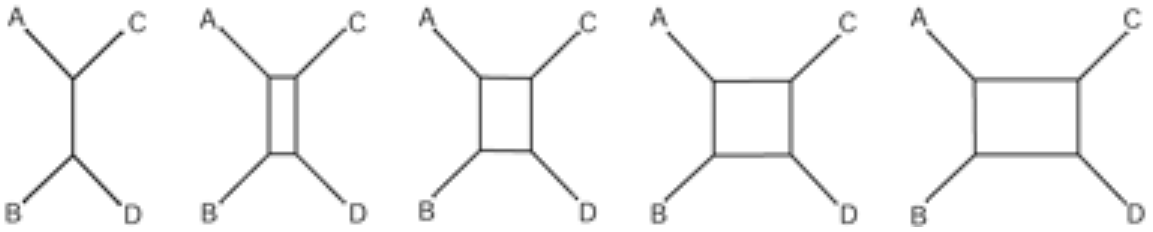


Fig. 3.2 Splits networks. Alignments contain increasingly conflicting phylogenetic information from left to right, and the networks produced are further from a bifurcating tree.

Table 3.2 Comparison of average nucleotide diversities calculated from nucleotide sequence data. CMV-sat=cucumber mosaic virus satellite RNA, CTV=citrus tristeza virus, CYSDV=cucurbit yellow stunting disorder virus, MCMV=maize chlorotic mottle virus, OLV1=olive latent virus-1, OMMV=olive mild mosaic virus, RTSV=rice tungro spherical virus, RYMV=rice yellow mottle virus, TNV-D=tobacco necrosis virus D, TMGMV=tobacco mild green mosaic virus.

Virus	Family	Genetic material	Isolate number	Sequence type	Number of continents	Average nucleotide diversity	Ref
MCMV	*Tombusviridae*	+ve RNA	49	Genome	4	0.01	This study
MCMV	*Tombusviridae*	+ve RNA	49	CP	4	0.01	This study
OLV1	*Tombusviridae*	+ve RNA	21	CP	3	0.02	Varanda et al. (2014a)
OMMV	*Tombusviridae*	+ve RNA	21	CP, in 81nt winows	1	0-0.02	Varanda et al. (2014b)
TNV-D	*Tombusviridae*	+ve RNA	14	CP, in 81nt winows	1	0-0.02	Varanda et al. (2014b)
TMGMV	*Virgaviridae*	+ve RNA	53	183K ORF	3	0.057	Fraile et al. (1996)
CTV	*Closteroviridae*	+ve RNA	30	1a, CP, 20k ORF	2	0.068	Rubio et al. (2001b)
CYSDV	*Closteroviridae*	+ve RNA	71	CP	3	0.002	Rubio et al. (2001a)
RYMV	*Sobemovirus*	+ve RNA	40	CP	1	0.0194	Pinel et al. (2000)
RTSV	*Sequiviridae*	+ve RNA	38	Two CP genes	1	0.09	Azzam et al. (2000)
CMV-sat	*Bromoviridae*	+ve RNA	25	Genome (satellite)	5	0.13	Fraile et al. (1991)

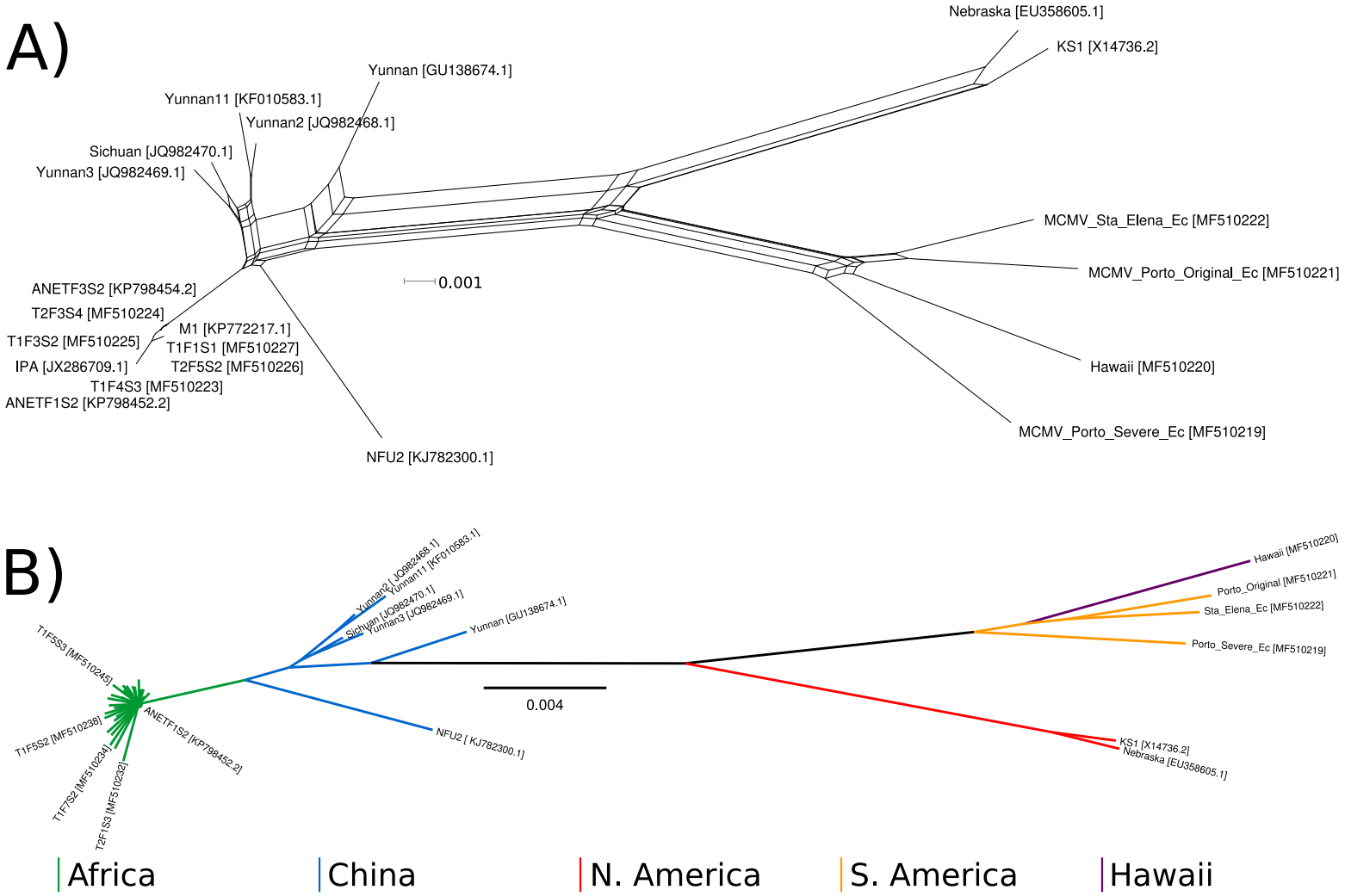


Fig. 3.3 A) Splits network of maize chlorotic mottle virus (MCMV) genomes, distances calculated with uncorrected P, and network generated by neighbour-net in Splitstree V4.6. B) MCMV phylogeny generated using Bayesian inference in Mr Bayes 3.2.

3.6 Maize chlorotic mottle virus genomic variation

3.6.1 Structural variation in maize chlorotic mottle virus

Structural variation within genomes was limited to the genomic termini (fig. 3.4). We found small 5' extensions of the genome – 18 African sequences had 1-3nt extensions beyond the 5' terminal A. All African isolates and the Taiwanese sweetcorn isolate (KJ782300.1) had a G insertion after the 5' terminal A to give a 5' terminal AGGG, as opposed to AGG (Deng et al., 2014). The G_9 sequence at the 3' terminal of the genome has a G insertion in North American isolate X14736.2, while the first reported Kenyan sequence, JX286709.1, has a G insertion immediately to the 3' of the TAAT sequence that follows G_9.

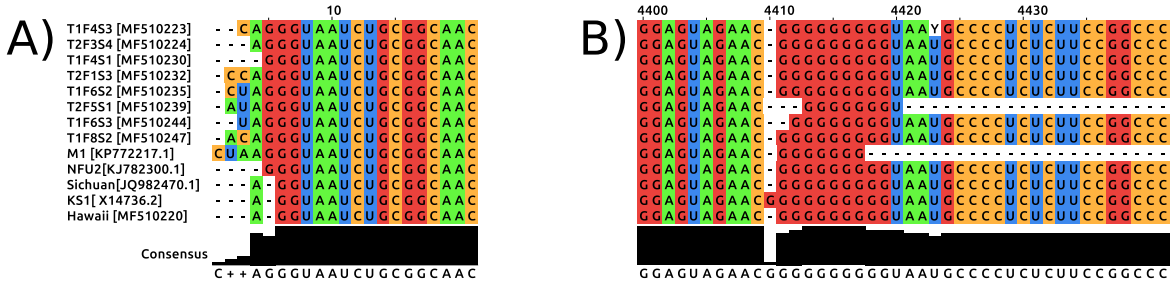


Fig. 3.4 Genomic termini of the maize chlorotic mottle virus genome, 5' (A) and 3' (B). Accession numbers are in square brackets. Sequences with identical termini have been removed for clarity.

3.6.2 Maize chlorotic mottle virus SNPs

There were 419 mutations at 388 polymorphic sites across the MCMV genome in our dataset (fig. 3.5). SNP density was highest 160bp from the 5' of the genome. This region contains the P50 and P32 ORFs, and the high SNP density is a mixture of ambiguous base calls in African isolates and SNPs between different geographic regions. Areas of low SNP density include known functional and selectively constrained sites, such as: the readthrough site of P111 (the RdRP), the GDD motif of P111, and the 3' non-coding region, which in *Tombusviridae* members contain a 3'CITE (Jiwan and White, 2011). Additional areas of low SNP density may represent functional sites, either in RNA or amino acid sequence, e.g. 600, 1800, and 2160bp into the genome.

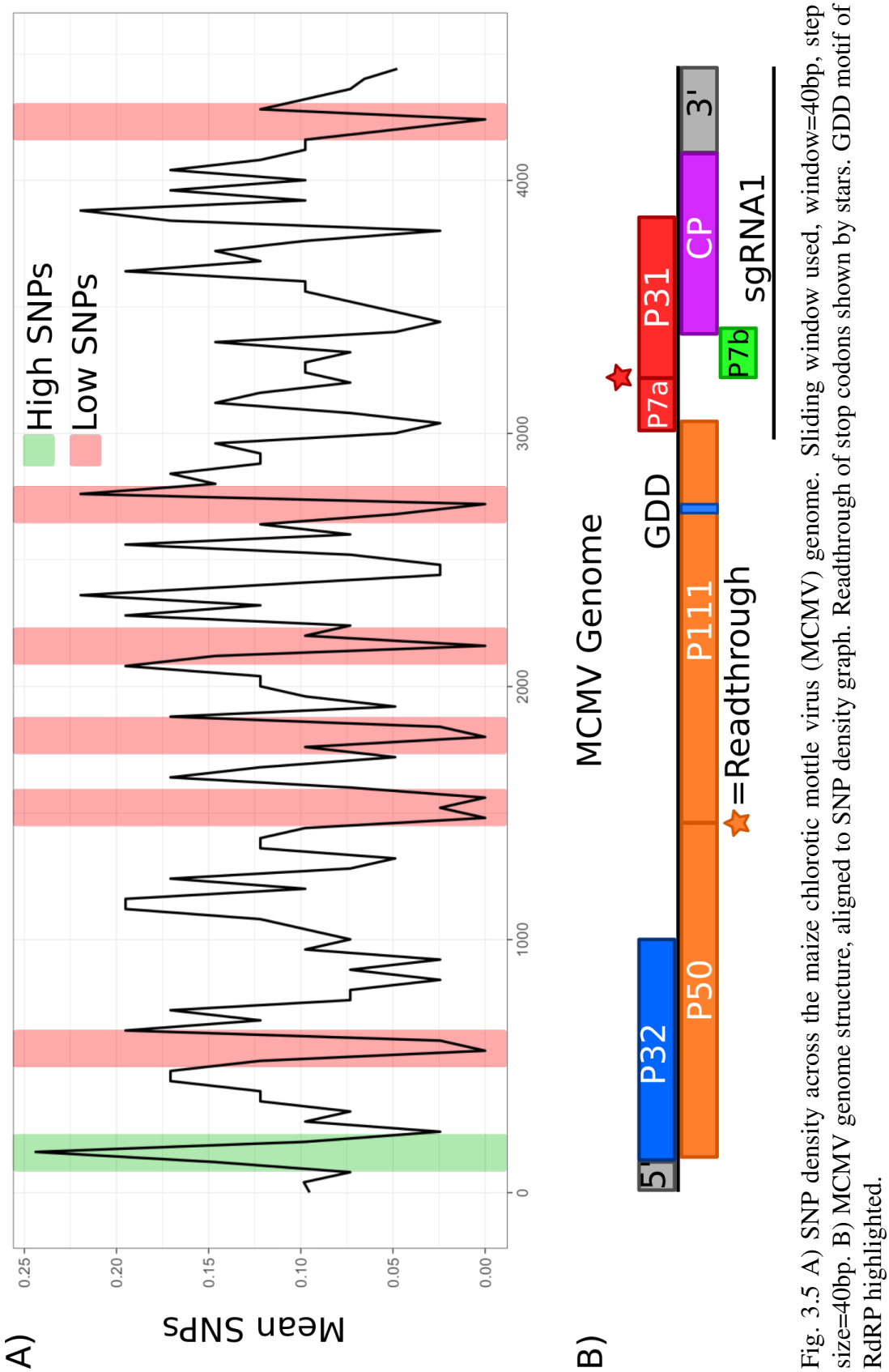


Fig. 3.5 A) SNP density across the maize chlorotic mottle virus (MCMV) genome. Sliding window used, window=40bp, step size=40bp. B) MCMV genome structure, aligned to SNP density graph. Readthrough of stop codons shown by stars. GDD motif of RdRP highlighted.

To identify geographic region-specific genomic variation I used the adjusted Rand index (ARI) (Hubert and Arabie, 1985). This is a chance-adjusted derivation of the Rand index, which measures the level of clustering between two datasets, and is defined as follows. Given a set, S, containing n elements, which is split into two alternative partitions of subsets, X and Y, then:

$a =$ no. of pairs of elements that are in the same subset in both X and Y

$b =$ no. of pairs of elements that are in different subsets in both X and Y

$c =$ no. of pairs of elements that are in the same subset in X and different subsets in Y

$d =$ no. of pairs of elements that are in different subsets in X and the same subset in Y

And the Rand index is given by:

$$R = \frac{a+b}{a+b+c+d}$$

The Rand index is 0 if X and Y agree on no pairs of points, and 1 if the clusters are identical. The ARI is a chance-adjusted measure of the similarity between two data clusterings, and instead varies between -1 (dissimilar) and 1 (similar). The ARI uses a contingency table to adjust for chance. Thereby:

$a_i =$ sum of elements present in cluster X_i that are present in each cluster of Y

$b_j =$ sum of elements present in cluster Y_j that are present in each cluster of X

And the adjusted Rand index is given by:

$$ARI = \frac{\sum_{ij}\binom{n_{ij}}{2} - [\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}]/\binom{n}{2}}{\frac{1}{2}[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}] - [\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}]/\binom{n}{2}}$$

Alternatively:

$$ARI = \frac{Index - Expectedindex}{Maxindex - Expectedindex}$$

Where Index is the total number of pairs shared between X and Y, ExpectedIndex is the expected number of pairs if they were spread randomly between the same size and number of clusters, and Maxindex the average of the maximum possible number of pairs for each cluster (i.e. X=Y with either X or Y reflecting the other). The phylogenetic clades (Africa, China, North America, and South America with Hawaii) were used as one data cluster, and the other was one site in the nucleotide alignment. Therefore, ARI values above zero are produced when SNPs segregated according to the phylogenetic clades. Invariant sites have an ARI of 0

and are not of interest so were removed from the alignment. To identify sites which clustered significantly better than chance, I randomised the members of each geographic clade and recorded the ARI across the genome 1000 times. Sites above the 95% level of the remaining randomised ARI scores were taken to be significant (fig. 3.6). Significant sites were extracted, identifying candidate SNPs for (currently uncharacterised) phenotypic variation between clades, which can be used to design diagnostics appropriate for differentiating isolates from different regions.

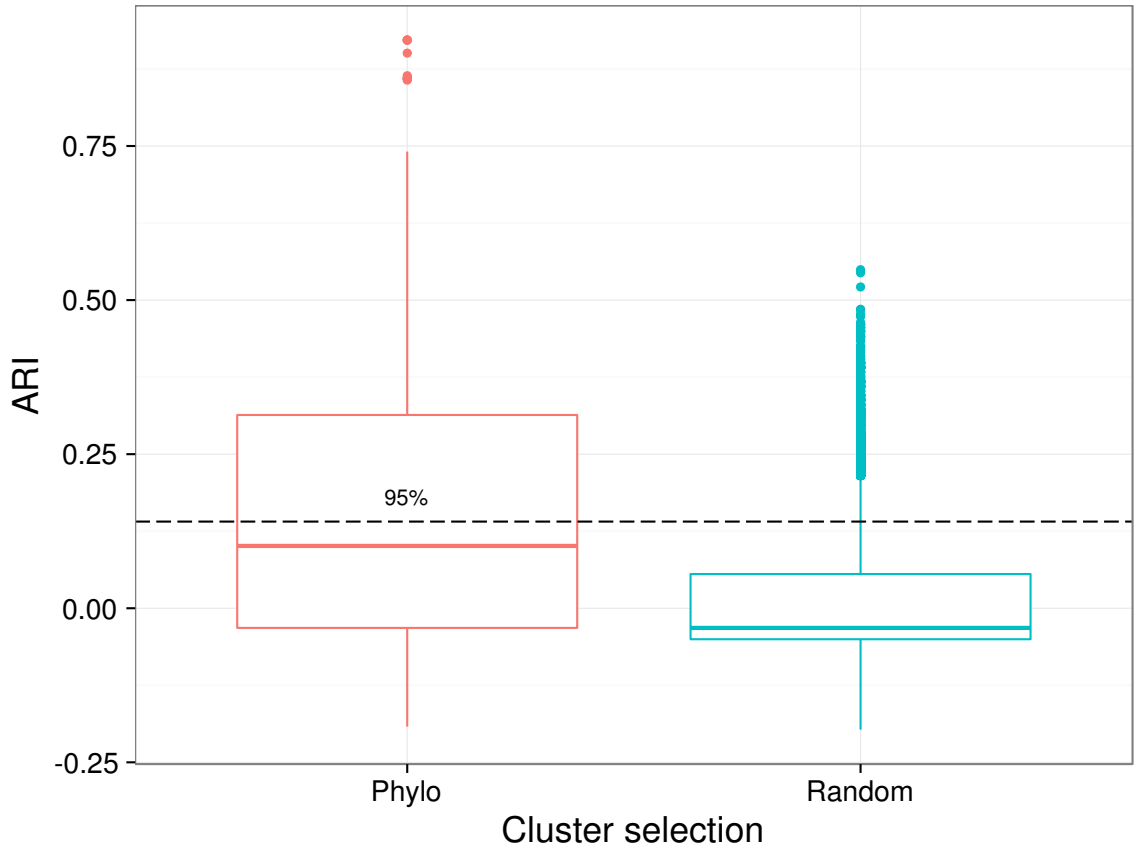


Fig. 3.6 Adjusted Rand Index (ARI) distribution for sites in the maize chlorotic mottle virus nucleotide alignment, generated using the phylogenetic and 1000 random clusters with the same size as the phylogenetic groups. Horizontal line indicates 95% significance of randomly sampled ARI values.

3.6.3 Maize chlorotic mottle virus amino acid variation

To identify protein coding changes within the clades I repeated the ARI analysis on translated alignments for each MCMV ORF (supplementary tables 8.1-8.8). The most significant variation was observed in the coat protein gene and the P7a/P31 region in which the systemic movement protein P31 has been proposed to be expressed by readthrough of the P7a ORF UGA stop codon. We observed unreported variation in an exposed loop of the MCMV capsid - Phe76Leu in an isolate from Chinese sugarcane (KF010583.1).

P31 enhances systemic MCMV movement via an unknown mechanism (Scheets, 2016). However, in Ecuadorian and Hawaiian isolates there is an early stop codon 18bp downstream of the P7a UGA stop codon, truncating the majority of P31 (fig. 3.7). This unexpected stop codon is also present in the two previously reported partially sequenced Ecuadorian MCMV isolates, in which there are also 47 and 36bp deletions in the P7a and P31 coding sequences upstream of the stop codon (Quito-Avila et al., 2016). This early stop codon could either be real variation or PCR error. The Sanger sequencing data for the Hawaii isolate was obtained by sequencing six PCR products subcloned into pCR4-TOPO TA vector. The five Ecuadorian isolates were all sequenced using PCR reactions in triplicate. Therefore, the stop codon appears in PCRs derived from six different isolates, from two different countries and performed by two different researchers (who were unaware of the stop codon until I notified them). Therefore, I conclude that the stop codon represents true variation rather than an artifact.

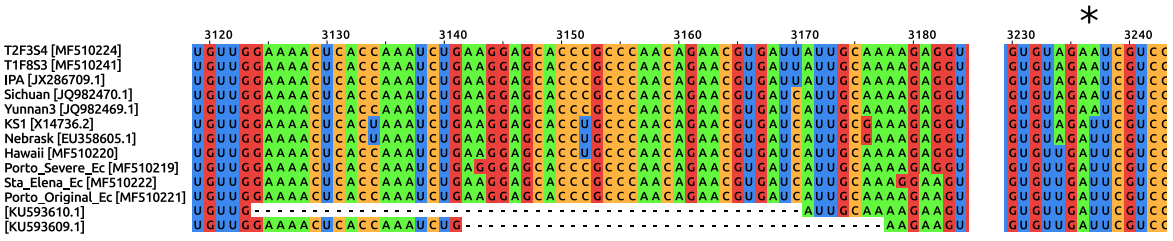


Fig. 3.7 Natural variation in systemic movement protein P31. Deletion present in the first reported Ecuadorian isolates, and stop codon represented by asterisk in all Ecuadorian and Hawaiian isolates.

3.6.4 Selection across the maize chlorotic mottle virus genome

The HYPHY server (www.datamonkey.org) was used to test for selection using alignments corresponding to each ORF in the MCMV genome. MEME was used to detect episodic

positive selection, with significant ($p<0.01$) codons found in P50, P32, P31, and P111 (Murrell et al., 2012). IFEL was used for detection of positive and negative selection in each MCMV ORF. Significant ($p<0.01$) positive selection on codons was detected in P50, P32, and P31, with significant negative (purifying) selection found at sites in all ORFs.

Positive selection has not previously been detected in the MCMV genome, likely due to the small number of sequences available previously (Boulila, 2011). This positive selection in three proteins of unknown function may represent adaptation to specific host interactions across different maize lines, or to different environments. Purifying selection was detected across the coding regions of the MCMV genome, especially in the post-readthrough region of the RdRP, which is highly conserved amongst *Tombusviridae* members, and the C-terminus of the CP (fig. 3.8). This is the location of an asymmetric unit important for interaction between capsid monomers in the assembled viral coat (Wang et al., 2015a). dN/dS analyses of selection in viral genomes may be limited by selection at the RNA sequence level to preserve functional RNA structures, for example the *Tombusviridae* subgenomic RNA promoters or 3'CITEs (Newburn and White, 2015). Additionally, dN/dS analysis of regions with overlapping ORFs is limited by selection on the two ORFs - mutations may be due to drift in one ORF or selection in the other.

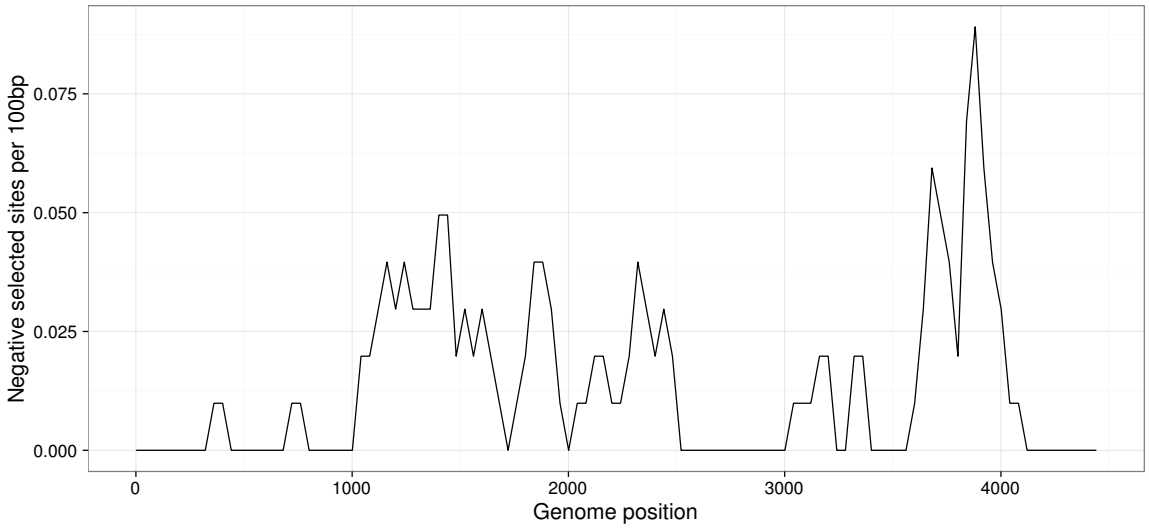


Fig. 3.8 Density of sites undergoing significant purifying selection, determined using IFEL, across the maize chlorotic mottle virus genome. Regions with a high density of negatively selected sites include the RNA-dependent RNA-polymerase (RdRP) readthrough region (around 1500bp), and the C-terminus of the coat protein (around 3800bp).

3.7 Population genetics of maize chlorotic mottle virus

Across all MCMV genomes sequence diversity was low, with 45.5 nucleotide differences between sequences on average (table 3.3). Haplotype diversity was high, 0.99, with 42/49 haplotypes unique, although these are separated by a low number of SNPs. For subpopulation analysis, China and Africa were considered separately to allow estimation of gene flow between the populations. In terms of nucleotide diversity, the most diverse clade was South American and Hawaiian (pi = 0.015), while the least was Africa (pi = 0.0017), with only 7.51 nucleotide differences between African sequences on average. Variation in nucleotide identity was extremely limited for both Chinese and African sequences, with >99% sequence identity within each group.

Genetic differentiation between subpopulations was tested using a number of test statistics (H_{st}, K_s, K_s*, Z, Z*, S_{nn}, N_{st}, F_{st}) (table 3.4), with all statistics indicating differentiation with a high degree of significance (table 3.5). S_{nn} is powerful at all sample sizes and diversities, and so is most appropriate in this case, due to uneven sample sizes (Hudson, 2000). Hudson's H_s tests differentiation based on haplotype diversity, and was not significant, presumably due to the high proportion of unique haplotypes across all populations. Likewise, F_{st} and N_{st} values of 0.74 indicate that a high proportion of the genetic variation in MCMV is explained by population structure (Hudson et al., 1992b; Lynch and Crease, 1990). F_{st} and N_{st} values generated from pairwise comparisons between populations (table 3.6) illustrate that the sub-populations with least genetic variation explained by population structure are China and Africa, which can be expected given their phylogenetic proximity.

Table 3.3 Maize chlorotic mottle virus genome diversity globally and within phylogenetic groups.

Population	Genomes	Segregating sites	No. of haplotypes	Haplotype diversity	Average no. of SNPs	Nucleotide diversity
Global	49	388	42	0.986	45.5	0.0104
South America and Hawaii	4	123	4	1	64.7	0.0148
North America	2	22	2	1	22	0.00505
China	6	98	6	1	36.1	0.00829
Africa	37	114	30	0.974	7.51	0.00172

Table 3.4 Population differentiation metrics

Metric	Data type	Notes	Ref
H_{st}	Haplotype	1- (weighted average of estimated subpop. hap. div./ estimated population hap. div.)	Hudson et al. (1992a)
K_s	Sequence	Weighted average of estimated subpop. nuc. div.	Hudson et al. (1992a)
K_s*	Sequence	K_s with log transformation of differences between sequences, gives less weighting to pairs with large differences	Hudson et al. (1992a)
Z	Sequence	Sequence differences between pairs are ranked, Z is the weighted sum of subpop. div. (judged by rank)	Hudson et al. (1992a)
Z*	Sequence	Like Ks*, but uses log transformation of ranks	Hudson et al. (1992a)
S_{nn}	Sequence	Average fraction of sequences whose nearest neighbours in sequence space are from the same region	Hudson (2000)
N_{st}	Sequence	Average substitutions per site between pop.s/ (Average within pop. substitutions + between pop. substitutions)	Lynch and Crease (1990)
F_{st}	Multiple (sequence in this study)	(Average pairwise differences between pop.s- Average pairwise differences within pop.s)/ Average pairwise differences between pop.s	Wright (1951)

Table 3.5 Maize chlorotic mottle virus population differentiation

Population statistic	Value	P-value
H_s	0.978	0.072
K_s	16.3	0
K_s*	2.24	0
Z	400	0
Z*	5.69	0
S_{nn}	1	0

Table 3.6 Maize chlorotic mottle virus subpopulation differentiation

Population 1	Population 2	H_s	K_s	N_{st}	F_{st}
South America and Hawaii	North America	1	50.4	0.684	0.680
South America and Hawaii	China	1	47.5	0.638	0.633
South America and Hawaii	Africa	0.976	13.1	0.747	0.744
North America	China	1	32.6	0.790	0.786
North America	Africa	0.974	8.25	0.900	0.898
China	Africa	0.977	11.5	0.511	0.511

Discussion

3.8 Maize chlorotic mottle virus epidemiology

3.8.1 Maize chlorotic mottle virus phylogeny and epidemiology

Bayesian phylogenetic analysis shows clear separation between geographic regions, with the exception of Hawaii and Ecuador. Phylogenetic confidence estimates (fig. 3.3) and measures of population diversity (table 3.5) illustrate that the majority of MCMV genome diversity occurs between sub-populations. For China, East Africa, and Ecuador, intra-group diversity is minimal, suggesting either (separate) single introduction events or repeated introductions from the same sub-populations. In terms of sequence identity there are three major clusters: A) Hawaii and Ecuador, B) North America, and C) China and Africa. The similarity of Ecuadorian sequences to the Hawaiian sequence, and African to Chinese sequences, could represent the epidemiological routes of MCMV across the globe. If that is the case, China would be the source of the East African outbreak, and Hawaii the source of the Ecuadorian; however corroborating evidence and more complete sampling of global MCMV genomes would be required to confirm this. In particular, sequence data from other Central and South American countries with MCMV presence could determine whether the MCMV isolates in Ecuador are more closely related to those found in neighbouring countries or Hawaii. It would also determine whether the extremely limited sequence divergence amongst MCMV isolates is universal, suggesting either recent evolution and/or intense purifying selection, or whether undocumented diversity is present in its presumed ancestral range – the ancestral ranges of *Teosinte* and maize. Finally I highlight two general conclusions from RNA virus experimental evolution studies: firstly, that advantageous mutations with large fitness benefit are fixed early on in adaptation to new conditions, and secondly, that genetic diversity remains low during the period of maximum fitness increase, then increases as fitness levels off (Moya et al., 2004). In the case of MCMV, I could infer that the small number of SNPs present are likely to increase fitness in their specific geographic regions, and that MCMV fitness is still increasing. However, experimental evidence using different MCMV strains and mutational analysis would be required to confirm this.

3.8.2 Seed transmission of maize chlorotic mottle virus

Seed transmission of MCMV has been reported at a rate of 0-0.33% (Jensen, 1991). However, this estimate was derived from a naturally infected population (i.e. likely that <100% of plants were infected), so may be an underestimate. 12 of 26 ten-seed samples bought at Kenyan markets and tested by RT-PCR were positive for MCMV, and 72% of kernels from a single plant, which demonstrates MCMV presence in Kenyan maize seed, but not its transmission to seedlings (Mahuku et al., 2015). In Ecuador, seeds from two commercial hybrids planted in the region were grown in sterile soil inside insect-proof growth chambers, and 8% and 12% of seedlings were MCMV-infected (Quito-Avila et al., 2016). These isolates, and the Ecuadorian genomes collected in this study, were most closely related to the Hawaiian genome. Hawaii is an important region for maize seed production for a number of major agricultural companies, although it is unknown whether the infected seeds from Ecuador were produced in Hawaii. Regardless, it is vital that the ability of MCMV to be transmitted via maize seed is investigated further, especially given Hawaii's central role in maize seed production and the rapid spread of MCMV across East Africa. However, given the low rate of seed transmission, and the high populations of thrip vectors in the *Frankliniella* genus at affected sites, I would suggest that although seed transmission may be important for long distance dispersal, once arriving in an area vector transmission and soil transmission will likely be more important for local spread (Deng et al., 2014; Mahuku et al., 2015; Wangai et al., 2012).

3.9 Coding variation in maize chlorotic mottle virus

3.9.1 P31 variation

MCMV P31 has unknown function, a unique carboxy-terminal extension, and mutagenesis experiments suggest that it promotes systemic movement (Scheets, 2016). All Hawaiian and Ecuadorian isolates contain an early stop codon in P31, truncating the ORF six amino acids after the readthrough stop codon of the movement protein P7a. This leaves 162nt of non-coding RNA in the Hawaiian and Ecuadorian isolates, before the capsid gene initiates. Interestingly, the partial Ecuadorian sequences reported previously also contain this stop codon, as well as deletions upstream within the P7a coding region (Quito-Avila et al., 2016). This could represent the result of selective pressure for smaller genomes once the function of P31 was lost, which is commonly observed in RNA viruses (e.g. Zwart et al. (2014)).

Scheets used a mutant cDNA clone with an early stop codon mutation (p7bQ12N) to investigate P31 function, similar to the natural mutants we report from Hawaii and

Ecuador. Inoculation with p7bQ12N resulted in slower systemic spread of MCMV, and sequencing showed that the systemically infected leaves contained either a mixture of p7bQ12N with pseudorevertant mutants (the stop codon removed, but not returned to its original sequence), or exclusively pseudorevertant genomes (Scheets, 2016). Mixtures were predominantly p7bQ12N, raising the possibility that the wild isolates from this study contained low-frequency genomes with functional P31 which complemented the consensus genomes with mutated P31. Complementation of this form has been observed in tomato aspermy cucumovirus - 76% of genomes had a mutated movement protein, which were complemented by a minority of genomes with functional movement proteins (Moreno et al., 1997). Experimentally mixing mutant and wild-type MCMV *in-vitro* transcripts for innoculation could establish whether this is a possibility for MCMV. Alternatively, it could represent transcomplementation by a common partner virus, such as SCMV, although this seems less likely given that we do not currently know them to use the same insect vectors (though both use seed transmission, and aphids [SCMV] and thrips [MCMV] are generally common).

3.9.2 CP variation

Exposed capsid loops can function in vector transmission, and Wang et al. (2015a) previously identified two variable residues in the exposed loops of the MCMV capsid, Pro81Ser in the Nebraska (NC003627) and Kansas (EU358605.1) isolates, and Ala62Asp in the 2012 Kenyan isolate (JX286709.1). We identified another variable residue, with Phe76Leu in an isolate from Chinese sugarcane (KF010583.1). Different species of thrips and chrysomelid beetles have been shown to vector MCMV, and it would be interesting to compare transmission rates via different thrip and chrysomelid beetle species between strains with variation in exposed CP residues.

Conclusions

The aim of this analysis was to investigate global MCMV epidemiology and diversity, and use this information to inform a sequence-mediated resistance approach. MCMV phylogeny suggests links between the Chinese and African MCMV outbreaks, and the Hawaiian and Ecuadorian outbreaks. Growing evidence about seed transmission of MCMV, combined with its rapid spread to distant geographic regions, highlights the importance of future work in this area.

I established the use of the ARI for identification of SNPs and amino acids that vary between clades, but this variation can't be connected to phenotypes until studies of variation between strains have been performed. ARI analysis is appropriate for identifying SNPs which correlate with non-phylogenetic clusters as well, for example it could be used to highlight SNPs that vary between viral strains that infect different hosts.

MCMV in East Africa, and globally, has low sequence diversity, similar to other tombusviridae members, when compared to RNA viruses more generally. The variation present is predominantly explained by population structure between geographically isolated populations. The extremely low diversity of MCMV makes it a promising candidate for engineered sequence-mediated resistance.

Acknowledgements

I would like to thank a number of colleagues at KALRO for their help in sampling maize in Kenya, especially Jane Wamaitha, who acted as my guide, and a number of KALRO field officers who aided in locating farms, communicating with farmers, and translation. Additionally I would like to thank the Kenyan farmers who allowed to enter their fields and collect samples, and Bramwel Wanjala for organising shipment of the samples.

Chapter 4

Characterisation of maize chlorotic mottle virus partner viruses in East Africa

4.1 Summary and objectives

MLN is caused by the interaction of MCMV with *Potyviridae* members. MCMV often interacts with locally present viruses, complicating predictions of MLN incidence and spread. The partner viruses active in East Africa are unknown, although the first two reports find SCMV in complex with MCMV (Adams et al., 2013; Wangai et al., 2012). I set out to characterise the partner viruses in complex with MCMV using the samples described in the previous chapter. I show that the only known MLN partner virus in my samples is SCMV. SCMV in East Africa, as elsewhere, is diverse in terms of nucleotide identity, and recombination analysis provides strong evidence of extensive and complex historical recombination. In addition, I use *de novo* assembly to investigate the presence of additional RNA viruses, and find three novel virus-like sequences, which I confirm by RT-PCR.

Methodology

4.2 Sugarcane mosaic virus analysis

4.2.1 Sugarcane mosaic virus consensus sequence generation

The methodology for maize sampling, library preparation, deduplication and quality trimming can be found in the methodology for chapter 3. To generate SCMV genome sequences, I aligned libraries to a bowtie2 reference containing all SCMV genome sequences available in March 2016 (parameters: *-D 20 -R 2 -N 1 -L 20 -i S,0,2.50 –phred64 –maxins 1000 –fr*) (Langmead and Salzberg, 2012)). Next I extracted SCMV-aligning reads and performed *de novo* assembly using Trinity (v2.0.2), extracted contigs above 2kb in length, then inspected and curated (if necessary) SCMV contigs (Grabherr et al., 2011). To generate SCMV consensus sequences, I aligned each library to its respective Trinity contig using bowtie2, generated pileups using samtools, and called sequences using the QUASR script *pileup_consensus.py*, with a threshold of zero or ten % of reads for the calling of ambiguity codes (parameters: *–ambiguity 0|10 –dependent –cutoff 25 –lowcoverage 20*) (Li, 2011).

4.2.2 Sugarcane mosaic virus alignment and diversity analysis

SCMV genomes were combined with those available from NCBI and aligned using MUSCLE (gap extension cost: 800, other settings default) in MEGA6 (Tamura et al., 2013). I checked the alignment manually in JALview and refined it where necessary (Clamp et al., 2004). I obtained diversity metrics using the alignment without ambiguous base calls in DnaSP v5 (Librado and Rozas, 2009).

4.2.3 Sugarcane mosaic virus recombination analysis

As with the MCMV analysis I used SplitsTree4 to generate splits networks, using default settings - distances calculated by uncorrected P (match option for ambiguous bases), and network generated by neighbour-net (Huson and Bryant, 2006). To generate more specific predictions of recombination, I used Recombination Detection Programme 4 (RDP4), using the algorithms RDP, GENECONV, MaxChi, BootScan, and SiScan (all default settings), and reviewing all breakpoints manually.

4.3 Virus-like sequence detection and characterisation

4.3.1 Alternative preprocessing of libraries

I de-multiplexed all libraries allowing one error within the index sequence using a custom python script. Then, in order to optimise the Trinity assembly of NGS reads, I used three alternative preprocessing strategies:

1. No deduplication, followed by adaptor removal with trimmomatic, and quality trimming with trimmomatic (parameters: *LEADING:20 TRAILING:20 SLIDINGWINDOW:4:20 MINLEN:40* - trimming bases with a PHRED score <20 at the 5' and 3' ends of the read, then moving from 5' to 3' and cutting the read once the average PHRED score drops below 20 over a 4bp window).

2. Adaptor removal using Trim galore! (parameters: *–phred64 –fastQC –illumina –length 30 –output_dir adaptors_removed_trimgalore_06_01_16 –paired -retain_unpaired input_1.fq input_2.fq* - removes bases at 3' of reads with PHRED <20), followed by use of QUASR scripts: for string-matching deduplication (script: *fastq_duplicate_remover.py*) and quality trimming (script: *quality_control.py*, parameters: *-m 30 -l 50* - removing bases from the 3' of reads until median PHRED score of the read is 30, with a minimum length of 50).

3. As for number 2, but for each specific virus-like sequence (VLS) being sought, the libraries containing it had their reads combined prior to Trinity assembly.

4.3.2 *De novo* assembly and virus-like sequence extraction

I took reads not aligning to the maize genome (B73 Refgen V2) and assembled them using Trinity (v2.0.2), then extracted contigs over 2kb in length and performed blast search (blastn) against the NCBI nt database (Grabherr et al., 2011). I then checked the blast results and selected contigs with viral hits for further analysis.

4.3.3 RT-PCR testing of virus-like sequences, and characterisation

To confirm the presence of VLS I reverse transcribed the RNA samples used for library preparation using Superscript III (Thermo-Fisher), then performed PCR using Phusion polymerase (NEB). VLS were characterised using the NCBI ORF finder (https://www.ncbi.nlm.nih.gov/orffinder/) and conserved domain search (https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml).

Results

4.4 Sugarcane mosaic virus is present with maize chlorotic mottle virus across East Africa

The aim of this analysis was to identify and characterise the viruses acting together with MCMV to cause MLN in East Africa, and survey RNA viruses in East African maize more generally. The known partners of MCMV in MLN are SCMV, MDMV and WSMV (Goldberg and Brakke, 1987; Uyemoto, 1983). I used the same samples and RNA-seq NGS libraries as in chapter 3, maize leaf samples with MLN symptoms from Kenya, Ethiopia, and Rwanda (Kenya and Ethiopia sites can be seen in fig. 4.17). I had wanted to collect symptom-less plants as part of my survey, but this was impossible on most sites due to high MLN occurrence, and no negative samples were taken. I aligned the libraries to a reference containing the maize transcriptome and all viruses known to be involved in MLN. This showed that only MCMV and SCMV are present in East African maize samples (fig. 4.1). MCMV has been reported to have an extremely high viral titre, and this is consistent with my data - MCMV reads make up the majority of reads in most of my libraries (fig. 4.1a) (Scheets, 1998). In 13 (ribo-zero treated) libraries, >90% of reads align to the MCMV genome, showing that in many MLN-infected plants, MCMV RNA vastly outnumbers both that of the host transcriptome and its partner virus.

The low percentage of reads aligning to SCMV in most libraries could represent either SCMV presence at low concentration, or reads aligning to similar RNA molecules, which could be derived from the host transcriptome or other RNA viruses present in East Africa. To differentiate between these two options, I checked the alignment of reads in igv (fig. 4.1c). I would expect that reads would be distributed approximately evenly across the SCMV genome if SCMV is present, whereas if they are reads from another potyvirus I would expect most of the reads in the alignment to be aligning the conserved P3 region of the genome, but not elsewhere. The alignments in igv showed that reads aligning to SCMV were distributed continuously along the SCMV genome, confirming that SCMV is present in all libraries. Figure 4.1c shows reads aligning to SCMV in the library with the lowest proportion of SCMV-aligning reads, ANETF4S2.

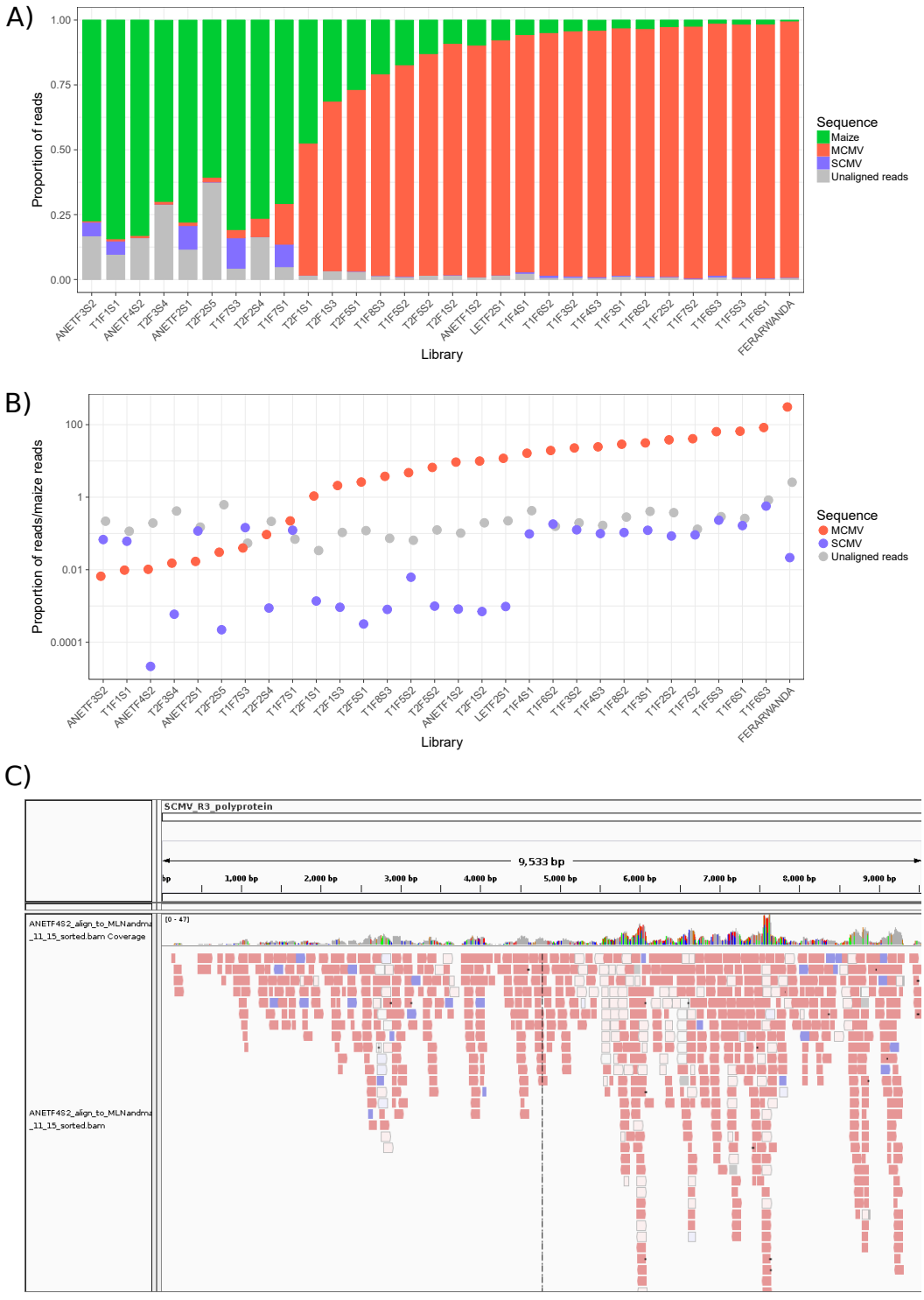


Fig. 4.1 A) Proportion of reads aligning to maize genome, maize chlorotic mottle virus (MCMV), and sugarcane mosaic virus (SCMV). B) Reads aligning to MCMV and SCMV, relative to maize genome reads, with $\log_{10}$ Y axis. Samples on X axis in both graphs are ordered in order of ascending MCMV load. C) Screengrab from igv showing alignment across the SCMV R3 genome (KF744390.1) for library ANETF4S2. The main window shows aligning reads (predominantly red), the small window above displays coverage (range: 0-47X). Multiple reads align with high confidence across almost the entire genome, confirming the presence of SCMV.

4.5 Sugarcane mosaic virus recombination and phylogeny

4.5.1 Sugarcane mosaic virus alignments patterns suggest recombination

The first NGS study of MLN in East Africa used sequencing of a pooled sample, and the partial SCMV consensus sequences obtained were most similar to the divergent strain of BD Chinese isolates (JN021933.1 and JX047381-91) (Adams et al., 2013). The BD strain is a recently discovered, highly virulent group Chinese SCMV isolates, with low (79-81%) nucleotide similarity to all other isolates on NCBI in 2013 (Gao et al., 2011). I used an alignment reference containing two SCMV genomes due to the divergence of BD8: BD8 and the Rwandan isolate R3 (KF744390.1) ((Adams et al., 2014)). Bowtie2 aligns reads to the most similar location in the reference, and the majority of libraries aligned predominantly to R3. BD8 alignment was higher than R3 in five libraries (T1F5S3, T1F6S1, T1F6S2, T1F6S3, T1F8S2), while in five others (T1F1S1, T1F4S3, T1F6S1, T1F7S2, T1F8S2) alignment was high for both BD8 and R3 at different locations in the genome (fig. 4.2). This pattern of alignment is suggestive of recombination, which has been reported previously in potyviruses and SCMV (Chare and Holmes, 2006; Li et al., 2013; Moradi et al., 2016; Xie et al., 2016).

4.5.2 Splits-network analysis of sugarcane mosaic virus

To investigate recombination in SCMV further I performed a splits network analysis. Using a similar approach to my analysis of MCMV (chapter 3) I carried out the following operations: extracted SCMV consensus sequences with alignment to all NCBI genomes, *de novo* assembly of aligning reads, manual curation of contigs, then alignment to these contigs followed by consensus calling. For recombination analysis, I used SCMV sequences with ambiguous bases called at the 10% threshold. The lower coverage of the SCMV genome compared to MCMV in my libraries meant that only 23 samples produced long (>2k) SCMV contigs. There are a large number of SCMV genomes available on NCBI (93), and in-depth recombination analyses generally use RDP, which requires manual checking of each recombination event (Martin et al., 2015).

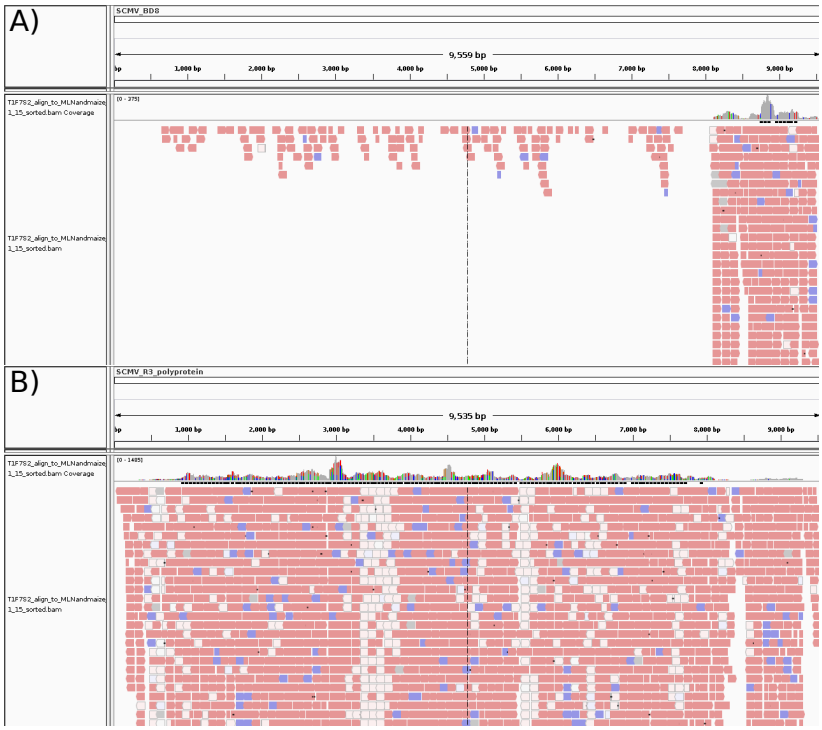


Fig. 4.2 Evidence for recombination in sugarcane mosaic virus (SCMV). Screengrabs from igv showing T1F7S2 alignment to SCMV genomes A) BD8 and B) R3. Note coverage panel (immediately above the red reads) to see alignment switch at 8000bp.

To simplify the analysis whilst preserving the most diversity, which is the information used for both recombination and phylogenetic analyses, I generated a nucleotide identity matrix for all 116 sequences (Electronic supplemental data). With this information I identified pairs of sequences with >99% nucleotide similarity and deleted the shorter sequence in each case, generating a final dataset of 55 SCMV genomes/contigs, including 13 from my NGS libraries. The splits network of this alignment is highly reticulate, and very distinct from the bifurcating tree produced by the MCMV analysis (fig. 3.3), indicating evidence for extensive recombination (fig. 4.3). In contrast to MCMV, sequences do not cluster by geographic region, suggesting historical global movement of SCMV.

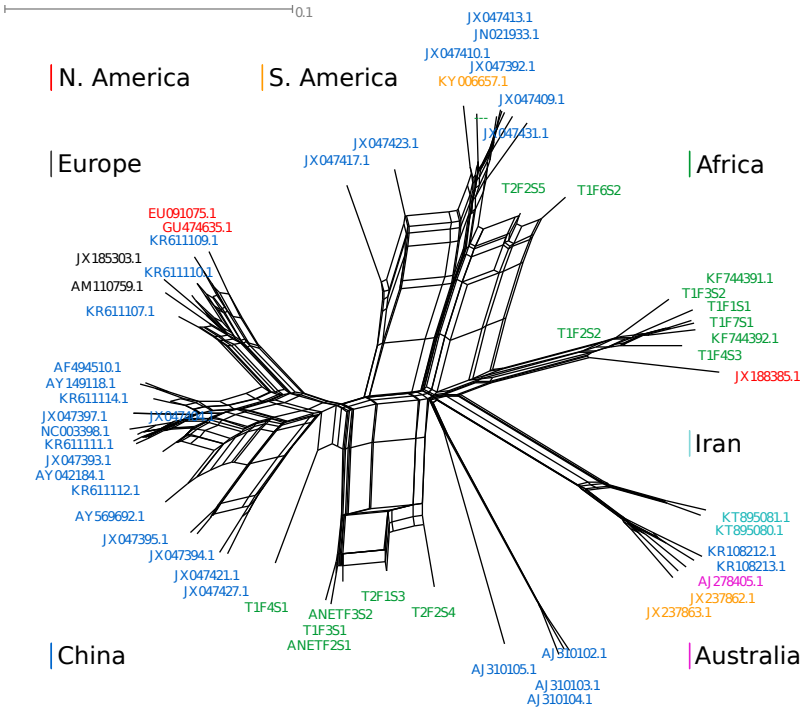


Fig. 4.3 Splits network of sugarcane mosaic virus genomes, distances calculated with uncorrected P, and network generated by neighbour-net in Splitstree V4.6. The reticulate network indicates conflicting phylogenetic signals within the alignment, suggesting recombination.

4.5.3 Evidence for specific recombination events in sugarcane mosaic virus

To estimate the number and location of recombination breakpoints I used RDP4 to generate a hypothesis for the historical recombination events in SCMV. RDP4 combines multiple recombination detection algorithms, estimating confidence depending on the number of algorithms supporting each event. I used five algorithms (see methodology) to predict recombination events, and checked each event manually. The resulting recombination hypothesis is extremely complex (fig. 4.4), as would be expected given fig. 4.3. The recombination hypothesis should not be seen as an accurate prediction of historical recombination events - RDP4 has (unavoidable) problems such as the weighting of different algorithms to predict which sequence in a triplet is the recombinant, for example, and is susceptible to making errors when recombination events are stacked, which seems very likely given both fig. 4.3 and fig. 4.4. However, it is suitable for illustrating the pervasive and complex history of recombination in SCMV.

Fig. 4.4 Recombination is ubiquitous across sugarcane mosaic virus (SCMV) phylogeny. Recombination hypothesis for SCMV genomes generated using RDP4 V4.94 shows evidence for recombination in most SCMV genomes, with evidence for multiple events seen in many genomes (Martin et al., 2015).

4.5.4 Recombination hot- and cold-spots in sugarcane mosaic virus

To search for regions with an over- or under- representation of recombination, I used RDP4 to generate sliding windows of breakpoint number across the SCMV genome (fig. 4.5a). I generated p-values with a permutation test using identical recombinant fragment lengths, randomly distributed, and 1000 permutations. Globally significant (i.e. window contains higher breakpoint density than 95% of all windows in the permutation test) recombination hot-spots are present at the 5' and 3' genomic termini. Significant local recombination hot-spots (higher density of breakpoints than 95% of permutations in that specific window) are seen around 800, 2800, 5100, 8010, 8120, 8460, 8550, 8800, 9300bp into the ANETF3S2 genome (fig. 4.5b). Significant local cold-spots are seen around 600, 1100, 1600, 2400, 3050, 3400, 3630, 3850, 6080, 6800, 7160, 7500bp (fig. 4.5b). This is the first statistically supported report of recombination hot-spots in SCMV, as far as I am aware, with previous reports finding breakpoints in CI, NIb, NIa-VPg, and NIa-prot (Achon et al., 2007; Gell et al., 2015; Li et al., 2013; Moradi et al., 2016; Padhi and Ramu, 2011; Xie et al., 2016; Zhong et al., 2005), and the first report of SCMV recombination cold-spots.

4.5.5 Sugarcane mosaic virus phylogeny

The issue of SCMV phylogeny, and with it the identification of SCMV groups has attracted much attention and many publications, with helpful summaries of changing clade names elsewhere (Francisca et al., 2012; Gao et al., 2011). The two main methods of generating SCMV phylogenies has been through CP and whole genome alignments. CP phylogenies generate clades that separate isolates roughly according to host - e.g. clades of maize-specific and *Saccharum officinarum*-specific, but there are clades with mixed hosts, and separation of some clades is low-confidence (Francisca et al., 2012; Gao et al., 2011; Wang et al., 2010a). CP phylogenies are often used in virology, and the CP region of SCMV has been reported to be non-recombinogenic, to justify its use. However, Li et al. (2013) finds recombination in the CP, and fig. 4.5 suggests there is a local hot-spot in CP, as assessed by the permutation test.

Whole genome phylogenies of SCMV typically group isolates into four strains (I, II, III, IV), which share around 80% nucleotide similarity between strains (Gao et al., 2011; Xie et al., 2016). Figure 4.6 shows these clades superimposed upon the splits network I generated earlier, with the clades AI and AII added for two groups of African sequences which don't cluster with previously reported clades. The novel sequences generated in this study undermine the separation between previously reported strains. To visualise this, I sampled four sequences per clade, generated a nucleotide identity matrix, and used this

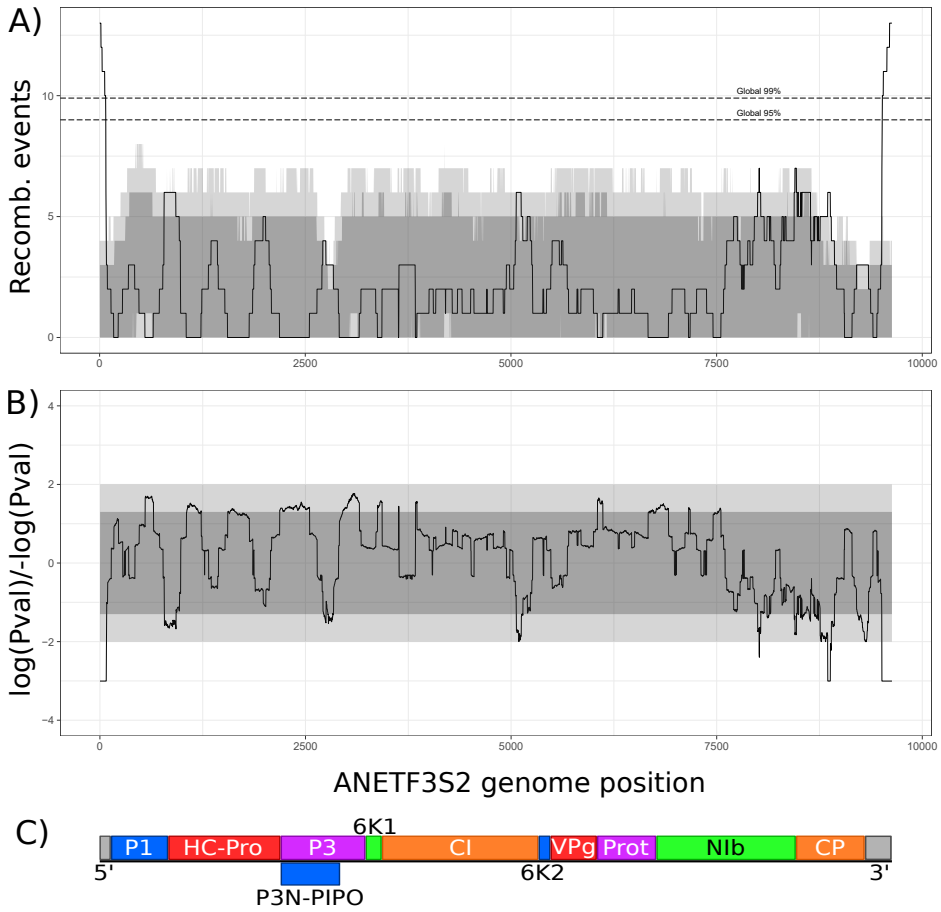


Fig. 4.5 A) Distribution of recombination events estimated using RDP4 in a 200bp sliding window, shown on the ANETF3S2 sugarcane mosaic virus (SCMV) genome. B) P-value distribution across the ANETF3S2 SCMV genome. Dark ribbon in graphs represents the local 95% limits of the permutation test, and the lighter ribbon the 99% limits. C) SCMV genome diagram of ANETF3S2, with final proteins shown as coloured rectangles. All proteins are expressed as one polyprotein which is cleaved into active proteins by viral proteases.

to generate a heat-map with rows and columns ordered by a dendrogram calculated using nucleotide distance (fig. 4.7). This analysis shows that the isolates do not cluster according to previously reported strains. Phylogenetic analyses assume a single evolutionary history, and although trees can be drawn using an alignment of SCMV isolates, it is not appropriate given the extent of recombination. For example, see Handley et al. (1998) for SCMV phylogenies calculated using different genomic regions, which are highly conflicting.

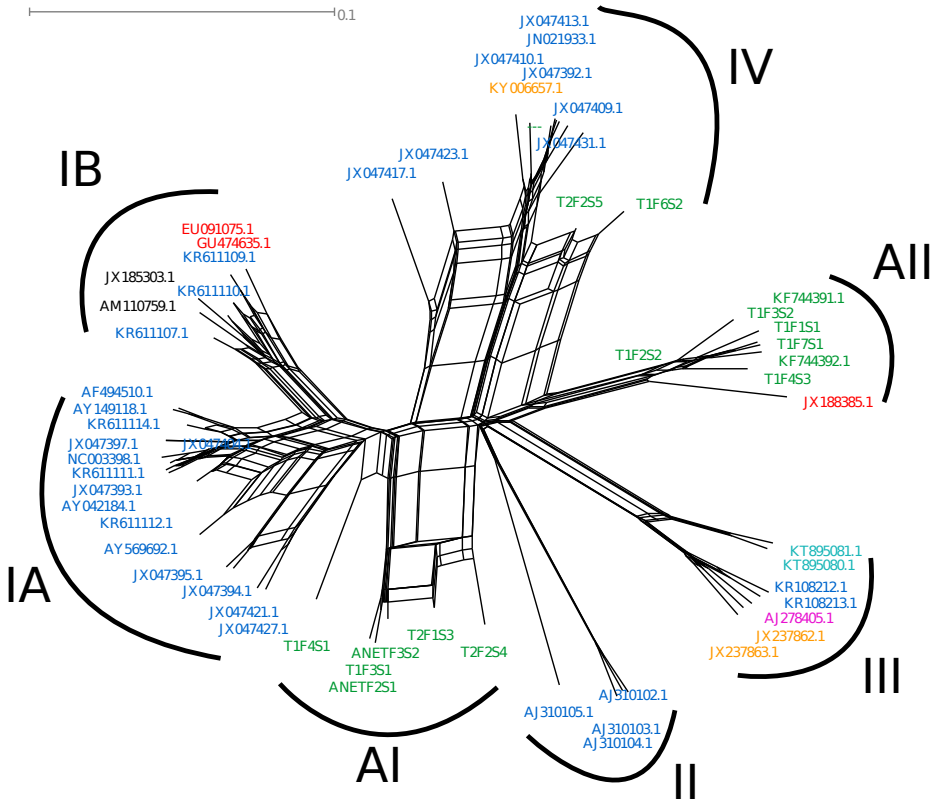


Fig. 4.6 Splits network from fig. 4.3 with strains previously reported in literature indicated. Clades AI and AII represent groups of African sugarcane mosaic virus isolates from this study.

4.6 Sugarcane mosaic virus variation

4.6.1 Sugarcane mosaic virus structural variation

I inspected the pre-trim SCMV alignment of 116 genomes (electronic supplement) for structural variation in the SCMV genome. This revealed a 10bp insertion in the Chinese HY8 isolate (JX047421.1) within the 5' UTR, a complex pattern of indels within the CP ORF, and indels within the 3' UTR (figs. 4.8 and 4.9). The CP indels feature two large insertions: a 51bp insertion into the Ecuadorian isolate MO1 (KY006657.1), and a 39bp insertion into nine SCMV isolates from my NGS survey, as well as isolates from Ohio (JX188385.1), Mexico (GU474635.1), Ethiopia (KP772216.1), and three from Rwanda (KF744390-2) (fig. 4.8b). The following 170bp of alignment contains a complex pattern of smaller indels across many isolates (fig. 4.9a). The 3' UTR contains a 6bp insertion in one German (JX185303.1) and two Mexican isolates (EU091075.1 and GU474635.1). The 13bp 3'UTR insertion is present in one Spanish isolate (AM110759.1) and 14 Chinese isolates.

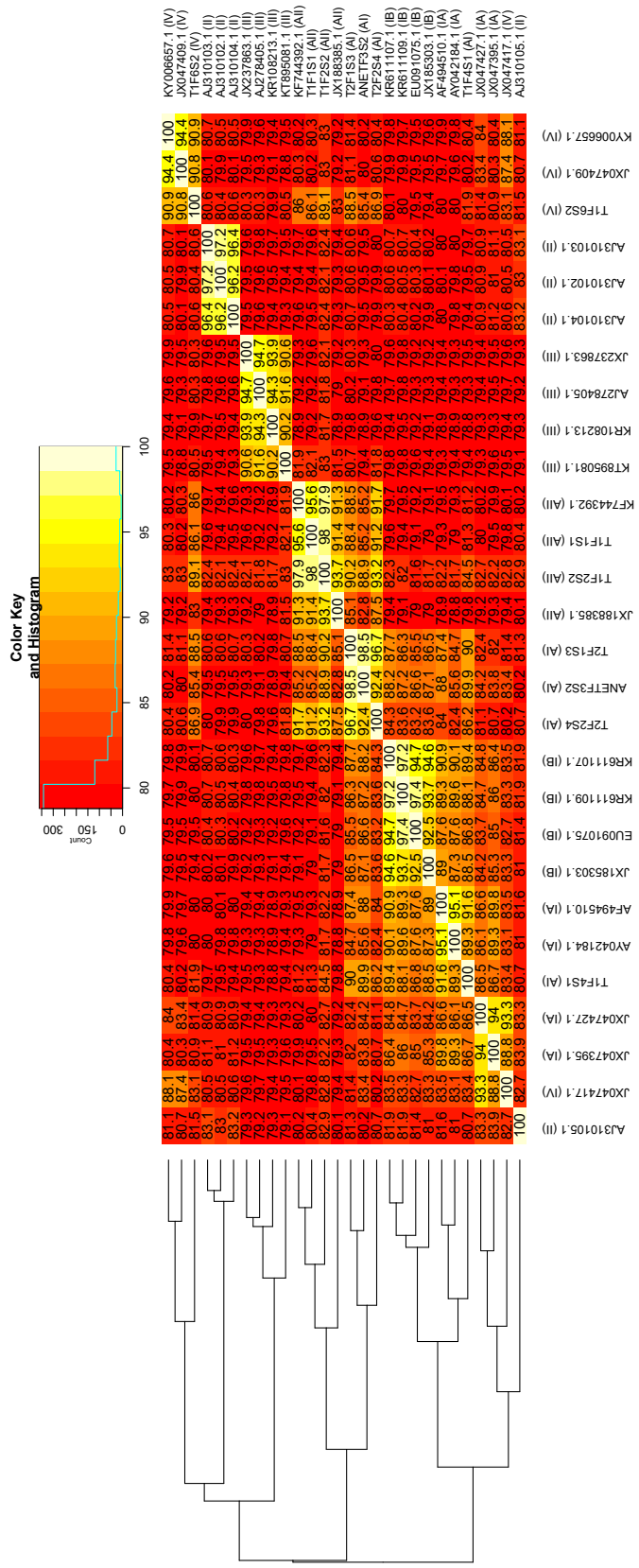


Fig. 4.7 Heatmap showing an identity dendrogram of sugarcane mosaic virus genomes calculated using four isolates per clade shown in fig. 4.6. Clades are shown in brackets after sequence IDs. Reading the rownames shows that sequences are not clustered according to clades. Additionally, regions of high similarity away from the diagonal show that these sequences do not fit neatly into clusters.

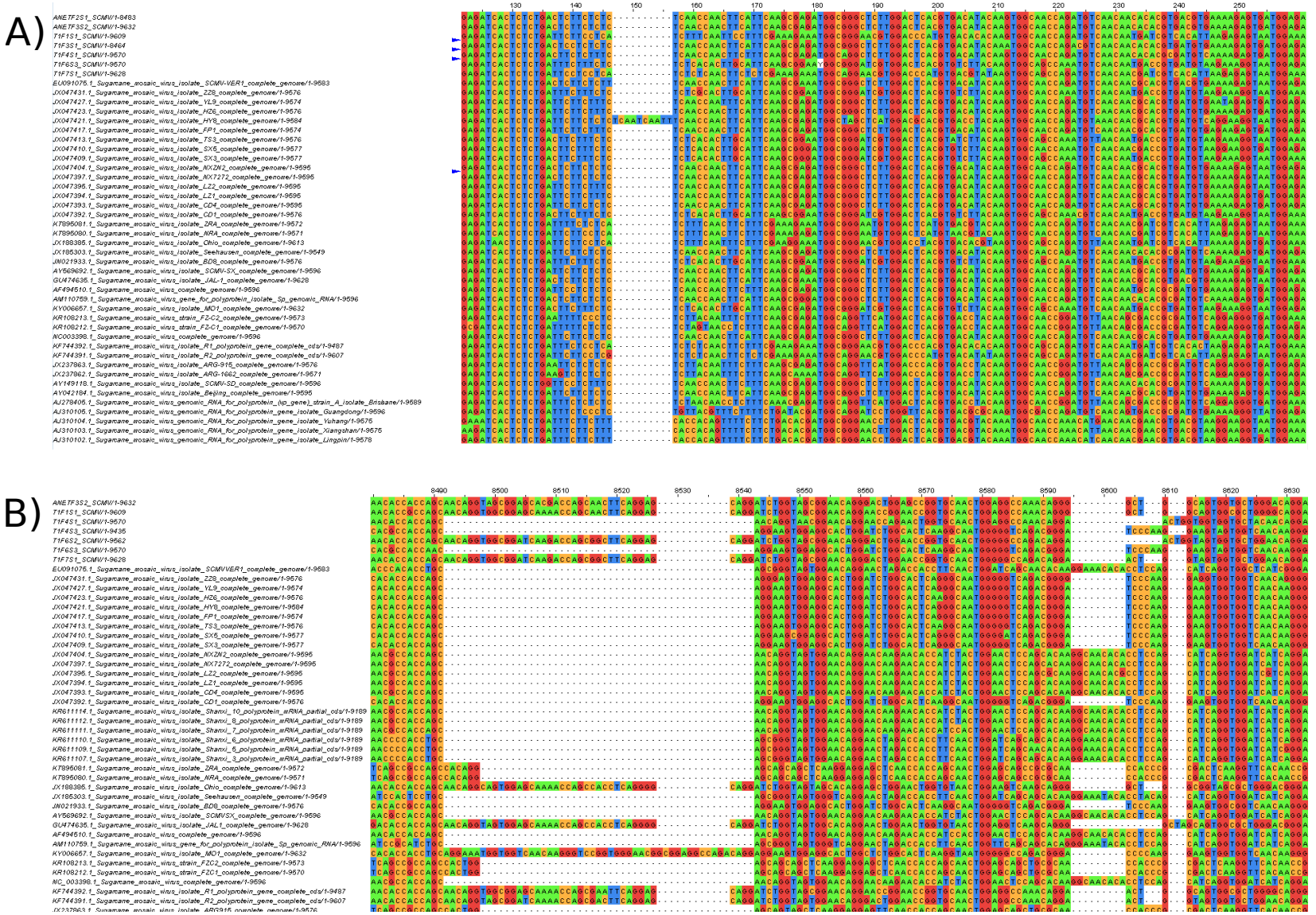


Fig. 4.8 Sugarcane mosaic virus (SCMV) alignment showing A) insertion at 5' UTR and B) large indels in the CP sequence of SCMV.

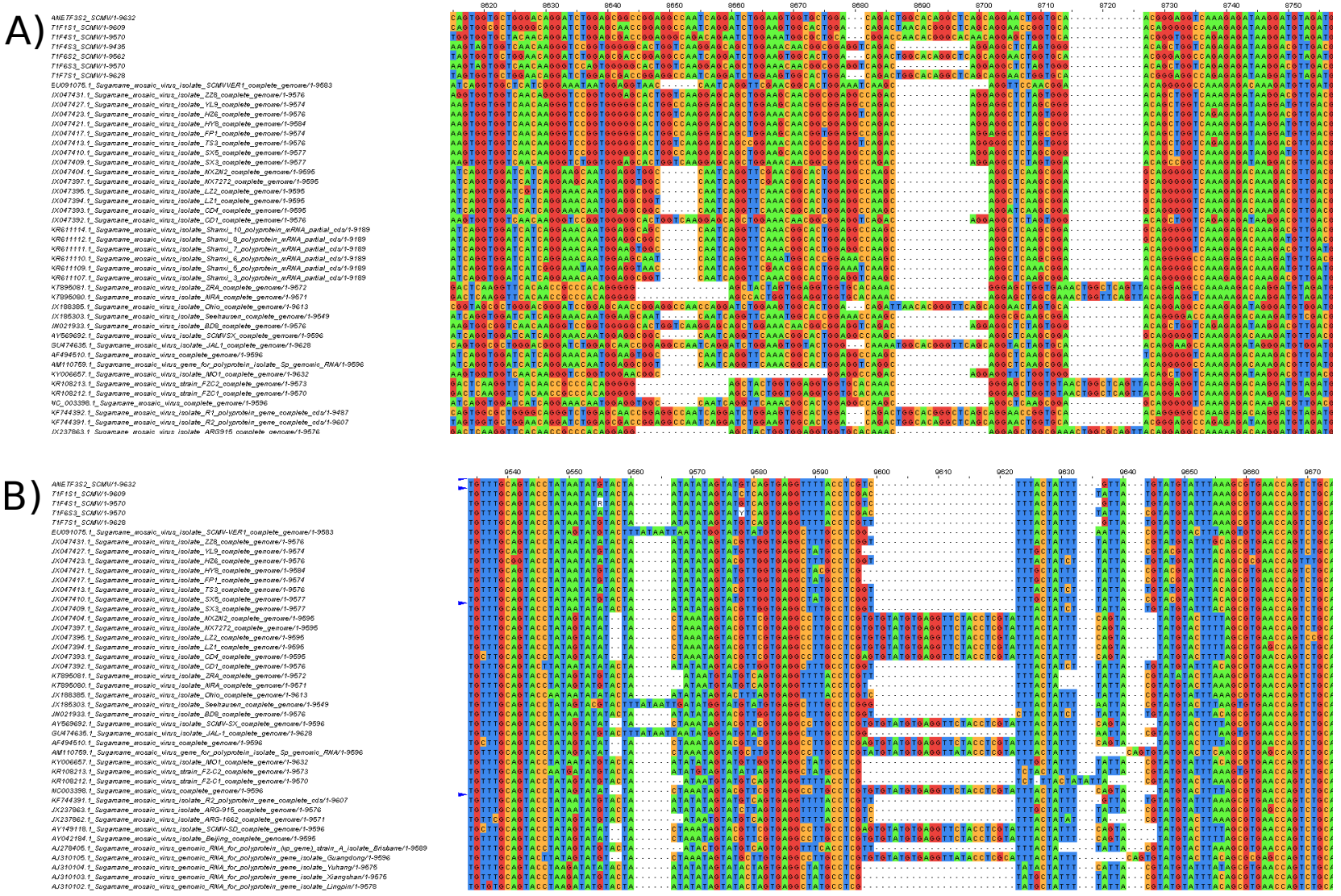


Fig. 4.9 Sugarcane mosaic virus alignment showing A) complex pattern of indels within the CP ORF and B) indels in the 3' UTR.

4.6.2 Sugarcane mosaic virus genomic variation

To investigate the level of variation across the SCMV genome I used the trimmed alignment used for recombination analysis, with SCMV genomes that did not have ambiguous bases called (for compatibility with DNAsp5). Diversity was high, with 4289 mutations spread over 2831 sites, and an average of 1121.1 nucleotide differences between sequences. Nucleotide diversity across the genome is 0.17, which is higher than most RNA viruses, but within the range previously reported for SCMV (Li et al., 2013; Xie et al., 2016). Plotting nucleotide diversity across the genome shows notable highs in the N-termini of P1 and CP, and notable lows in central P3 and the 3' UTR (fig. 4.10). Due to the impossibility of generating an accurate SCMV phylogeny, and the difficulty of appropriately clustering sequences by sequence identity, I concluded my analysis of SCMV in East Africa.

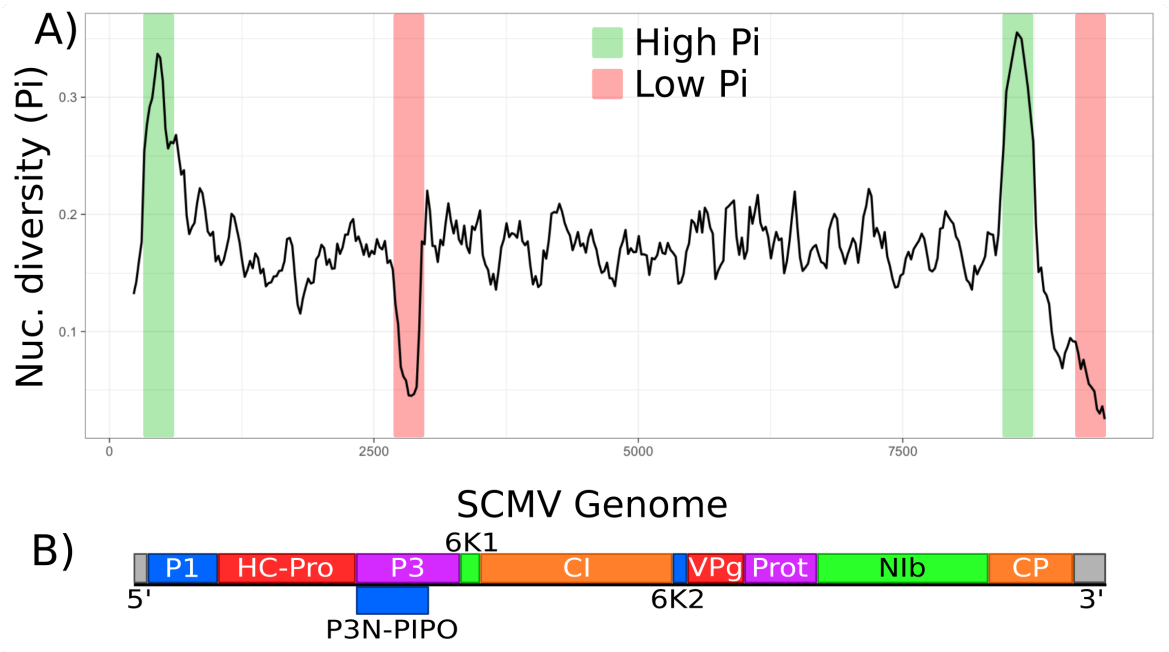


Fig. 4.10 A) Nucleotide diversity plotted across the SCMV genome, with a window size of 100bp and step size of 25bp. B) SCMV genome with mature peptide (rather than polyprotein) positions shown, aligned with graph above.

4.7 Investigation of RNA viruses in East African maize

4.7.1 Survey of additional RNA viruses in East African maize

Given the power of NGS to capture genetic data, independently of sequence, and the moderate proportion of unaligned reads in a number of libraries (fig. 4.1b), I set out to determine whether additional RNA viruses are present in East African maize. Additional *Potyviridae* viruses would represent possible partners for MCMV, whilst the limited amount of research performed in East African agriculture means that novel viruses are likely to be present.

4.7.2 Optimisation of Trinity assembly

I selected Trinity for *de novo* assembly, a tool for transcriptome assembly, due to its ease of use and high performance (Grabherr et al., 2011; Honaas et al., 2016). One problem with assessing the efficacy of *de novo* assembly is that it can't be assessed unless the underlying information is already known, for example using high-quality transcriptomes to benchmark performance. Even in these cases, optimising assembly is difficult, for example coverage can be too low *or* too high for optimal performance on different mRNAs within the same sample (Honaas et al., 2016). For this reason, I initially compared the trinity assemblies produced on NGS data preprocessed in two different ways: one with no deduplication and less stringent quality trimming with trimmomatic (remove starting and ending bases with PHRED score <20, cut read once average PHRED over 4bp window <20, minimum length 40bp) to maximise coverage, and one with deduplication performed by string matching with QUASR and more stringent quality trimming using trim galore! and QUASR (trim reads from 3' until median quality of read is >PHRED 30, minimum length 50bp) to maximise read quality and minimise redundant reads.

For both pipelines, I then aligned reads to the maize B73 genome, extracted unaligned reads, and assembled them using Trinity. I extracted contigs over 2kb then blasted them against the NCBI nt database, and selected those with top hits against viruses. Comparing the length of these contigs showed that deduplication and more stringent quality trimming generally resulted in a larger number of longer viral contigs being detected (table 4.1). To try and increase the length of homologous contigs that were present in multiple libraries further, I then pooled unaligned (to the maize genome) reads for every library containing a specific type of contig. These were assembled and blasted as before, which in most cases did not improve contig length over the best individual library assembly, although maximum contig length for maize yellow dwarf virus (MYMV) increased slightly (table 4.1).

Table 4.1 Trinity assembly comparison. Processing method in bold produces longest contig. NP=not present.

Blast hit	Processing	Libraries (number)	Longest contig
Maize yellow mosaic virus	No dedup.	ANETF3S2, ANETF4S2, T1F1S1, T1F2S2, T1F5S3, T1F7S1, T1F7S2, T1F7S3, T2F2S4, T2F2S, T2F3S, T2F5S1 (12)	5682
	Dedup.	ANETF1S2, ANETF2S1, ANETF3S2, ANETF4S2, FERARWANDA, LETF2S1, T1F1S1, T1F2S2, T1F3S2, T1F4S3, T1F5S2, T1F5S3, T1F6S2, T1F7S1, T1F7S2, T1F7S3, T1F8S2, T1F8S3, T2F2S4, T2F2S5, T2F3S4, T2F5S1, T2F5S2 (23)	5682
	Pooled	As for dedup., combined	5749
Soybean-associated bicistronic virus	No dedup.	ANETF4S2, T2F2S4, T2F2S5, T2F3S4 (4)	5852
	Dedup.	ANETF4S2, T1F7S1, T2F2S4, T2F2S5, T2F3S4 (5)	6776
	Pooled	As for dedup., combined	3723
Maize streak virus - A[Km]	**No dedup.**	T1F1S1 (1)	4183
	Dedup.	T1F1S1 (1)	3133

Table 4.1 Trinity assembly comparison. Processing method in bold produces longest contig. NP=not present.

Blast hit	Processing	Libraries (number)	Longest contig
Aphid lethal paralysis virus	**No dedup.**	T1F4S3, T1F5S2, T2F3S4 (3)	9828
	Dedup.	ANETF2S1, T1F4S3, T1F5S2, T2F3S4 (4)	9828
	Pooled	As for dedup., combined	9828
Maize white line mosaic virus	**No dedup.**	T1F7S1, T1F7S3 (2)	4241
	Dedup.	T1F1S1, T1F7S1, T1F7S3 (3)	4220
	Pooled	As for dedup., combined	4220
Acyrthosiphon pisum virus	No dedup.	T2F3S4 (1)	9832
	Dedup.	T2F3S4 (1)	9869
Rhopalosiphum padi virus	No dedup.	T2F3S4 (1)	6816
	Dedup.	T2F2S4, T2F2S5, T2F3S4 (3)	8806
	Pooled	As for dedup., combined	8798
Black raspberry virus F	No dedup.	NP	NA

Table 4.1 Trinity assembly comparison. Processing method in bold produces longest contig. NP=not present.

Blast hit	Processing	Libraries (number)	Longest contig
	Dedup.	T2F1S2 (1)	4426
Citrus yellow vein-associated virus	No dedup.	NP	NA
	Dedup.	LETF2S1 (1)	2985
Nilaparvata lugens honeydew virus-3	No dedup.	NP	NA
	Dedup.	T2F3S4 (1)	9075
Ribes virus F	No dedup.	NP	NA
	Dedup.	ANETF1S2, ANETF3S2, ANETF4S2, T1F5S2, T1F6S1, T1F6S3, T2F2S4 (7)	4990
	Pooled	As for dedup., combined	4990

4.7.3 RT-PCR confirmation of virus-like sequences

To confirm that the contigs attained through Trinity assembly were real, rather than unconnected reads assembled through regions of high sequence homology, I designed two primer pairs for each VLS, each spanning approximately 1kb. I converted the RNA samples used for library preparation to cDNA using reverse transcription, then performed PCR on samples whose libraries were predicted to contain each VLS. Most of the VLSs assembled by Trinity produced amplicons of the predicted size in at least one sample (table 4.2). To confirm the sequence, I Sanger sequenced one of each different amplicon and aligned these back to the appropriate VLS. These confirmed the identity of the amplicons, and I proceeded to characterisation of the VLSs confirmed by RT-PCR.

Table 4.2 RT-PCR results for virus-like sequences predicted by Trinity *de novo* assembly.

Closest BLAST hit	Amplicon	Samples tested	Positive
Acyrthosiphon pisum virus (APV)	A	1	0
	B	1	0
Aphid lethal paralysis virus (ALPV)	A	3	2
	B	3	2
Black raspberry virus F (BRVF)	A	1	1
	B	1	1
Citrus yellow vein-associated virus (CYVA)	A	1	0
	B	1	0
Maize white line mosaic virus (MWLMV)	A	1	1
	B	1	0
Maize yellow mosaic virus-RMV (MYMV)	A	14	12
	B	14	11
Nilaparvata lugens honeydew virus-3 (NLHV)	A	1	0
	B	1	0
Rhopalosiphum padi virus (RPV)	A	2	1
	B	2	1
Ribes virus F (RVF)	A	4	4
	B	4	2
Soybean-associated bicistronic virus (SABV)	A	3	2
	B	3	1

4.7.4 Characterisation of novel virus-like sequences

To investigate the similarity of these sequences to actual viruses, I used two approaches - ORF prediction and conserved protein domain searches. The presence of long ORFs suggests that these sequences have been selected to suppress the presence of stop codons, implying that they code for protein sequences with a biological function that is maintained through evolutionary time. The arrangement of ORFs across a genome is also useful for classifying novel VLS. The majority of VLSs have very similar ORF organisation to their nearest blast hits, with differences observed in the 3' region of the MWLMV-VLS genome, and an apparent 5' extension in the MYMV-VLS genome (figs. 4.11 and 4.12). Likewise, the presence of conserved viral protein domains aids in classification and provides support for these sequences representing real viruses. The VLSs contain many protein domains in common with their nearest blast hits (figs. 4.13 and 4.14). Notable differences are the MWLMV-VLS lacking the *Tombusviridae* P33 domain and the tomato bushy stunt virus P22 domain possessed by MWLMV, and MYMV-rel seeming to have Polerovirus ORF 5 domain split between its 3' and 5' genomic termini, in contrast to MYMV which has the domain at the 3' of the genome. The implications of this are discussed below, with updated blast results (the blast analysis above was performed in March 2016).

ALPV-VLS

I blasted the ALPV-VLS contig and the top hit had 98% coverage and 97% identity to ALPV (KX883690.1, Dicistroviridae). This, combined with the same organisation of ORFs and protein domains, leads me to conclude that ALPV-VLS is ALPV. This, and a metagenomics study of Kenyan aphids performed by a colleague in the Department of Plant Sciences (Francis Wamonje, University of Cambridge) represent the first report of ALPV in East Africa.

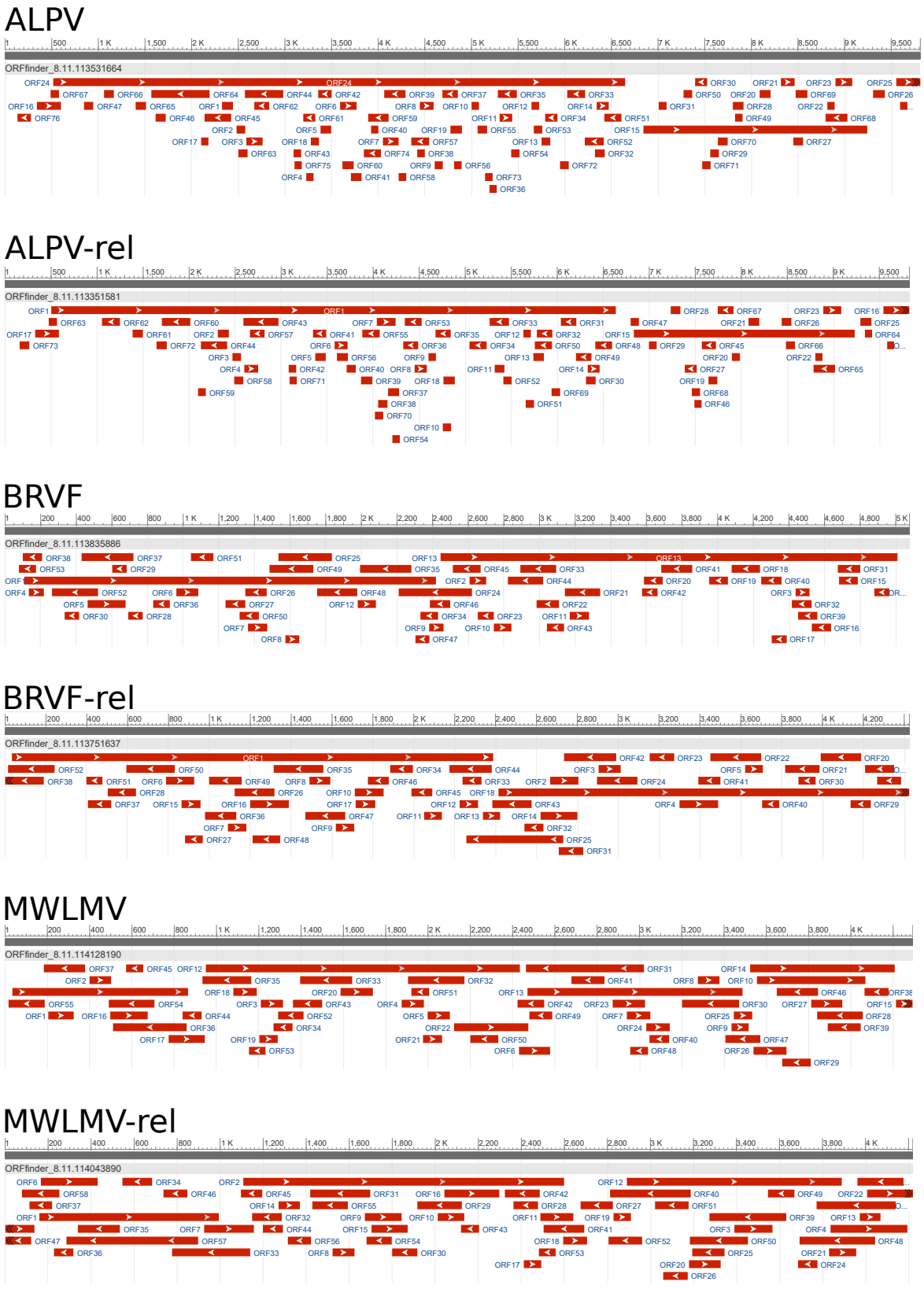


Fig. 4.11 Open reading frames identified by the NCBI ORF finder in virus-like sequences generated by trinity and their nearest blast hits. ALPV=Aphid lethal paralysis virus, BRVF=Black raspberry virus F, MWLMV=Maize white line mosaic virus. Rel denotes the novel virus-like sequences identified in this work.

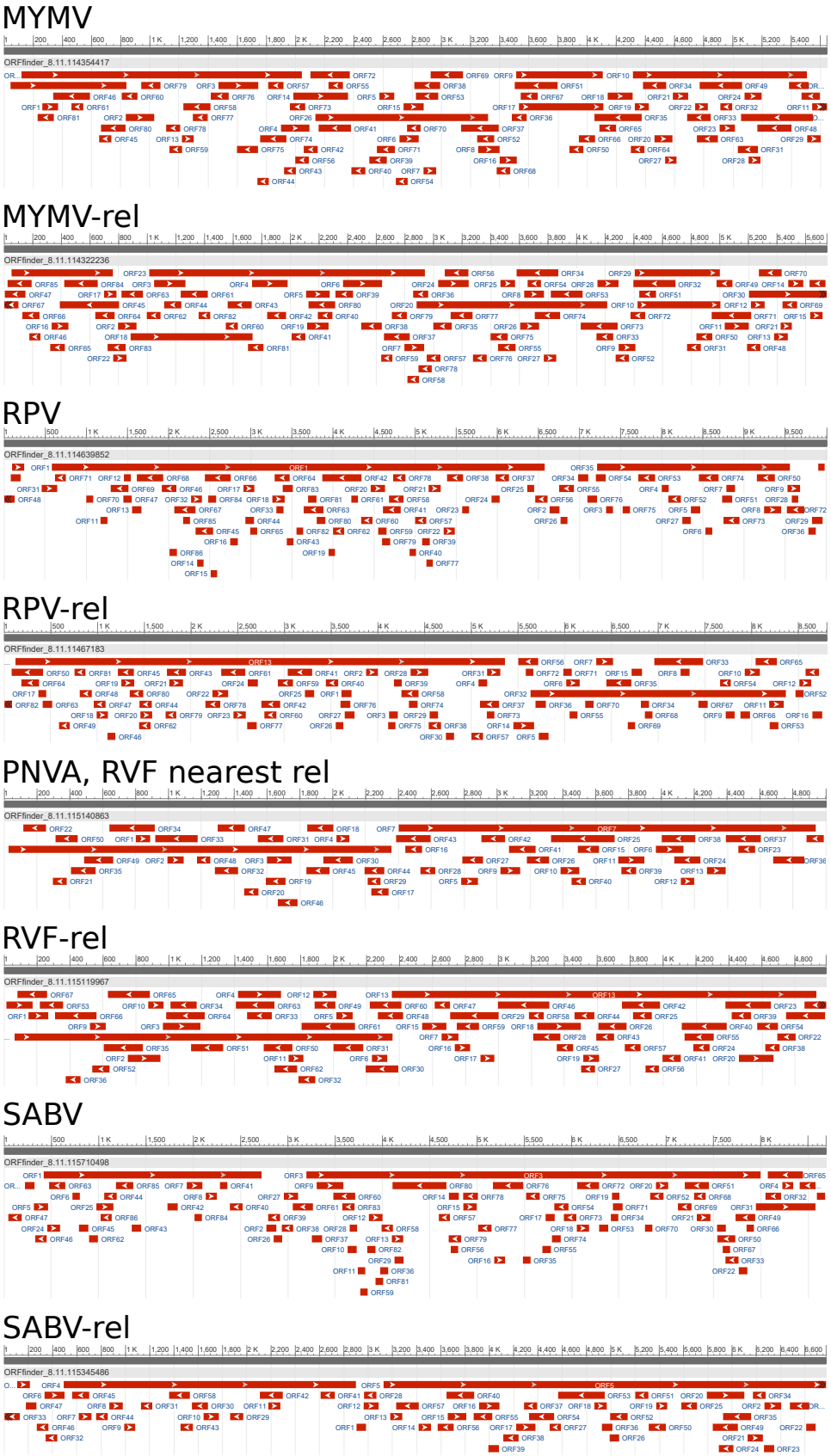


Fig. 4.12 Open reading frames identified by the NCBI ORF finder in virus-like sequences generated by trinity and their nearest blast hits. MYMV=Maize yellow mosaic virus, RPV=Rhopalosiphum padi virus, SABV=Soybean-associated bicistronic virus, PNVA=Panax notoginseng virus A, RVF=Ribes virus F. PNVA used instead of RVF for domain search as RVF is not a full length viral genome. Rel denotes the novel virus-like sequences identified in this work.

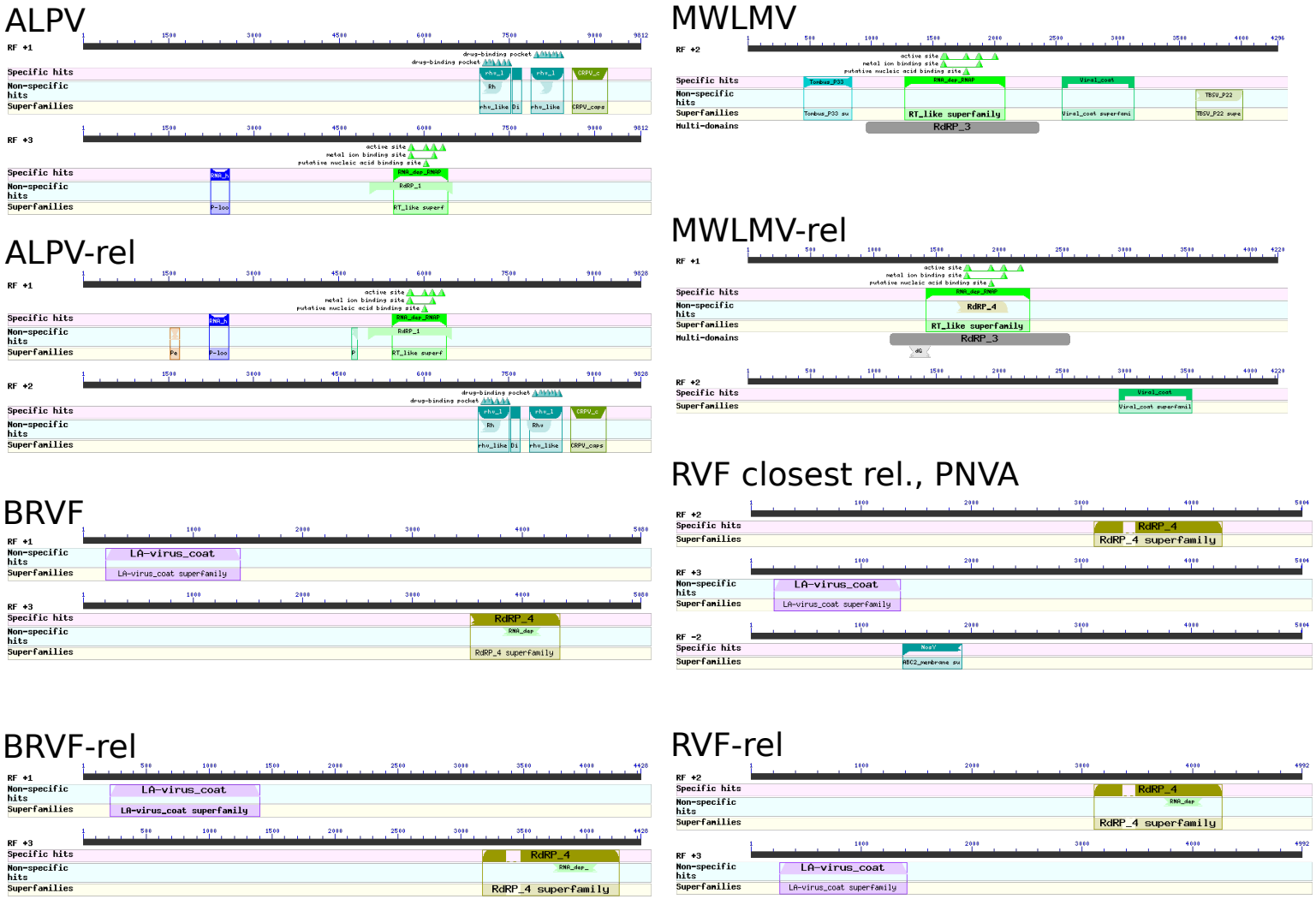


Fig. 4.13 Conserved protein domains found by the NCBI conserved domain finder in virus-like sequences generated by trinity and their nearest blast hits. ALPV=Aphid lethal paralysis virus, BRVF=Black raspberry virus F, MWLMV=Maize white line mosaic virus, PNVA=Panax notoginseng virus A, RVF=Ribes virus F. PNVA used instead of RVF for domain search as RVF is not a full length viral genome.

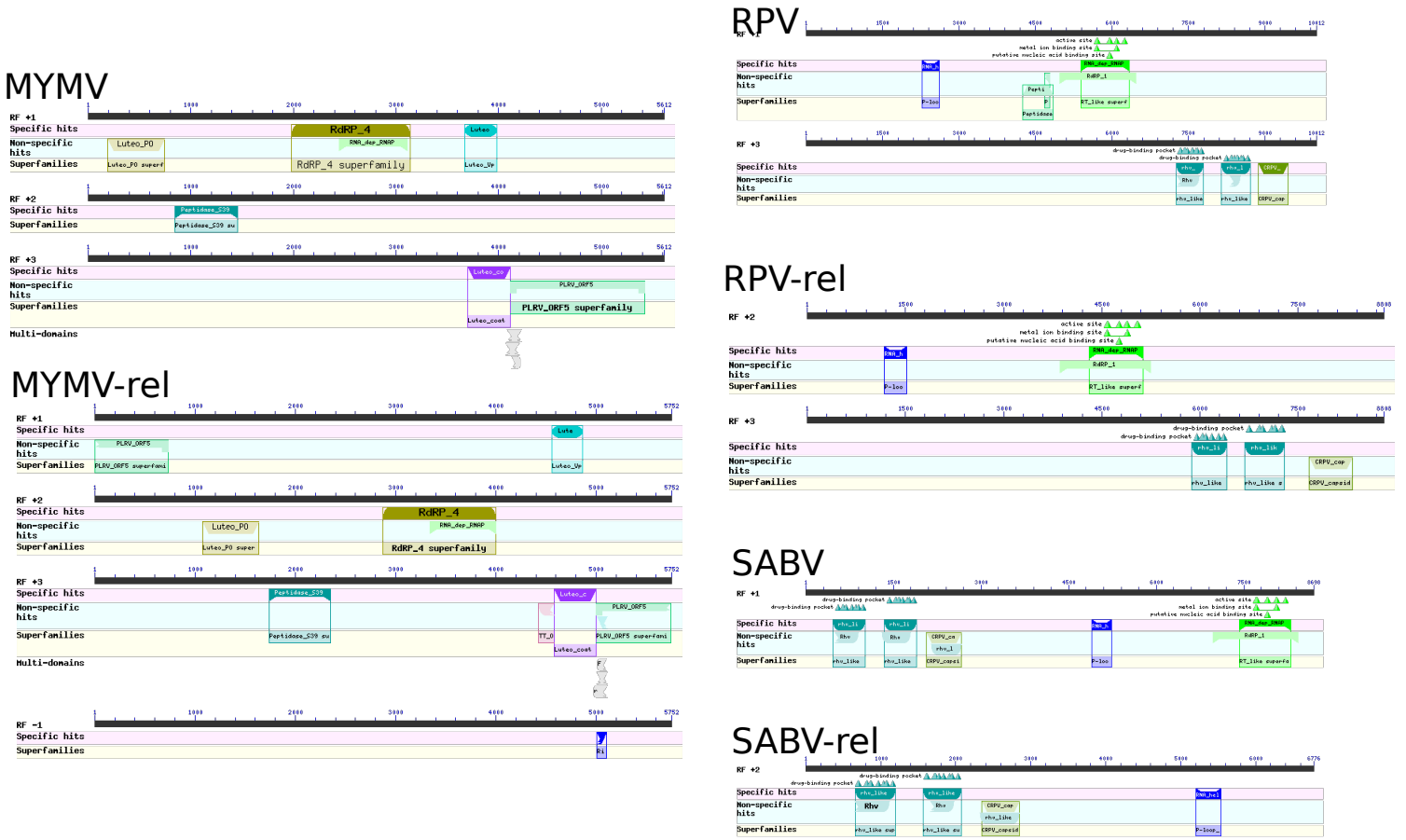


Fig. 4.14 Conserved protein domains found by the NCBI conserved domain finder in virus-like sequences generated by trinity and their nearest blast hits. MYMV=Maize yellow mosaic virus, RPV=Rhopalosiphum padi virus, SABV=Soybean-associated bicistronic virus.

BRVF-VLS and RVF-VLS

Blasting BRVF-VLS returns a top hit with 99% coverage and 99% identity to maize-associated totivirus-Ec (MATV-Ec) (KT722800, *Totiviridae*), a recently reported variant of the Chinese MATV with an insertion in the capsid ORF (Alvarez-Quinto et al., 2017). Given the similarity in sequence identity, I conclude that BRVF-VLS is an isolate of MATV-Ec.

Four libraries contained RVF-VLS, which has a top blast hit to MATV-Ec as well (KT722800), with 99% coverage but only 81% nucleotide identity. Similarities in ORFs and protein domains between BRVF-VLS, RVF-VLS, and their nearest blast hits in the above analysis can be seen in figs. 4.11 to 4.13. Conventions vary between viral families, and RVF-VLS may represent either a novel species of totivirus or a divergent strain of MATV-Ec (the difference being somewhat academic in viruses). These sequences represent the first report of MATV-Ec in Africa, and a novel strain/species of MATV-Ec/Totivirus respectively.

MWLMV-VLS

The top (discontinuous) blast hit for MWLMV-VLS remains MWLMV (EF589670.1, *Tombusviridae*), with 20% coverage and 69% nucleotide identity. Overall, ORFs in MWLMV-VLS are consistent with the *Tombusviridae*, and removing the first ORF's stop codon demonstrates the present of the characteristic (for the *Tombusviridae* family) read-through expression of the RNA-dependent RNA polymerase (RdRP) (fig. 4.15). The presence of a *Tombusviridae* genome organisation, protein domains, and high divergence from previously reported *Tombusviridae* members, leads me to conclude that MWLMV-VLS is a novel species in the *Tombusviridae* family, likely in a novel genus, that I name maize-associated tombusviridae.

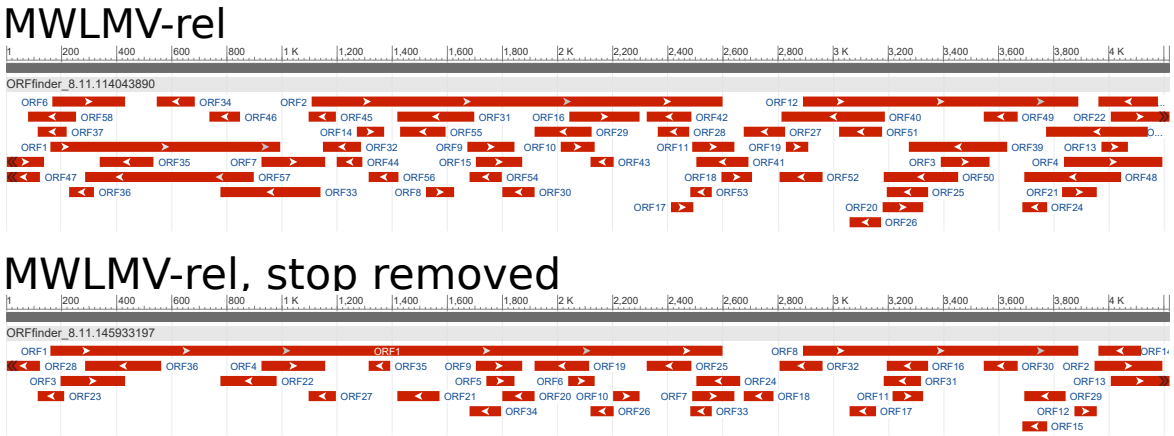


Fig. 4.15 Removing the stop codon at the end of ORF1 in MWLMV-VLS increases the length of ORF1 to that expected in the *Tombusviridae*, strongly suggesting that MWLMV-VLS is a member of the *Tombusviridae*.

RPV-VLS

The top (discontinuous) blast hit for RPV-VLS remains RPV (AF022937.1, Dicistroviridae), with 2% coverage and 72% identity. RPV has two major ORFs, 5997 and 2184bp long, separated by 801bp, while the two major RPV-VLS ORFs are 5421 and 2733bp long, separated by 275bp. As well as Dicistroviridae-like genome organisation, RPV-VLS contains all of the conserved protein domains present in RPV. Given the similarities in genome organisation and protein domains, despite the large sequence divergence from all known viruses, I conclude that RPV-VLS represents a novel member of the Dicistroviridae, which I name kenyan maize-associated Dicistoviridae.

Dicistroviridae sequences are not generated by insect contamination of samples

Two of the VLS identified (ALPV and kenyan maize-associated dicistoviridae)are members of the *Dicistroviridae*, which have insect hosts. To investigate whether these viruses are present due to insect contamination, I used MEGAN to visualise blast results for each library (Huson et al., 2011). The hypothesis would be that libraries contaminated with an insect would have a large proportion of reads blasting to insect sequences. This analysis showed that there were a very small proportion of reads blasting to insect sequences in almost all libraries containing *Dicistroviridae* sequences, with the exception of T2F2S5, which appears to contain a mite (*Tetranychus*), and serves to illustrate the efficacy of this analysis (fig. 4.16). I conclude that both *Dicistroviridae* VLS are present in samples without significant blast hits for insects, suggesting either that maize is acting as a host, or viral particles are present on the maize surface. If the viruses are present within the host, this would likely be retained in host tissues directly exposed to insect feeding, for example the phloem for aphid-transmitted viruses. The phylogenetic distance between insects and plants means that insect viruses are unlikely to be competent for replication in maize, due to the high number of host factors which are co-opted in RNA virus replication - these factors are likely to have diverged between plants and insects.

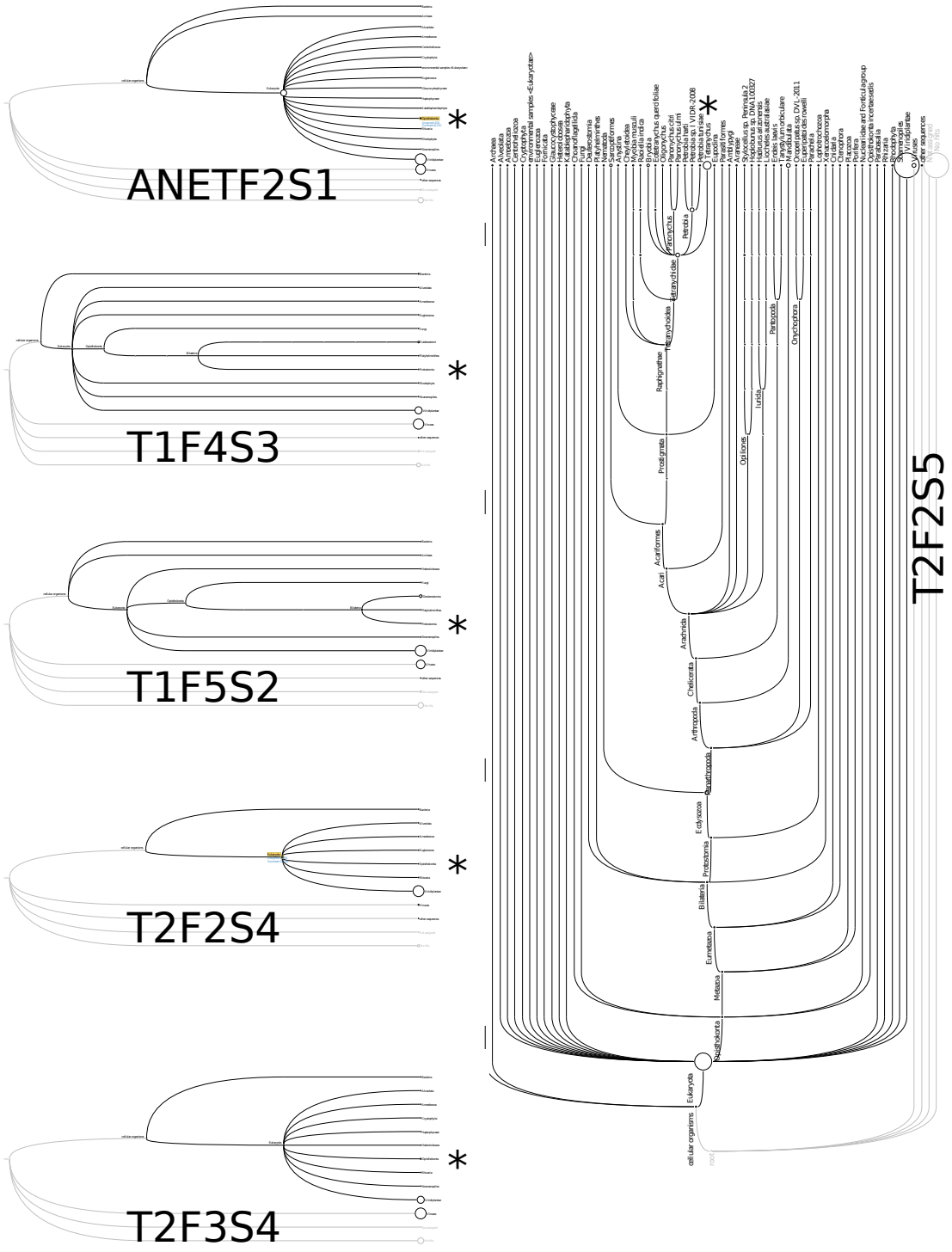


Fig. 4.16 Libraries containing *Dicistroviridae* sequences do not contain a high proportion of insect reads. This suggests that the presence of the *Dicistroviridae* is due to viral presence inside maize cells or on the plant surface. The taxa containing insects is indicated by an asterisk for each sample. T2F2S5 appears to be contaminated by mites (*Tetranychus*), indicated by asterisk, verifying that this analysis is able to detect contaminants.

SABV-VLS

The top blast hit for SABV-VLS remains SABV (KM015260.1, *Picornavirales*), with 38% coverage and 76% identity. SABV-rel has a similar genome organisation and the same protein domains as SABV (figs. 4.12 and 4.14), with the exception of RNA-polymerase, which is present after the SABV-VLS contig ends. Given these data, I would suggest that SABV-VLS represents a novel member of the *Picornavirales*, which I name maize-associated Picornavirales.

MYMV-VLS

The top blast hit for MYMV-VLS remains MYMV (KU291103.1, *Luteoviridae*), with 100% coverage and 98% identity. However, as noted above, the MYMV-VLS contig appears to be shifted relative to previously reported MYMV sequences. This could either represent a mutation in which the 3' terminus of the genome has been transposed to the 5' end, or error in *de novo* assembly. The more likely case is an error in assembly, because the shift splits both an ORF and a protein domain, the *Polerovirus* ORF5 domain (figs. 4.12 and 4.14). 5 and 3' RACE reactions would determine the true termini of the MYMV present in East Africa.

MYMV-VLS was the most widespread virus in our survey of East African maize, with the exception of MCMV and SCMV. It was widespread within the survey area (fig. 4.17). This represents the first report of MYMV in East Africa, which is an emerging virus worldwide, with first reports in China (2016), Brazil (2017), Nigeria (2017) (Chen et al., 2016; Gonçalves et al., 2017; Yahaya et al., 2017). The economic impact of MYMV is unknown, but it is reported to be associated with chlorosis, dwarfing, and commonly found in mixed infections with other viruses (Chen et al., 2016; Gonçalves et al., 2017)

4.7.5 VLS presence across libraries

To investigate the proportion of reads within the NGS libraries corresponding to the RNA viruses described above, I generated a bowtie2 reference using the longest contigs for each VLS and aligned NGS libraries to them. For most libraries, a small proportion of reads aligned to VLS sequences, typically less than to SCMV (fig. 4.18). In two samples with a higher proportion of VLS alignment (T2F2S5, T1F4S3) this is due to alignment to ALPV, while in FERARWANDA it is due to alignment with MYMV.

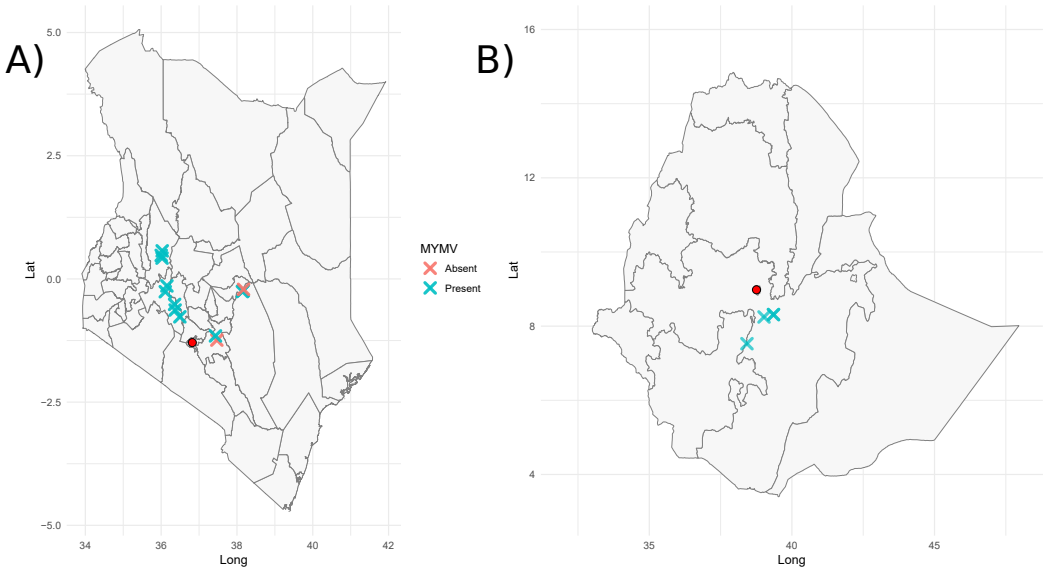


Fig. 4.17 Map showing the presence of maize yellow mosaic virus in Kenya (A) and Ethiopia (B).

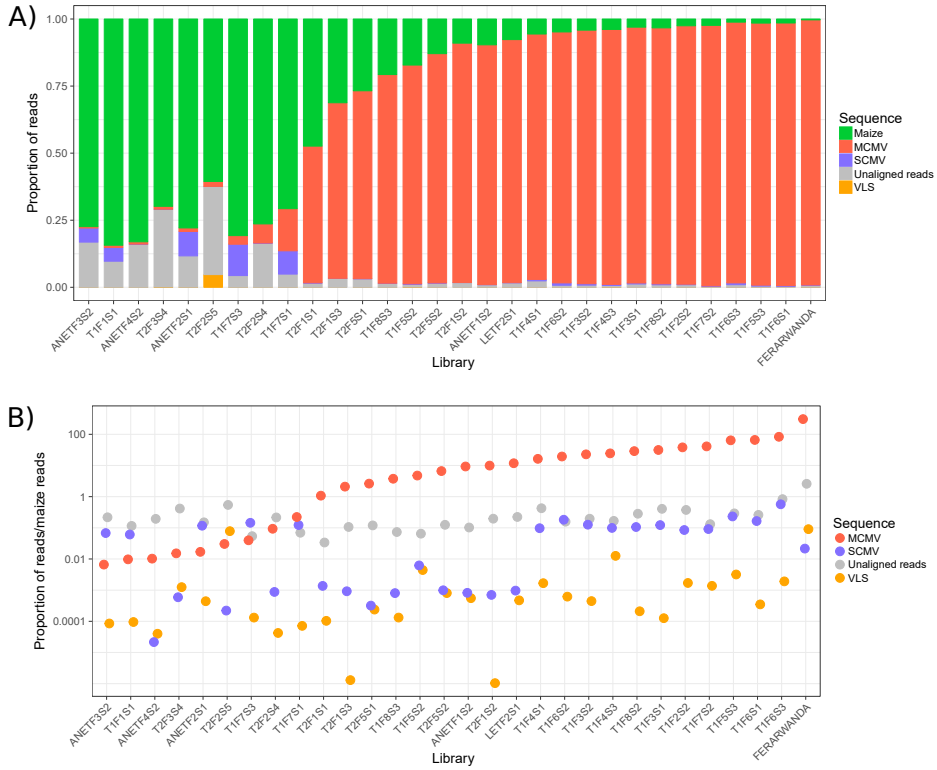


Fig. 4.18 Proportion of reads mapping to maize genome, maize lethal necrosis-related viruses, and virus-like sequences in NGS libraries (A), and shown on a log-scale relative to number of reads aligning to maize genome.

Discussion

4.8 Sugarcane mosaic virus in East Africa

4.8.1 Sugarcane mosaic virus is the major partner of maize chlorotic mottle virus in East Africa

The purpose of my NGS survey in East Africa was to characterise MCMV and its partner viruses in MLN-symptomatic maize plants. The only partner virus found in my samples and known to be involved in MLN was SCMV, which is the most common MCMV-partner globally, being the only reported partner in China, Ecuador, and East Africa (Quito-Avila et al., 2016; Wang et al., 2017).

My survey data are only correlational - they do not prove that the combination of SCMV and MCMV in East African maize reproduce the symptoms observed in field maize there. However, this is the most parsimonious explanation given the absence of other known partner viruses or *Potyviridae* members in my samples, and the higher titre of SCMV compared to other RNA viruses in most of my samples (fig. 4.18). Experimental inoculation experiments performed by colleagues at KALRO using MCMV and SCMV isolated in Kenya did reproduce field symptoms, confirming that MCMV and SCMV are causing MLN in East Africa (fig. 4.19).

4.8.2 Sugarcane mosaic virus diversity and recombination

East African SCMV isolates are very diverse in terms of nucleotide sequence, as previously reported for SCMV (Xie et al., 2016). For example, I found more SCMV nucleotide sequence divergence on a single farm (T1F3, 84% nt ID) than there is MCMV sequence divergence globally (96%). The most diverse genomic regions were the P1 protein and the N-terminus of the CP. P1 is a serine protease which autocatalytically cleaves itself from HC-Pro, and has a hyper-variable region. In *Plum pox virus*, this hyper-variable region was shown to be intrinsically disordered, and to negatively regulate protease activity, which in turn negatively regulates virulence, leading the authors to hypothesise a role in modulating symptom severity in order to maximise transmission (Pasin et al., 2014). N-terminal variability in the CP, as well as structural variation (figs. 4.8 and 4.9) matches the region of SCMV CP previously reported to have variation in amino acid length, and low conservation compared to the *Potyvirus* MDMV CP, compared to the rest of the CP gene (Frenkel et al., 1991; Xiao et al., 1993). The N-terminus is surface located in potyviruses, raising the possibility that variation in this region may alter interactions, either with host or vector proteins (López-Moya et al.,

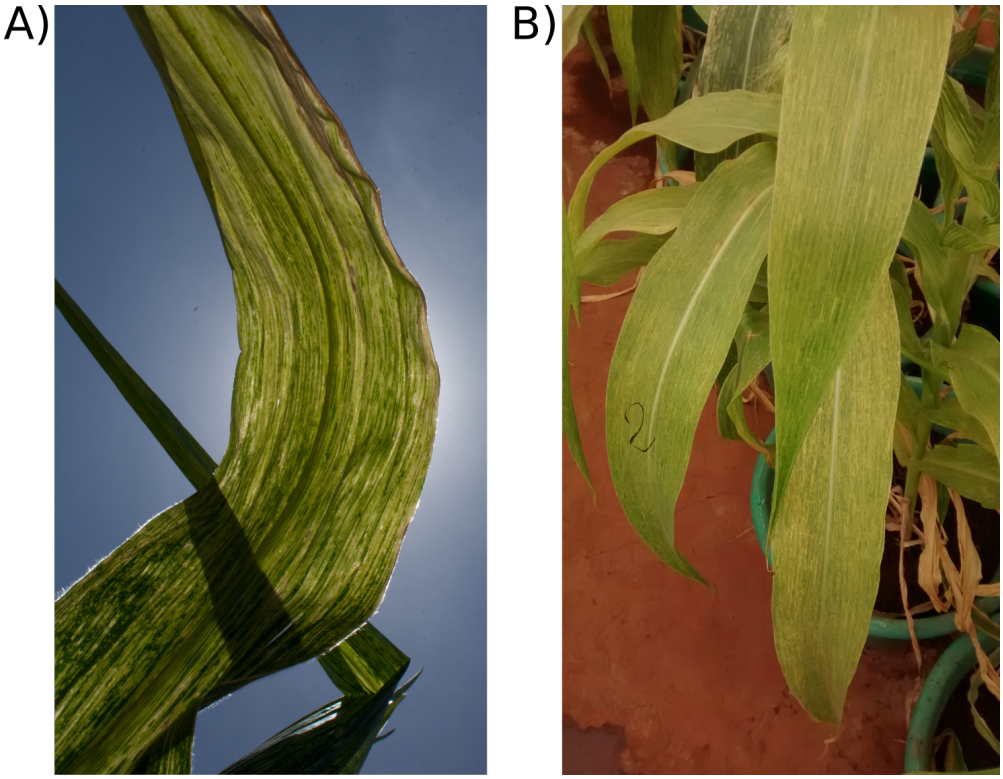


Fig. 4.19 Maize lethal necrosis symptoms in field (A), and produced in greenhouse by inoculation with Kenyan isolates of maize chlorotic mottle virus and sugarcane mosaic virus.

1999; Shukla et al., 1988). The regions of low diversity in the SCMV genome correspond to P3 and the 3' UTR. P3 is essential for potyviral replication, and its C-terminal region has recently been shown to be essential for P3 targeting to viral RCs, with P3 essential for RC function in turnip mosaic virus (TuMV). P3 is also the site of the cryptic fusion protein P3N-PIPO, a movement protein, although it seems more likely that the high conservation in this region is related to P3 function, as P3N-PIPO contains an early stop codon in maize-infecting SCMV isolates (Chung et al., 2008; Hillung et al., 2013). The PIPO site also coincides with a recombination hotspot (fig. 4.5). The 3' UTR of potyviruses contains a poly-A tail to promote genome stability and translation, which is completely conserved.

I observed very little correlation between SCMV sequence similarity and isolate origin (figs. 4.3 and 4.7). This is probably due to the historical cosmopolitan distribution of SCMV - it has been present in East Africa, China, North and South America for decades (Chen et al., 2002; Francisca et al., 2012; Louie, 1980), which can be explained by the long history of global trade in SCMV hosts. Recombination is clearly a major force in SCMV evolution (fig. 4.4), as it is in the *Potyviridae* more generally (Chare and Holmes, 2006; Revers et al., 1996). The geographically nonsensical distribution of similar indels amongst SCMV isolates

globally provides independent evidence of extensive recombination in SCMV (figs. 4.8 and 4.9). Recombination hotspots have been reported elsewhere in the potyviruses, in the P1 region of TuMV, and in a number of species in the CI-NIa-protease region of the genome (Bousalem et al., 2000; Moreno et al., 2004; Ohshima et al., 2007; Seo et al., 2009; Tugume et al., 2010). SCMV does not contain a recombination hot spot in central P1 like TuMV, but it does in the C-terminal region of CI (fig. 4.5). An interesting observation made on TuMV recombination is that breakpoints tend to have higher GC content to the 5' of the breakpoint, and higher AU content downstream, similar to a pattern reported in *Tombusviridae* genera (Ohshima et al., 2007).

I decided not to perform a phylogenetic analysis on SCMV genome sequences. Simulations have demonstrated that phylogenetic analyses are most affected by recombination when it occurs near the centre of genomes, and has occurred recently amongst diverged taxa, as appears to be the case in SCMV (fig. 4.4) (Posada and Crandall, 2002). Therefore, current methods of phylogenetic inference are inappropriate, and a more productive direction of future research may be to correlate genomic variation with variation in host preference and virulence, in order to classify SCMV isolates according to disease risk potential.

4.9 Value of NGS in characterising viral conditions

The rise of metagenomics has demonstrated that, even in asymptomatic plants, viral infection is extremely common. In addition to MCMV and SCMV, I found a diverse range of insect (*Dicistroviridae*) and plant RNA viruses, including what appears to be three novel viral species (table 4.2). None of these are strong candidates for possible partners of MCMV (i.e. not *Potyviridae*). However, the widespread distribution of MYMV in combination with MCMV and SCMV means it is an interesting candidate for future study, both as an emerging viral disease, and for potential involvement in the MLN disease complex (fig. 4.17). Recently published work on MLN in East Africa has shown that the *Potyvirus* member johnsongrass mosaic virus (JGMV) is present in complex with MCMV in Uganda and Kenya, and that this combination causes MLN (Stewart et al., 2017). This represents the first report of JGMV in complex with MCMV anywhere in the world, and therefore a novel partner virus.

NGS can detect diverse and novel viral sequences, in comparison to alternatives such as PCR and ELISA. This is valuable in the context of MLN due to the potential diversity of MCMV partner viruses. NGS of maize can be used to monitor for novel viruses in maize, as I did in this chapter, and may identify novel partner viruses such as JGMV. Another use of NGS would be monitoring common plants and companion crops to monitor for potential emergent partners of MCMV in the future.

Conclusions

MLN-symptomatic maize in East Africa contains MCMV, SCMV, and a number of RNA viruses not known to be involved in MLN, including novel species. The prevalence of viruses in these samples highlights that multiple viral infection may be the rule, rather than the exception for plants, and that NGS is an ideal tool for characterising these complex and variable viral mixes.

MLN in East Africa is predominantly caused by MCMV interacting with SCMV. Compared to MCMV, SCMV is extremely diverse in terms of nucleotide sequence, both globally and locally, with extensive recombination. Historic movement is likely responsible for the high level of sequence variation, even within the same geographic region. Therefore, SCMV is a far less appealing candidate for the engineering of sequence-mediated resistance, as for most possible target sites there are already SNPs present in populations, which would rapidly lead to the fixation of viral resistance-breaking variants.

Chapter 5

Engineering resistance to maize lethal necrosis

5.1 Summary and objectives

The global spread of MCMV has led to the emergence of MLN as a widespread threat to maize production. The aim of this chapter was to engineer maize lines that are resistant to MCMV, and thereby MLN with the aim of making them available for use in East Africa where maize is the most important food crop. To generate resistance which is robust despite MCMV evolution my approach was to use transgenic amiRNAs to target multiple, conserved regions of the MCMV genome. Using NGS data I found that intra-sample variability is reproducible between samples, i.e. low-frequency viral variants are visible within a single sample in NGS data, and this low-frequency variation occurs in similar locations between samples. To use this information, I developed a novel pipeline to use intra-sample variation in amiRNA design for the first time. I tested processing of amiRNAs using transient expression in *Nicotiana benthamiana*, and colleagues at Kenyatta University (KU, Kenya), transformed two amiRNA constructs into tropical maize lines. Molecular characterisation of the lines is ongoing.

Introduction

5.2 Resistance-breaking by RNA viruses

amiRNA mediated resistance relies on Watson-Crick base pairing between the amiRNA and its target sequence, in this case a section of viral genome. Mutations in amiRNA target sites decrease the efficiency of amiRNA binding to the viral genome, and thereby decrease the efficiency of the translational repression/cleavage mediated by RISC activity. The decreased efficiency of amiRNA targeting on target sites with mismatches generates a selective pressure on the rapidly evolving RNA virus, which can lead to viral variants which break resistance rising to prominence in the population, thereby rendering the engineered resistance obsolete. The selection of resistance breaking strains of virus has been demonstrated experimentally in *Arabidopsis* infected with turnip mosaic virus (TuMV). Lafforgue et al. (2011) performed two experiments, in one they passaged TuMV through *Arabidopsis* plants with sub-inhibitory and variable expression of an anti-TuMV miRNA every two weeks. At each passage, they used sap to inoculate 20 additional *Arabidopsis* plants with high expression of an amiRNA targeting TuMV, and counted resistance as broken if one or more of these plants developed TuMV symptoms. 25 lineages were passaged in this way. Viral lineages evolving under this selection pressure rapidly broke resistance (1-8 passages), with mutations observed in the amiRNA target site. The second experiment was identical, but with passaging between wild-type *Arabidopsis* plants rather than the low amiRNA expression line, to test whether viruses can evolve amiRNA resistance through drift, which could be relevant in pathogen management efforts, if non-resistant reservoir plants can also generate resistance-breaking viral lineages. This resulted in slower resistance-breaking (6-25 passages), but illustrated that drift is also capable of producing resistant variants (fig. 5.4). NGS analysis of viral samples at each passage showed that 21 potential escape alleles were present at the start of one wild-type passaging lineage (only one was sequenced) (Martínez et al., 2012). This shows that low frequency amiRNA escape variants are likely to be present in any given individual host, due to the high number of viral genomes and error rate of RNA polymerase. Together these experiments highlight the evolutionary power of viruses, and the importance of taking this into account if engineered resistance is to last for a useful length of time.

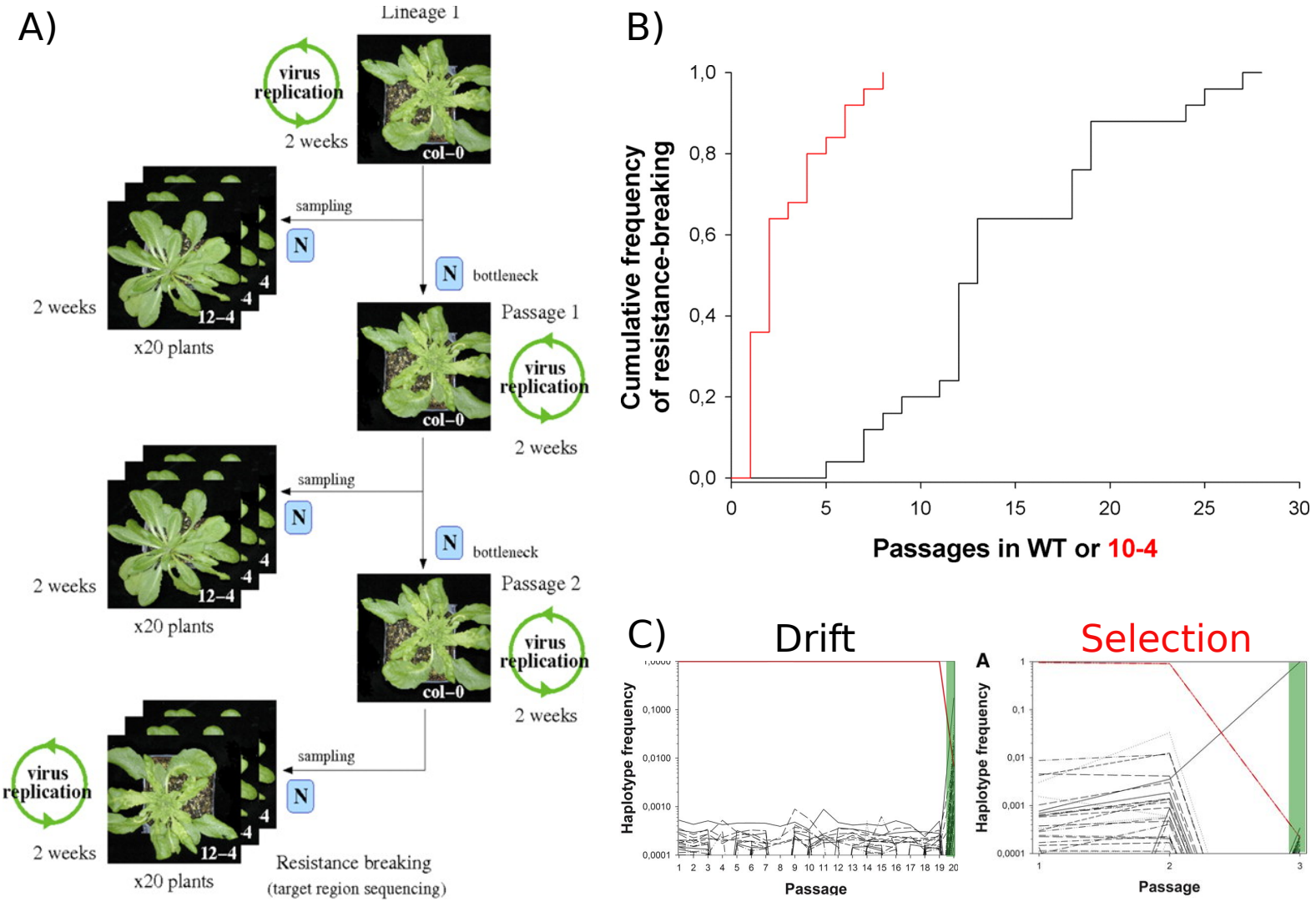


Fig. 5.1 A) Experimental design from Lafforgue et al. (2011). B) Cumulative frequency graph of viral lineages showing resistance during passage through wild-type *Arabidopsis* or plants expressing low levels of a turnip mosaic virus (TuMV) amiRNA (10-4). C) TuMV haplotype frequency in lineages from B) showing selective sweeps in wild-type (drift) and amiRNA (selection) plants. A) and B) adapted from Lafforgue et al. (2011), C) adapted from Martínez et al. (2012).

Methodology

5.3 amiRNA design and synthesis

5.3.1 Maize chlorotic mottle virus consensus extraction and alignment

The initial demultiplexing, adaptor removal, and quality trimming is as described in the methodology of chapter 3. As the *de novo* assembly performed for MCMV in chapter 3 did not show indels in coding regions, I extracted reads aligning to the MCMV Sichuan isolate (JQ982470.1) from the bowtie2 alignment performed to maize and MLN viruses performed in chapter 4. I generated pileups to a maximum coverage of 50,000 across the MCMV Sichuan genome, then called consensus sequences using the QUASR script *pileup_consensus.py*, with ambiguous bases called at the 1% level (parameters: *–ambiguity 0.01 –dependent –cutoff 20 –lowcoverage 25*) (Li, 2011; Watson et al., 2013). I then aligned these MCMV sequences using MUSCLE together with the MCMV genome sequences available in Genbank (gap extension cost of 1000, other settings default) in MEGA6 (Tamura et al., 2013).

5.3.2 amiRNA design

I took the regions of the MCMV genome which were completely conserved across the alignment and 20bp or longer and combined them into a fasta file (electronic supplement). I then used two amiRNA design programmes, WMD3 (http://wmd3.weigelworld.org/cgi-bin/webapp.cgi) and P-SAMS (http://p-sams.carringtonlab.org/) on each of these conserved regions (Fahlgren et al., 2015; Ossowski et al., 2008). The highest scoring WMD3 amiRNA for each site was used for further analysis (P-SAMS provides one amiRNA per target). I trimmed the P-SAMS amiRNAs with three or more predicted (by P-SAMS) off-targets in the maize transcriptome. I then removed amiRNAs targeting the region covered by MCMV sgRNA1.

5.3.3 Screening amiRNAs with heterozygosity

To narrow down the number of potential amiRNAs for quality control checking, I used within-sample heterozygosity data in order to select the amiRNAs targeting the most conserved sites, including low-frequency viral variants. I generated pileups from bam alignment files and loaded them into R using the script *pattern_search_pileup.r*, a wrapper for the Rsamtools script *PileupParam* (parameters: *max_depth=10000000, min_base_quality=25, min_mapq=5, min_nucleotide_depth=20, min_minor_allele_depth=0, distinguish_strands=FALSE,*

distinguish_nucleotides=TRUE, ignore_query_Ns=TRUE,include_deletions=TRUE, include_insertions=TRUE). This list of pileups contains multiple rows per site, so I parsed them and calculated heterozygosity using the script *pileup_to_base_freqs_and_heterozygosity.R*. Sites with coverage below 20 were not imported, so I filled missing sites with NA using the custom script *pad_data_frame.R*. These data (and scripts) can be found in the electronic supplement. I then calculated mean unbiased heterozygosity for each site across the genome (Tajima, 1993). Nucleotide positions in amiRNAs have different importance for target binding, and I used the three tiers described in Lin et al. (2009). I selected target sites in which tier 1 positions avoided the top 40%/30% (P-SAMS/WMD3) most heterozygous sites in the genome, and other tiers avoided the top 20% using the script *amiRNAranking_from_hetvec.R*. I then took the amiRNAs targeting these sites onto quality control.

5.3.4 amiRNA quality control

To obtain amiRNAs with high activity and low probability of off-target effects, I used the following steps for quality control (QC):

- Search for off-target matches in the maize transcriptome:
  - Collected plus/minus blast hits against the maize Refseq RNA and EST databases.
  - Collected hits predicted by the plant small RNA target analysis server (http://plantgrn .noble.org/psRNATarget/, Dai and Zhao (2011)) against the following databases: CDS Plant GDB genomic project, and Release 5a of the NSF funded maize genome project.
- Calculation of amiRNA-target hybridisation energies. I used RNAhybrid (https://bibiserv.cebitec.uni-bielefeld.de/rnahybrid/, Rehmsmeier et al. (2004)) to calculate the minimum free energy of binding for each amiRNA to the MCMV Sichuan genome (JQ982470.1).
- Calculation of self-complementarity. I used the RNALfold programme from the ViennaRNA package (v2.1.9, Lorenz et al. (2011)) to calculate the tendency of amiRNA sequences to form secondary structures.
- Calculation of within-sample heterozygosity at target sites. Detailed below.

P-SAMs and WMD3 designed amiRNAs often introduce one or more mismatches between amiRNA and target sequence to discourage RdRP priming, which generates transitive silencing (Ossowski et al., 2008). However this is desirable for antiviral application, as it will

effectively increase the number of virus-targeting amiRNAs within infected cells. Therefore, I put the direct revcom of each selected amiRNA's target sequence, with an initiating U, through the same QC process as above. I then performed pairwise comparison for each target reverse complement with the original amiRNA, selecting the one with lower numbers of predicted off-targets. The final amiRNAs selected were assembled into two constructs, each containing five amiRNAs, called LABami08 and LABami09.

5.3.5 amiRNA* design

The reverse complements of chosen amiRNA sequences were taken as the basis of the amiRNA* sequence. I then inspected the predicted structure of the wild-type form of the *OsaMIR395* transcript using mfold (http://unafold.rna.albany.edu/?q=mfold, Zuker (2003)), and inserted mismatches into each amiRNA* sequence such that the structure of each amiRNA loop matched the corresponding wild-type loop. Base pairing at position 1 and 19 was then checked to ensure 5' stability was lower in the amiRNA strand compared to the amiRNA* strand.

5.3.6 Negative control design

To generate a negative control amiRNA construct I randomised the amiRNA sequences of the LABami08 construct and subjected them to the QC process described above, randomising again until each sequence passed. I then designed amiRNA* sequences as described above.

5.3.7 Construct cloning

I ordered the amiRNA constructs to be synthesised and provided within the pUC57 cloning vector (Biomatik). I amplified the maize *pUbi1-1* promoter, including intron, from the plasmid EC15455. I gel purified the *pUbi1-1* amplicon, TA cloned it into pGEM T easy, and Sanger sequenced to confirm the promoter sequence was correct. To connect LABami constructs to the *pUbi1-1* promoter I amplified both using primers designed for overlap-extension PCR (OE-PCR), then used OE-PCR to join *pUbi1-1* to each construct (Heckman and Pease, 2007). I gel-purified the correct sized products, A-tailed using PCR Bio taq, and TA cloned into p-GEM T easy, and used Sanger sequencing to confirm the construct sequence was correct. To move inserts into a maize tranformation vector (pTF101), I double restriction digested *pUbi1-1*-LABami inserts and pTF101, gel purified the desired fragments, and ligated using a 2:1 insert:vector ratio. pTF101 plasmids containing LABami constructs had their insert and backbone completely sequenced by Sanger sequencing, which confirmed

that insert sequences were correct and that there were three SNPs in the 35S promoter driving *bar* expression, and one single base deletion in the 3' terminator of the soybean vegetative storage protein gene which terminates *bar* expression. I concluded that these were unlikely to abolish *bar* expression and proceeded to transformation.

5.4 amiRNA transient expression, maize transformation and characterisation

5.4.1 Transient amiRNA expression in *Nicotiana benthamiana*

To express amiRNA constructs transiently in *N. benthamiana* I cloned the *pUbi1-1* LABamiRNA inserts from the pGEM t easy-LABami plasmids into the pAGMA-4723 amiRNA expression vector using restriction enzymes and ligation. amiRNA expression from pAGMA-4723 is driven by the *CmpS* promoter. I transformed *A. tumefaciens* strain C58C1 by electroporation with each pAGMA-4723 LABami construct. For infiltration, I grew *A. tumefaciens* strains in 5ml LB for 48 hours under antibiotic selection, spun cultures down and resuspended in infiltration buffer, then used a spectrophotometer to adjust concentration of each culture to OD^{600} of 0.5. The two youngest leaves per 21 day old *N. benthamiana* plant were fully inoculated with the appropriate strain, with a GFP-expressing strain (C58C1 containing pBin61[GFP]) used as a negative control. Two days post inoculation I harvested inoculated leaves and snap froze them in liquid nitrogen, then extracted RNA using Trizol (Ambion) according to manufacturer's instructions, with a spin step before phase separation to remove excess polysaccharides.

5.4.2 sRNA Northerns to confirm amiRNA expression

For sRNA Northern blots I ran 12μg of denatured total RNA per lane on precast 15% TBE-Urea gels (Invitrogen) which were pre-run for 50 minutes at 100 V. To separate RNA molecules, I ran at 100 V for 30 minutes, and then 150 V until the bromophenol blue dye reached the end of the gel. I stained gels with SYBR gold to confirm correct gel running and loading. For sRNA transfer, I soaked gels in 20X SSC for 10 minutes, then laid them on a clean glass plate in 2 ml 20x SSC. I then placed one layer of Hybond-N+ membrane (GE healthcare) on top of the gel, followed by two layers of Whatman filter paper, moistened with 20x SSC, then added blotting paper on top and weighed down the stack and left for 16 hours. I cross-linked RNA to the membrane with UV, then to probe for amiRNA I used DNA oligos of the reverse complement that I 5' labeled with P^{32} using T4 polynucleotide kinase (NEB).

To remove excess unincorporated radioactive nucleotides I cleaned the probe on illustra Microspin G25 spin columns (GE healthcare). As a loading control I probed for the maize U6 sequence using a DNA oligo with a 5' IRD 700 fluorophore from IDT. For hybridisation I used hybridisation tubes with 10 ml Ambion Ultrahyb-Oligo at 38°C overnight. To wash membranes I used 10 ml sRNA wash buffer for three 10 minute washes at 38°C. I imaged U6 labeling with an Licor Odyssey infrared imager, and miRNAs using phosphorimaging screens (Kodak) visualised on a Licor Odyssey infrared imager.

5.4.3 Maize transformation

Growth of maize for generation of explants

The maize transformation work was performed at KU by Joel Masanga (JM), with technical assistance from KU support staff. JM used maize inbred lines CML 144 (CIMMYT) and Namba nane (KALRO) for production of explants for transformation. Seeds were planted and maintained in potted soil with regular watering. At silking, JM self-pollinated the maize plants and covered them with plastic bags to prevent cross pollination, and harvested ears the 10th and 14th day after pollination, storing them at 4ºC to await transformation.

Surface sterilization and embryo excision

A protocol by Ishida et al. (2007) was adapted for tropical maize transformation. Briefly, JM harvested maize ears with 1-2mm long embryos, removed husks, and surface sterilized ears using 3.85% (v/v) NaOCl in Jik (bleach), together with 2 drops of tween 20 for 20 minutes, rinsing 3 times with autoclaved distilled water. To isolate embryos, JM used a sterile scalpel blade to chop off the top of kernels whilst still attached to the cob and an excisor to remove them. JM suspended the embryos in 5ml of filter sterilized infection medium (pH 5.2) inside a sterile petri dish. For infection cultures, JM grew *A. tumefaciens* containing the transformation constructs on selective LB for three days, then placed one colony in 20ml of infection medium containing 100µM acetosyringone in sterile 50ml centrifuge tubes. The cultures were incubated with shaking at 28ºC in darkness for 3 hours on a shaker. To infect embryos, JM added 5ml of the *A. tumefaciens* suspension to each petri dish containing embryos and covered the plate with aluminum foil to provide darkness for 5 minutes for efficient infection before co-cultivation.

Co-cultivation and callus induction

To initiate callus formation, JM transferred infected embryos onto co-cultivation medium, removing excess *A. tumefaciens* suspension with a sterile pipette tip and orientating embryos scutella-side upwards. As described by Ishida et al. (2007), plates were sealed with parafilm and incubated at 22ºC in the dark for 3 days. JM then transferred responding embryos onto resting medium and incubated at 28ºC in the dark for 10 days before taking them to selection.

Selection and regeneration of putative transformants

To initiate selection of putatively transformed tissues, JM transferred explants forming calli to initial selection media (250mg/l carbenicillin, 1.5mg/l Basta) for 2 weeks. The cultures were covered with 2 layers of aluminum foil to provide darkness and incubated at 28ºC. For second selection, JM transferred surviving calli to second selection medium (3mg/l Basta) for 28 days in darkness, subculturing at 2 weeks. For embryo maturation, JM transferred basta-resistant calli (now with somatic embryos) onto embryo maturation medium for 2 weeks in darkness. JM then inspected calli and transferred those showing presence of somatic embryos to regeneration medium, and incubated in light. JM and myself selected regenerated plants with well-developed roots for acclimatisation and hardening on peat moss for ten days, then transferred plantlets to potting soil in the glasshouse where they grew to maturity.

5.4.4 Maize T0 characterisation

We took tissue samples from T0 plants at KU, placed samples for RNA in RNA Later, and sent them to Cambridge, where I extracted DNA using CTAB and RNA using Trizol according to manufacturer's instructions.

Genomic PCRs

To amplify transgene amplicons from maize transformant T0 DNA I initially used Phusion polymerase (NEB), and varied annealing temperature, primer pairs, template concentration, and additives (DMSO, betaine). Later I compared KOD Hot start Xtreme (MilliporeSigma) and Extaq (TaKaRa) with Phusion, and decided to use KOD hot start in later PCRs due to its higher efficiency amplification from maize genomic DNA.

sRNA Northern blots

I performed sRNA Northern blots of T0 transformants using 10μg of denatured total RNA for loading, and the protocol described above for *N. benthamiana* expression.

Southern blots

To digest DNA, I placed 10μg of maize genomic DNA from T0 transformants into a 200μl restriction digest reaction containing 25 units each of *BamHI*-HF and *HindIII*-HF (NEB). Digestions were incubated overnight at 37ºC, then I added 10 units more of both enzymes in the morning, and allowed the reaction to proceed for two more hours. I precipitated the DNA with isopropanol, washed with 70% ethanol, dried and resuspended in 10μl TE buffer. To separate DNA fragments, I loaded samples into a 5mm thick TBE agarose (1%) gel and ran them overnight at 30V. To visualise DNA separation and loading I used a transilluminator, then washed the gel for 7 minutes in depurination solution, 30 minutes in denaturing solution, and 30 minutes in neutralising solution, rinsing the gel with Milli-Q between washes. I placed the gel in 10X SSC before transfer. I performed capillary transfer by placing a long sheet of pre-wet Whatman 3MM filter paper over a glass plate with its ends in a reservoir of 10X SSC below. Then I placed on top of the filter paper: the agarose gel, pre-wet Hybond-N+ membrane (GE healthcare), two pieces of pre-wet Whatman 3MM filter paper, a stack of blotting paper, then a weight. I let transfers proceed overnight, and checked successful transfer by again imaging the agarose gel using a transilluminator. To cross-link DNA to the membrane I used UV. For probes, I used gel purified PCR products 500-1000bp in length, labeled with P^{32} CTP using the Rediprime labelling II kit (Amersham). Unincorporated radioactive nucleotides were removed by using illustra Microspin G25 spin columns (GE healthcare). To hybridise probes to genomic DNA I used hybridisation tubes with 10ml of UltraHyb Ultrasensitive Hybridisation buffer (42ºC) or Church Gilbert buffer (65ºC), hybridising overnight. Membranes were pre-hybridised for at least one hour. The initial, more stringent washing I used was:

1) 2X SSC, 0.1% SDS for 20 minutes.

2) 1X SSC, 0.1% SDS for 20 minutes.

3) 0.1X SSC, 0.1% SDS for 5 minutes.

which was altered after the first Southern to the same washes but with 10, 5 and 5 minutes incubation respectively. I imaged the Southerns using phosphorimaging screens (Kodak) visualised on a Typhoon 8610.

Results

5.5 amiRNA construct design

5.5.1 Promoter choice

High expression is desirable for antiviral amiRNA applications, because higher amiRNA expression is correlated with lower viral load and resistance breaking (Ai et al., 2011; Lafforgue et al., 2011). I decided to use the maize Ubiquitin promoter (*pUbi1*) to drive amiRNA expression as it is native (less objectionable to regulators and general public), produces high expression in monocots, and also contains an intron in the 5' UTR (Christensen and Quail, 1996; Christensen et al., 1992). Intron presence has recently been implicated in the avoidance of transgene silencing through RNA-dependent DNA methylation (RdDM). Transgene silencing in *Arabidopsis* depends on RDR6 activity and the production of siRNAs which stimulate RdDM of the transgene locus, and therefore transcriptional silencing (Christie et al., 2011; Dalmay et al., 2000). However, transgenes typically lack introns and examination of *Arabidopsis* sRNA libraries shows that native intronless genes have a far higher sRNA density than those with one or more introns (fig. 5.2) (Christie et al., 2011). Introducing introns to 35S:GFP transgenes in *Arabidopsis* reduced transgene silencing in the T1 approximately threefold (Christie et al., 2011). RdDM in maize requires orthologs and paralogs of those involved in RdDM in *Arabidopsis*, suggesting the presence of a divergent but functionally similar system of transgene silencing in maize (Haag et al., 2014). Therefore, *pUbi1* was also attractive because the downstream 5' UTR contains an intron immediately before the coding sequence, which may help limit the level of transgene silencing in transformed maize lines. This is not a certainty in maize, due to phylogenetic distance, and *pUbi1* has been silenced in barley lines due to RdDM. However, the scarcity of monocot promoters with characterised expression in maize, and the other benefits mentioned above meant that I selected the *pUbi1* promoter.

5.5.2 amiRNA design rationale

The introduction covers my reasoning for selecting amiRNA as the strategy for engineering sequence-mediated resistance to MLN. Chapters 3 and 4 illustrate that MLN in East Africa is caused predominantly by a mixture of MCMV and SCMV, and that the much lower genetic variability of MCMV means it is a much more attractive candidate for targeting with sequence-mediated resistance. SCMV has far higher nucleotide sequence variation than MCMV, so it also has many fewer suitable (conserved) potential targeting sites. Therefore, I

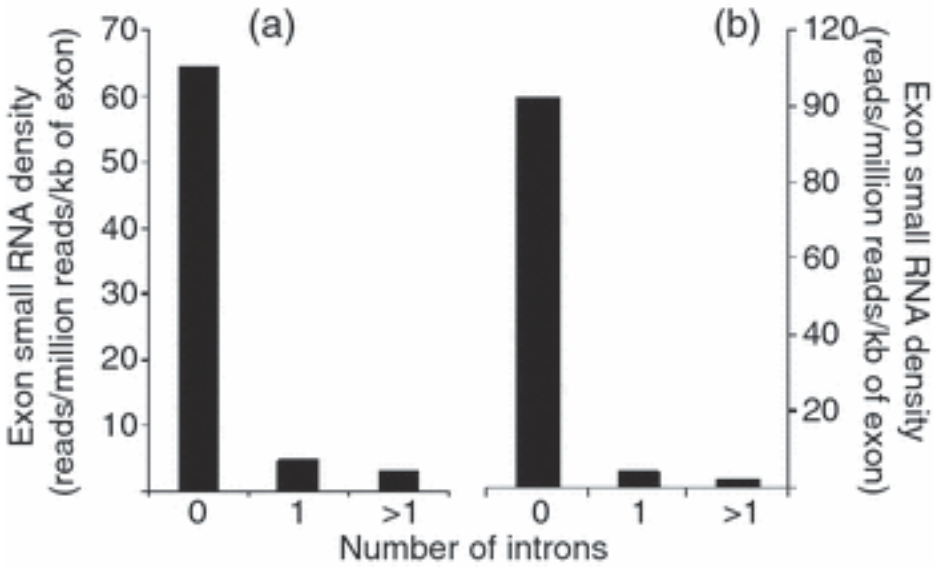


Fig. 5.2 Density of siRNAs targeting *Arabidopsis* exons in wild-type seedlings shows genes with no exons are more commonly targeted by siRNAs. Graphs without (A) and with (B) Repeatmasker filtering for transposons. Figure adapted from Christie et al. (2011).

decided to target MCMV with amiRNAs because it is much less feasible to produce robust sequence mediated immunity against SCMV.

Two general principles are used to design amiRNAs intended to be more robust against viral evolution of resistance breaking: targeting of conserved regions, and targeting of multiple regions. Conserved regions are more likely to be under purifying selection than variable regions, implying preservation of a functional motif, whether in RNA or amino acid sequence. Therefore, mutations in these sites are more likely to decrease viral fitness by impacting the function of said motif. Targeting multiple regions means that viral variants need multiple, independent, mutations to fully escape amiRNA targeting. This exponentially decreases the probability that sufficient escape mutations will occur simultaneously during genome replication. The efficacy of both strategies has been demonstrated experimentally in plants, again using TuMV in *N. benthamiana*, with Lafforgue et al. (2013) testing the efficacy of either two amiRNAs, or one amiRNA targeting the 3' region of the CP ORF, which may be a functional RNA motif in TuMV. For both, there were 0 resistance-breaking events in around 1000 inoculations.

The purpose of this part of the project is to generate maize lines with robust sequence-mediated resistance against MCMV, with a long-term view to applying these as part of a MLN disease management strategy in the field. Given the long timescales and large budgets required for field applications, I decided to combine both of the above strategies to minimise

the probability of MCMV breaking resistance, i.e. to target multiple, conserved regions of the MCMV genome.

5.5.3 amiRNA backbone selection

In order to target multiple sites in the MCMV genome, I decided to use a modified polycistronic (multiple miRNAs from one transcript) miRNA gene. To maximise the chance of efficient processing, I wanted to use a native polycistronic maize miRNA gene. However, the only polycistronic maize miRNAs genes confirmed by RACE are *miR156c/b* and *miR166k/m*, both of which only contain two miRNAs (Zhang et al., 2009). Fahim et al. (2012) modified a rice (*Oryza sativa*) miRNA cluster of *miR395* to target five sites in the WSMV genome to engineer resistance to WSMV in wheat. There are also clusters of miR395 miRNAs in the maize genome, with slightly wider spacing than in rice, but their cotranscription has not been experimentally confirmed and miR395 expression is low, which could represent inefficient pri-miRNA processing (Zhang et al., 2009). Therefore, I decided to use the rice *miR395a-g* cluster, truncated to 800bp containing five miRNA sites.

5.5.4 Feasibility of using intra-sample variation in amiRNA design

Given that NGS can be used to understand how resistance-breaking occurs through its ability to track low-frequency variants of viral genomes, as in Martínez et al. (2012), I wanted to investigate whether it is possible to instead use NGS data in the design process to generate amiRNAs more robust against the viral evolution of resistance breaking. The concept is to sample low-frequency genomic variants in the wild, and avoid regions of high heterozygosity which are not represented by consensus sequences (e.g. see fig. 5.3).

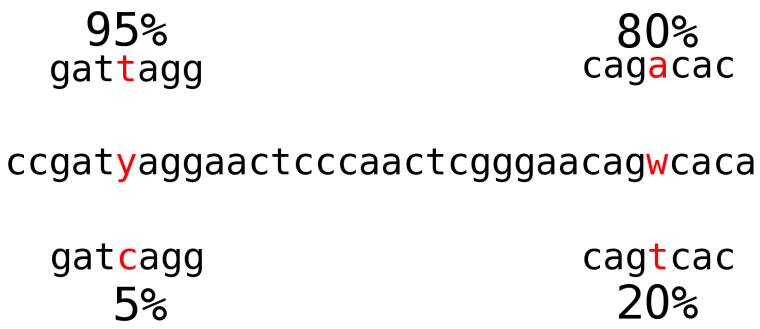


Fig. 5.3 Example of hypothetical intra-sample variation at an amiRNA target site. This variation would be undetected using conventional consensus sequences.

There are two conditions that must be met for intra-sample variation to be useful in amiRNA design. Firstly, variability must vary significantly across the MCMV genome, or there will be little benefit to avoiding the most variable regions. To measure variability I used Tajima's unbiased estimate of heterozygosity:

$$h_{\mathrm{i}} = \frac{n(1 - \sum_{j=1}^{4} {x_{\mathrm{ij}}}^2)}{n-1}$$

Where:

h_{i} = heterozygosity at site i.

n = number of sequences sampled.

x_{ij} = frequency of base j at site i, j being 1-4 to represent each base.

In words, 1 - the sum of each base frequency squared, multiplied by $\frac{n}{n-1}$ (Tajima, 1993). Plotting the mean heterozygosity across the MCMV genome (across all samples) shows that variability changes significantly across the MCMV genome, with most sites exhibiting very low heterozygosity, but the distribution rising to over 0.5 at the most variable sites. (fig. 5.4).

The second requirement is that within-sample variation is similar between samples. otherwise a variable site in one sample may actually be conserved in another. Plotting individual sample heterozygosity within a 20bp window (approximately amiRNA target size) shows that peaks of heterozygosity are conserved between samples, even though some of the samples were collected hundreds of miles apart (fig. 5.4c). Therefore, I conclude that the use of within-sample variation is a useful additional factor for optimising the design of robust sequence-mediated resistance to MCMV.

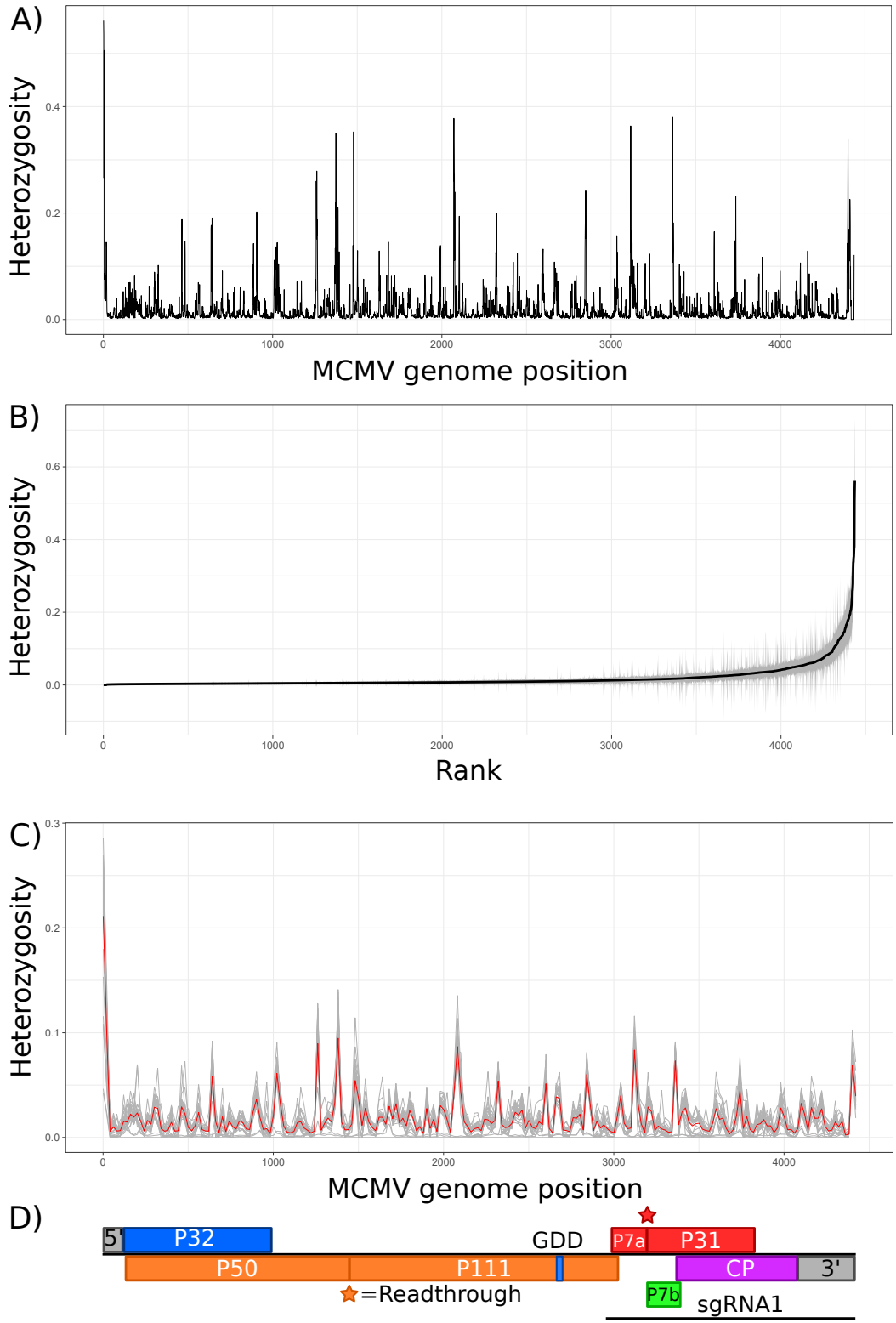


Fig. 5.4 Heterozygosity (variability) varies significantly across the maize chlorotic mottle virus (MCMV) genome. A) Tajima's unbiased estimate of heterozygosity across the MCMV genome, averaged across all MCMV NGS libraries. B) As for A), but with heterozygosity scores ranked, and standard deviation visible in grey. C) Intra-sample heterozygosity for each sample, plotted with a 20bp sliding windows, showing consistency between samples. D) Schematic of MCMV genome, aligned with above graphs.

5.5.5 amiRNA target site selection

In order to target the most conserved regions of the MCMV genome, I extracted MCMV consensus sequences for each of my NGS libraries with a lower threshold of 1% for ambiguous base calls. I then aligned these genomes together with the 15 MCMV genomes available on NCBI in January 2016, using MUSCLE (gap extension cost, -1000, other settings default) in MEGA6 (Tamura et al., 2013). From this alignment I took the 43 completely conserved regions of 20bp or more in length across the alignment. To design amiRNAs targeting these regions I used two open source amiRNA design algorithms, WMD3 and P-SAMS (Fahlgren et al., 2015; Ossowski et al., 2008). The design process for WMD3 can be seen in fig. 5.5, while for P-SAMS all possible target sites are examined, and amiRNA designed which are complementary to the target but with:

- 5' terminal U
- 19th nucleotide as C
- Mismatch at position 21

These design tools take into account features of native plant miRNAs by looking at base conservation in plant miRNAs, the distribution of mismatches in validated miRNA-target pairs, and targeting criteria developed for mammalian siRNAs (Fahlgren and Carrington, 2010; Fahlgren et al., 2015; Ossowski et al., 2008).

5.5.6 Screening of amiRNAs with sgRNA1 and heterozygosity

This produced 383 suggested amiRNAs from WMD3, split over 96 target sites in the MCMV genome, and 89 suggested amiRNAs from P-SAMS (which designs one amiRNA per target site). P-SAMS has a function for checking the host-transcriptome for potential off-targets, so I excluded amiRNAs with three or more predicted off-targets in the maize transcriptome, which cut the number to 56. This is still too many to take through quality control checking, so I decided to use the location of MCMV sgRNA1 and intra-sample heterozygosity to reduce the number of amiRNAs for quality control. MCMV produces the 1.47kb sgRNA1, which covers the 3' region of the genome (fig. 5.4d), and is required for CP production and is typically present at a higher concentration than the full length MCMV genome (Scheets, 2000, 2016). Additionally, none of the proteins present on this sgRNA are required for MCMV genome replication (Scheets, 2016). Therefore, because this region of the genome is non-essential for genome replication, and the high concentration of sgRNA1 makes it

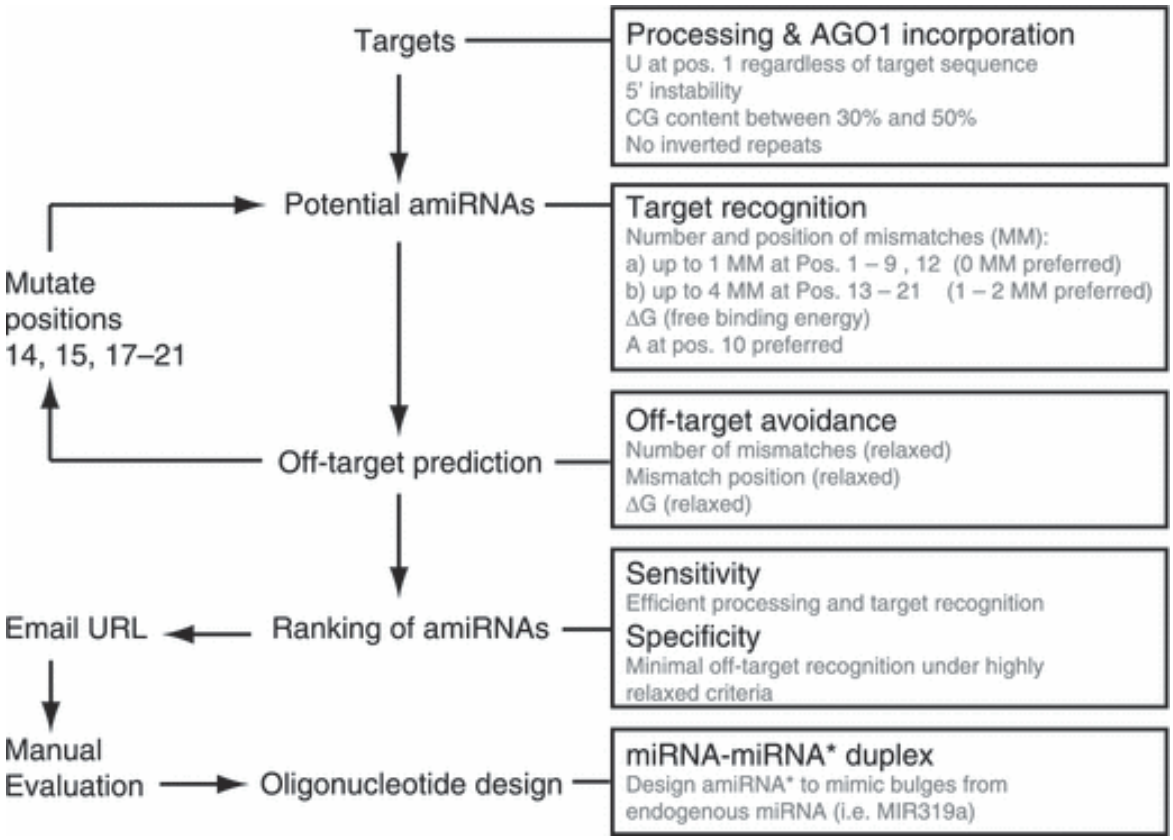


Fig. 5.5 Flowchart of the design process used by the open source WMD3 amiRNA design algorithm (Ossowski et al., 2008)

possible that it would titrate away amiRNA molecules, I decided to avoid targeting this region of the genome, cutting WMD3 down to 33 target sites, and P-SAMS to 42.

To use heterozygosity as a filter of target sites, I calculated heterozygosity across the genome for each library (see methodology). The aim was to avoid target sites containing the regions of highest intra-sample variation. However, different nucleotide positions in amiRNAs have different importance for efficient target binding. I wanted my filter to reflect this, and decided to use data from Lin et al. (2009) to inform my strategy. This study used a chimaeric TuMV genome, with an inserted non-functional amiRNA target site, to inoculate *N. benthamiana* plants expressing the corresponding amiRNA, providing resistance. Systematic mutagenesis across the amiRNA target site showed that there were three tiers of importance for resistance breaking:

- Tier 1: sites 3, 4, 5, 6, 9, 12.
- Tier 2: sites 2, 10, 11, 13, 15, 18.
- Tier 3: sites 1, 7, 8, 14, 16, 17, 19, 20, 21.

Inoculation of *N. benthamiana* plants expressing the amiRNA with chimaeric TuMV resulted in 0% of plants developing viral disease symptoms, while inoculating the amiRNA expressing plants with wild-type (no amiRNA target site) TuMV resulted in 100% of plants developing symptoms. Tier 1 mutations resulted in 60-90% of plants developing symptoms, tier 2 mutations resulted in 20-55% of plants developing symptoms, and there was 5-20% of plants developing symptoms when tier 3 sites were mutated (Lin et al., 2009). I used a custom R script to only allow through sites which avoided 40% most variable sites in tier 1 sites, and avoided the 20% most variable sites in tier 2 and 3 sites for P-SAMS amiRNAs, and thresholds of 30%, 20%, and 20% for tiers 1, 2, and 3 respectively for WMD3 amiRNAs (fig. 5.6). These thresholds were chosen to produce around 10 amiRNAs from each programme for quality control checking - there were 9 for each taken onto quality control.

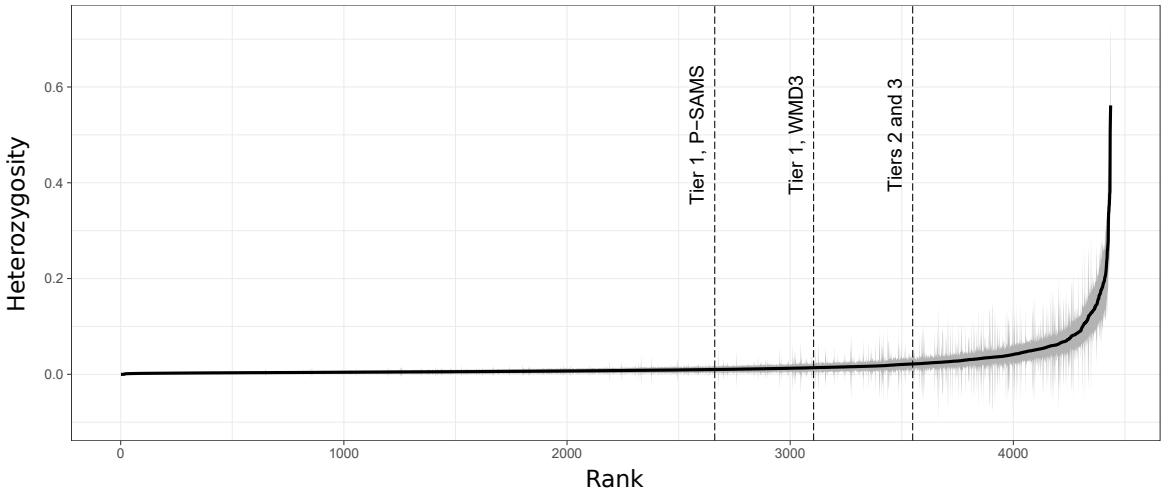


Fig. 5.6 amiRNA target site selection thresholds. Tajima's unbiased estimate of heterozygosity (Tajima, 1993) across the maize chlorotic mottle virus (MCMV) genome, averaged across all MCMV NGS libraries, and sorted. Vertical lines show variability thresholds for sites in tiers of differing importance for resistance breaking, as established in Lin et al. (2009). amiRNA target sites selected are to the left of the thresholds, minimising intra-sample variability.

The heterozygosity analysis I performed has some limitations. I used string-matching deduplication (i.e. deleting all reads with exactly the same sequence) because it is more appropriate for viral analyses than alignment-based deduplication, which was designed for alignment to large genomes, in which two reads with the same alignment position is unlikely to occur by chance. Alignment-based deduplication removes reads which have the same start and end position when aligned to a reference genome, regardless of sequence identity. However, when aligning to short viral genomes with high coverage, as with MCMV in this study, reads are bound to share the same alignment position. String matching also

has a drawback: it will remove real, independent reads which derive from different, but identical fragments of MCMV genome in samples. This is likely to happen to common MCMV genome variants with a very high titre in samples. Another problem with using low frequency variation is that some fraction of the intra-sample variation will be caused by sequencing error, rather than true variation. However, I do not believe these limitations invalidate the analysis. The occurrence of independent, identical reads should occur at roughly the same probability across the MCMV genome. Removal of these independent reads is unavoidable using current deduplication methods, and deduplication is required to prevent skewing of data by PCR duplicates. Sequencing errors will increase heterozygosity estimates, but stringent quality trimming was used so that the median PHRED score of reads is >30, i.e. 99.9% accuracy. Additionally, the real purpose of this analysis is not to precisely quantify MCMV heterozygosity across the MCMV genome, but to identify the sites with highest heterozygosity so that they can be avoided. At these sites, due to deduplication, high heterozygosity values are generated by independent reads, each with a low probability of being incorrect due to sequencing error. Therefore, the sites with highest heterozygosity are those most likely to have this value predominantly explained by true variation, rather than sequencing error.

5.5.7 Quality control checking

The factors considered in my quality control procedure were:

- Off-targets for amiRNAs in the maize transcriptome, to avoid altering maize gene expression.
- amiRNA-target hybridisation energies, to select amiRNAs with more energetically favourable binding.
- Self-complementarity, to avoid amiRNAs with a tendancy to form secondary structures, which would inhibit amiRNA-target binding.

The tools used to collect the above data are detailed in the methodology. Plus/minus hits only were collected from blast, as these represent base pairing between amiRNA and the positive-sense form of the mRNA. Plus/plus hits would represent amiRNA* binding. amiRNA* are mostly degraded without being loaded into RISC, in general, and when expressed from the backbone I selected for these constructs (Carbonell et al., 2014; Fahim et al., 2012). The following criteria were used for selection of amiRNAs:

- Fewer than 8 targets in total predicted by the plant small RNA target analysis server.

- Maximum blast hit lengths of 16nt for the EST database, and 16nt for Refseq.
- Lower than -30kcal/mol amiRNA-target binding energy change.
- Higher than -5kcal/mol minimum free energy of self-complementarity.

This resulted in seven amiRNAs passing quality control checking (table 5.1). The selected target sites are shown in fig. 5.7.

WMD3 introduces mismatches to amiRNA sequences to reduce the probability of transitive silencing - miRNAs acting as a primer to plant RDRs, resulting in the production of long dsRNA molecules and siRNA from cleavage of these dsRNAs by DCL proteins (Ossowski et al., 2008). This is undesirable for targeted knockdowns of native genes due to the increased risk of off-target effects. However, in a viral defence scenario (whether natural or engineered) it is beneficial, as it increases the number of sRNA species targeting the viral genome, both decreasing the probability of successful escape variants, and spreading selection for escape variants across the genome. Therefore, I decided to take the direct reverse-complement of the seven selected amiRNA target sites (with a 5' U to promote AGO1 loading, Takeda et al. (2008)), and put them through the same quality control process as above. I compared each target site reverse-complement to its equivalent amiRNA, and selected the sequence performing better in terms of quality control, shown in bold in table 5.2. The overall design process is summarised in fig. 5.8.

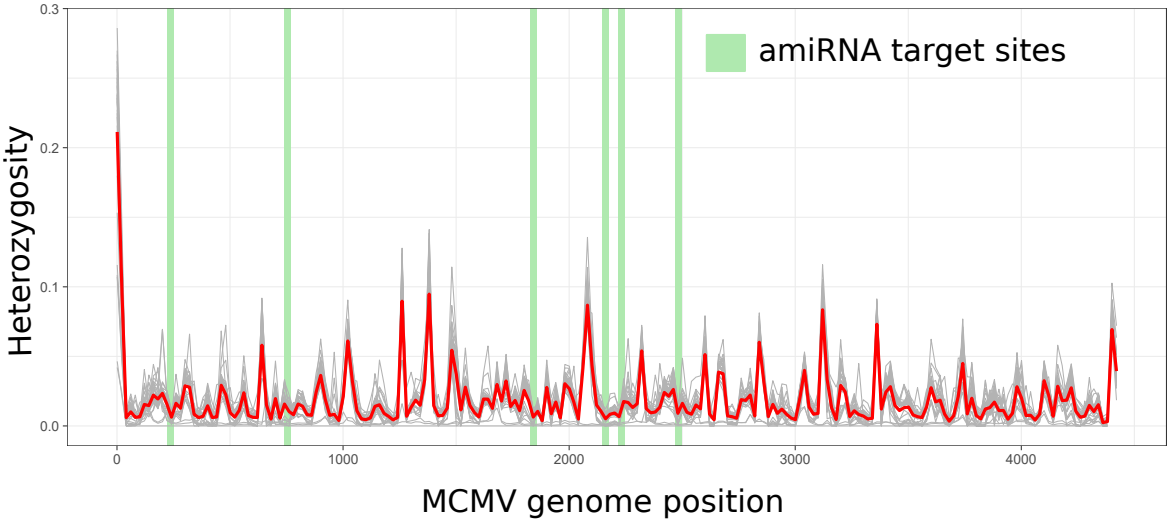


Fig. 5.7 Tajima's unbiased estimate of heterozygosity across the maize chlorotic mottle virus (MCMV) genome, averaged across all MCMV NGS libraries (Tajima, 1993). Final amiRNA target sites shown in green. Third bar from the left represents both 1830 and 1831bp target sites.

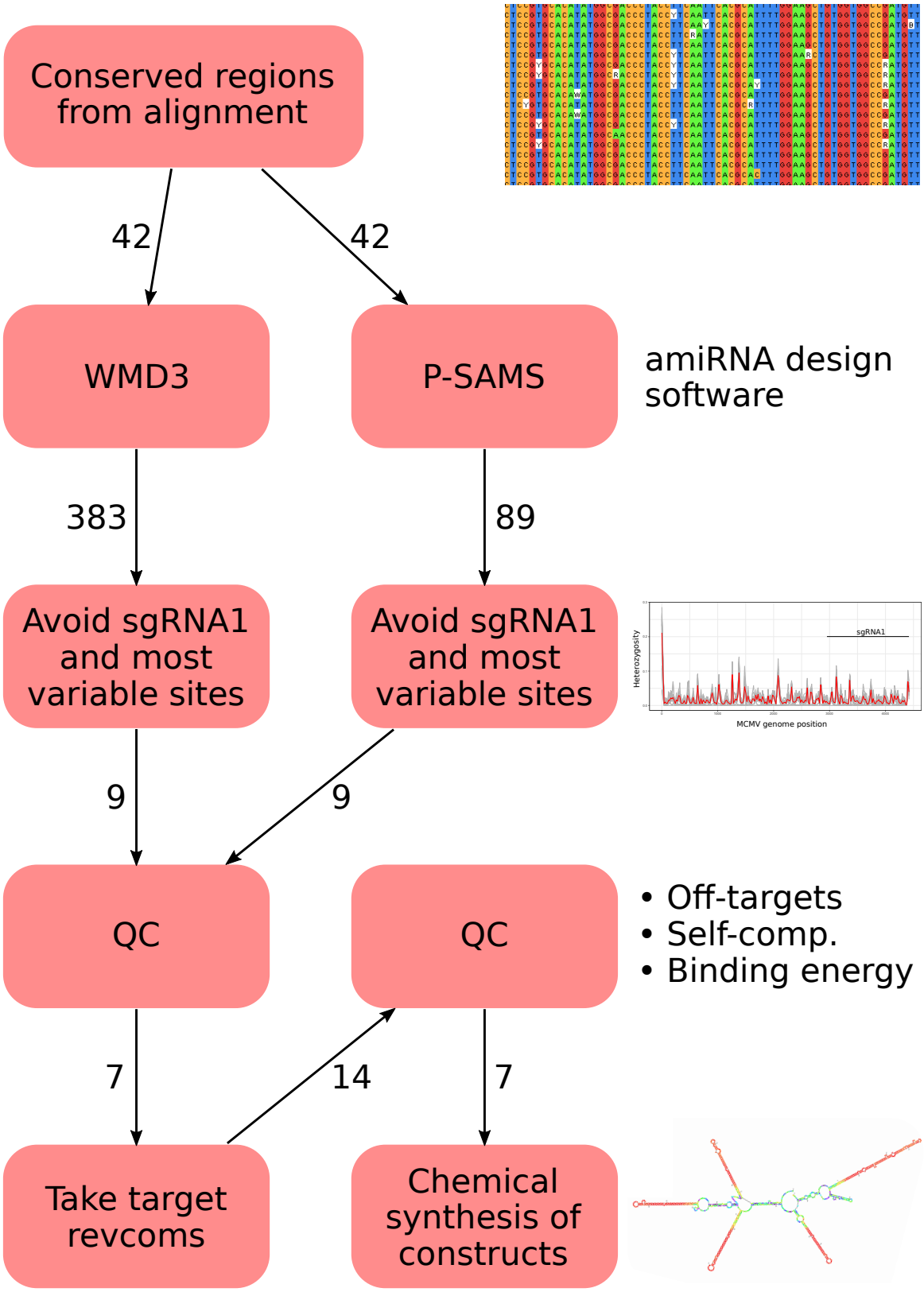


Fig. 5.8 Summary of amiRNA design process used to select amiRNAs for generating maize chlorotic mottle virus resistance.

Table 5.1 Metrics used for quality control filtering of amiRNAs. Selected amiRNAs shown in bold. Table shows the start position of amiRNA target site in maize chlorotic mottle virus (MCMV) genomes, longest blast hits to EST and Refseq databases, targets predicted by the psRNA (Plant small RNA target analysis server) against cds and NSF maize databases. amiRNA-target hybridisation energy shown under Hyb. energy, and finally the minimum free energy of self-complementarity.

Start pos.	Design	amiRNA sequence	Longest blast hit, EST	Longest blast hit, Refseq	psRNA off targets, cds	psRNA off targets, NSF	Total psRNA targets	Hyb. energy, kcal/mol	Self-comp. mfe
745	**WMD3**	**TAGAAGCGAACTGCTACCCTA**	14	14	0	0	0	-33.8	-1.6
746	WMD3	TGAGAAGCGAACTGCTACCCG	14	14	0	0	0	-37.7	-1.6
923	WMD3	TTTGAGTTTCATAACCATCGT	15	15	9	8	17	-34.4	0
1830	**WMD3**	**TGTAGAGGGAATTTATCGCTT**	14	13	2	2	4	-38.5	-2.4
1835	WMD3	TATTTCGTAGAGGAAACTCAC	14	14	1	4	5	-27.4	-2.5
1832	WMD3	TTCGTAGAGGGAATTCACCCC	14	14	4	4	8	-39.6	-3.8
2151	**WMD3**	**TAACACTCACGTGTTGGTCAA**	14	14	0	0	0	-38.4	-4.4
2220	**WMD3**	**TCAATTGTGTGAGACGCTCTG**	16	13	2	5	7	-34.4	-0.9
2474	**WMD3**	**TCAATGTCTGTACAGTCCCTA**	16	15	0	0	0	-34.5	0
226	**PSAMS**	**TTGCGCGTCAAACCATTCCTG**	14	13	6	1	7	-42.4	0
930	PSAMS	TCGGACCATTGAGTTTCACAT	16	16	4	4	8	-36.5	0
931	PSAMS	TTCGGACCATTGAGTTTCCTT	17	16	0	0	0	-34.6	0
1828	PSAMS	TAGAGGGAATTCACCGCTCTG	14	13	6	6	12	-38.6	-5.6
1831	**PSAMS**	**TCGTAGAGGGAATTCACCCCA**	14	14	4	4	8	-39.1	-3.8
1870	PSAMS	TATGTTTTCAGTTTAGCACCC	15	18	9	9	18	-28.7	-0.9
1905	PSAMS	TTTTCTTAGTGAGATTGACCA	16	13	8	11	19	-30.2	-0.7
2150	PSAMS	TACACTCACGTGTTGATCCAT	18	18	2	3	5	-34.2	-2.6
2159	PSAMS	TAGAGCTTCAACACTCACCTC	18	14	5	4	9	-36.6	-1.2

Table 5.2 Metrics used for quality control filtering of amiRNAs and reverse complements of their target sites (-RC suffix). Selected amiRNAs shown in bold. Table shows the start position of amiRNA target site in MCMV genomes, longest blast hits to EST and Refseq databases, targets predicted by the psRNA (Plant small RNA target analysis server) against cds and NSF maize databases. amiRNA-target hybridisation energy shown under Hyb. energy, and finally the minimum free energy of self-complementarity.

Start pos.	amiRNA sequence	Longest blast hit, EST	Longest blast hit, Refseq	psRNA off targets, cds	psRNA off targets, NSF	Total psRNA targets	Hyb. energy, kcal/mol	Self-comp. mfe
745-RC	TAGAAGCGAACTGCTAGCCGA	15	14	0	4	4	-43.9	-1.6
745	**TAGAAGCGAACTGCTACCCTA**	14	14	0	0	0	-33.8	-1.6
2151-RC	**TAACACTCACGTGTTGATCAA**	15	14	0	0	0	-39	-4.4
2151	TAACACTCACGTGTTGGTCAA	14	14	0	0	0	-38.4	-4.4
2474-RC	**TCAATGTCTGTACAGTTCCTG**	15	15	4	3	7	-40.5	0
2474	TCAATGTCTGTACAGTCCCTA	16	15	0	0	0	-34.5	0
2220-RC	**TCAATTGTGTGAGAAGCTCTG**	0	13	3	5	8	-38.5	0
2220	TCAATTGTGTGAGACGCTCTG	16	13	2	5	7	-34.4	-0.9
1830-RC	TGTAGAGGGAATTCACCGCTT	14	14	8	10	18	-43.3	-2.3
1830	**TGTAGAGGGAATTTATCGCTT**	14	13	2	2	4	-38.5	-2.4
226-RC	**TTGCGCGTCAAACCATTCCTC**	14	13	6	1	7	-43.7	0
226	TTGCGCGTCAAACCATTCCTG	14	13	6	1	7	-42.4	0
1831-RC	**TCGTAGAGGGAATTCACCGCT**	14	13	6	6	12	-45.8	-1.7
1831	TCGTAGAGGGAATTCACCCCA	14	14	4	4	8	-39.1	-3.8

5.5.8 Construct and amiRNA* design

I then assembled the final amiRNAs into the rice *miR395* backbone, placing amiRNAs with the lowest number of predicted off-targets in the positions shown to have higher expression when expressed in wheat (fig. 5.9) (Fahim et al., 2012). Structural RNA features are likely to influence miRNA processing in plants. At the time of construct design, the main known determinant of efficient processing to mature miRNAs was the presence of a mismatched bulge in the pri-miRNA stem loop, about 15nt below the miRNA sequence, and I observe these loops in the predicted secondary structure of the rice *miR395* cluster transcript (fig. 5.10). However, in case of unknown structural signals, I preserved mismatches and bulges found in the miRNA-miRNA* duplexes of the wild-type rice *miR395* transcript, in order to produce stem-loops and overall transcripts with the same predicted structure as wild-type (figs. 5.10 and 5.11). The amiRNA sequences and design process for the two test constructs, LABami08 and LABami09, can be seen in table 5.3. I also generated a negative control construct by randomising the amiRNA sequences of LABami08, and testing them using the same quality control requirements as above, repeating the randomisation until the control sequence passed quality control. amiRNA* sequences for the negative control were designed as described above. The amiRNA sequences for all constructs can be seen in table 5.3.

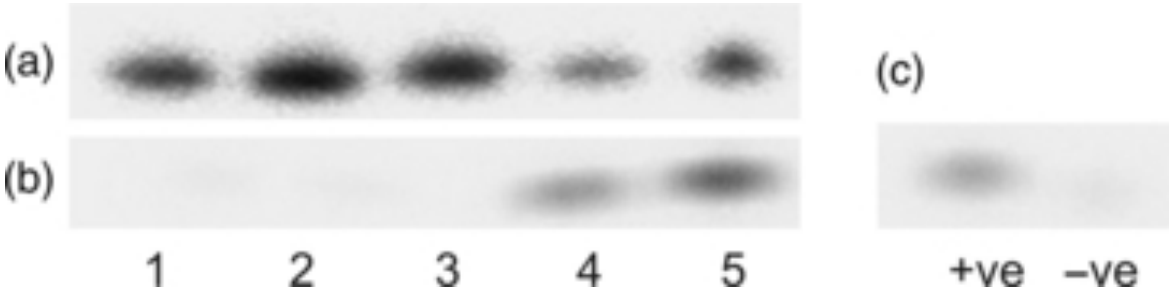


Fig. 5.9 Different locations in the *miR395* transcript are processed with different efficiencies. amiRNA (A) and amiRNA* (B) processing, with positive and negative controls (C), detected using MiRtect IT splinted ligation assay. Equivalent positions shown in figs. 5.10 and 5.11. Figure reproduced from Fahim et al. (2012).

Table 5.3 amiRNA sequences and design process for final maize chlorotic mottle virus-resistance and negative control constructs. Loop numbers are as in figs. 5.9 and 5.10. Revcom = reverse complement.

Loop	Construct	Design	Sequence
1	LABami08	WMD3	UAGAAGCGAACUGCUACCCUA
	LABami09	Target revcom	UAGAAGCGAACUGCUACCCUA
	LABamineg	Random	UUGCCGCGAUAACAUAAGCCA
2	LABami08	Target revcom	UAACACUCACGUGUUGAUCAA
	LABami09	Target revcom	UAACACUCACGUGUUGAUCAA
	LABamineg	Random	UGAGUCUCGUUAAUCCACAAA
3	LABami08	Target revcom	UCAAUGUCUGUACAGUUCCUG
	LABami09	Target revcom	UCAAUGUCUGUACAGUUCCUG
	LABamineg	Random	UGAUUUCGUCUCUACAUCGAG
4	LABami08	Target revcom	UCAAUUGUGUGAGAAGCUCUG
	LABami09	Target revcom	UUGCGCGUCAAACCAUUCCUC
	LABamineg	Random	UCGAAGUCGCUGAAUUAUUGG
5	LABami08	WMD3	UGUAGAGGGAAUUUAUCGCUU
	LABami09	Target revcom	UCGUAGAGGGAAUUCACCGCU
	LABamineg	Random	UUUAAUGCGCGGUUAGAAGUU

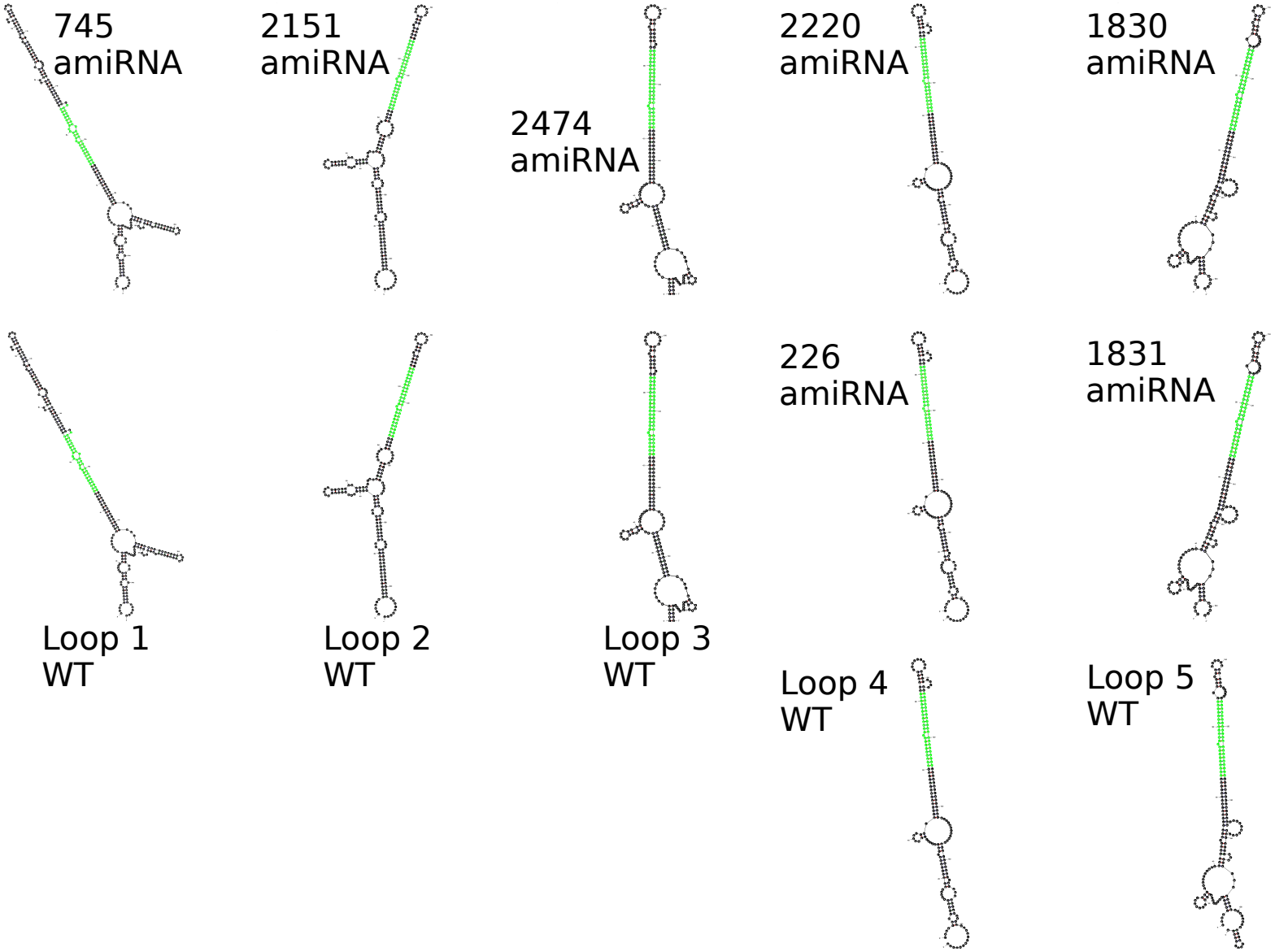


Fig. 5.10 amiRNA* sequences chosen preserve wild-type secondary structures. Secondary structures of wild-type and amiRNA loops, as predicted by mfold (Zuker, 2003). Labeling of loops is the same as fig. 5.9, and loops in the same column occupy the same position in constructs. amiRNA and amiRNA* sequences highlighted in green.

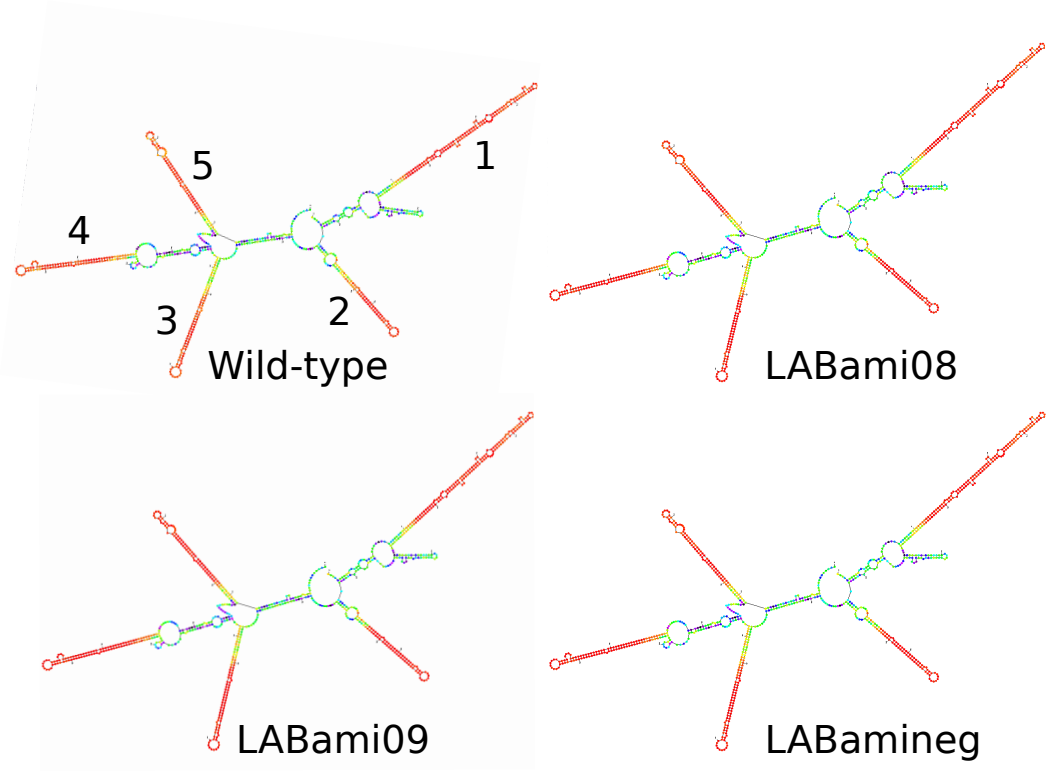


Fig. 5.11 amiRNA constructs preserve wild-type secondary structure. Secondary structures of wild-type rice *miR395* transcript, LABami08 and LABami09 amiRNA construcs, and LABamineg negative control construct, as predicted by mfold (Zuker, 2003). Loops from fig. 5.9 indicated on wild-type. Base colours indicate confidence, from red (high confidence of correct structure) to purple/black (low confidence).

5.6 Transient construct expression in *N. benthamiana*

Maize transformation is a time-consuming protocol, taking approximately 6 months from start to finish. Before committing to this, I wanted to test the processing of the LABami constructs. To do this, I transiently expressed each LABami construct in *N. benthamiana*, then performed sRNA Northerns to examine sRNA processing and accumulation. This demonstrated that the majority of sRNAs are successfully processed, although 2151 and 1830 were present at extremely low concentration (fig. 5.12). Additionally, amiRNAs 1831 and 226 were processed into what appears to be 21, 22, and 24bp sRNAs (fig. 5.12c and d). The majority of sRNAs were processed successfully, and I expect that processing will be more efficient in maize, as the backbone (rice) and maize are both monocot, whereas *N. benthamiana* is a dicot. Therefore, I decided to proceed to maize transformation with the current constructs.

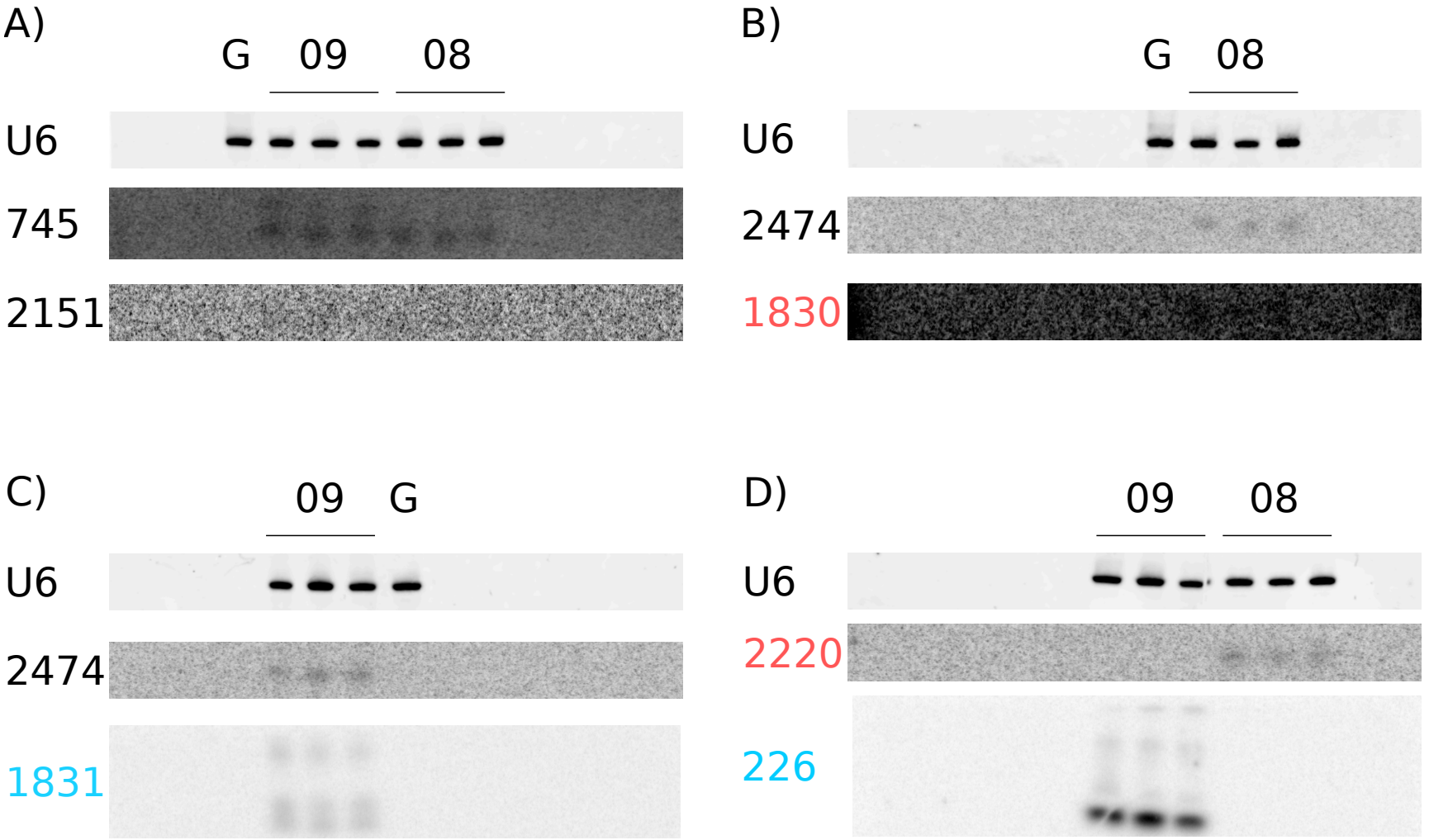


Fig. 5.12 LABami constructs are processed into sRNAs in *N. benthamiana* leaves. sRNA Northerns, with four membranes (A-D) probed for sRNA expression. Fluorophore tagged DNA oligos were used to label *N. benthamiana* U6 small nucleolar RNA as a loading control. The top amiRNA panel for each gel was exposed to a phosphorimaging screen for 60 hours, the lower panel for 120 hours. Contrast enhanced for 2151 and 1830 amiRNAs, others as recorded. amiRNAs labelled in black are present in LABami08 and LABami09, those in red are only in LABami08, and those in blue only in LABami09. G=GFP expression as negative control.

5.7 Tropical maize transformation

I wanted to transform tropical maize in order to generate MCMV-resistant maize lines which could be suitable for use in East Africa. There are multiple methods for maize transformation, the most common being particle bombardment, and *Agrobacterium tumefaciens* mediated transformation of immature embryos. I decided to use the latter as it is the most likely to produce stable, low copy number transgenic maize lines (Zhang et al., 2005). To perform *A. tumefaciens* mediated transformation of tropical maize lines I collaborated with colleagues at KU (Kenya). We selected two tropical lines with established transformation protocols at KU: CML144 and Namba nane.

To control for tissue culture, we decided to transform each line with LABami08, LABami09, and LABamineg. LABamineg will act as an amiRNA negative control and tissue culture control i.e. to demonstrate that neither tissue culture nor randomised amiRNAs generate MCMV resistance. We also included non-transformed embryos as a control for the antibiotic selection step, which showed that 0% of non-transformed embryos passed through basta selection successfully. This process produced 13 LABami08 lines (1 CML144, 13 Namba nane), 17 LABami09 lines (3 CML144, 14 Namba nane), and 7 LABamineg lines (all Namba nane) (table 5.4). Transformation efficiencies ranged from 0.3-3.2%, which is low for *Agrobacterium*-mediated transformation, and was caused by the loss of approximately 66% of rooted explants due to asynchronous production of ears and pollen (tables 5.4 and 5.5).

Table 5.4 Tropical maize transformation, showing numbers at each stage. JM performed maize transformation at Kenyatta University. EMM=Embryo maturation medium, RM=Regeneration medium, embryo.=embryogenic.

Line	Construct	Embryos infected	Calli at start, selection 1	Calli past selection 1	Calli past selection 2	Embryo. calli after EMM	Calli with shoots after RM	T1 seed
CML144	LABami08	330	234	48	38	31	20	1
CML144	LABami09	179	138	43	32	22	11	3
Namba nane	LABami08	401	354	81	71	45	30	12
Namba nane	LABami09	469	400	177	70	51	37	14
Namba nane	LABamineg	217	178	84	37	17	11	7

Table 5.5 Efficiency of tropical maize transformation. Percent selection is the proportion of calli passing second selection compared to those introduced to first selection. Percent regeneration is the proportion of embryogenic calli which generated shoots. Transformation efficiency is taken as the proportion of embryos infected that produced T1 seed.

Line	Construct	Percent selection	Regeneration %	Transformation efficiency
CML144	LABami08	16.24	64.52	0.30
CML144	LABami09	23.19	50.00	1.68
Namba nane	LABami08	20.06	66.67	2.99
Namba nane	LABami09	17.50	72.55	2.99
Namba nane	LABamineg	20.79	64.71	3.23

5.8 Characterisation of T0 maize transformants

5.8.1 PCR of transgene using T0 maize DNA

In order to confirm transgene presence within the genome of T0 transformants, and to confirm the correct sequence of transgenes, I decided to perform PCR on the T0 DNA. PCR confirmed the presence of the basta resistance gene in five lines of transformants. Other lines were negative, despite optimisation (see methodology), and for future work I will be using a polymerase with higher processivity which may be more successful with the high complexity template of maize genomic DNA.

5.8.2 sRNA Northerns of T0 maize RNA

I performed an sRNA Northern to investigate processing of sRNAs from the LABami constructs in T0 plants. Initial probing with a mixture of P^{32} labeled DNA oligos complementary to the amiRNA sequences was unsuccessful (fig. 5.13). Therefore, to test whether the samples were intact enough to preserve miRNAs, I also probed for the highly expressed native miR159a, which showed that small RNAs were present in the samples (fig. 5.13). I conclude that the sRNAs are either not present, or present at a lower concentration than the highly expressed miR159a. Labeling of miR159a is still reasonably weak, so if LABami sRNA expression is lower than miR159a, protocol optimisation will be required to detect them.

5.8.3 Southern blotting of T0 maize DNA

I performed Southern blots in order to check for transgene presence, and determine the copy number in transgenic lines. I initially performed Southern blots with 10μg DNA per sample, with 50 pg of cut plasmid (pTF101[LABami09]) loaded as a positive control, probing with a P^{32} labeled 1 kb amplicon of the LABami09 construct (primers: pTFseq_R1 and LABami_probe_F1). With a 5411 bp transgene in a 2500Mb genome ($2.5x10^9$bp), loading 10μg of genomic DNA results in approximately 20 pg of transgene per sample. This blot resulted in only one sample with a visible band, and the positive control was not visible (fig. 5.14a). Therefore, I performed two more Southerns, again with 50 pg of positive control, but changed the washing conditions to be less stringent. 24 hour exposures of these blots were negative, so I exposed one for seven days instead. The amiRNA construct has extensive self-complementarity in order to form hairpins (fig. 5.11), which could interfere with the probe binding. Therefore, I probed the second gel with a 430 bp amplicon from the basta resistance gene (primers: bar_F1+R1) present for selection of transformants, and exposed

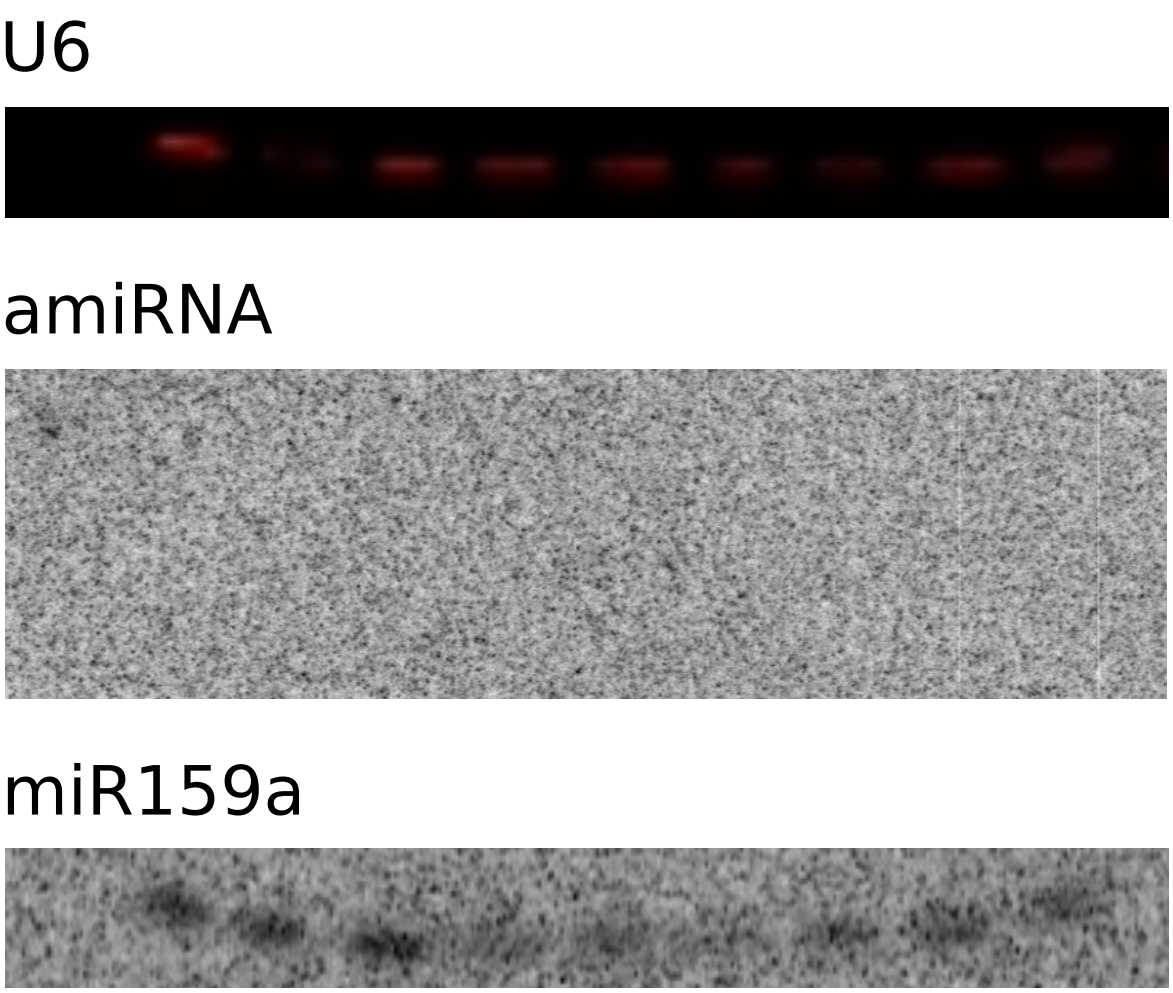


Fig. 5.13 sRNAs were not visible in initial sRNA blot of T0 RNA. Fluorophore tagged DNA oligo was used to probe U6 snRNA as a loading control. amiRNA were probed with a mix of complementary DNA oligos, and miR159a with a complementary oligo, both end labeled with P^{32} using T4 PNK.

for one week. Both blots were completely negative, with the exception of line 17 from the LABami09 probed gel (fig. 5.14c). The visible DNA size ladder is a known phenomenon with NEB DNA markers, and may represent non-specific binding - this is seen in blots probed with LABami constructs (see fig. 5.15a for blot without LABami09 probing and ladder staining).

To test my protocol, I ran a Southern with 10μg of wild-type maize gDNA mixed with either 100x (4 ng) and 500x (20 ng) the predicted transgene concentration. First, I probed with a labeled 26S rRNA amplicon (primers: Zm_23S_F1+R1) which was positive, but subsequent LABami09 probing, this time hybridising under a temperature gradient in an attempt to minimise the effects of secondary structure formation, was negative (fig. 5.15a

and b). The wild-type gDNA used was harvested, stored, and extracted in the same way as transformed lines, and the rRNA gene staining does suggest moderate degradation of the sample, as evidenced by the smear (fig. 5.15a). Next, I mixed gDNA with 50 ng, 5 ng, or 0.5 ng of cut plasmid, and loaded it together with a concentration gradient of uncut plasmid. This loading was repeated across the gel, the blot cut in half, and each half hybridised either in Ultra-Hyb buffer (which I had used up to this point), or modified Gilbert-Church buffer. After hybridisation with the LABami09 probe, I could see that hybridisation with UltraHyb buffer was more efficient than with modified Gilbert-Church, and that staining of the 0.5 ng band was very weak (fig. 5.15c and d). 0.5 ng is approximately 25x higher concentration than would be expected of the transgene, suggesting that I will need to either load more gDNA or refine my blotting protocol to get definitive experimental results.

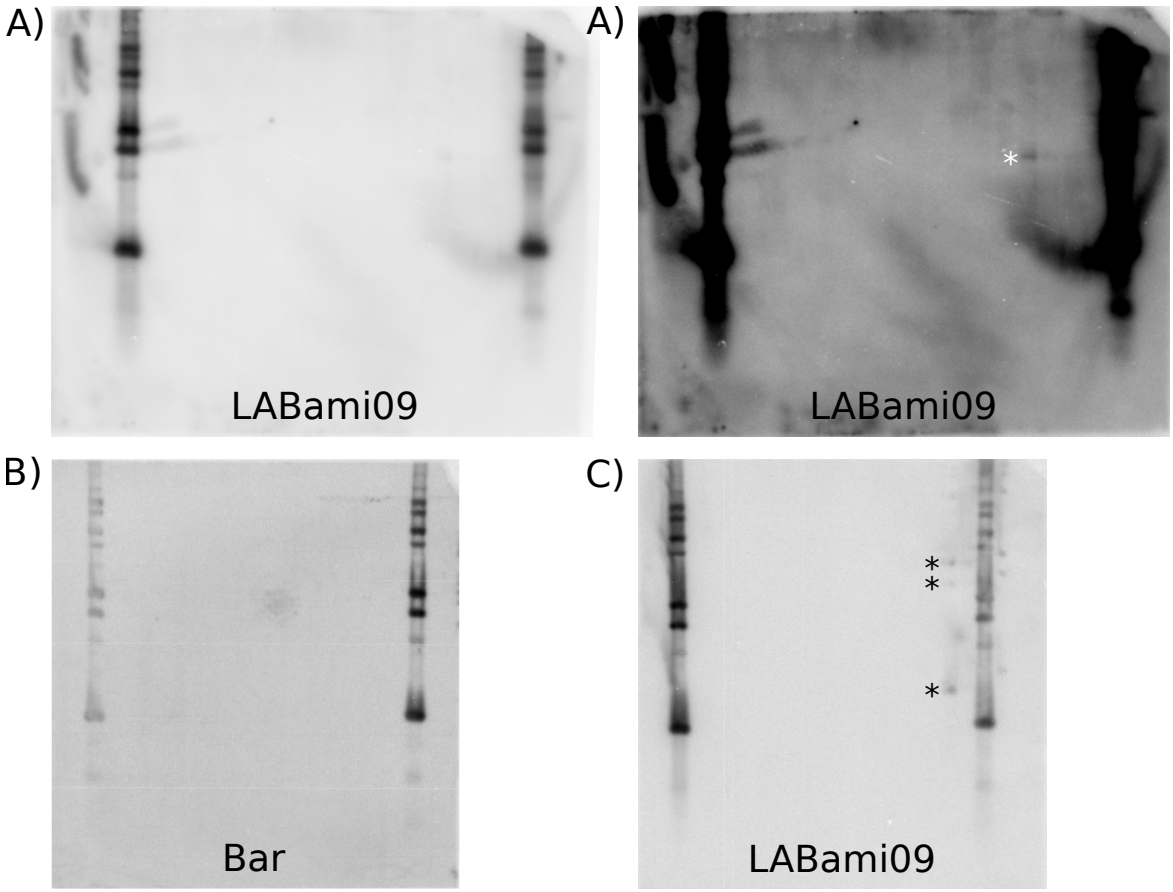


Fig. 5.14 Southern blots of T0 maize transformant DNA samples are negative. Probe identities shown on blots. A) Initial blot, with contrast boosted on right and positive shown by asterisk, exposed for 48 hours. B) Second blot, exposed for seven days with Basta probe. C) Third blot, exposed for seven days with LABami09 probe. Asterisks correspond to bands at 4000, 3200, and 1200bp.

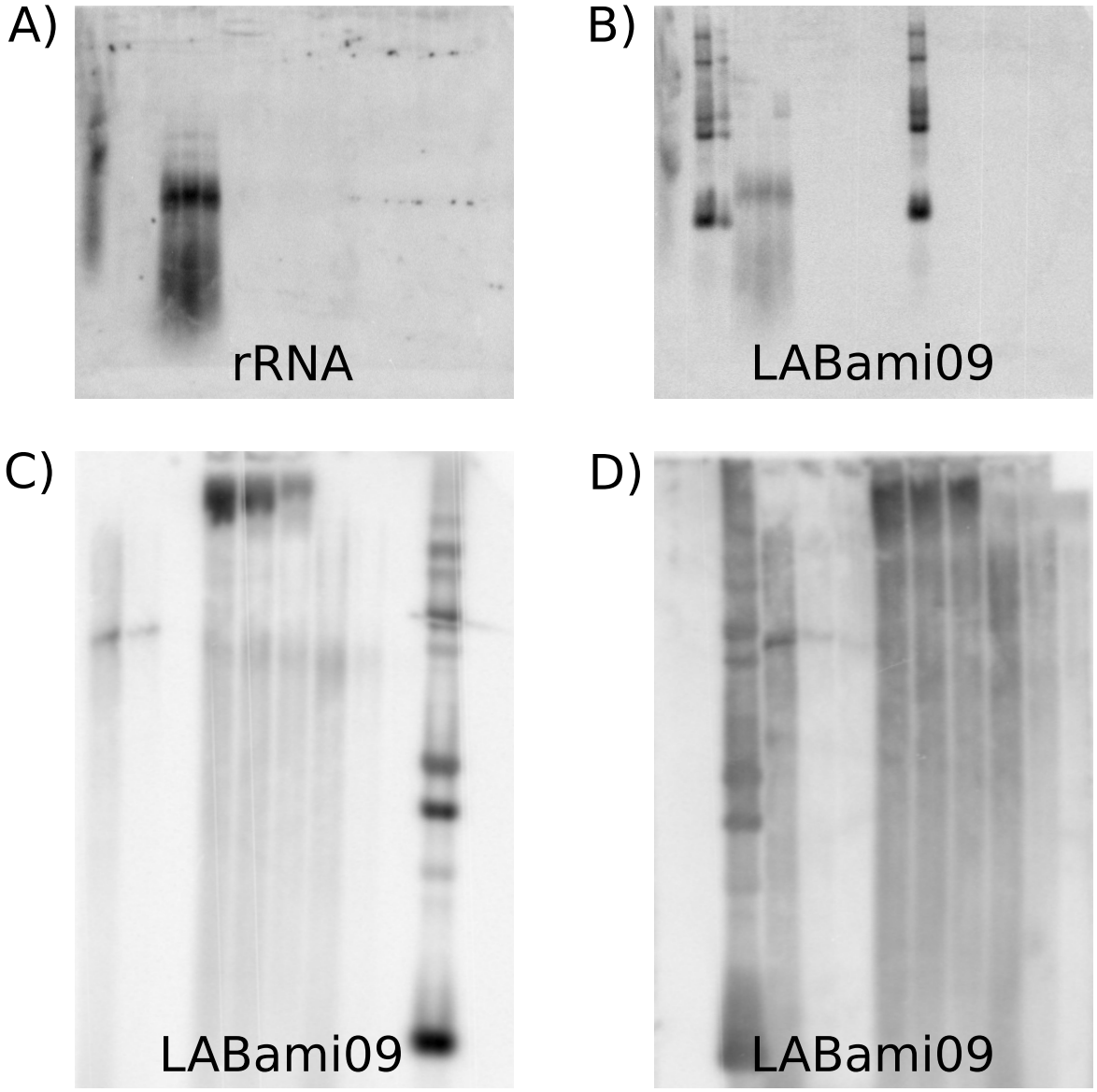


Fig. 5.15 Investigating sensitivity of Southern blots. Probe identities shown on blots. rRNA probe is 1000bp fragment of 26S. A) Maize gDNA mixed with 50, 5, or 0.5ng of LABami09 cut plasmid (left to right). B) As A) but with LABami09 probing, exposed for one week.C) Blot with, from left to right, 50, 5, 0.5ng of cut LABami09 plasmid with 10ug of maize gDNA, undetectable at 0.5ng. Then uncut plasmid at 750, 500, 250ng, and cut plasmid at 50, 5, 0.5ng without gDNA. Hybridised in modified Gilbert-Church buffer. D) As for C), but hybridised in Ultra-Hyb buffer.

Discussion

5.9 amiRNA design

5.9.1 Viral quasispecies

RNA viruses are often described as quasispecies: mutant clouds generated by mutations introduced by an error-prone RdRP (typically $\geqslant$1 error per genome replication), and selection pressures acting on viral genomes. I will use a modern definition: a group of closely related viral genomes, subject to the continuous processes of genetic variation, competition amongst variants, and selection of the most fit distributions (Domingo et al., 2012). I will highlight two features of quasispecies, before discussing how this relates to amiRNA design. Firstly, there are positive and negative interactions between genome variants. Positive interactions are demonstrated by observations that viral clones with a given consensus sequence can be less fit than assemblages with the same consensus which include mutants (Domingo et al., 1978; Vignuzzi et al., 2006). Negative interactions can be seen in the failure of fitter haplotypes to rise to dominance unless above a concentration threshold (Carlos et al., 1990; Martínez et al., 2012). Secondly, the diversity present in a viral population is determined by the interaction between mutation and selection, but the force of selection will depend on the mutational robustness of the virus. Mutational robustness can be seen as the ability of a genetic sequence to be mutated without large changes in fitness, which is vital for survival in RNA viruses, where genetic sequences are shortest and mutation rates highest. Figure 5.16 shows two different viral populations, one with a high fitness peak, the other with a low, flat peak. Inevitable mutation ensures that at a population level, the population with the flatter fitness landscape, and higher mutational robustness, is fittest. This concept is known as survival of the flattest.

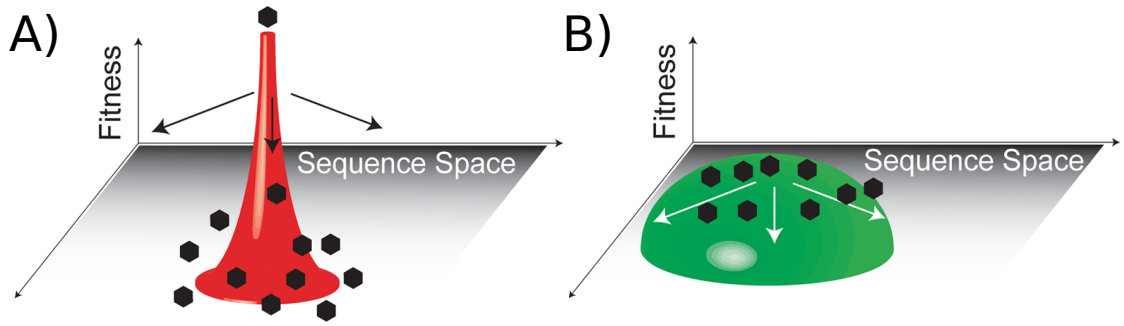


Fig. 5.16 Alternative viral fitness landscapes, with A) high peak, and lower average fitness or B) flat, low peak, and higher average fitness. Adapted from Lauring and Andino (2010).

5.9.2 Viral quasispecies and amiRNA design

NGS has made it possible to examine a large number of low frequency variants within multiple viral samples. Sequence-mediated resistance functions through Watson-Crick base-pairing, and therefore provides a strong selection pressure for mutations within the target sequence. I investigated intra-sample variation in MCMV, and found that peaks of heterozygosity across the genome were replicated in all or almost all samples (fig. 5.4c). This is despite high mutation rates, high multiplicity of infection, and the population bottlenecks of transmission events, over an area of hundreds of miles squared and a timespan of at least five years in East Africa (Scheets, 2000; Wangai et al., 2012). Therefore, taking into account the location of these fairly stable variability peaks during the design process may be expected to improve the robustness of resistance in the face of MCMV evolution. These peaks could have a number of explanations: positive interactions between genome variants, memory of previously dominant variants, or they could represent sites of the genome at which the fitness landscape is flatter. An initial investigation into what causes these peaks could examine whether heterozygosity peaks are consistent in a wider variety of MCMV isolates, and in samples inoculated with homogeneous IVT clones of MCMV. It would also be interesting to investigate whether consistent variability peaks are observed in other RNA viruses. Regardless of the explanation, avoiding these peaks should improve resistance durability, and quasispecies dynamics may prove beneficial - given that positive selection has been insufficient to cause selective sweeps for fitter haplotypes while they are at low proportions, including one example where this selection is generated by an amiRNA (Carlos et al., 1990; Martínez et al., 2012). A model system, such as passaging TuMV through *N. benthamiana* would be ideal for measuring whether avoiding conserved intra-sample variability maxima increases the number of passages required for resistance breaking.

5.9.3 amiRNA design

In addition to heterozygosity, I used a number of other factors known to be important in amiRNA efficiency to select the final amiRNAs: target-hybridisation energy, and low self-complementarity. Additionally, to minimise the probability of amiRNAs modifying the host transcriptome, and thereby modifying the phenotype of uninfected maize plants, I filtered for off-targets. All of these factors are important, but the relative importance of each for overall amiRNA efficiency is unknown. My approach was to examine the distribution of each factor, and try to pick amiRNAs that avoided large values in any of them. However, this approach would be improved by a quantitative understanding of the contribution of each factor. To begin with, a model system (for example *A. thaliana*) containing a reporter

with a non-functional terminal extension sequence targeted by an amiRNA could be used. Target and amiRNA sequence could then be altered to test the effect on reporter intensity of varied levels of binding energy and self-complementarity, along the length of the amiRNA (as different sites have varying importance for targeting). These findings could then be validated in relevant crop plants, and then the value of avoiding viral variability investigated. Mathematical modeling of viral escape mutation probabilities could also be helpful in this regard, however linked experimental work would likely be required to parameterise it effectively - e.g. genome replication is inhibited by what factor at different levels of expression of the same amiRNA.

5.10 Maize transformation and characterisation

Maize transformation generated 1 (LABami08) and 3 (LABami09) CML144 amiRNA lines, and 10+ lines for each amiRNA construct in Namba nane. A good number of transformed lines is desirable as the chromatin structure of the insertion site, or transgene silencing, may limit or prevent expression. The inclusion of the *Ubi1* promoter with intron may decrease the probability of transgene silencing (Christie et al., 2011). Unexpectedly, asymmetric production of tassels and ears meant that 72 lines were lost, as unfortunately there were no wild-type lines at reproductive age at KU to make heterozygous crosses. Tassels emerged one to two weeks before ears with silks were present. Reported causes of asynchrony in pollen production and silk receptivity are tissue culture, and drought, with the observed extreme asynchrony making tissue culture effects more likely (Bohorova et al., 1999). This effect has not previously been observed at KU, using the same protocol and lines, but heat stress was also high due to the weather during T0 growth for these transformants.

The aim of molecular characterisation was to identify maize lines with single or low-copy transgene insertion and high expression of sRNAs, which could be taken forward to contained resistance testing. Genomic PCR of transgenes had only partial success, despite optimisation of annealing temperature and template concentration for multiple primer pairs. Maize gDNA is a high complexity template, and my next approach will be to utilise a polymerase with higher processivity such as TaKaRa Ex Taq or MilliporeSigma KOD Xtreme. Detection of single copy genes using comparable loading of maize gDNA to mine is reported in the literature, although detailed methodologies are rare, and I will be using techniques from reported protocols in optimisation (Della Vedova et al., 2005; Dong et al., 2015). An alternative would be to load far more material, as in Ishida et al. (1996), but material is limiting. Overall, molecular characterisation of maize transformants needs optimisation to provide definitive results. I am confident that there should be transformants containing the

transgene, as the selection controls which were not transformed had no escapes from basta selection, and the *LABami* genes are present beside the right border (which is inserted into gDNA first) in my constructs, so lines with basta resistance should also contain the transgene.

Conclusions

NGS data is very valuable for identifying the causes of emerging pathogen outbreaks due to the small amount of pre-required knowledge. Additionally, this same data can be used to assess the most viable target for sequence mediated resistance (in this case, MCMV), and the most conserved genomic regions for targeting. I have found that intra-sample MCMV variability is consistent between samples over a wide area. This demonstrates that it is feasible to use low frequency viral variants as part of a novel design pipeline for robust sequence-mediated resistance. We generated over ten transformed lines per resistance construct, which will be characterised at the molecular level, and the most promising lines taken on to functional resistance testing.

Acknowledgements

The work in this chapter included the work of Joel Masanga, who individually performed the tropical maize transformation work, prior to characterisation. I would also like to thank Steven Runo for his help in setting up our collaboration and welcoming me to his lab.

Chapter 6

Investigation of RNA silencing suppression by maize chlorotic mottle virus

6.1 Summary and objectives

Both natural and engineered sequence-mediated immunity rely on the RNA silencing system to degrade the virus's genetic material. However, many viruses produce viral suppressors of silencing (VSRs) to combat the antiviral role of RNA silencing. VSRs could interact with efforts to engineer resistance, and additionally, VSRs are often the basis for synergistic viral interactions and could be the basis of MLN (Syller, 2012). There are a number of VSRs in the *Tombusviridae* with independent evolutionary origins, and no known VSRs in the MCMV genome. Therefore I investigated local and systemic VSR activity of MCMV ORFs by coinfiltration experiments using *N. benthamiana*. There was no clear evidence for local or systemic VSR activity, although the work on systemic VSRs is incomplete due to difficulties with reproducibility. While investigating this variability, I discovered that long day conditions (in my growth conditions) prevent the spread of systemic silencing, revealing a novel link between photoperiod and systemic silencing.

Methodology

6.2 Cloning

To generate a Gateway library of MCMV ORFs I amplified sequences from the cDNA clone pMCM41 using Phusion polymerase (Scheets et al., 1993). To remove unwanted start codons (for other ORFs) and stop codons (for readthrough proteins) I introduced errors using mismatched primers. To join modified sequences as necessary I used OE-PCR (Heckman and Pease, 2007). I predicted protein structures with RaptorX (http://raptorx.uchicago.edu/) to inform the choice replacement amino acids for the stop codon in readthrough proteins (Källberg et al., 2012). All ORFs were cloned into pENTR by D-TOPO cloning, transformed into *E. coli* strain *DH5*α, and inserts checked with diagnostic digests and Sanger sequencing. For VSR assays, I moved MCMV ORFs into the destination vector pGWB402Ω using the Gateway LR reaction and checked plasmids using diagnostic digests.

6.3 Local silencing suppressor assays

To investigate the local VSR activity of MCMV ORFs I used coinfiltration of *N. benthamiana* plants. Plant growth conditions were 22ºC, 65% humidity, 160μmol, 16 hours light, and I inoculated 21 days after germination. For inoculum, I grew *A. tumefaciens* strains containing plasmids for GFP or MCMV ORF expression under antibiotic selection in LB for 48 hours, then resuspended in infiltration buffer and adjusted the OD^{600} to 0.5. For each condition, I mixed an ORF strain with the GFP strain 50:50, then infiltrated the two youngest expanded leaves. Test mixtures were inoculated to the left of the leaf axis (tip upwards), and control mixtures to the right. I observed that younger leaves tend to have higher GFP expression so I used the youngest expanded leaf for the comparison between the test ORF and negative control. To image fluorescence, I illuminated plants using a UVP Blak-Ray B100-AP UV lamp, and imaged them using a Nikon D7100.

6.4 Systemic silencing suppressor assays

To test systemic silencing, I grew *N. benthamiana* plants in 22ºC, 60% humidity, 160μmol, 8 hours light, and inoculated at the four leaf stage, inoculating the first two true leaves. I prepared inoculum as described for local VSR assays, but adjusted OD^{600} to 1. Plants were imaged as described above. To quantify the area of leaves silenced, I coloured silenced leaf patches in GIMP, then converted images to binary and measuring the silenced area in FIJI.

Results

6.5 Investigating local silencing suppressors in maize chlorotic mottle virus

6.5.1 Silencing suppressors in the *Tombusviridae*

Due to the importance of RNA silencing in antiviral plant defence, most, or maybe all viruses, have evolved proteins which interfere with these pathways in order to inhibit the degradation of their own genomes. The *Tombusviridae* family contains a number of viral suppressors of RNA silencing (VSRs), which are summarised in table 6.1. There appear to be at least three independently evolved forms of VSR in the *Tombusviridae*: suppressor proteins (P19, P14), coat proteins (P38, P37), and polymerases (RdRP) (fig. 6.1). MCMV ORFs are all modified when compared to the known *Tombusviridae* VSRs - MCMV RdRP has a unique 5' extension, the reading frame with homology to P19 (P31) has been proposed to be expressed as a readthrough of a movement protein, and its coat protein lacks the protruding (P) domain of the *Carmovirus* coat protein. Therefore, given the interesting range of diversity in VSR ancestry and action across the *Tombusviridae*, I decided to search for VSRs in the MCMV genome.

6.5.2 Construction of a Gateway library of maize chlorotic mottle virus open-reading frames

In order to express MCMV ORFs in order to test their VSR activity, first I needed to clone them. I decided to clone them into Gateway entry vectors to facilitate rapid tagging with epitopes or N/C-terminal proteins in case of downstream analysis. ORFs were cloned with stop codons to prevent readthrough, and without, to facilitate C-terminal tagging. Cloning was complicated by the presence of overlapping ORFs and readthrough proteins in the MCMV genome (fig. 6.1). Therefore, I used mismatched primers to introduce mutations in order to remove unwanted translational start sites (ATGs to ACGs), whilst keeping the amino acid sequence of the desired ORF unchanged. Readthrough expression of the larger RdRP component (P111 in MCMV) is conserved across the *Tombusviridae*, but the mechanism for this readthrough is unknown. I decided to substitute the stop codon with an amino acid, and to used predicted protein structure to inform my choice. I used RaptorX to predict the structure of P111, and the readthrough site was predicted to lie within an unstructured linker region between two domains (the pre- and post- readthrough coding sequence) (fig. 6.2a). Therefore,

I replaced the stop codon with either alanine (small, neutral) or glutamine (small, hydrophilic) to minimise the probability of altering P111 function. P31 has been proposed to be expressed through readthrough of the movement protein P7a, again with an unknown mechanism. I submitted P31 to RaptorX and found that the readthrough site again corresponded to an unstructured region (fig. 6.2), and replaced it with a glycine (small, hydrophobic). In case of future work aiming to establish the roles of the pre- and post-stop codon fragments of readthrough proteins, I also cloned the 3' post-readthrough portion of P111 and P31 with a start codon in place of the readthrough stop codon. Proteins with 3' readthrough extensions (P50 and P7a) had their stop codon switched for alternatives to minimise the probability of unwanted readthrough. Mutations were introduced at the 5' and 3' of PCR fragments covering the remainder of the coding sequence, and these fragments joined together using overlap-extension PCR (fig. 6.2c), (Heckman and Pease, 2007). The library of ORFs cloned into entry vectors can be seen in table 6.2.

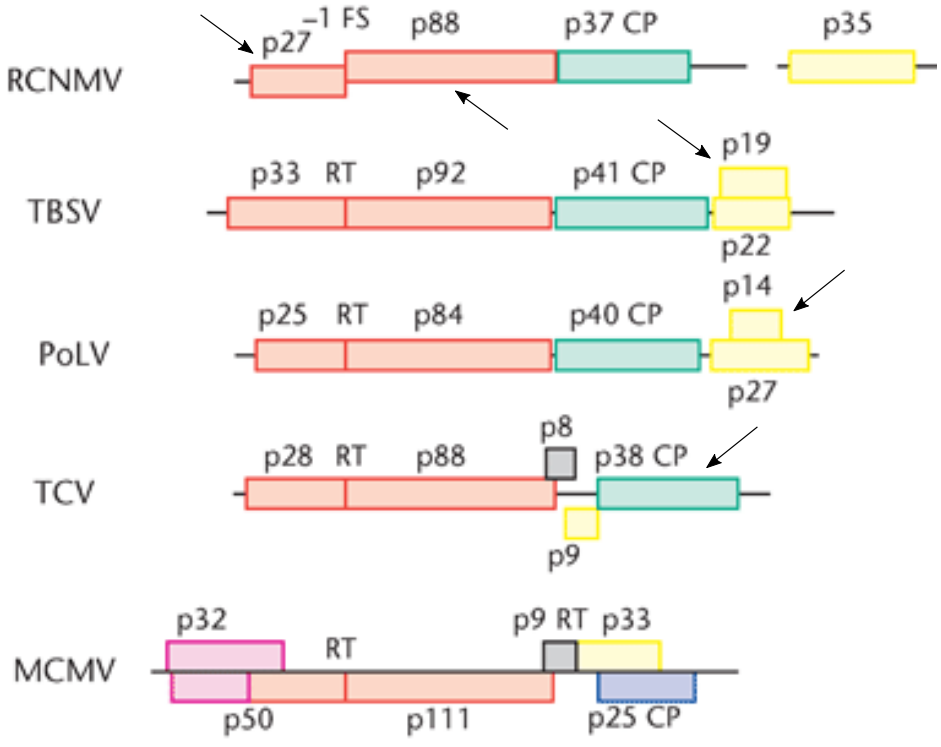


Fig. 6.1 There are a variety of known silencing suppressors in the *Tombusviridae*. VSRs indicated with arrows. Colour in ORFs indicates homology. Figure adapted from Sit and Lommel (2010). RCNMV=red clover necrotic mottle virus, TBSV=tomato bushy stunt virus, PolV=pothos latent virus, TCV=turnip crinkle virus, MCMV=maize chlorotic mottle virus.

Table 6.1 Viral suppressors of silencing reported across the *Tombusviridae*, highlighting the diversity of ancestry and action. TBSV=tomato bushy stunt virus, CymRSV=cymbidium ringspot virus, TCV=turnip crinkle virus, HCRSV=hibiscus chlorotic ringspot virus, PFBV=pelargonium flower break virus, PLPV=pelargonium line pattern virus, PoLV=pothos latent virus, RCNMV=red clover necrotic mottle virus.

VSR	Virus	Genus	Notes	References
P19	TBSV,	*Tombusvirus*	Short dsRNA duplex binding.	Scholthof (2006)
P19	CymRSV	*Tombusvirus*	CymRSV has been shown to preferentially bind vsiRNA duplexes over miRNA in infected plants.	Kontra et al. (2016)
P38	TCV	*Carmovirus*	Binds both long and short dsRNA. Reported to inhibit both dsRNA processing into sRNAs, and sRNA loading into AGO1.	Iki et al. (2017); Mérai et al. (2006)
P38	HCRSV	*Carmovirus*	Unlike TCV P38 can only suppress ssRNA induced silencing, not dsRNA.	Meng et al. (2006)
P37	PFBV	*Carmovirus*	Binds siRNA, doesn't interfere with siRNA formation (unlike TCV P38).	Martínez-Turiño and Hernández (2009)
P37	PLPV	*Carmovirus*	Short dsRNA duplex binding.	Perez-Canamas and Hernandez (2015)
P14	PoLV	*Aureusvirus*	Binds both long and short dsRNA.	Mérai et al. (2005)
RdRP	RCNMV	*Dianthovirus*	RdRP subunits and RNA substrate must be present for VSR activity to occur.	Takeda et al. (2005)
MP	RCNMV	*Dianthovirus*	Reported, but I believe should be discounted. MP does not function in coinfiltration assays as a VSR, and the assay used measures viral movement. MP is a movement protein, and mutations which abolish its movement protein activity also abolish its 'VSR activity'.	Powers et al. (2008)

Table 6.2 Library of Gateway entry vectors containing MCMV ORFs, with interfering overlapping ORFs removed, and readthrough stop codons replaced to allow sole expression of readthrough proteins

Plasmid	Description
pLAE01	pENTR with MCMV P50 ORF (stop codon switched to dissuade readthrough)
pLAE02	pENTR with MCMV P50 ORF (no stop codon)
pLAE03	pENTR with second half of MCMV p111 ORF and start codon in place of normal read-through stop codon
pLAE04	pENTR with MCMV P7a ORF (with stop)
pLAE05	pENTR with second half of MCMV p31 and start codon in place of normal read-through stop codon
pLAE06	pENTR with MCMV coat protein ORF (with stop)
pLAE07	pENTR with full length MCMV from pMCM41
pLAE08	pENTR with MCMV P111 ORF with alanine substitution for read-through stop codon and and terminal stop
pLAE09	pENTR with MCMV P111 ORF with glutamine substitution for read-through stop codon and terminal stop
pLAE10	pENTR with MCMV P111 ORF with alanine substitution for read-through stop codon and no terminal stop
pLAE11	pENTR with MCMV P111 ORF with glutamine substitution for read-through stop codon and no terminal stop
pLAE12	pENTR with MCMV p31 ORF (with stop)
pLAE13	pENTR with MCMV p31 ORF (without stop)
pLAE14	pENTR with MCMV P7a ORF (without stop)
pLAE15	pENTR with MCMV coat protein ORF (without stop)
pLAE16	pENTR with MCMV P32 ORF (with stop)
pLAE17	pENTR with MCMV P32 ORF (without stop)

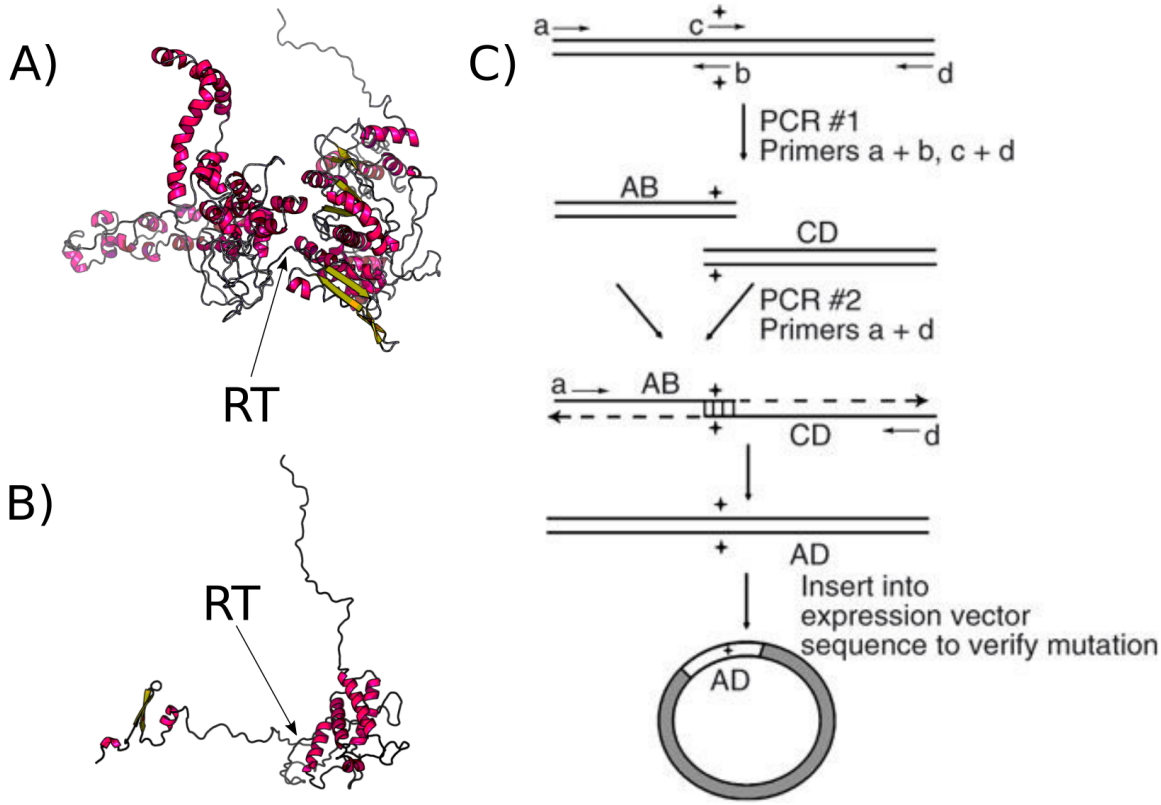


Fig. 6.2 Information on cloning of maize chlorotic mottle virus (MCMV) ORFs into Gateway library. Readthrough proteins had their stop codons replaced with small amino acids, as readthrough locations were predicted to be in unstructured regions for A) P111 and B) P31. C) shows how mismatched primers can be used to introduce mutations (+), and overlap-extension PCR used to connect mutated fragments. RT=readthrough. C) adapted from Heckman and Pease (2007).

6.5.3 Local silencing suppressor assays using *Nicotiana benthamiana*

To express MCMV ORFs transiently I used the LR recombination reaction to transfer each ORF (with stop codon) into pGWB402Ω, an improved binary Gateway destination vector, driving expression with a duplicated CaMV35S promoter and tobacco etch virus Ω translational enhancer (Nakagawa et al., 2007). I also transferred the β-glucuronidase gene into the same destination vector to act as a negative control. The assay I selected for measuring local VSR activity was coinfiltration of wild-type *N. benthamiana*. In this assay, leaves are coinfiltrated with two *A. tumefaciens* strains, with their T-DNAs containing either a GFP gene or a test (or control) ORF. Transient expression of the GFP occurs for 3-5 days, before transgene silencing is initiated and the level of GFP expression (and fluorescence) drops markedly. If the test ORF interferes with local RNA silencing, then GFP fluorescence will remain for a longer time period, as transgene silencing is not initiated efficiently. To

determine an appropriate sampling point for detection of silencing suppression using my experimental setup, I performed a timecourse of *N. benthamiana* coinfiltrated with GFP and either GUS (negative control) or one of two known silencing suppressors as potential positive controls: P19 (tomato bushy stunt virus) and P1b (cucumber vein yellowing ipomovirus) (Valli et al., 2006). This demonstrated that both P1b and P19 are suitable positive controls and that clear differences are visible between the negative and positive controls by 5 days post-inoculation (dpi) (fig. 6.3).

To test the VSR activity of individual MCMV ORFs I then coinfiltrated *N. benthamiana* plants with each MCMV ORF together with GFP, with GUS + GFP as a negative control, and P19 or P1b + GFP as a positive control. I found that younger leaves tend to have stronger GFP expression, and as comparison between the test ORF and negative control is the most important part of the assay, I used the youngest expanded leaf to compare test ORF with negative control, and the next youngest to compare test ORF with positive control. There was no observable difference in GFP fluorescence compared to the negative control for any single MCMV ORFs (figs. 6.4 and 6.5). Next, to test whether full-length MCMV is required for VSR activity, which could be due to a requirement for multiple proteins and/or viral genomic RNA, I coinfiltrated *N. benthamiana* with pLG07 (pGWB402Ω containing the full length MCMV transcript), but again could not observe a difference in GFP fluorescence between the test condition and negative control (fig. 6.5).

6.5.4 Testing combinations of maize chlorotic mottle virus ORFs for local silencing suppression

The *Tombusviridae* replication complex has two subunits, and both are required for local RNA silencing suppression by red clover necrotic mottle virus (Takeda et al., 2005). To investigate whether MCMV ORFs act in combination to suppress local RNA silencing, I performed coinfiltration with GFP, and combinations of MCMV ORFs. The components of the MCMV RdRP (P50 and P111) were expressed either with the unique P32 ORF, which has unknown function and promotes high viral titre, or the P31 ORF which is required for systemic movement of MCMV (Scheets, 2016). Additionally, I inoculated leaves with a mixture of all MCMV ORFs, with the intention of allowing all possible interactions to occur, although likely in different subsets of cells. I did not observe differences between any combination of MCMV ORF and the negative control (fig. 6.6).

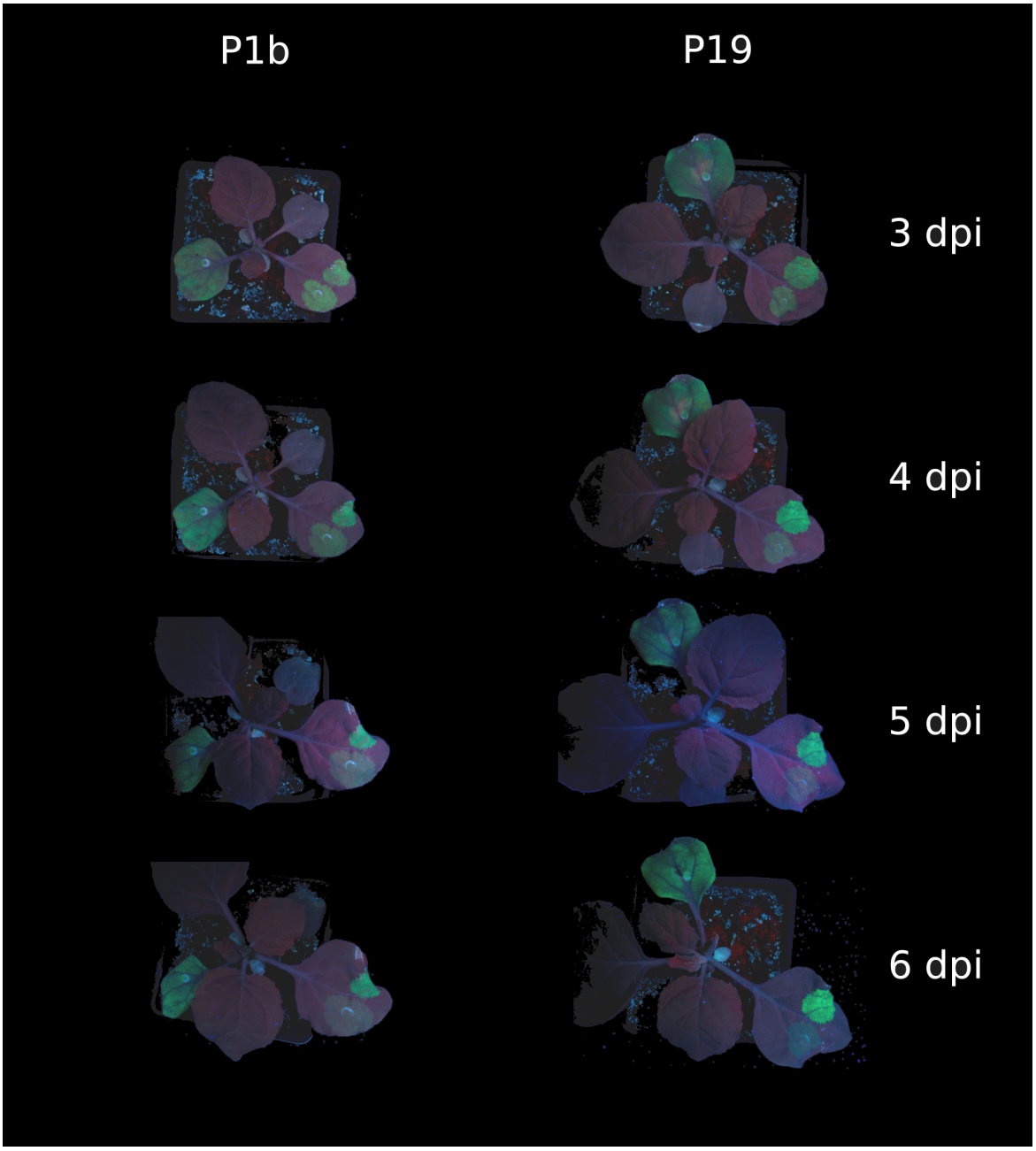


Fig. 6.3 Transgene silencing of GFP, with or without known VSRs in *N. benthamiana*. Youngest expanded leaf was inoculated with GFP + positive control (top patch) or GFP + negative control (GUS, bottom patch). Oldest leaf was infected with GFP + positive control. Clear differences were visible between positive and negative control by 5 days post inoculation (dpi).

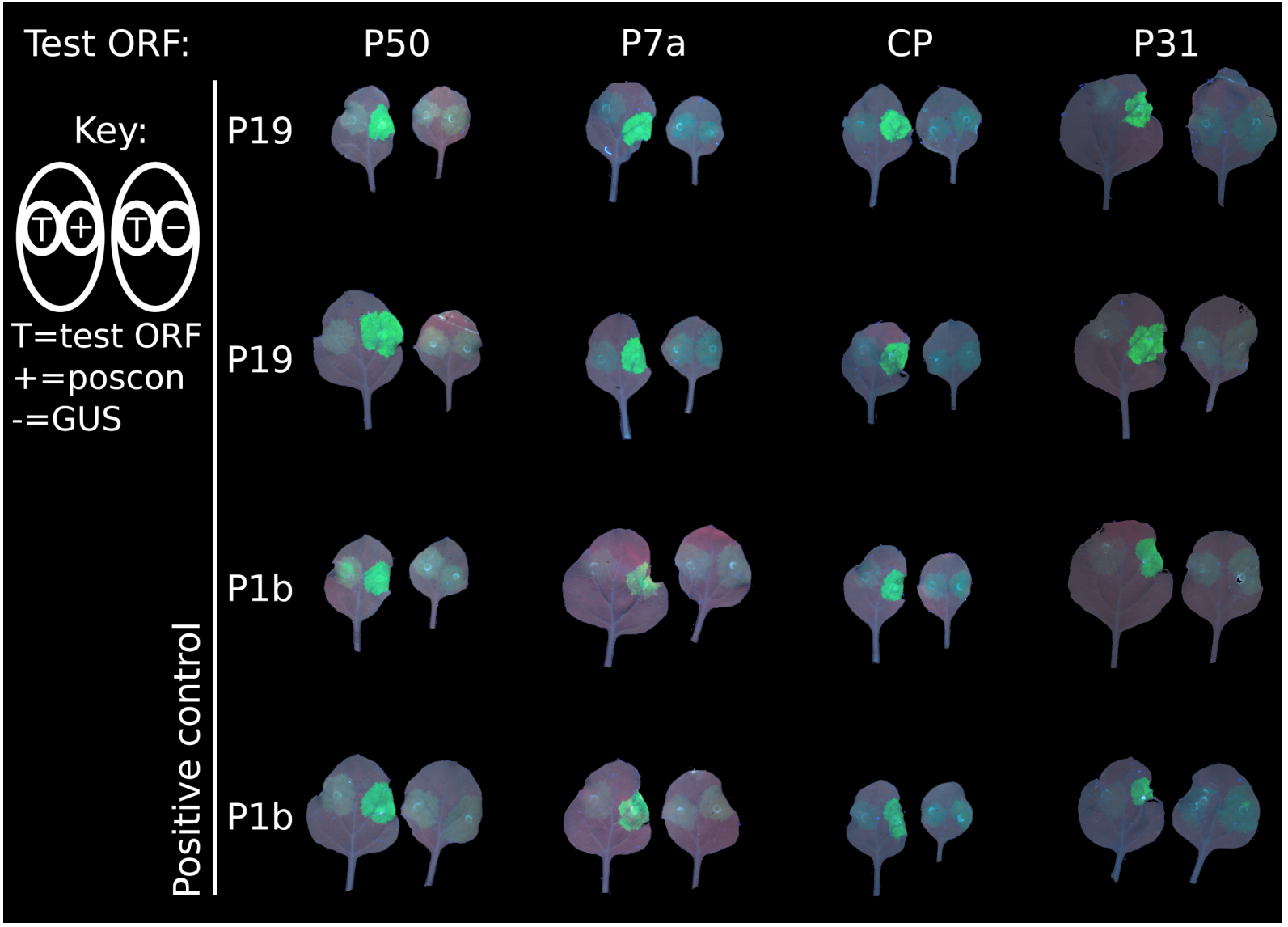


Fig. 6.4 Maize chlorotic mottle virus ORFs do not show local VSR activity. Adjacent leaves are taken from the same plant. Left hand leaves contain test ORF and positive control (P19 or P1b), right hand leaves contain test ORF and negative control. Left hand patch on leaves is test ORF, right hand patch is control (see key). There is no observable difference between the test ORF and negative control, suggesting a lack of local VSR activity.

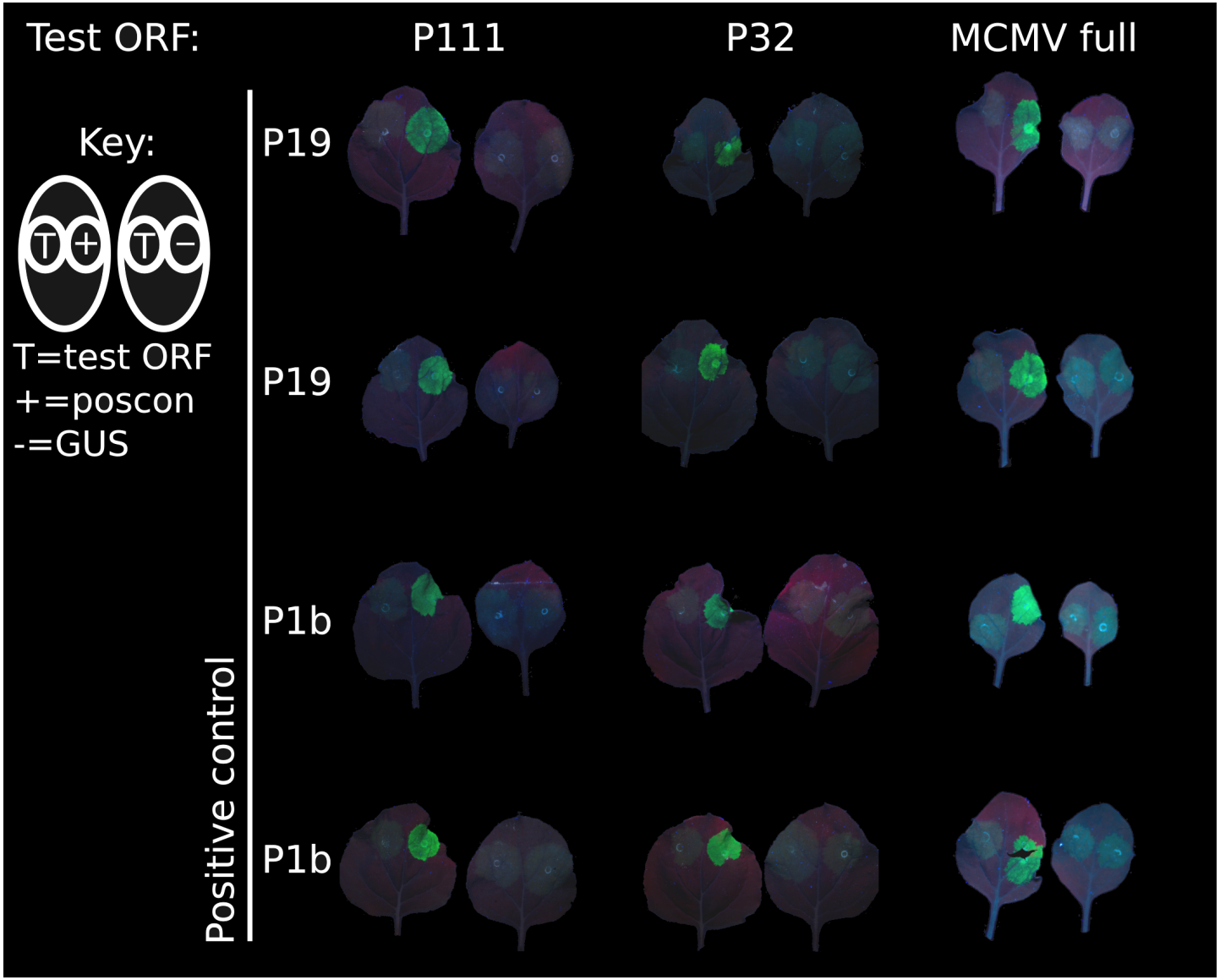


Fig. 6.5 Maize chlorotic mottle virus ORFs do not show local VSR activity. Adjacent leaves are taken from the same plant. Left hand leaves contain test ORF and positive control, right hand leaves contain test ORF and negative control. Left hand patch on leaves is test ORF, right hand patch is control (see key). There is no observable difference between the test ORF and negative control, suggesting a lack of local VSR activity.

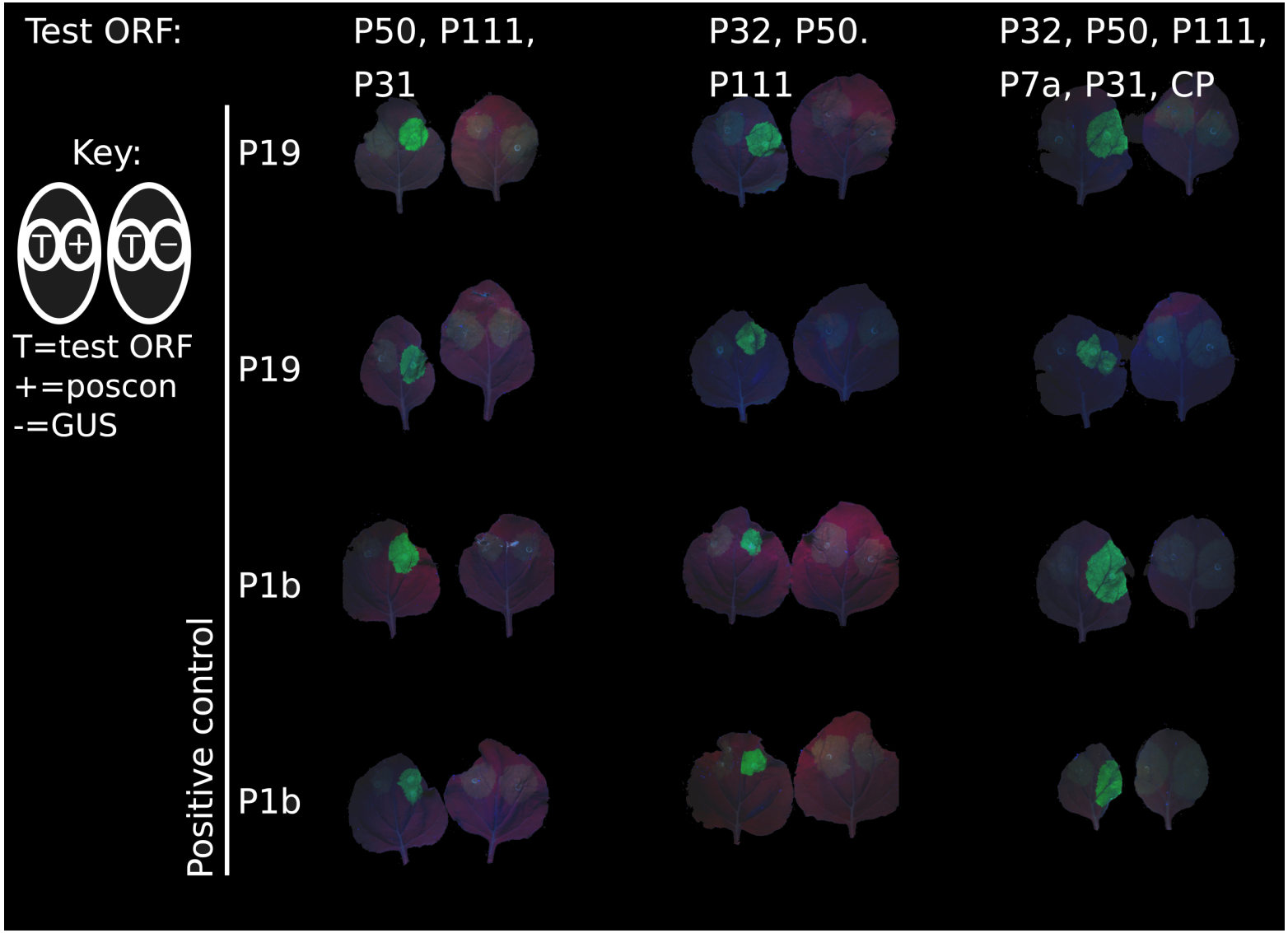


Fig. 6.6 Combinations of maize chlorotic mottle virus ORFs do not show local VSR activity. Adjacent leaves taken from the same plant. Left hand leaves contain test ORF and positive control (P19 or P1b), right hand leaves contain test ORF and negative control. Left hand patch is test ORF, right hand patch is control (see key). There is no observable difference between the test ORF combinations and negative control, suggesting a lack of local VSR activity.

6.6 Investigation of systemic silencing suppressors in maize chlorotic mottle virus

6.6.1 Systemic silencing

The sRNAs generated from viral genomes upon infection target viral material in local cells, but also spread systemically, predominantly to sink tissues, inhibiting the systemic spread of the virus and producing the phenomenon of recovery, in which younger leaves emerge free of viral symptoms (Melnyk et al., 2011; Mlotshwa et al., 2008; Xie and Guo, 2006). Many VSRs, such as TBSV P19, suppress both local and systemic RNA silencing, but others, such as the P6 protein of rice yellow stunt rhabdovirus, specifically interfere with the systemic spread of silencing (Guo et al., 2013; Hamilton et al., 2002).

6.6.2 Systemic silencing suppressor assays using *Nicotiana benthamiana*

To assay for systemic silencing activity I decided to use induction of systemic silencing in *N. benthamiana* and coinfiltration with test ORFs. In these assays, young (4 leaf stage) 16C *N. benthamiana* seedlings (constitutively expressing GFP) are inoculated with *A. tumefaciens* strains containing T-DNAs with GFP and test ORFs. After local silencing is induced, the silencing signal spreads to upper leaves, resulting in reduced fluorescence, especially around leaf veins (fig. 6.7). However, if the test ORF has systemic VSR activity then there will be increased fluorescence compared to the negative control. I have found systemic silencing to be variable, and difficult to compare qualitatively by eye (fig. 6.8). Therefore, I used three different metrics, % area of the third leaf silenced, proportion of leaves displaying silencing, and proportion of petioles displaying silencing. The first systemic VSR assay I performed with effective silencing (see section below) was on the MCMV ORFs P111 and P32 (fig. 6.9), and appeared to show a general trend of silencing suppression by P32 across the three metrics. To validate this result, I repeated the experiment with P32 and included the 2b protein of cucumber mosaic virus, a known systemic VSR. 2b shows a trend to lower silencing than the negative controls, but this is not significant statistically (fig. 6.10). In contrast to the earlier experiment, P32 had a trend towards higher levels of silencing than the negative control, although not significantly, despite using the same protocol and materials as previously. Comparing the data between experiments shows that leaf area silencing at 15 dpi is similar in both experiments in P32 inoculated plants, but the negative control group has far lower silencing in the second experiment. However, by the other metrics, silencing

is higher in the second experiment in P32 inoculated plants, with no statistically significant differences. The results of the second experiment do not provide support for P32 being considered a systemic VSR. I also plotted the silencing metrics for individual plants, to investigate how silencing tends to vary between time points for individual plants (fig. 6.11). This shows that the proportion of petioles silenced increases through time for all treatment groups, while proportion of leaves silenced and % of third leaf silenced commonly both increase and decrease between time points. These results overall highlight the variability I found to be present in the induction of systemic silencing, which may be partially explained by its sensitivity to one environmental factor: light.

6.7 Effect of photoperiod on systemic silencing

6.7.1 Optimising systemic silencing suppressor assays

My initial attempts at performing systemic silencing assays were unsuccessful due to the extremely limited induction of silencing I observed (fig. 6.12a). To test whether growth conditions were responsible, I grew 14 *N. benthamiana* plants in each of the following conditions:

A) 22ºC, 65% humidity, 160μmol, 16 hours light.

B) 22ºC, 60% hum, 160μmol, 8 hours light.

C) 22ºC, 60% hum, 200μmol, 16 hours light.

I inoculated plants with *A. tumefaciens* expressing GFP at the four leaf stage. Plants grown in the short day (8 hours light) conditions had a different morphology to those grown in long day (16 hours light), with thinner petioles and smaller leaf lamina visible on short day plants (fig. 6.12b). Additionally, short day plants have a redder appearance in areas with silenced GFP, as can be seen around inoculation sites, suggesting that they have a higher concentration of a substance which fluoresces red compared to long day plants (fig. 6.12c). This could represent elements of the photosynthetic apparatus. Systemic silencing was much more visible and extensive in short day than long day plants (fig. 6.12d). Therefore, I proceeded with the systemic silencing experiments described above in short day conditions.

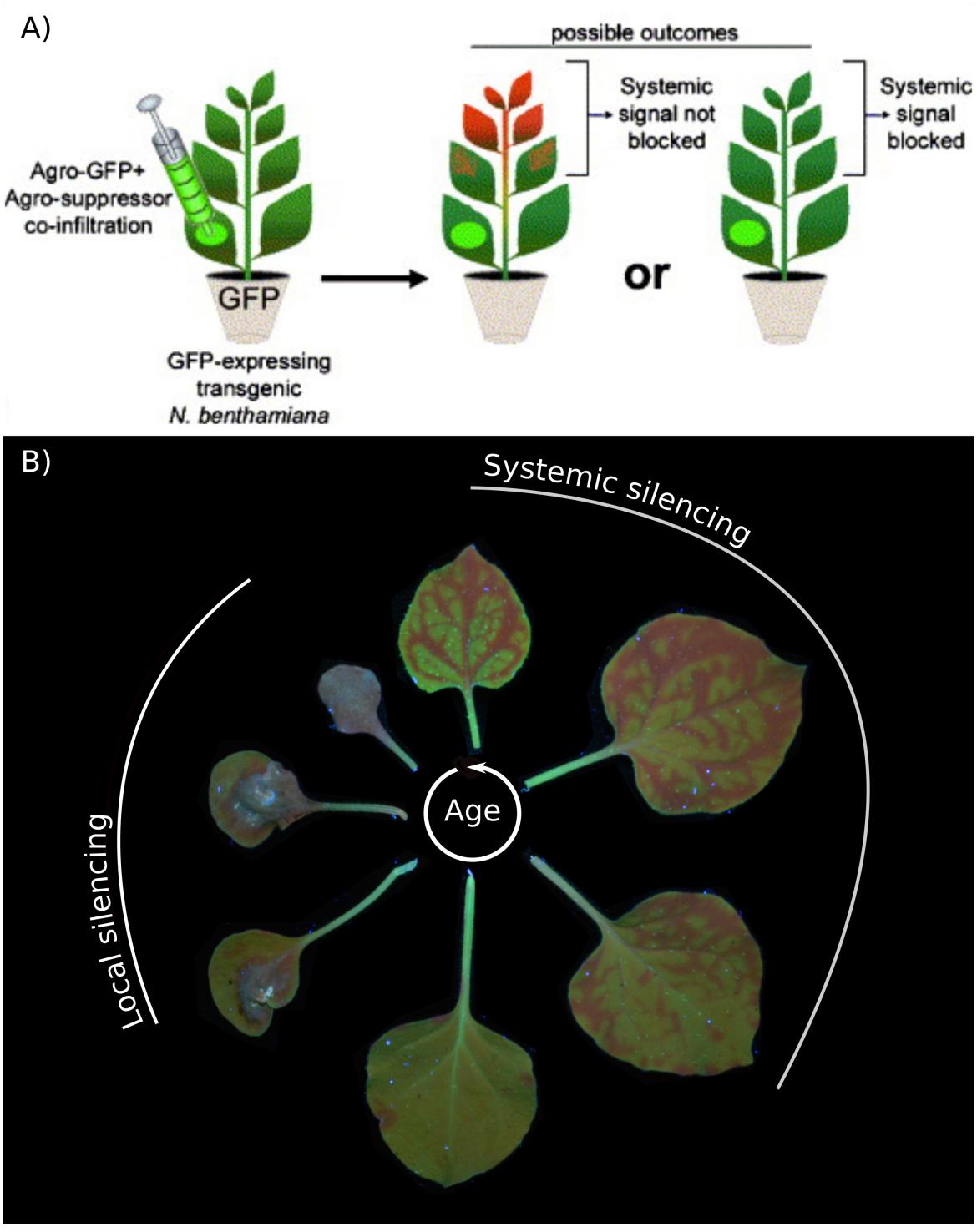


Fig. 6.7 Coinfiltration of *N. benthamiana* can be used to test the systemic silencing activity of viral ORFs. A) Schematic of coinfiltration assay, maintenance of fluorescence indicates systemic VSR activity of test ORF. B) Silencing caused by coinfiltration of a single *N. benthamiana* plant with GFP and GUS (negative control), illustrating local and systemic silencing. Part A) adapted from Roth et al. (2004). Part B) taken by me.

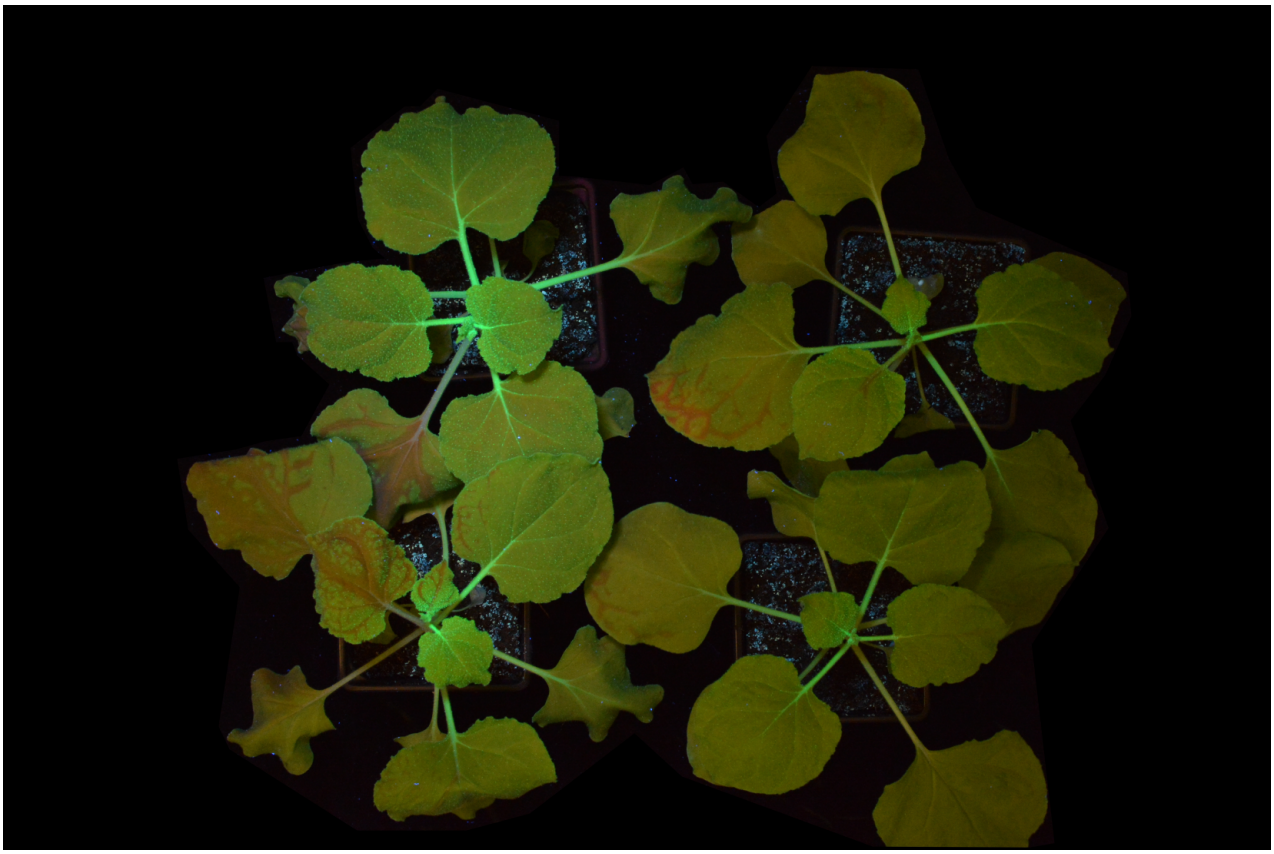


Fig. 6.8 Systemic silencing induction is highly variable. Four 16C *N. benthamiana* plants, all grown in identical conditions at same time, inoculated with GFP + GUS, at 25 days post inoculation.

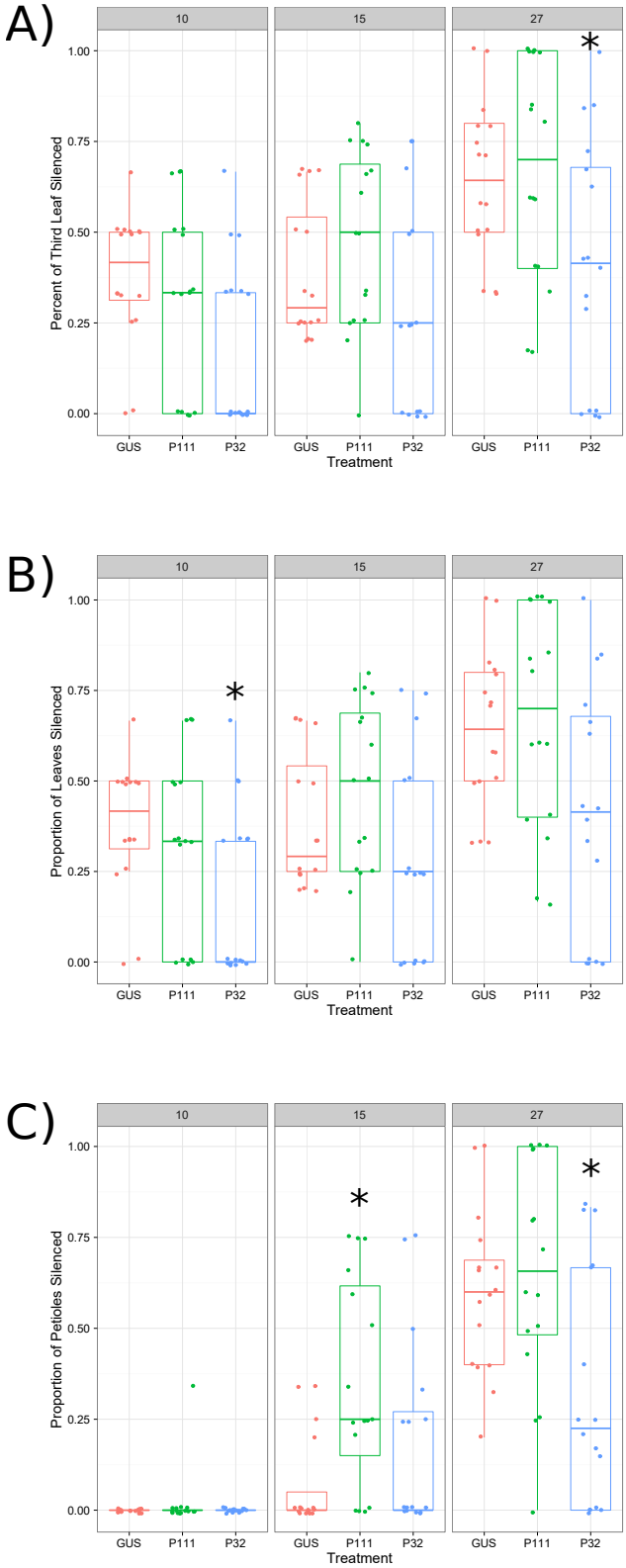


Fig. 6.9 P32 appears to have mild systemic VSR activity. Coinfiltration of *N. benthamiana* with negative control (GUS) and the maize chlorotic mottle virus ORFs P111 and P32. Days post inoculation is shown at the top of panels. Level of systemic silencing is assessed using A) % of third leaf silenced, B) proportion of leaves displaying silencing, and C) proportion of petioles displaying silencing. Statistical significance is indicated by an asterisk, as assessed by a Mann-Whitney U test comparing GUS with each test ORF.

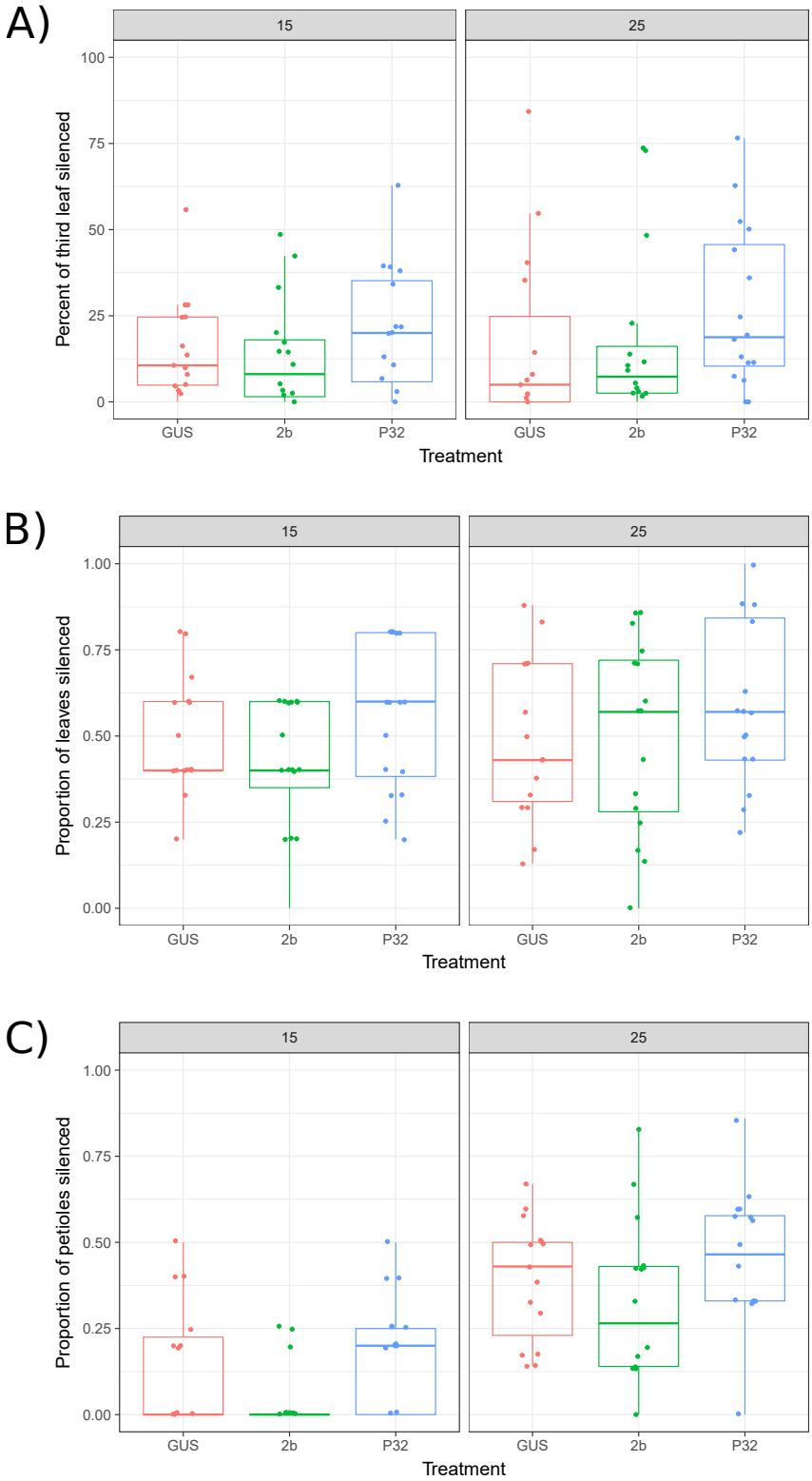


Fig. 6.10 P32 mild systemic VSR activity is not visible. Coinfiltration of *N. benthamiana* with negative control (GUS), positive control (CMV 2b), and the maize chlorotic mottle virus ORFs P32. Days post inoculation is shown at the top of panels. Level of systemic silencing is assessed using A) % of third leaf silenced, B) proportion of leaves displaying silencing, and C) proportion of petioles displaying silencing. There were no statistically significant differences between GUS and test ORFs, as assessed with a Mann-Whitney U test.

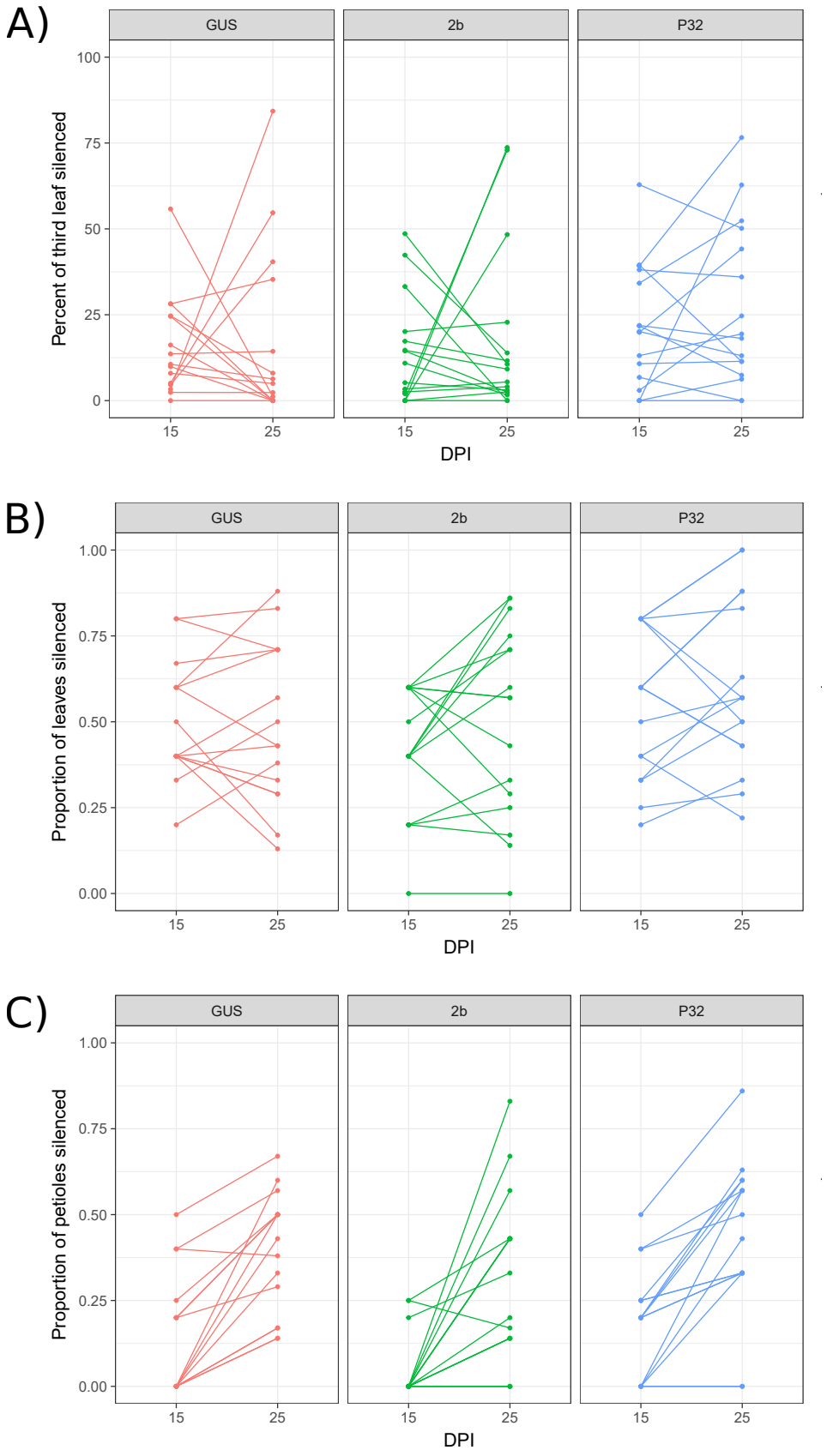


Fig. 6.11 Systemic silencing assessed for individual plants. Coinfiltration of *N. benthamiana* with negative control (GUS), positive control (CMV 2b), and the maize chlorotic mottle virus ORF P32. Level of systemic silencing is assessed using A) % of third leaf silenced, B) proportion of leaves displaying silencing, and C) proportion of petioles displaying silencing. Line represent individual plants, and highlight variability between time points in A) and B).

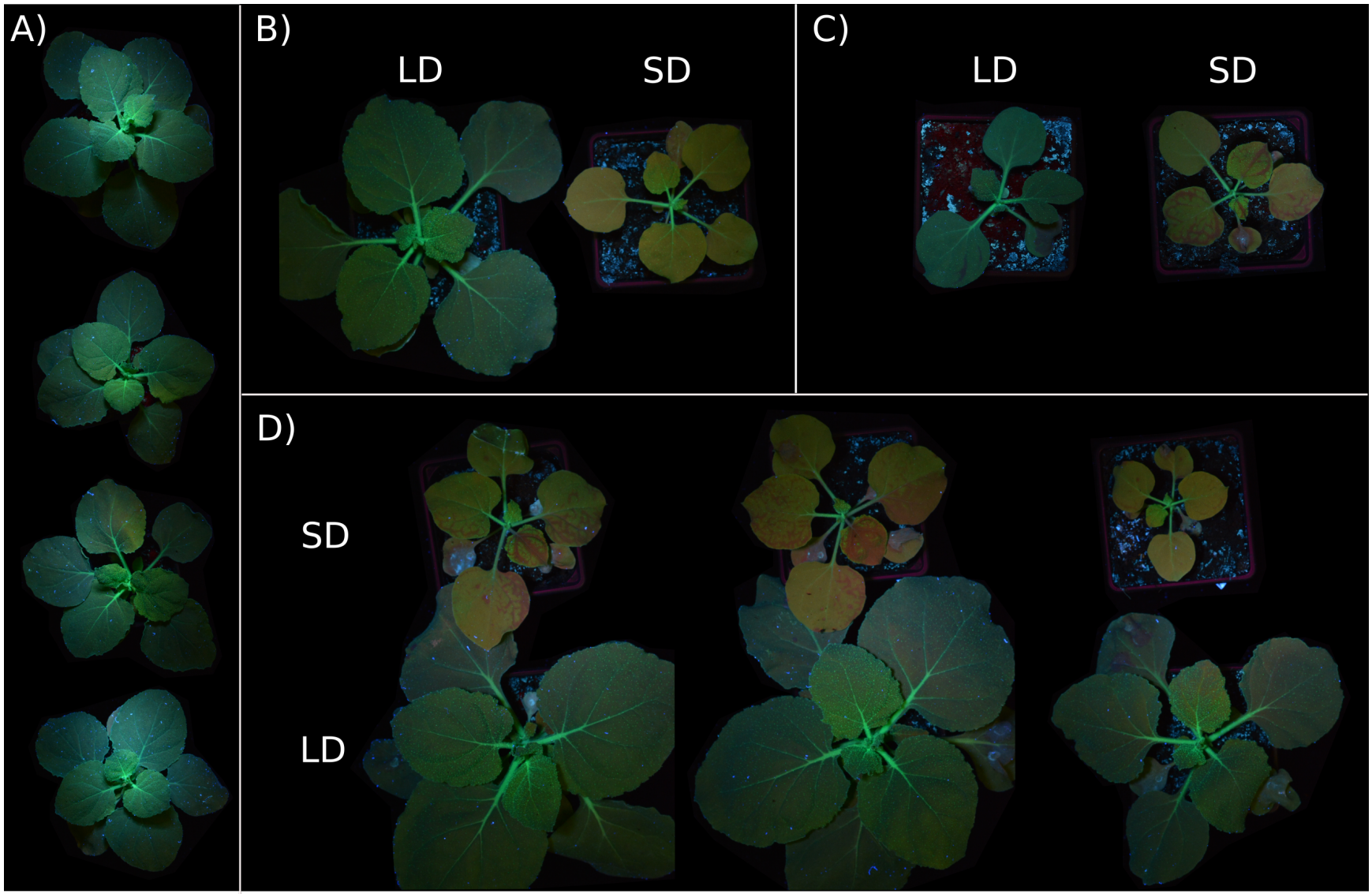


Fig. 6.12 Effects of photoperiod on *N. benthamiana* growth and systemic silencing. A) Plants grown in initial long day (LD) conditions do not show systemic silencing. B) Short day (SD) plants have smaller leaf laminas and thinner petioles. C) Systemic silencing in LD (bottom left leaf) and SD conditions, with redder colouration in SD plants. D) Systemic silencing is suppressed in plants grown in LD conditions.

6.7.2 Photoperiod can suppress systemic silencing

In order to quantify the effect of photoperiod on systemic silencing induction, when performing the experiment seen in Figure 6.10, I included 15 *N. benthamiana* plants grown in the same conditions but with long instead of short day photoperiod. Quantifying this difference as previously shows a clear difference between long and short day conditions, with plants grown in long day displaying almost no signs of systemic silencing, consistent with the earlier observations (figs. 6.12 and 6.13). This is far greater than the impact I observed using either MCMV ORFs or the known systemic silencing suppressor 2b. Examining plants confirmed that local silencing occurred as expected, but that the silencing signal did not spread efficiently through the rest of the plant (fig. 6.14).

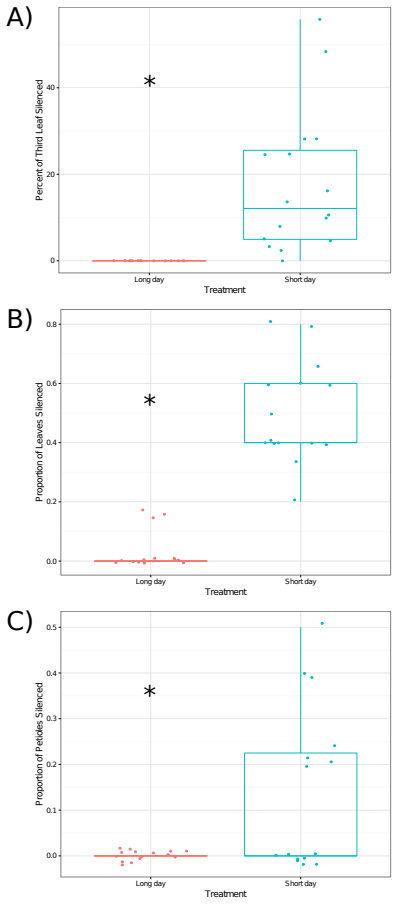


Fig. 6.13 Long day conditions suppress systemic silencing. *N. benthamiana* plants were grown in short and long day conditions and inoculated with *A. tumefaciens* expressing GFP to induce silencing. Measurements were taken 15 days post inoculation, with silencing assessed by A) % of third leaf silenced, B) proportion of leaves silenced, and C) proportion of petioles silenced. Asterisks indicate statistically significant differences ($p<0.05$), as assessed by Mann-Whitney U tests.

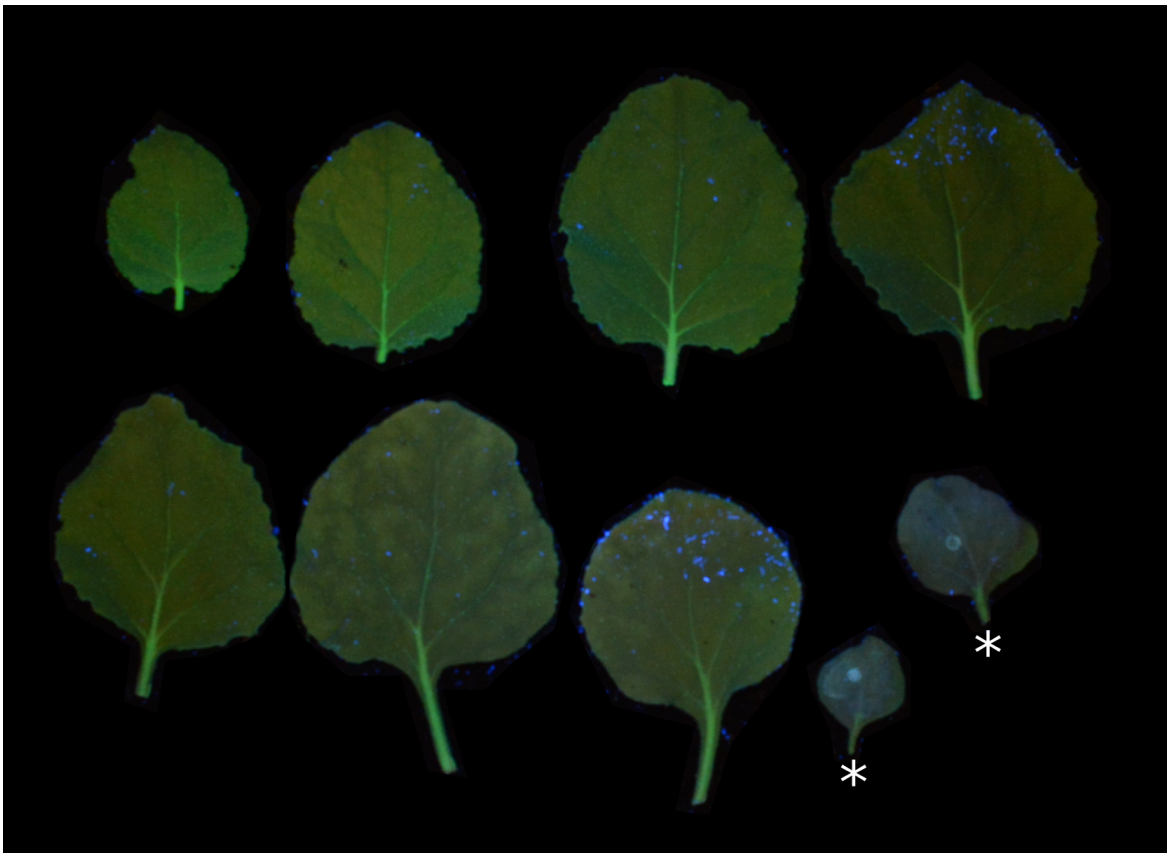


Fig. 6.14 Long day conditions do not suppress local silencing. Leaves from an individual *N. benthamiana* plant grown in long day conditions. Systemic silencing is suppressed, but local silencing on inoculated leaves, as indicated by asterisks, remains.

Discussion

6.8 Local silencing suppression by maize chlorotic mottle virus

Interference with the antiviral RNA silencing machinery is extremely common amongst viruses; many phytoviruses have been found to encode a broad range of VSRs (Wu et al., 2010). VSRs typically work in one (or more) of three ways:

- Binding long double-stranded RNA to inhibit sRNA production.
- Binding short double-stranded molecules to inhibit RISC loading.
- Interacting with RNA silencing proteins to interfere with their function.

The *Tombusviridae* includes a wide variety of local VSRs (table 6.1), which typically function by RNA-binding mechanisms. However, in a number of cases the relation between RNA-binding ability and VSR activity has not been proven using mutational analysis (TBSV P19, CymRSV P19, PoLV P14). The highly multifunctional nature of viral proteins means that careful mutational analysis is required to demonstrate that specific biochemical properties (such as RNA binding) are required for VSR activity, as exemplified in Martínez-Turiño and Hernández (2009); Perez-Canamas and Hernandez (2015). *Tombusviridae* phylogeny varies depending on the portion of the genome used to construct trees, but MCMV is always present as the sister clade to the panicoviruses. However, panicoviruses have not been investigated for VSR activity, and VSRs have also not been reported in the alpha- or betanecroviruses, which are often grouped with MCMV and the panicoviruses (Scheets, 2016). This, combined with the lack of homologs to known *Tombusviridae* VSRs (fig. 6.1), stimulated my interest in possible VSRs within the MCMV genome.

6.8.1 No evidence for silencing suppressors in the maize chlorotic mottle virus genome

I used two alternative coinfiltration assays of *N. benthamiana* to test for local and systemic silencing in MCMV ORFs. All MCMV ORFs were assayed for local VSR activity, with none showing noticeable differences from the negative control (figs. 6.4 to 6.6). Additionally, there was not clear evidence for systemic VSR activity in the MCMV ORFs that I assayed (P111, P32) (figs. 6.9 to 6.11). There is no good evidence for local or systemic VSR activity in the experiments I have performed. There are a number of reasons that my assays could have been unsuccessful. *N. benthamiana* is a dicot, and used as a model in VSR assays. However maize is a monocot, and therefore a MCMV-encoded VSR could interact with a protein site, or protein, not present in *N. benthamiana*. For example, maize has a very high number of AGO proteins (estimated 18), including AGO18, a monocot-specific AGO clade, with AGO18 shown to provide broad-spectrum viral resistance by sequestering miR168, which targets the antiviral AGO1 (Wu et al., 2015). *AGO18* expression is induced by viral infection in rice, and maize *AGO18a* shows a significant increase in expression in response MCMV, SCMV, or double infection (Wu et al., 2015; Xia et al., 2016). If (hypothetically) a MCMV VSR targeted AGO18a in maize, this effect would likely not be visible in *N. benthamiana*-based assays. Additionally, I did not assay all MCMV ORFs for systemic VSR activity, due to the difficulty I had in obtaining consistent results (fig. 6.8). The Gateway library can be used to epitope tag the MCMV ORFs and confirm their expression in *N. benthamiana*, which would be another explanation for the lack of local VSR activity observed.

6.8.2 Does maize chlorotic mottle virus encode a silencing suppressor?

I will now review evidence for the presence or absence of a VSR in the MCMV genome. There have been other investigations of VSR activity by MCMV, although none with conclusive positive results. Scheets (2016) reported inoculating 16C *N. benthamiana* plants with *A. tumefaciens* strains with T-DNAs containing the MCMV ORFs P32, P31, and CP, together with GFP. She observed silencing of GFP, suggesting that there was no VSR activity. It seems possible that MCMV lacks a local silencing suppressor. My own results in *N. benthamiana* are clear (figs. 6.4 to 6.6), and additionally, MCMV infection does not relieve the silencing of chalcone synthase in maize (Della Vedova et al., 2005), unlike maize necrotic streak virus, another *Tombusviridae* member. Chalcone synthase silencing appears to occur post-transcriptionally, as transcripts can be detected in nuclear run-off experiments (Della Vedova et al., 2005). Therefore, MCMV has been reported to be unable to inhibit local post-transcriptional gene silencing in its natural host maize. However, coinfection of maize with MCMV and WSMV (but not SCMV or MDMV) does result in an increase in WSMV titre, which could be due to a VSR within the MCMV genome, but alternatively another WSMV-beneficial protein, such as a movement protein (Scheets, 1998). Local VSR activity has been reported for MCMV ORFs P7a, P50, P111, and the post-readthrough portion of P111, but the data show a small effect (much less than HC-Pro, the positive control), and the plant pictures are unconvincing, suggesting that the inoculation has not been performed well (many injection sites, implying either poor technique or leaves which are too old) (Bacheller, 2017).

In terms of candidates for VSR activity within the MCMV genome, P32 contains a WG motif, P31 contains two GW motifs, P50 contains one GW motif, and P111 contains a WG and a GW motif. WG/GW motifs are often present in proteins with AGO binding activity, which includes VSRs (Csorba et al., 2015). However WG/GW motifs are not unique to VSRs, and there is an example of GW being required for VSR action through its impact on sRNA binding rather than AGO binding (Perez-Canamas and Hernandez, 2015). There are a number of candidates for VSR activity in the MCMV genome, given the presence of WG/GW motifs and VSRs across the *Tombusviridae* (table 6.1), but *in vivo* evidence in maize and my own assays do not support the presence of a local VSR in the MCMV genome (Della Vedova et al., 2005; Scheets, 2016). If MCMV does not produce a local silencing suppressor, it would be interesting to investigate how it is able to achieve such high titres in host cells, in particular whether the membrane bound viral replication complexes of MCMV are capable of blocking the majority of RNA silencing activity. Another possibility is that MCMV's ability to partner with *Potyviridae* viruses with strong VSR activity, combined with

the frequency with which it is found in partnership, means that VSR activity is predominantly provided to MCMV by trans-complementation from its infection partners.

The data generated from my systemic VSR assays was very variable (figs. 6.8 to 6.10). Systemic VSR assays have not been reported for MCMV previously, and my principle difficulty in these experiments was the variability of systemic silencing induction, including in the negative control group. Due to the difficulty in assessing effects visually, I developed three different metrics for assessing systemic silencing, which were also highly variable (e.g. see fig. 6.9). Unfortunately, quantitative data on systemic silencing effects is rare/absent in the literature (comparisons of photos of single plants is most common), which makes it difficult to evaluate whether my own results are typical, or whether they are more variable than those obtained by other scientists, which would suggest a need for me to further optimise my growth conditions. For example, CMV 2b is widely accepted to be a systemic VSR, but only one paper has reported data on this, and although the sample sizes used are suitable (50-100), and the effects reported are very strong (100% systemic silencing in negative controls, 4% systemic silencing when coinfiltrating with CMV 2b), the single plant images provided are not (Guo and Ding, 2002). Growth conditions are not reported in Guo and Ding (2002), so accurate replication is not possible. Useful next steps would be to determine whether other labs are able to induce systemic silencing reliably, and build on my work linking environmental conditions to systemic silencing to minimise variation.

6.9 Photoperiod and systemic silencing

I found that photoperiod had a striking impact on systemic RNA silencing, with long day conditions (16 hours light) essentially eliminating visible signs of systemic silencing spread in my growth conditions (fig. 6.13). There is a growing body of evidence linking systemic silencing with environmental factors. Given the variability in silencing activity seen in *N. benthamiana* repeatedly transformed with identical GFP constructs, it seems likely that the chromatin context of transgene insertion affects the systemic silencing signal, through an unknown mechanism (Kalantidis et al., 2006). Silencing of a nitrate reductase transgene varied between winter (less silencing) and summer (more silencing) (Vaucheret et al., 1997). However, this comparison includes variation in light intensity, photoperiod, and temperature. Comparisons of *N. benthamiana* plants showed that siRNA production, but not miRNA production, is reduced with lower temperatures (Szittya et al., 2003). Kotakis et al. (2010) found that *N. benthamiana* plants grown in high light ($130\mu mol/m^2/s^1$) had higher spontaneous induction of systemic silencing of a GFP transgene than those grown in low light ($35\mu mol/m^2/s^1$), and increased expression of a number of *DCL* genes, as well as *RDR6*.

Using VIGs to knockdown *DCL3* or *DCL4* resulted in decreased systemic silencing spread in both light conditions (Kotakis et al., 2011). More intense light conditions ($450\mu mol/m^2/s^1$ +) will also inhibit systemic silencing (Patil and Fauquet, 2015).

The systemic silencing signal flows with the movement of phloem, as evidenced by radiolabelling experiments, and appears to involve 24nt siRNAs (Melnyk et al., 2011; Patil and Fauquet, 2015; Tournier et al., 2006). The spread of systemic silencing signals is consistent with the silencing signal spreading according to source-sink relationships between tissues (Tournier et al., 2006). This could explain my results, if the young leaves in long day conditions are only briefly sink tissues, with the long photoperiod allowing them to grow rapidly and become source tissues shortly after emergence from the meristem. To investigate whether this is the case for photoperiod impact on systemic silencing, experiments removing leaves and covering them to alter the source-sink relationship will be informative. Likewise transfer experiments between different growth conditions will be valuable to establish the dynamics of the process. This, in combination with knockdown mutants (or VIGs) of relevant genes (*DCL3* and *RDR6* to start) should provide details on the mechanism by which photoperiod has such a drastic effect on systemic silencing. Investigating the effect of photoperiod on symptoms generated by viral challenges would determine whether these effects are biologically relevant in widespread pathogen-host interactions.

Conclusion

Silencing suppression is crucial to viral reproduction, as it protects their genetic material from cleavage or translational repression. It is also a typical basis for viral synergy, for example the widely observed potyviral synergy is mediated by the silencing suppressor HC-Pro. I decided to investigate silencing suppression by MCMV, and found no evidence for local VSR activity, in agreement with previous work (Della Vedova et al., 2005; Scheets, 2016). My data on systemic VSR activity did not provide reproducible evidence for a systemic VSR, but the reproducibility problems I had led me to discover a link between photoperiod and the spread of systemic silencing. This striking effect may be explained by source-sink relationships, as described previously, or a novel mechanism.

Chapter 7

Discussion and conclusions

7.1 Controlling maize lethal necrosis

MLN is an aggressive viral condition which is emerging globally (fig 1.1). Over the last decade MLN has been spreading across East Africa, where maize is a major employer, source of calories, and component of the regional economy. Accordingly, the severity of this threat has galvanised action on multiple fronts, with contributions from the commercial sector, not-for-profits, and the academic community. The multi-modal response includes conventional breeding, genetic engineering, and agronomic practice.

Initial screening of commercial maize lines for MLN resistance was broadly unsuccessful (CIMMYT). However, later efforts included more diverse breeding panel lines (rather than commercial lines) from the drought tolerant maize for Africa and improved maize for African soils programmes, and the use of genotyping-by-sequencing allowed the rapid identification of minor and medium effect QTLs for MLN resistance (Gowda et al., 2015). Ongoing breeding efforts at CIMMYT's MLN screening facility at KALRO Naivasha have identified a number of lines which are tolerant enough to produce ears despite artificial inoculation with MLN (CIMMYT, personal communication). These lines are very promising for relieving MLN pressure in East Africa, especially given the lower regulatory burden on conventionally bred crop lines. Another approach to MLN control is vector management through the application of pesticides, which was trialled by Bayer CropScience. Kibaki and Francis (2013) trialled the efficacy of seed treatment with the insect neurotoxin imidacloprid and different spraying regimes using imidacloprid and the insect neurotoxin β-cyfluthrin in an attempt to control the thrip (MCMV) and aphid (SCMV) vectors. Lowering vector numbers decreases the rate at which the viruses spread to uninfected hosts, thereby raising the average age of infection and sparing more individuals from infection. At eight weeks after planting

Kibaki and Francis (2013) saw MLN incidence halved from 80% to 40% between the control and most intensely sprayed groups, and lower symptom severity.

In this thesis I surveyed the East African MLN outbreak using NGS, which is an invaluable technology in characterising novel conditions. The sequencing data revealed the co-occurrence of two viruses with drastically different patterns of variation. MCMV has extremely low sequence diversity globally, with a maximum of 4% difference in nucleotide identity, and most of the diversity present is explained by population structure - most differences are between geographic regions rather than within them. In contrast, SCMV is extremely diverse, with a maximum of 23% difference in nucleotide identity, less of this variation explained by population structure, and recombination so extensive that I decided against constructing a phylogeny. NGS data is also ideal for informing the design of sequence-mediated resistance to viral conditions, as it provides consensus sequences as well as low-frequency variants within samples. Therefore, NGS can be used to identify an emerging disease agent, and simultaneously provide data on how to engineer sequence mediated immunity. In this case, MCMV was a far more attractive candidate for sequence mediated resistance due to its lower diversity. Molecular characterisation of lines transformed with amiRNAs targeting MCMV is ongoing, and functional testing through artificial inoculation of transformed lines will reveal the efficacy of this approach.

Effectively controlling MLN in East Africa will likely require a combination of technologies and management strategies that vary depending on the farming system and local severity of MLN. For example, although pesticide spraying is moderately effective in slowing the spread of MLN, the cost of these inputs will be beyond the reach of the majority of smallholder farmers. Conventionally bred tolerant lines will reduce yield losses, however may not drastically reduce the concentration of virions in vectors and soil water, and therefore would act as a source of inoculum for neighbouring farms. If the lines we engineered are resistant, they would be a valuable addition to a toolkit for dealing with MLN, and would likely have most value in regions with a high concentration of both smallholders (small farms lessen the benefits of crop rotation and make it harder to implement) and MLN. Crop rotation has been shown to be effective in reducing MLN disease pressure, but large-scale implementation in areas with a tradition of relay cropping and a high concentration of smallholders will require significant investment in education and training. Mobile technologies may provide a lower cost means to disseminate information and coordinate behaviour, although research would be needed to maximise engagement, ease-of-use, and compliance. Another crucial component of controlling MLN will be securing a MLN-free seed supply. I personally witnessed seed production fields in Kenya with heavy MLN infection, and recent results from Ecuador suggest that seed transmission can occur at a far higher rate than previously

reported (fig. 7.1) (Quito-Avila et al., 2016). The rapid global spread of MCMV is most easily explained by seed transmission, and RT-PCR has shown the presence of MCMV in 50% of seed lots purchased at market in Kenya (Mahuku et al., 2015). Therefore, the MCMV screening capacity being developed for seed by the Kenyan Plant Health Inspectorate Service will be vital to reduce the amount of MCMV present in the Kenyan maize system. There is no simple, one-size-fits-all solution to MLN, but we can maximise our chances of success by simultaneously pursuing multiple complementary responses, and crucially coordinating their implementation at the landscape scale.

Fig. 7.1 Maize lethal necrosis is present in commercial seed production fields in Kenya. Commercial Kenyan seed production field, with MLN symptoms visible - leaf margins drying and extensive chlorosis.

7.2 Resistance engineering - thoughts for the future

The first attempt we made to engineer resistance to MLN used amiRNAs because this technique has been demonstrated to be effective across a wide range of hosts and viruses. This project was started with the end goal of application in mind, and therefore we decided to use an established technique, albeit utilising a novel design pipeline. Using an established technique is attractive in maize due to its relatively (compared to models) slow transformation and maturation. Using a novel technique may require multiple test-modification cycles, and each of these would take approximately a year in maize. However, there are now some interesting new approaches which have been demonstrated and could be attempted in maize.

New genetic engineering approaches

R genes are most commonly intracellular NB-LRR proteins which detect pathogen effectors or the modifications to host proteins induced by pathogen effectors. Modifying the specificity of *R* genes would allow engineering of monogenic dominant resistance when it is not naturally present. For example, in *Arabidopsis* the *R* gene *RPS5* encodes an NB-LRR which detects cleavage of the host kinase PBS1, which is caused by the *Pseudomonas syringae* effector AvrPphB. Kim et al. (2016) replaced the cleavage site in PBS1 recognised by AvrPphB with a cleavage site recognised by a different *P. syringae* effector, AvrRpt2. This produced plants which were less susceptible to *P. syringae* strains expressing AvrRpt2 than control plants. This elegant approach could be adapted to other systems, although lower knowledge of *R* gene mechanisms in maize make applying it here more difficult. There are no known guards detecting the cleavage of a host protein in maize, so possible approaches to try would be:

- Transform maize with *Arabidopsis RPS5* and a *PBS1* gene modified to contain an SCMV polyprotein cleavage site.
- Transform maize with the modified *PBS1* gene as above, and a chimaeric *Arabidopsis RPS5* containing the LRR domain from *RPS5* and the NB domain from a native maize *R* gene.
- Engineer the SCMV cleavage site into a synthetic circuit resulting in cell death upon cleavage. For example a plant-specific toxin activated by proteolysis, or a transcription factor activated by cleavage which drives expression of transgenes which trigger cell death.

The success of the first two approaches would depend on whether the *Arabidopsis* NB-LRR-guard circuit functions in maize - this same system does function in *N. benthamiana*, but maize is more distantly related, being a monocot (Ade et al., 2007).

To engineer a more general resistance to viruses, it would be interesting to test the effects of increasing expression of AGO18 isoforms in maize. In rice, AGO18 sequesters miR168 to inhibit its targeting of the antiviral AGO1 mRNA, and insertion of an additional constitutively expressed AGO18 generates broad spectrum resistance to viruses Wu et al. (2015).

Sequence mediated immunity selects for mutations in the target sites in viral genomes. The higher the number of sites in the viral genome targeted, the lower the probability of viral genome variants emerging with mutations in all target sites. In designing my own amiRNA constructs, I removed amiRNA:target mismatches because perfect pairing promotes transitivity - RDR activity on the amiRNA:target duplex followed by siRNA production from the dsRNA molecule. This is equivalent to the production of another form of sRNA: trans-acting small interfering RNAs (tasiRNAs), which are naturally generated from *TAS* genes. *TAS* genes are non-coding transcripts targeted by one or two miRNAs, which induce cleavage and in doing so RDR6 activity, generating dsRNA (fig. 7.2a). This dsRNA is then cleaved by DCL4, producing tasiRNAs which target genes with homology to the non-coding *TAS* transcript (Eckardt, 2013). Artificial *TAS* genes with miRNA target sites can be modified so the tasiRNAs target whatever sequence is desired. This technique has now been used to engineer resistance against tomato leaf curl Gujarat virus in tomato (*Solanum lycopersicum*), and simultaneous resistance to TuMV and CMV in *Arabidopsis* (fig. 7.2b) (Chen et al., 2016; Singh et al., 2015). The tasiRNAs produced are phased predictably, so biosafety concerns regarding long viral sequences can also be addressed by designing artificial *TAS* genes which contain discontinuous stretches of the viral genome, as demonstrated by Chen et al. (2016).

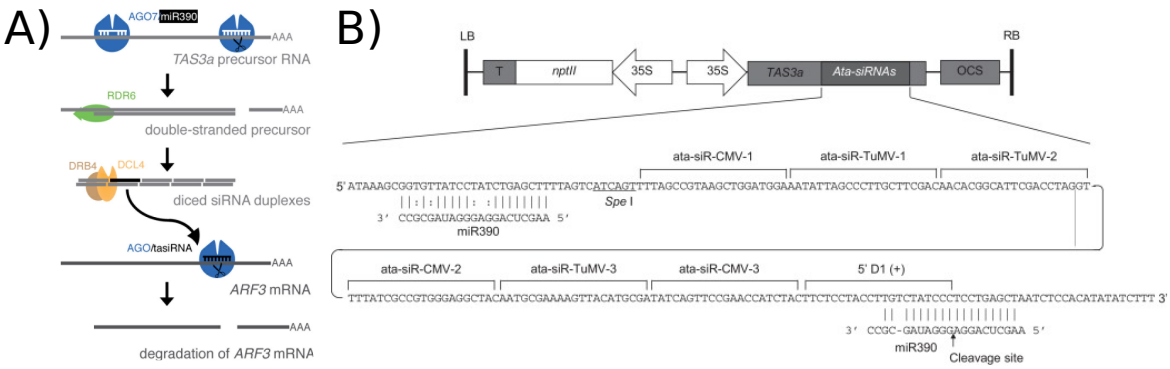


Fig. 7.2 A) Production of trans-acting small interfering RNAs (tasiRNAs) in *Arabidopsis*, showing non-coding transcript targetted by two miRNAs, followed by RDR action and siRNA production from the resulting dsRNA. B) Artificial *TAS* construct used by Chen et al. (2016) to engineer resistance to cucumber mosaic virus and turnip mosaic virus in *Arabidopsis*. A) adapted from Schwab et al. (2009), B) adapted from Chen et al. (2016).

Non-engineered artificial resistance

Public opposition and legal inhibition of genetic modification has hampered its application in many countries. There are now additional techniques available for generating artificial resistance which are either not classified as genetically modified, or not universally classed as genetically modified. Topical sprays of dsRNA are sufficient to induce temporary resistance (usually $\leqslant 7$ days) to viral inoculation, and can either be purified RNA or bacterial extracts modified to express the dsRNA (Robinson et al., 2014). Bacterial extracts expressing dsRNA against SCMV have been shown to reduce SCMV titre and symptoms in maize (Gan et al., 2010). Typically work in this field uses mechanical inoculation of viruses, which would also make cell interiors more accessible to dsRNA. Mitter et al. (2017) showed that delivering RNA in clay nanosheets increased the protection window to around 20 days, and crucially that the dsRNA could downregulate a reporter gene (without mechanical inoculation disrupting the leaf surface), and provide resistance in a newly emerged leaf, suggesting that the RNAi signal had been absorbed by the plant and spread systemically. This is convincing evidence that topically applied dsRNA can provide antiviral protection without cuticle disruption in *Arabidopsis* and *N. tabacum*, but cuticles from different plant families, poeacae for example, could have different uptake properties. Regardless, the high synthesis costs of RNA and requirement for repeat application means this technology, although exciting, is unlikely to be suitable for the majority of farmers in East Africa.

Interest and investment in genome editing has increased enormously over the past five years with the rise of CRISPR-Cas9. Cas9 is a bacterial RNA-guided DNA endonuclease, which is naturally guided by viral sequences held in the bacterial genome within clustered regularly interspaced palindromic repeats (CRISPRs) (Doudna and Charpentier, 2014). Artificial single-guide RNAs can be used to target the Cas9 nuclease to complementary 20 nucleotide regions of DNA, resulting in cleavage or nicking depending on the system used. This system can be used to generate transgenic plants with Cas9 nucleases targeted to DNA virus genomes, resulting in antiviral resistance (Ali et al., 2015). However, this is classified as a genetically modified plant as it contains a transgene. Cas9 can instead be programmed to target specific sites in the host genome using single-guide RNAs, resulting in mutations at that site. Offspring can then be selected which contain the desired mutation but lack the Cas9 transgene, making this type of modification more comparable to a mutation derived through undirected mutagenesis, which is not regulated as genetic modification. The regulatory status of genome-edited crops is not currently clear, and will likely vary between jurisdictions with process-based and product-based regulation of transgenic organisms. Regardless, Cas9 could be targeted to translation factors in the maize genome, such as eIF4E, to induce mutations and then screen for resistance against multiple maize viruses (many viruses interact with

the translation initiation complex), with the aim of generating recessive resistance genes against one or more maize viruses. This approach has been successful in both *Arabidopsis* and cucumber (*Cucumis sativus*) (Chandrasekaran et al., 2016; Pyott et al., 2016).

7.3 Building pyramids

Viruses have huge populations, short generation time, and for RNA viruses an especially error-prone polymerase. This means that they exist as a cloud of mutations around a consensus sequence, and can therefore rapidly respond to changing selection pressures through the selection of beneficial mutations. Efforts to engineer resistance can rapidly be overcome by viral evolution if it is not taken into account during the design of antiviral constructs (Lafforgue et al., 2011). Pyramiding refers to the insertion of multiple sources of resistance into a single cultivar, and can be performed through conventional breeding or genetic engineering. A virus targeted with multiple different forms of resistance is less likely to possess the multiple mutations required simultaneously to break resistance (Fuchs, 2017). My own work on engineering resistance to MCMV in maize used a pyramid of five amiRNAs expressed from a polycistronic precursor. Another benefit of pyramiding resistance genes is that it can be used to provide resistance against multiple viruses. Mixed infection is common in plant species, and my own work has demonstrated that East African maize is regularly subject to mixed viral infections. In the case of my transformed lines expressing anti-MCMV amiRNAs, a concern is that VSR activity from SCMV HC-Pro could weaken sequence mediated resistance to MCMV in the case of MCMV superinfection. This will be tested, but we are also planning to design a second generation of constructs including resistance to both MCMV and SCMV. To generate SCMV resistance, there are a number of options, which aren't mutually exclusive, such as inserting the cloned dominant *Scmv1* gene, using genome editing to try and generate recessive SCMV resistance, and using amiRNAs to target the short region of the SCMV RdRP with little to no sequence variation.

7.4 Conclusion

This work demonstrates the value of using an integrated NGS pipeline to characterise an emerging viral disease and use this data in designing sequence mediated resistance. Streamlining the process would result in the production of engineered T1 seed approximately one year after sampling, far shorter than the time required for a conventional breeding response to a disease with absent or rare resistance alleles. Likewise the resources required are far smaller than required for large scale conventional screening and breeding programmes. This approach would be valuable in forming part of a multi-pronged response against future emerging pathogens, and quickly provide genes which generate resistance. These genes can then ideally be combined with other sources of resistance to produce robust resistance in crop lines. These crop lines can then be combined with beneficial agronomic practices such as crop rotation and vector management to form an integrated and rapid response. Rapid and effective response to emerging crop pathogens will be vital if we are to maintain and increase crop yields sustainably in the face of both pathogen outbreaks and a changing climate.

Chapter 8

Supplementary data

8.1 Summary

This chapter contains supplementary data, specifically the regions of MCMV amino acid alignments with region-specific variation (as judged by high ARI values), and the primers used in this study. Additional information can be found in the electronic supplement.

8.2 Chapter 3 supplementary data

Table 8.1 Site in the CP ORF found to have an ARI index significantly greater than 95% of instances generated by a permutation test with randomised (rather than geographic) clustering of sequences. Column headers are codon numbers within the ORF. Regions are separated by horizontal lines. Accession numbers and clades of isolates are given, and remain the same for following tables

Isolate	Accession	Clade	81
KS1	X14736.2	N. America	s
Nebraska	EU358605.1	N. America	s
Porto_Severe_Ec	MF510219	S. America/Hawaii	p
Porto_Original_Ec	MF510221	S. America/Hawaii	p
Sta_Elena_Ec	MF510222	S. America/Hawaii	p
Hawaii	MF510220	S. America/Hawaii	p
Yunnan	GU138674.1	China/Taiwan	p
Yunnan2	JQ982468.1	China/Taiwan	p
Yunnan11	KF010583.1	China/Taiwan	p
Yunnan3	JQ982470.1	China/Taiwan	p
Sichuan	JQ982470.1	China/Taiwan	p
NFU2	KJ782300.1	China/Taiwan	p
T1F4S3	MF510223	Africa	p
T2F3S4	MF510224	Africa	p
T1F3S2	MF510225	Africa	p
T2F5S2	MF510226	Africa	p
T1F1S1	MF510227	Africa	p
T2F2S5	MF510228	Africa	p
T1F7S3	MF510229	Africa	p
T1F4S1	MF510230	Africa	p
T1F7S1	MF510231	Africa	p
T2F1S3	MF510232	Africa	p
T2F2S4	MF510233	Africa	p
T1F7S2	MF510234	Africa	p
T1F6S2	MF510235	Africa	p
T1F2S2	MF510236	Africa	p

Table 8.1 Site in the CP ORF found to have an ARI index significantly greater than 95% of instances generated by a permutation test with randomised (rather than geographic) clustering of sequences. Column headers are codon numbers within the ORF. Regions are separated by horizontal lines. Accession numbers and clades of isolates are given, and remain the same for following tables

Isolate	Accession	Clade	81
T1F3S1	MF510237	Africa	p
T1F5S2	MF510238	Africa	p
T2F5S1	MF510239	Africa	p
T2F1S1	MF510240	Africa	p
T1F8S3	MF510241	Africa	p
T1F6S3	MF510244	Africa	p
T1F5S3	MF510245	Africa	p
T2F1S2	MF510246	Africa	p
T1F8S2	MF510247	Africa	p
IPA	JX286709.1	Africa	p
T1F6S1	MF510250	Africa	p
ANETF1S2	KP798452.2	Africa	p
ANETF3S2	KP798454.2	Africa	p
ANETF4S2	MF510243	Africa	p
ANETF2S1	KP798453.2	Africa	p
LETF2S1	KP798455.2	Africa	p
M1	KP772217.1	Africa	p
B1_S1	MF510242	Africa	p
B2_S2	MF510248	Africa	p
FERARWANDA	MF510249	Africa	p
IPARwanda3	KF744394.1	Africa	p
B3_S3	MF510251	Africa	p
Rwanda1	KP851970.3	Africa	p

Table 8.2 Site in the P7b ORF found to have an ARI index significantly greater than 95% of instances generated by a permutation test with randomised (rather than geographic) clustering of sequences. Column headers are codon numbers within the ORF. Regions are separated by horizontal lines

Isolate	7
KS1	d
Nebraska	d
Porto_Severe_Ec	d
Porto_Original_Ec	d
Sta_Elena_Ec	d
Hawaii	d
Yunnan	d
Yunnan2	e
Yunnan11	e
Yunnan3	e
Sichuan	e
NFU2	e
T1F4S3	e
T2F3S4	e
T1F3S2	e
T2F5S2	e
T1F1S1	e
T2F2S5	e
T1F7S3	e
T1F4S1	e
T1F7S1	e
T2F1S3	e
T2F2S4	e
T1F7S2	e
T1F6S2	e
T1F2S2	e
T1F3S1	e
T1F5S2	e
T2F5S1	e

Table 8.2 Site in the P7b ORF found to have an ARI index significantly greater than 95% of instances generated by a permutation test with randomised (rather than geographic) clustering of sequences. Column headers are codon numbers within the ORF. Regions are separated by horizontal lines

Isolate	7
T2F1S1	e
T1F8S3	e
T1F6S3	e
T1F5S3	e
T2F1S2	e
T1F8S2	e
IPA	e
T1F6S1	e
ANETF1S2	e
ANETF3S2	e
ANETF4S2	e
ANETF2S1	e
LETF2S1	e
M1	e
B1_S1	e
B2_S2	e
FERARWANDA	e
IPARwanda3	e
B3_S3	e
Rwanda1	e

Table 8.3 Sites in the P32 ORF found to have an ARI index significantly greater than 95% of instances generated by a permutation test with randomised (rather than geographic) clustering of sequences. Column headers are codon numbers within the ORF. Regions are separated by horizontal lines

Isolate	60	65	88	110	124	128	209	224	231	247	248	254	285
KS1	v	n	v	i	f	n	c	y	e	g	a	i	s
Nebraska	v	n	v	i	s	n	c	y	e	g	a	i	s
Porto_Severe_Ec	v	n	i	v	a	d	r	f	e	g	a	i	s
Porto_Original_Ec	v	n	i	i	a	d	r	f	e	g	a	i	s
Sta_Elena_Ec	v	n	i	i	a	d	r	f	e	g	a	i	s
Hawaii	v	n	i	v	a	d	r	f	e	g	a	i	s
Yunnan	v	s	v	i	a	d	r	f	e	g	t	i	s
Yunnan2	v	s	v	i	a	d	r	f	k	r	t	v	s
Yunnan11	v	s	v	i	a	d	r	f	k	r	t	v	s
Yunnan3	v	n	v	i	a	d	r	f	e	g	t	v	p
Sichuan	v	s	v	i	a	d	r	f	e	g	t	v	p
NFU2	v	n	v	i	a	d	r	f	e	g	a	v	s
T1F4S3	a	n	v	i	a	d	r	f	e	g	t	v	s
T2F3S4	a	n	v	i	a	d	r	f	e	g	t	v	s
T1F3S2	a	n	v	i	a	d	r	f	e	g	t	v	s
T2F5S2	a	n	v	i	a	d	r	f	e	g	t	v	s
T1F1S1	a	n	v	i	a	d	r	f	e	g	t	v	s
T2F2S5	a	n	v	i	a	d	r	f	e	g	t	v	s
T1F7S3	a	n	v	i	a	d	-	f	e	g	t	v	s
T1F4S1	a	n	v	i	a	d	r	f	e	g	t	v	s
T1F7S1	a	n	v	i	a	d	-	f	e	g	t	v	s
T2F1S3	a	n	v	i	a	d	r	f	e	g	t	v	s
T2F2S4	a	n	v	i	a	d	r	f	e	g	t	v	s
T1F7S2	a	n	v	i	a	d	r	f	e	g	t	v	s
T1F6S2	a	n	v	i	a	d	r	f	e	g	t	v	s
T1F2S2	a	n	v	i	a	d	r	f	e	g	t	v	s
T1F3S1	a	n	v	i	a	d	r	f	e	g	t	v	s
T1F5S2	a	n	v	i	a	d	r	f	e	g	t	v	s
T2F5S1	a	n	v	i	a	d	r	f	e	g	t	v	s

Table 8.3 Sites in the P32 ORF found to have an ARI index significantly greater than 95% of instances generated by a permutation test with randomised (rather than geographic) clustering of sequences. Column headers are codon numbers within the ORF. Regions are separated by horizontal lines

Isolate	60	65	88	110	124	128	209	224	231	247	248	254	285
T2F1S1	a	n	v	i	a	d	r	f	e	g	t	v	s
T1F8S3	a	n	v	i	a	d	r	f	e	g	t	v	s
T1F6S3	a	n	v	i	a	d	r	f	e	g	t	v	s
T1F5S3	a	n	v	i	a	d	r	f	e	g	t	v	s
T2F1S2	a	n	v	i	a	d	r	f	e	g	t	v	s
T1F8S2	a	n	v	i	a	d	r	f	e	g	t	v	s
IPA	a	n	v	i	a	d	r	f	e	g	t	v	s
T1F6S1	a	n	v	i	a	d	r	f	e	g	t	v	s
ANETF1S2	a	n	v	i	a	-	r	f	e	g	t	v	s
ANETF3S2	a	n	v	i	a	d	r	f	e	g	t	v	s
ANETF4S2	a	n	v	i	a	d	r	f	e	g	t	v	s
ANETF2S1	a	n	v	i	a	d	r	f	e	g	t	v	s
LETF2S1	a	n	v	i	a	d	r	f	e	g	t	v	s
M1	a	n	v	i	a	d	r	f	e	g	t	v	s
B1_S1	a	n	v	i	a	d	r	f	e	g	t	v	s
B2_S2	a	n	v	i	a	d	r	f	e	g	t	v	s
FERARWANDA	a	n	v	i	a	d	r	f	e	g	t	v	s
IPARwanda3	a	n	v	i	a	d	r	f	e	g	t	v	s
B3_S3	a	n	v	i	a	d	r	f	e	g	t	v	s
Rwanda1	a	n	v	i	a	d	r	f	e	g	t	v	s

Table 8.4 Sites in the first portion of the P50 ORF found to have an ARI index significantly greater than 95% of instances generated by a permutation test with randomised (rather than geographic) clustering of sequences. Column headers are codon numbers within the ORF. Regions are separated by horizontal lines

Isolate	17	43	54	59	65	69	80	83	103	117
KS1	t	q	y	m	q	k	a	r	k	n
Nebraska	t	q	y	m	q	k	a	r	k	n
Porto_Severe_Ec	t	q	y	m	q	k	v	k	r	k
Porto_Original_Ec	t	q	y	m	q	k	v	k	r	k
Sta_Elena_Ec	t	q	y	m	q	r	v	r	r	r
Hawaii	t	q	y	m	q	k	v	k	r	k
Yunnan	t	r	y	v	r	k	v	k	r	r
Yunnan2	t	r	y	v	r	e	v	k	r	r
Yunnan11	t	r	y	v	r	e	v	k	r	r
Yunnan3	t	r	y	m	r	k	v	k	r	r
Sichuan	t	r	y	v	r	k	v	k	r	r
NFU2	t	r	y	m	r	k	v	r	r	r
T1F4S3	m	r	h	m	r	k	v	k	r	r
T2F3S4	m	r	h	m	r	k	v	k	r	r
T1F3S2	m	r	h	m	r	k	v	k	r	r
T2F5S2	m	r	h	m	-	k	-	k	r	r
T1F1S1	m	r	h	m	r	k	v	k	r	r
T2F2S5	m	r	h	m	r	k	v	k	r	r
T1F7S3	m	r	h	m	r	k	v	k	r	r
T1F4S1	m	r	h	m	r	k	v	k	r	r
T1F7S1	m	r	h	m	r	k	v	k	r	r
T2F1S3	m	r	h	m	r	k	v	k	r	r
T2F2S4	m	r	h	m	r	k	v	k	r	r
T1F7S2	m	r	h	m	r	k	v	k	r	r
T1F6S2	m	r	h	m	r	k	v	k	r	r
T1F2S2	m	r	h	m	r	k	v	k	r	r
T1F3S1	m	r	h	m	r	k	v	k	r	r
T1F5S2	m	r	h	m	r	e	v	k	r	r
T2F5S1	m	r	h	m	r	k	v	k	r	r

Table 8.4 Sites in the first portion of the P50 ORF found to have an ARI index significantly greater than 95% of instances generated by a permutation test with randomised (rather than geographic) clustering of sequences. Column headers are codon numbers within the ORF. Regions are separated by horizontal lines

Isolate	17	43	54	59	65	69	80	83	103	117
T2F1S1	m	r	h	m	r	k	v	k	r	r
T1F8S3	m	r	h	m	r	k	v	k	r	r
T1F6S3	m	r	h	m	r	k	v	k	r	r
T1F5S3	m	r	h	m	r	k	v	k	r	r
T2F1S2	m	r	h	m	r	k	v	k	r	r
T1F8S2	m	r	h	m	r	k	v	k	r	r
IPA	m	r	h	m	r	k	v	k	r	r
T1F6S1	m	r	h	m	r	k	v	k	r	r
ANETF1S2	m	r	h	m	r	-	v	k	r	r
ANETF3S2	m	r	h	m	r	k	v	k	r	r
ANETF4S2	m	r	h	m	r	k	v	k	r	r
ANETF2S1	m	r	h	m	r	k	v	k	r	r
LETF2S1	m	r	h	m	r	k	v	k	r	r
M1	m	r	h	m	r	k	v	k	r	r
B1_S1	m	r	h	m	r	k	v	k	r	r
B2_S2	m	r	h	m	r	k	v	k	r	r
FERARWANDA	m	r	h	m	r	k	v	k	r	r
IPARwanda3	m	r	h	m	r	k	v	k	r	r
B3_S3	m	r	h	m	r	k	v	k	r	r
Rwanda1	m	r	h	m	r	k	v	k	r	r

Table 8.5 Sites in the second portion of the P50 ORF found to have an ARI index significantly greater than 95% of instances generated by a permutation test with randomised (rather than geographic) clustering of sequences. Column headers are codon numbers within the ORF. Regions are separated by horizontal lines

Isolate	118	124	170	190	195	218	350	390
KS1	f	r	i	r	a	i	h	e
Nebraska	h	r	i	r	a	i	h	e
Porto_Severe_Ec	h	r	t	q	v	f	h	e
Porto_Original_Ec	h	r	t	q	v	f	h	d
Sta_Elena_Ec	h	r	t	q	v	f	h	d
Hawaii	h	r	t	q	v	f	h	e
Yunnan	h	k	t	q	v	f	q	e
Yunnan2	h	k	t	q	v	f	q	e
Yunnan11	h	k	t	q	v	f	q	e
Yunnan3	h	k	t	q	v	f	q	e
Sichuan	h	k	t	q	v	f	q	e
NFU2	r	r	t	q	v	f	q	e
T1F4S3	h	k	t	q	v	f	q	e
T2F3S4	h	k	t	q	v	f	q	e
T1F3S2	h	k	t	q	v	f	q	e
T2F5S2	h	k	t	q	v	f	q	e
T1F1S1	h	k	t	q	v	f	q	e
T2F2S5	h	k	t	q	v	f	q	e
T1F7S3	h	k	t	q	v	f	q	e
T1F4S1	h	k	t	q	v	f	q	e
T1F7S1	h	k	t	q	v	f	q	e
T2F1S3	h	k	t	q	v	f	q	e
T2F2S4	h	k	t	q	v	f	q	e
T1F7S2	h	k	t	q	v	f	q	e
T1F6S2	h	k	t	q	v	f	q	e
T1F2S2	h	k	t	q	v	f	q	e
T1F3S1	h	k	t	q	v	f	q	e
T1F5S2	h	k	t	q	v	f	q	e
T2F5S1	h	k	t	q	v	f	q	e

Table 8.5 Sites in the second portion of the P50 ORF found to have an ARI index significantly greater than 95% of instances generated by a permutation test with randomised (rather than geographic) clustering of sequences. Column headers are codon numbers within the ORF. Regions are separated by horizontal lines

Isolate	118	124	170	190	195	218	350	390
T2F1S1	h	k	t	q	v	f	q	e
T1F8S3	h	k	t	q	v	f	q	e
T1F6S3	h	k	t	q	v	f	q	e
T1F5S3	h	k	t	q	v	f	q	e
T2F1S2	h	k	t	q	v	f	q	e
T1F8S2	h	k	t	q	v	f	q	e
IPA	h	k	t	q	v	f	q	e
T1F6S1	h	k	t	q	v	f	q	e
ANETF1S2	h	k	t	q	v	f	q	e
ANETF3S2	h	k	t	q	v	f	q	e
ANETF4S2	h	k	t	q	v	f	q	e
ANETF2S1	h	k	t	q	v	f	q	e
LETF2S1	h	k	t	q	v	f	q	e
M1	h	k	t	q	v	f	q	e
B1_S1	h	k	t	q	v	f	q	e
B2_S2	h	k	t	q	v	f	q	e
FERARWANDA	h	k	t	q	v	f	q	e
IPARwanda3	h	k	t	q	v	f	q	e
B3_S3	h	k	t	q	v	f	q	e
Rwanda1	h	k	t	q	v	f	q	e

Table 8.6 Site in the post-readthrough region of the P111 ORF found to have an ARI index significantly greater than 95% of instances generated by a permutation test with randomised (rather than geographic) clustering of sequences. Column headers are codon numbers within the ORF. Regions are separated by horizontal lines

Isolate	508	533	688	767	870	933
KS1	d	n	s	l	t	k
Nebraska	d	n	s	l	t	k
Porto_Severe_Ec	n	d	s	l	a	r
Porto_Original_Ec	n	n	s	l	t	r
Sta_Elena_Ec	n	n	s	l	t	r
Hawaii	n	d	s	l	t	r
Yunnan	n	n	t	m	a	r
Yunnan2	n	n	t	l	a	r
Yunnan11	n	n	t	m	a	r
Yunnan3	n	n	t	m	a	r
Sichuan	n	n	t	l	a	r
NFU2	n	n	t	l	s	r
T1F4S3	n	n	t	l	a	r
T2F3S4	n	n	t	l	a	r
T1F3S2	n	n	t	l	a	r
T2F5S2	n	n	t	l	a	r
T1F1S1	n	n	t	l	a	r
T2F2S5	n	n	t	l	a	r
T1F7S3	n	n	t	l	a	r
T1F4S1	n	n	t	l	a	r
T1F7S1	n	n	t	l	a	r
T2F1S3	n	n	t	l	a	r
T2F2S4	n	n	t	l	a	r
T1F7S2	n	n	t	l	a	r
T1F6S2	n	n	t	l	a	r
T1F2S2	n	n	t	l	a	r
T1F3S1	n	n	t	l	a	r
T1F5S2	n	n	t	l	a	r
T2F5S1	n	n	t	l	a	r

Table 8.6 Site in the post-readthrough region of the P111 ORF found to have an ARI index significantly greater than 95% of instances generated by a permutation test with randomised (rather than geographic) clustering of sequences. Column headers are codon numbers within the ORF. Regions are separated by horizontal lines

Isolate	508	533	688	767	870	933
T2F1S1	n	n	t	l	a	r
T1F8S3	n	n	t	l	a	r
T1F6S3	n	n	t	l	a	r
T1F5S3	n	n	t	l	a	r
T2F1S2	n	n	t	l	a	r
T1F8S2	n	n	t	l	a	r
IPA	n	n	t	l	a	r
T1F6S1	n	n	t	l	a	r
ANETF1S2	n	n	t	l	a	r
ANETF3S2	n	n	t	l	a	r
ANETF4S2	n	n	t	l	a	r
ANETF2S1	n	n	t	l	a	r
LETF2S1	n	n	t	l	a	r
M1	n	n	t	l	a	r
B1_S1	n	n	t	l	a	r
B2_S2	n	n	t	l	a	r
FERARWANDA	n	n	t	l	a	r
IPARwanda3	n	n	t	l	a	r
B3_S3	n	n	t	l	a	r
Rwanda1	n	n	t	l	a	r

Table 8.7 Sites in the first portion of the P31 ORF found to have an ARI index significantly greater than 95% of instances generated by a permutation test with randomised (rather than geographic) clustering of sequences. Column headers are codon numbers within the ORF. Regions are separated by horizontal lines. Asterisk indicates stop codon.

Isolate	75	76	101	119	120	148	171	180	193	200	206	210	217
KS1	r	f	d	d	t	k	p	p	r	v	r	r	p
Nebraska	r	f	d	d	t	k	p	p	r	v	r	r	p
Porto_Severe_Ec	*	f	d	n	t	r	h	p	k	a	q	h	l
Porto_Original_Ec	*	f	d	d	t	r	h	p	k	a	q	h	p
Sta_Elena_Ec	*	f	d	d	t	r	h	p	k	a	q	h	p
Hawaii	*	f	d	d	t	r	h	p	k	a	q	h	l
Yunnan	r	f	d	v	t	k	h	p	k	a	q	h	p
Yunnan2	r	i	n	n	a	k	h	p	k	a	q	h	p
Yunnan11	r	i	n	n	a	k	h	p	k	a	q	h	p
Yunnan3	r	i	n	n	a	k	h	p	k	a	q	h	p
Sichuan	r	i	n	n	a	k	h	p	k	a	q	h	p
NFU2	r	i	n	n	a	k	h	p	k	a	q	h	l
T1F4S3	r	i	n	n	t	k	h	-	k	a	q	h	p
T2F3S4	r	i	n	n	t	k	h	-	k	a	q	h	p
T1F3S2	r	i	n	n	t	k	h	l	k	a	q	h	p
T2F5S2	r	i	n	n	t	k	h	l	k	a	q	h	p
T1F1S1	r	i	n	n	t	k	h	l	k	a	q	h	p
T2F2S5	r	i	n	n	t	k	h	l	k	a	q	h	p
T1F7S3	r	i	n	n	t	k	h	l	k	a	q	h	p
T1F4S1	r	i	n	n	t	k	h	l	k	a	q	h	p
T1F7S1	r	i	n	n	t	k	h	l	k	a	q	h	p
T2F1S3	r	i	n	n	t	k	h	l	k	a	q	h	p
T2F2S4	r	i	n	n	t	k	h	l	k	a	q	h	p
T1F7S2	r	i	n	n	t	k	h	l	k	a	q	h	p
T1F6S2	r	i	n	n	t	k	h	l	k	a	q	h	p
T1F2S2	r	i	n	n	t	k	h	l	k	a	q	h	p
T1F3S1	r	i	n	n	t	k	h	l	k	a	q	h	p
T1F5S2	r	i	n	n	t	k	h	l	k	a	q	h	p
T2F5S1	r	i	n	n	t	k	h	p	k	a	q	h	p

Table 8.7 Sites in the first portion of the P31 ORF found to have an ARI index significantly greater than 95% of instances generated by a permutation test with randomised (rather than geographic) clustering of sequences. Column headers are codon numbers within the ORF. Regions are separated by horizontal lines. Asterisk indicates stop codon.

Isolate	75	76	101	119	120	148	171	180	193	200	206	210	217
T2F1S1	r	i	n	n	t	k	h	l	k	a	q	h	p
T1F8S3	r	i	n	n	t	k	h	l	k	a	q	h	p
T1F6S3	r	i	n	n	t	k	h	l	k	a	q	h	p
T1F5S3	r	i	n	n	t	k	h	l	k	a	q	h	p
T2F1S2	r	i	n	n	t	k	h	l	k	a	q	h	p
T1F8S2	r	i	n	n	t	k	h	l	k	a	q	h	p
IPA	r	i	n	n	t	k	h	l	k	a	q	h	p
T1F6S1	r	i	n	n	t	k	h	l	k	a	q	h	p
ANETF1S2	r	i	n	n	t	k	h	l	k	a	q	h	p
ANETF3S2	r	i	n	n	t	k	h	l	k	a	q	h	p
ANETF4S2	r	i	n	n	t	k	h	l	k	a	q	h	p
ANETF2S1	r	i	n	n	t	k	h	l	k	a	q	h	p
LETF2S1	r	i	n	n	t	k	h	l	k	a	q	h	p
M1	r	i	n	n	t	k	h	l	k	a	q	h	p
B1_S1	r	i	n	n	t	k	h	l	k	a	q	h	p
B2_S2	r	i	n	n	t	k	h	l	k	a	q	h	p
FERARWANDA	r	i	n	n	t	k	h	l	k	a	q	h	p
IPARwanda3	r	i	n	n	t	k	h	l	k	a	q	h	p
B3_S3	r	i	n	n	t	k	h	l	k	a	q	h	p
Rwanda1	r	i	n	n	t	k	h	l	k	a	q	h	p

Table 8.8 Sites in the second portion of the P31 ORF found to have an ARI index significantly greater than 95% of instances generated by a permutation test with randomised (rather than geographic) clustering of sequences. Column headers are codon numbers within the ORF. Regions are separated by horizontal lines.

Isolate	218	223	228	232	236	238	240	251	252	257
KS1	n	n	i	s	a	l	r	f	p	p
Nebraska	n	n	i	s	v	l	r	f	p	p
Porto_Severe_Ec	n	s	t	s	a	p	r	f	p	p
Porto_Original_Ec	n	s	t	s	a	p	r	f	p	p
Sta_Elena_Ec	n	s	t	l	a	p	r	f	p	p
Hawaii	n	n	t	l	a	p	r	f	p	p
Yunnan	n	n	t	s	v	l	k	f	p	l
Yunnan2	n	n	t	s	a	l	k	s	p	l
Yunnan11	n	n	t	s	a	l	k	f	p	l
Yunnan3	n	n	t	s	a	l	k	f	p	l
Sichuan	n	n	t	s	a	l	k	f	p	l
NFU2	n	n	t	s	a	l	k	s	p	l
T1F4S3	s	n	t	s	a	l	-	s	-	l
T2F3S4	s	n	t	s	a	l	-	s	q	l
T1F3S2	s	n	t	s	a	l	k	s	q	l
T2F5S2	s	n	t	s	a	l	k	s	q	l
T1F1S1	s	n	t	s	a	l	k	s	q	l
T2F2S5	s	n	t	s	a	l	k	s	q	l
T1F7S3	s	n	t	s	a	l	k	s	q	l
T1F4S1	s	n	t	s	a	l	k	s	q	l
T1F7S1	s	n	t	s	a	l	k	s	q	l
T2F1S3	s	n	t	s	a	l	k	s	q	l
T2F2S4	s	n	t	s	a	l	k	s	q	l
T1F7S2	s	n	t	s	a	l	r	s	q	l
T1F6S2	s	n	t	s	a	l	k	s	q	l
T1F2S2	s	n	t	s	a	l	k	s	q	l
T1F3S1	s	n	t	s	a	l	k	s	q	l
T1F5S2	s	n	t	s	a	l	k	s	q	l
T2F5S1	s	n	t	s	a	l	k	s	q	l

Table 8.8 Sites in the second portion of the P31 ORF found to have an ARI index significantly greater than 95% of instances generated by a permutation test with randomised (rather than geographic) clustering of sequences. Column headers are codon numbers within the ORF. Regions are separated by horizontal lines.

Isolate	218	223	228	232	236	238	240	251	252	257
T2F1S1	s	n	t	s	a	l	k	s	q	l
T1F8S3	s	n	t	s	a	l	k	s	q	l
T1F6S3	s	n	t	s	a	-	k	s	q	l
T1F5S3	s	n	t	s	a	l	k	s	q	l
T2F1S2	s	n	t	s	a	l	k	s	q	l
T1F8S2	s	n	t	s	a	l	k	s	q	l
IPA	s	n	t	s	a	l	k	s	q	l
T1F6S1	s	n	t	s	a	l	k	y	q	l
ANETF1S2	s	n	t	s	a	l	k	s	q	l
ANETF3S2	s	n	t	s	a	l	k	s	q	l
ANETF4S2	s	n	t	s	a	l	k	s	q	l
ANETF2S1	s	n	t	s	a	l	k	s	q	l
LETF2S1	s	n	t	s	a	l	k	s	q	l
M1	s	n	t	s	a	l	k	s	q	l
B1_S1	s	n	t	s	a	l	k	s	q	l
B2_S2	s	n	t	s	a	l	k	s	q	l
FERARWANDA	s	n	t	s	a	l	k	s	q	l
IPARwanda3	s	n	t	s	a	l	k	s	q	l
B3_S3	s	n	t	s	a	l	k	s	q	l
Rwanda1	s	n	t	s	a	l	k	s	q	l

8.3 Primers

Table 8.9 Primer sequences

Primer name	Sequence	Notes
MCMV_RDRP_R1	AGTTCCTGTCTGACTCTGCC	pMCM41 cloning, sequencing
MCMV_RDRP_F1	CTTCTCACACAATTGCTGCG	pMCM41 cloning, sequencing
MCMV_P32_R1	CCGTCGGACCATTGAGTTTC	pMCM41 cloning, sequencing
MCMV_P32_F1	TCAACGAGCCCACTGATGAC	pMCM41 cloning, sequencing
MCM41_J2R1	TGTGGAATTGTGAGCGGATA	pMCM41 cloning, sequencing
MCM41_J2F1	CTGGCAGCACAATCACAGAT	pMCM41 cloning, sequencing
MCM41_J1R1	GACCATGTGGGTTGGATAGG	pMCM41 cloning, sequencing
MCM41_J1F1	AGGGAAGAAAGCGAAAGGAG	pMCM41 cloning, sequencing
MCM41_500_R1	GTAGGGGTGTCTCCTCCACA	pMCM41 cloning, sequencing
MCM41_500_F1	CTTTCAGGCACGGTCTCTCT	pMCM41 cloning, sequencing
MCM41_1300_R1	TCGCACTGATGTTGGAAGAG	pMCM41 cloning, sequencing
MCM41_1300_F1	TCAAACCGACTTGCATCAAA	pMCM41 cloning, sequencing
MCM41_1200F	ACCCAACTGACTCCCGTACT	pMCM41 cloning, sequencing
MCM41_2200R	ACGCTATGTTCCCAGCGTAG	pMCM41 cloning, sequencing
MCM41_2900_R1	TCGGGTACACACCAGATGAA	pMCM41 cloning, sequencing
MCM41_2900_F1	ATCCGCTAGTGGTGTCTGCT	pMCM41 cloning, sequencing
MCM41_2000_R1	TTGGGGTGGTCCGACTATTA	pMCM41 cloning, sequencing
MCM41_2000_F1	TTCAAACGTAGGGGTCCAAG	pMCM41 cloning, sequencing
MCM41_amp_R1	ATTCACTGGCCGTCGTTT	pMCM41 cloning, sequencing

Table 8.9 Primer sequences

Primer name	Sequence	Notes
MCM41_amp_F1	ACACGTTCTGGTTCCAGAG	pMCM41 cloning, sequencing
pMCM41_J2R3	AGCCCTCCCGTATCGTAGTT	pMCM41 cloning, sequencing
pMCM41_J2R2	GCCTACATACCTCGCTCTGC	pMCM41 cloning, sequencing
pMCM41_J1F3	CTCTTCGCTATTACGCCAGC	pMCM41 cloning, sequencing
pMCM41_J1F2	GTTTTCCCAGTCACGACGTT	pMCM41 cloning, sequencing
Zm_ubi1_F1	GATGCAACAGCTTCACAGGA	pUbi1 cloning
Zm_ubi1_R1	CTGAGACGGAGCACAAGGT	pUbi1 cloning
Zm_Ubi1_promseq_F1	AGTGTGCATGTGTTCTCCTTTT	LABami construct sequencing
Zm_Ubi1_promseq_R1	TGCATATGCCATCATCCAAG	LABami construct sequencing
Zm_Ubi1_promseq_F2	ACCCTCTTTCCCCAACCTC	LABami construct sequencing
gib_amiRNA_pTF101_R	TCGAGCTCGGTACCCCATCTCATTCTTGTGGTTTAAG	LABami construct cloning
gib_pTF101_pUbi1_F	CTCTAGAGGATCCCCCTGCAGTGCAGCGTGACC	LABami construct cloning
gib_pUbi1_amiRNA_F	TTCTGCAGGAAGTAATCAGGTCTAGGGAG	LABami construct cloning
gib_pUbi1_amiRNA_R	GATTACTTCCTGCAGAAGTAACACCAAACAACAG	LABami construct cloning
Zm_Ubi1_promseq_R2	AGACATGCAATGCTCATTATCTC	LABami construct cloning
Zm_Ubi1_R2	CTGCAGAAGTAACACCAAACAACAGGGTGAG	LABami construct cloning
HindIII_pUbi1_F1	CGGGTAAAGCTTGTTGTATGGATC	LABami construct cloning
pUbi1_R3	CTGCAGAAGTAACACCAAACAACA	LABami construct cloning
LABami_XhoI_F1	CGGGTACTCGAGGTTTGGTGTTACTT	LABami construct cloning
LABami_XbaI_R1	TACCCGTCTAGATACCCGGAGCTC	LABami construct cloning
Zm_Ubi1_BamHI_F	GTTGTATGGATCCCTGCAGTGCAGCG	LABami construct cloning

Table 8.9 Primer sequences

Primer name	Sequence	Notes
Zm_Ubi1_BamHI_F2	GTTGTATGGATCCCTGCAGTGCAGCGTGAC	LABami construct cloning
MCM41_650_F1	CCTATCCAACCCACATGGTC	pMCM41 cloning, sequencing
MCM41_650_R1	TCATAACCATCATCGCTCCA	pMCM41 cloning, sequencing
P50_TOPO_step1_F1	CACCATGGCGACCCTACCTTCAA	MCMV Gateway library cloning
P33_TOPO_tail_step1_R1	ACAATGGACGGCCCTCTGTTCT	MCMV Gateway library cloning
P33_TOPO_tail_step1_Gsub_F1	TCAATTTCAACGGAGCTGGAGTGT	MCMV Gateway library cloning
P33_TOPO_step2_nostop_R1	TTGTAGCTGAGGGCACGATCC	MCMV Gateway library cloning
P33_TOPO_step2_fullstop_R1	TGGACGGCCCTCTGTTCTTCA	MCMV Gateway library cloning
P33_TOPO_step1_newstart_F1	CACCTTCAACATGGCTGGAGTGTGTG	MCMV Gateway library cloning
P33_TOPO_head_step1_Gsub_R1	ACACTCCAGCTCCGTTGAAATTGAAGT	MCMV Gateway library cloning
P33_TOPO_head_step1_F1	TGGCAGGAGGTCCAAGAGAGC	MCMV Gateway library cloning
P32_TOPO_tail_step1_nostart_F1	CCGTGCACATACGGCGACCC	MCMV Gateway library cloning
P32_TOPO_tail_step1_R1	CACTGTGCCTTCTTCGTCCACA	MCMV Gateway library cloning
P32_TOPO_step3_stop_R1	TCGTCCACAATCTCTGGAACCA	MCMV Gateway library cloning
P32_TOPO_step3_nostop_R1	AGAACGTGTTTCGTCAGACAGTCCTG	MCMV Gateway library cloning
P32_TOPO_step3_F1	CACCATGCCCTCTCCGTGCA	MCMV Gateway library cloning
P32_TOPO_step2_R1	TCGTCCACAATCTCTGGAACCA	MCMV Gateway library cloning
P32_TOPO_step2_F1	CCAACGCGCTAAACACGACCT	MCMV Gateway library cloning
P32_TOPO_head_step1_nostart_R1	GGGTCGCCGTATGTGCACGG	MCMV Gateway library cloning
P32_TOPO_head_step1_F1	ACCCCAACGCGCTAAACACG	MCMV Gateway library cloning
P111_TOPO_tail_step1_newstart_F1	CACCGGAGCTGAAAATGGGGTGTCT	MCMV Gateway library cloning

Table 8.9 Primer sequences

Primer name	Sequence	Notes
P111_TOPO_tail_step1_R1	TCTTCGGATCGCACGGCAGT	MCMV Gateway library cloning
P111_TOPO_tail_step1_Qsub_F1	AGGAGCTGAAACAGGGGTGTCTTGAA	MCMV Gateway library cloning
P111_TOPO_tail_step1_Asub_F1	AGGAGCTGAAAGCGGGGTGTCTTGAA	MCMV Gateway library cloning
P111_TOPO_step2_nostop_R1	CGTCGGTGGGAGGGGATTG	MCMV Gateway library cloning
P111_TOPO_step2_fullstop_R1	GGCAGTGCCGGTCTGTTGTC	MCMV Gateway library cloning
P111_TOPO_head_step1_Qsub_R1	AGACACCCCTGTTTCAGCTCCTGG	MCMV Gateway library cloning
P111_TOPO_head_step1_F1	TCATGCCCTCTCCGTGCACA	MCMV Gateway library cloning
P111_TOPO_head_step1_Asub_R1	AGACACCCCGCTTTCAGCTCCTGG	MCMV Gateway library cloning
MCMV_FULL_TOPO_R1	GGGCCGGAAGAGAGGGG	MCMV Gateway library cloning
MCMV_FULL_TOPO_F1	CACCAGGTAATCTGCGGCAAC	MCMV Gateway library cloning
CP_TOPO_nostop_R1	CGGCTCACCTTCTGCTCCATGATTTG	MCMV Gateway library cloning
CP_TOPO_R1	ACCTCATGCCGGCTCACCTT	MCMV Gateway library cloning
CP_TOPO_F1	CACCATGGCGGCAAGTAGC	MCMV Gateway library cloning
P7_TOPO_nostop_Gsub_R1	TCCAGCTCCGTTGAAATTGAAGTGG	MCMV Gateway library cloning
P7_TOPO_fullstop_R1	ACACTCCAGCTTAGTTGAAATTGAAGT	MCMV Gateway library cloning
P7_TOPO_F1	CACCATGTCTTCTTCTCAAACACAATCC	MCMV Gateway library cloning
P50_TOPO_step1_nostop_Qsub_R1	ACACCCCTGTTTCAGCTCCTGGA	MCMV Gateway library cloning
P50_TOPO_step1_fullstop_R1	AGACACCCTTATTTCAGCTCCTGG	MCMV Gateway library cloning
APV_400_F1	GCAGAACGAGGAGGAAGTGA	RT-PCR of VLSs
APV_1300_R1	CGAATCTCCCACGTGGCTTA	RT-PCR of VLSs
APV_1000_F1	GCAGTGTAGCGTAGAGTGGG	RT-PCR of VLSs

Table 8.9 Primer sequences

Primer name	Sequence	Notes
APV_1900_R1	TAACCCCGCGTCTTTCACAA	RT-PCR of VLSs
ALPV_350_F1	ACTGCGTACGTCCTTGATCG	RT-PCR of VLSs
ALPV_1250_R1	AAATGCGTTGCGTATGGTCG	RT-PCR of VLSs
ALPV_1200_F1	AATGACCAACAACCCGACCA	RT-PCR of VLSs
ALPV_2150_R1	GGTAGCTGGCAAGAGAGAGC	RT-PCR of VLSs
BRVF_400_F1	AGCAGTCTGGAGGCATGAAC	RT-PCR of VLSs
BRVF_1250_R1	CCGTATATGTCGGCAGGCAT	RT-PCR of VLSs
BRVF_1150_F1	GGGAGTACATTGGCGGGAAA	RT-PCR of VLSs
BRVF_2050_R1	GGAGGTAGCGGAACATCAGG	RT-PCR of VLSs
CYVA_200_F1	GAGTTTCTCCAACCGGCTGA	RT-PCR of VLSs
CYVA_1100_R1	TGCGACATGGAACGAACTCA	RT-PCR of VLSs
CYVA_1050_F1	TCCACATTCCCAGAACACCG	RT-PCR of VLSs
CYVA_1900_R1	AATACCTGGCCGAAGTTGGG	RT-PCR of VLSs
MWLMV_400_F1	GAGGGCATGTATGACGGGAG	RT-PCR of VLSs
MWLMV_1250_R1	CACCCTTCCCATCCTTCACC	RT-PCR of VLSs
MWLMV_1250_F1	GTGAAGGATGGGAAGGGTGG	RT-PCR of VLSs
MWLMV_2200_R1	GTCCTTGGCTATGGCAGTGT	RT-PCR of VLSs
MYDV_50_F1	GCGAGGTAAAATTCACCGGC	RT-PCR of VLSs
MYDV_900_R1	ATGAAGCGTAACCACCAGGG	RT-PCR of VLSs
MYDV_900_F1	ATAACGACGCCCTCAACTCC	RT-PCR of VLSs
MYDV_1800_R1	CGGTTGTTCTTTGGGCGTTT	RT-PCR of VLSs

Table 8.9 Primer sequences

Primer name	Sequence	Notes
NLHV_500_F1	AACTGCCGTGTCTGTTGGAA	RT-PCR of VLSs
NLHV_1300_R1	ACACCTTGCTTGGCTATCCC	RT-PCR of VLSs
NLHV_1150_F1	AGGTTGTGGTACAGGGGAGT	RT-PCR of VLSs
NLHV_2000_R1	AAACAAACTGCCGCAAACCA	RT-PCR of VLSs
RPV_200_F1	TTTGGAAGACGTGTGCGAGA	RT-PCR of VLSs
RPV_1100_R1	TCGTGCAGCTGAGAACGAAT	RT-PCR of VLSs
RPV_1350_F1	GATGGGTACACTGGACAGCC	RT-PCR of VLSs
RPV_2300_R1	CTCTCGCTCGCAGCAAATTC	RT-PCR of VLSs
RVF_100_F1	ATCGCGGAATGGACACCTTT	RT-PCR of VLSs
RVF_1000_R1	GGCAGACGAGAACTGGTCAA	RT-PCR of VLSs
RVF_800_F1	TCACAGACACGCTCCAGTTC	RT-PCR of VLSs
RVF_1800_R1	GGTAAGGCGTACACTGCGTA	RT-PCR of VLSs
SABV_100_F1	CGCATAGATGTGCAGTCCCT	RT-PCR of VLSs
SABV_1000_R1	CTGGATCAGATTCCGAAGCCA	RT-PCR of VLSs
SABV_1400_F1	CCGAATCGACCTCTTGGACC	RT-PCR of VLSs
SABV_2300_R1	TTCCTTCTCCTTCTCCGGGT	RT-PCR of VLSs
N_745_rc24	TAGAAGCGAACTGCTACCCTACAG	Probe for sRNA Northerns
N_745_rc21	TAGAAGCGAACTGCTACCCTA	Probe for sRNA Northerns
N_2474_rc24	TCAATGTCTGTACAGTTCCTGCAG	Probe for sRNA Northerns
N_2474_rc21	TCAATGTCTGTACAGTTCCTG	Probe for sRNA Northerns
N_226_rc24	TTGCGCGTCAAACCATTCCTCCAG	Probe for sRNA Northerns

Table 8.9 Primer sequences

Primer name	Sequence	Notes
N_226_rc21	TTGCGCGTCAAACCATTCCTC	Probe for sRNA Northerns
N_2220_rc24	TCAATTGTGTGAGAAGCTCTGCAG	Probe for sRNA Northerns
N_2220_rc21	TCAATTGTGTGAGAAGCTCTG	Probe for sRNA Northerns
N_2151_rc24	TAACACTCACGTGTTGATCAACAG	Probe for sRNA Northerns
N_2151_rc21	TAACACTCACGTGTTGATCAA	Probe for sRNA Northerns
N_1831_rc24	TCGTAGAGGGAATTCACCGCTCAG	Probe for sRNA Northerns
N_1831_rc21	TCGTAGAGGGAATTCACCGCT	Probe for sRNA Northerns
N_1830_rc24	TGTAGAGGGAATTTATCGCTTCAG	Probe for sRNA Northerns
N_1830_rc21	TGTAGAGGGAATTTATCGCTT	Probe for sRNA Northerns
pTF101_seq_F1	TCCTAAGTTACGCGACAGGC	pTF101 plasmid sequencing
pTF101_seq_F2	GAGCTGATCGACCAGGAAGG	pTF101 plasmid sequencing
pTF101_seq_F3	CGACCATCGCAACCCATCTA	pTF101 plasmid sequencing
pTF101_seq_F4	CGCTGGCCGCTGAAATTAAA	pTF101 plasmid sequencing
pTF101_seq_F5	CGCTGATCGAATCCGCAAAG	pTF101 plasmid sequencing
pTF101_seq_F6	ATCGGCCGTTTTCTCTACCG	pTF101 plasmid sequencing
pTF101_seq_F7	GCAAAAAGCGCCTACCCTTC	pTF101 plasmid sequencing
pTF101_seq_F8	GTGAGCAAAAGGCCAGCAAA	pTF101 plasmid sequencing
pTF101_seq_F9	TTTTCTACGGGGTCTGACGC	pTF101 plasmid sequencing
pTF101_seq_F10	ATCAAAGAGTTCCTCCGCCG	pTF101 plasmid sequencing
pTF101_seq_F11	AACAAGCCATGAAAACCGCC	pTF101 plasmid sequencing
pTF101_seq_F12	CCAGTGCCAAGCTAATTCGC	pTF101 plasmid sequencing

Table 8.9 Primer sequences

Primer name	Sequence	Notes
pTF101_seq_F13	CTGAAGTCCAGCTGCCAGAA	pTF101 plasmid sequencing
pTF101_seq_F14	AGTAGAATGCTTGATTGCTTGAGA	pTF101 plasmid sequencing
pTF101_seq_F15	GGCAGAGGCATCTTCAACGA	pTF101 plasmid sequencing
pTF101_seq_F16	GCTTGGATCAGATTGTCGTTTCC	pTF101 plasmid sequencing
pTF101_seq_F17	CCAGGCAATCTACCAGGGC	pTF101 plasmid sequencing
pTF101_seq_F18	ATCACTGTGTGGCTTCAGGC	pTF101 plasmid sequencing
pTF101_seq_F19	GCATCTTCAACGATGGCCTT	pTF101 plasmid sequencing
pTF101_seq_F20	GGTACCGAGCTCGAATTCGT	pTF101 plasmid sequencing
pTF101_seq_F21	TTAATGAATCGGCCAACGCG	pTF101 plasmid sequencing
pTF101_seq_R1	CAATACGCAAACCGCCTCTC	pTF101 plasmid sequencing
pTF101_seq_R2	ACTAAGCTGCCGGGTTTGAA	pTF101 plasmid sequencing
pTF101_seq_R3	CGTAGGTGGTCAAGCATCCT	pTF101 plasmid sequencing
pTF101_seq_R4	GCGACTAAAACACGCGACAA	pTF101 plasmid sequencing
LABami_probe_R1	CTTAAATGCTTGGCGACACT	Southern probe generation
LABami_probe_F1	GGGAGTCAAAATTTGGTTGG	Southern probe generation
bar_r1	AAGTCCAGCTGCCAGAAAC	Southern probe generation
bar_f1	TCAACCACTACATCGAGACA	Southern probe generation
Zm_23S_F1	TGTACCCGAAACCGACACAG	Southern probe generation
Zm_23S_R1	TCACGACGTTCTGAACCCAG	Southern probe generation

References

Abel, P. P., Nelson, R. S., De, B., Hoffmann, N., Rogers, S. G., Fraley, R. T., and Beachy, R. N. (1986). Disease Development in Transgenic Plants That Express the Tobacco Mosaic Virus Coat Protein Gene. *Science*, 232:738–743.

Achon, M., Serrano, L., Clemente-Orta, G., and Sossai, S. (2017). First Report of Maize chlorotic mottle virus on a Perennial Host, Sorghum halepense, and Maize in Spain. *Plant Disease*, 101(2):393.

Achon, M. A., Serrano, L., Alonso-Dueñas, N., and Porta, C. (2007). Complete genome sequences of Maize dwarf mosaic and Sugarcane mosaic virus isolates coinfecting maize in Spain. *Archives of Virology*, 152(11):2073–2078.

Adams, I. P., Harju, V. A., Hodges, T., Hany, U., Skelton, A., Rai, S., Deka, M. K., Smith, J., Fox, A., Uzayisenga, B., Ngaboyisonga, C., Uwumukiza, B., Rutikanga, A., Rutherford, M., Ricthis, B., Phiri, N., and Boonham, N. (2014). First report of maize lethal necrosis disease in Rwanda. *New Disease Reports*, 29:22.

Adams, I. P., Miano, D. W., Kinyua, Z. M., Wangai, A., Kimani, E., Phiri, N., Reeder, R., Harju, V., Glover, R., Hany, U., Souza-Richards, R., Deb Nath, P., Nixon, T., Fox, A., Barnes, A., Smith, J., Skelton, A., Thwaites, R., Mumford, R., and Boonham, N. (2013). Use of next-generation sequencing for the identification and characterization of Maize chlorotic mottle virus and Sugarcane mosaic virus causing maize lethal necrosis in Kenya. *Plant Pathology*, 62(4):741–749.

Ade, J., Deyoung, B. J., Golstein, C., and Innes, R. W. (2007). Indirect activation of a plant nucleotide binding site Ã¢â‚¬Œ leucine-rich repeat protein by a bacterial protease. *Proc. Nat. Acad. Sci*, 104(7):2531–2536.

Ai, T., Zhang, L., Gao, Z., Zhu, C. X., and Guo, X. (2011). Highly efficient virus resistance mediated by artificial microRNAs that target the suppressor of PVX and PVY in plants. *Plant Biology*, 13(2):304–316.

Ali, Z., Abulfaraj, A., Idris, A., Ali, S., Tashkandi, M., and Mahfouz, M. M. (2015). CRISPR / Cas9-mediated viral interference in plants. *Genome Biology*, pages 1–11.

Alvarez-Quinto, R. A., Espinoza-Lozano, R. F., Mora-Pinargote, C. A., and Quito-Avila, D. F. (2017). Complete genome sequence of a variant of maize-associated totivirus from Ecuador. *Archives of Virology*, 162(4):1083–1087.

Asurmendi, S., Berg, R. H., Smith, T. J., Bendahmane, M., and Beachy, R. N. (2007). Aggregation of TMV CP plays a role in CP functions and in coat-protein-mediated resistance. *Virology*, 366:98–106.

Azzam, O., Yambao, M. L. M., Muhsin, M., McNally, K. L., and Umadhay, K. M. L. (2000). Genetic diversity of rice tungro spherical virus in tungro-endemic provinces of the Philippines and Indonesia. *Archives of Virology*, 145(6):1183–1197.

Bacheller, N. E. (2017). Detecting , Cloning , and Screening for Suppressors of RNA Silencing in Maize Chlorotic Mottle Virus and Sugarcane Mosaic Virus. Masters thesis.

Baulcombe, D. (2004). RNA silencing in plants. *Nature*, 431(7006):356–363.

Baulcombe, D. C. (1996). Mechanisms of Pathogen-Derived Resistance to Viruses in Transgenic Plants. *The Plant cell*, 8(October):1833–1844.

Bebber, D. P., Holmes, T., and Gurr, S. J. (2014). The global spread of crop pests and pathogens. *Global Ecology and Biogeography*, 23(12):1398–1407.

Bendahmane, M., Chen, I., Asurmendi, S., Bazzini, A. A., Szecsi, J., and Beachy, R. N. (2007). Coat protein-mediated resistance to TMV infection of Nicotiana tabacum involves multiple modes of interference by coat protein. *Virology*, 366(1):107–116.

Bernstein, E., Caudy, A. A., Hammond, S. M., and Hannon, G. J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, 409(6818):363–366.

Bockelman, D. L. (1982). Host Range and Seed-Transmission Studies of Maize Chlorotic Mottle Virus in Grasses and Corn. *Plant Disease*, 66(1):216.

Bohorova, N., Fenell, S., McLean, S., Pellegrineschi, A., and Hoisington, D. (1999). *Laboratory protocols: CIMMYT Applied Genetic Engineering Laboratory*.

Bond, W. and Pirone, T. (1970). Evidence for soil transmission of sugarcane mosaic virus. *Phytopathology*, 60:437–440.

Bouché, N., Lauressergues, D., Gasciolli, V., and Vaucheret, H. (2006). An antagonistic function for Arabidopsis DCL2 in development and a new function for DCL4 in generating viral siRNAs. *The EMBO Journal*, 25(14):3347–3356.

Boulila, M. (2011). Positive selection, molecular recombination structure and phylogenetic reconstruction of members of the family Tombusviridae: Implication in virus taxonomy. *Genetics and Molecular Biology*, 34(4):647–660.

Bousalem, M., Bousalem, M., Douzery, E. J. P., Douzery, E. J. P., Fargette, D., and Fargette, D. (2000). High genetic diversity, distant phylogenetic relationships and intraspecies recombination events among natural populations of Yam mosaic virus: a contribution to understanding potyvirus evolution. *Journal of General Virology*, (2000):243–255.

Bulegeya, V. B. (2016). The effect of potyviurs resistance on maize lethal necrosis (MLN). pages 1–146.

Cabanas, D., Watanabe, S., Higashi, C., and Bressan, A. (2013). Dissecting the Mode of Maize Chlorotic Mottle Virus Transmission (Tombusviridae: Machlomovirus) by Frankliniella williamsi (Thysanoptera: Thripidae). *Journal of Economic Entomology*, 106(1):16–24.

Caplan, J. L., Mamillapalli, P., Burch-smith, T. M., and Czymmek, K. (2008). Chloroplastic Protein NRIP1 Mediates Innate Immune Receptor Recognition of a Viral Effector. pages 449–462.

Carbonell, A., Takeda, A., Fahlgren, N., Johnson, S. C., Cuperus, J. T., and Carrington, J. C. (2014). New generation of artificial MicroRNA and synthetic trans-acting small interfering RNA vectors for efficient gene silencing in Arabidopsis. *Plant physiology*, 165(1):15–29.

Carlos, J., Torret, D. E. L. a., and Holland, J. J. (1990). Quasispecies Populations. *Journal of virology*, 64(12):6278–6281.

Castillo, J. (1983). Present knowledge of virus and mollicute diseases of maize in Peru. In *Proceedings International Maize Virus Disease Colloquium and Workshop*, pages 87–92.

Castillo, J. and Hebert, T. (1974). Nueva enfermedad virosa afectando al maiz en el peru. *Fitopatologia*, 9(2):79–84.

Chandrasekaran, J., Brumin, M., Wolf, D., Leibman, D., Klap, C., Sherman, A., Arazi, T., and Gal-on, A. (2016). Development of broad virus resistance in non-transgenic cucumber using CRISPR / Cas9 technology. *Molecular plant pathology*, 7(20 16):1140–1153.

Chare, E. R. and Holmes, E. C. (2006). A phylogenetic survey of recombination frequency in plant RNA viruses. *Archives of Virology*, 151(5):933–946.

Chen, J., Chen, J., and Adams, M. J. (2002). Characterisation of potyviruses from sugarcane and maize in China. *Archives of Virology*, 147(6):1237–1246.

Chen, L., Cheng, X., Cai, J., Zhan, L., Wu, X., Liu, Q., and Wu, X. (2016). Multiple virus resistance using artificial trans-acting siRNAs. *Journal of Virological Methods*, 228:16–20.

Cheng, C.-P. and Nagy, P. D. (2003). Mechanism of RNA recombination in carmo- and tombusviruses: evidence for template switching by the RNA-dependent RNA polymerase in vitro. *Journal of virology*, 77(22):12033–12047.

Christensen, A. H. and Quail, P. H. (1996). Ubiquitin promoter-based vectors for high-level expression of selectable and/or screenable marker genes in monocotyledonous plants. *Transgenic Research*, 5(3):213–218.

Christensen, A. H., Sharrock, R. A., and Quail, P. H. (1992). Maize polyubiquitin genes: structure, thermal perturbation of expression and transcript splicing, and promoter activity following transfer to protoplasts by electroporation. *Plant Molecular Biology*, 18(4):675–689.

Christie, M., Croft, L. J., and Carroll, B. J. (2011). Intron splicing suppresses RNA silencing in Arabidopsis. *Plant Journal*, 68(1):159–167.

Chung, B. Y.-W., Miller, W. A., Atkins, J. F., and Firth, A. E. (2008). An overlapping essential gene in the Potyviridae. *Proceedings of the National Academy of Sciences of the United States of America*, 105(15):5897–5902.

Clamp, M., Cuff, J., Searle, S. M., and Barton, G. J. (2004). The Jalview Java alignment editor. *Bioinformatics*, 20(3):426–427.

Csorba, T., Kontra, L., and Burgyán, J. (2015). viral silencing suppressors: Tools forged to fine-tune host-pathogen coexistence. *Virology*, 479-480:85–103.

Curtin, S. J., Watson, J. M., Smith, N. A., Eamens, A. L., Blanchard, C. L., and Waterhouse, P. M. (2008). The roles of plant dsRNA-binding proteins in RNAi-like pathways. *FEBS letters*, 582(18):2753–60.

Czech, B. and Hannon, G. J. (2011). Small RNA sorting: matchmaking for Argonautes. *Nature Reviews Genetics*, 12(1):19–31.

Dai, X. and Zhao, P. X. (2011). PsRNATarget: A plant small RNA target analysis server. *Nucleic Acids Research*, 39(SUPPL. 2):155–159.

Dalmay, T., Hamilton, A., Rudd, S., Angell, S., and Baulcombe, D. C. (2000). An RNA-dependent RNA polymerase gene in Arabidopsis is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. *Cell*, 101(5):543–553.

Dalmay, T., Horsefield, R., Braunstein, T. H., and Baulcombe, D. C. (2001). SDE3 encodes an RNA helicase required for post-transcriptional gene silencing in Arabidopsis. *The EMBO Journal*, 20(8):2069–2077.

Dangl, J. L., Horvath, D. M., and Staskawicz, B. J. (2013). Pivoting the Plant Immune System. *Science*, 341(August):746–752.

De Groote, H., Oloo, F., Tongruksawattana, S., and Das, B. (2016). Community-survey based assessment of the geographic distribution and impact of maize lethal necrosis (MLN) disease in Kenya. *Crop Protection*, 82:30–35.

De Groote, H., Tongruksawattana, S., Oloo, F., and Das, B. (2015). Distribution and impact of maize lethal necrosis in Kenya. *CIMMYT*, (May):1–4.

Della Vedova, C. B., Lorbiecke, R., Kirsch, H., Schulte, M. B., Scheets, K., Borchert, L. M., Scheffler, B. E., Wienand, U., Cone, K. C., and Birchler, J. a. (2005). The dominant inhibitory chalcone synthase allele C2-Idf (inhibitor diffuse) from Zea mays (L.) acts via an endogenous RNA silencing mechanism. *Genetics*, 170(4):1989–2002.

Deng, T.-C., Chou, C.-M., Chen, C.-T., Tsai, C.-H., and Lin, F.-C. (2014). First Report of Maize chlorotic mottle virus on Sweet Corn in Taiwan. *Plant Disease*, 98(12):1748–1748.

Devert, A., Fabre, N., Floris, M., Canard, B., Robaglia, C., and Crété, P. (2015). Primer-Dependent and Primer-Independent Initiation of Double Stranded RNA Synthesis by Purified Arabidopsis RNA-Dependent RNA Polymerases RDR2 and RDR6. *PloS one*, 10(3):e0120100.

Dietrich, C. and Maiss, E. (2003). Fluorescent labelling reveals spatial separation of potyvirus populations in mixed infected Nicotiana benthamiana plants. *Journal of General Virology*, 84(10):2871–2876.

Domingo, E., Sabo, D., Taniguchi, T., and Weissman, C. (1978). Nucleotide Sequence Phage Population Heterogeneity of an RNA. *Cell*, 13(April):735–744.

Domingo, E., Sheldon, J., and Perales, C. (2012). Viral quasispecies evolution. *Microbiology and molecular biology reviews : MMBR*, 76(2):159–216.

Dong, Q., Jiang, H., Xu, Q., Li, X., Peng, X., Yu, H., Xiang, Y., and Cheng, B. (2015). Cloning and characterization of a multifunctional promoter from maize (Zea mays L.). *Applied biochemistry and biotechnology*, 175(3):1344–57.

Doudna, J. a. and Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. *Science*, 346(6213):1258096–1258096.

Duan, C.-G., Wang, C.-H., and Guo, H.-S. (2012). Application of RNA silencing to plant disease resistance. *Silence*, 3(1):5.

Eckardt, N. A. (2013). The Plant Cell Reviews Aspects of MicroRNA and PhasiRNA Regulatory Function The Plant Cell Reviews Aspects of MicroRNA and PhasiRNA Regulatory Function. *Plant Cell*, 25(July):2382.

Eskelin, K., Hafre, A., Rantalainen, K. I., and Ma, K. (2011). Potyviral VPg Enhances Viral RNA Translation and Inhibits Reporter mRNA Translation In Planta. *Journal of virology*, 85(17):9210–9221.

Fahim, M., Millar, A. A., Wood, C. C., and Larkin, P. J. (2012). Resistance to Wheat streak mosaic virus generated by expression of an artificial polycistronic microRNA in wheat. *Plant biotechnology journal*, 10(2):150–163.

Fahlgren, N. and Carrington, J. C. (2010). mirna target prediction in plants. *Plant MicroRNAs: Methods and Protocols*, pages 51–57.

Fahlgren, N., Hill, S. T., Carrington, J. C., and Carbonell, A. (2015). P-SAMS: a web site for plant artificial microRNA and synthetic trans-acting small interfering RNA design. *Bioinformatics (Oxford, England)*, 32(1):157–158.

Flasinski, S. and Cassidy, B. G. (1998). Potyvirus aphid transmission requires helper component and homologous coat protein for maximal efficiency. *Archives of Virology*, 143(11):2159–2172.

Foster, T. M., Lough, T. J., Emerson, S. J., Lee, R. H., Bowman, J. L., Forster, R. L. S., and Lucas, W. J. (2002). A Surveillance System Regulates Selective Entry of RNA into the Shoot Apex. *The Plant Cell Online*, 14(7):1497–1508.

Fraile, A., Malpica, J. M., Aranda, M. A., Rodríguez-Cerezo, E., and García-Arenal, F. (1996). Genetic diversity in tobacco mild green mosaic tobamovirus infecting the wild plant Nicotiana glauca. *Virology*, 223(1):148–55.

Francisca, M., María, P., Filippone, P., Sergio, A., Inés, M., Atilio, C., and Castagnaro, P. (2012). An Overview of the Sugarcane Mosaic Disease in South America.

Frenkel, M. J., Jilka, J. M., McKern, N. M., Strike, P. M., Clark, J. M., Shukla, D. D., and Ward, C. W. (1991). Unexpected sequence diversity in the amino-terminal ends of the coat proteins of strains of sugarcane mosaic virus. *Journal of General Virology*, 72(2):237–242.

Fuchs, M. (2017). Pyramiding resistance-conferring gene sequences in crops. *Current Opinion in Virology*, 26:36–42.

Gan, D., Zhang, J., Jiang, H., Jiang, T., Zhu, S., and Cheng, B. (2010). Bacterially expressed dsRNA protects maize against SCMV infection. *Plant Cell Reports*, 29(11):1261–1268.

Gao, B., Cui, X.-W., Li, X.-D., Zhang, C.-Q., and Miao, H.-Q. (2011). Complete genomic sequence analysis of a highly virulent isolate revealed a novel strain of Sugarcane mosaic virus. *Virus genes*, 43(3):390–7.

García-Arenal, F., Fraile, A., and Malpica, J. M. (2001). VARIABILITY AND GENETIC STRUCTURE OF PLANT VIRUS POPULATIONS. *Annual Review of Phytopathology*, 39:157–186.

Gell, G., Sebestyén, E., and Balázs, E. (2015). Recombination analysis of Maize dwarf mosaic virus (MDMV) in the Sugarcane mosaic virus (SCMV) subgroup of potyviruses. *Virus Genes*, 50(1):79–86.

Godfray, H., Crute, I., Haddad, L., Lawrence, D., Muir, J., Nisbett, N., Pretty, J., Robinson, S., Toulmin, C., and Whiteley, R. (2010). The future of the global food system. *Philosophical transactions of the Royal Society B: Biological sciences*, 365(1554):2769–2777.

Goldberg, K.-B. and Brakke, M. K. (1987). Concentration of maize chlorotic mottle virus increased in mixed infections with maize dwarf mosaic virus, strain B. *Phytopathology*, 77(2):162–167.

Gonçalves, M. C., Godinho, M. T., Alves-Freitas, D. M. T., Varsani, A., and Ribeiro, S. G. (2017). First Report of Maize yellow mosaic virus Infecting Maize in Brazil. *Plant Disease*, pages PDIS–04–17–0569–PDN.

González-Jara, P., Atencio, F. A., Martínez-García, B., Barajas, D., Tenllado, F., and Díaz-Ruíz, J. R. (2005). A Single Amino Acid Mutation in the Plum pox virus Helper Component-Proteinase Gene Abolishes Both Synergistic and RNA Silencing Suppression Activities. *Phytopathology*, 95(8):894–901.

Gonzalez-Jara, P., Fraile, A., Canto, T., and Garcia-Arenal, F. (2009). The Multiplicity of Infection of a Plant Virus Varies during Colonization of Its Eukaryotic Host. *Journal of Virology*, 83(15):7487–7494.

Gordon, D., Knoke, J., Nault, L., and Ritter, R. (1983). International maize virus disease colloquium and workshop; wooster, ohio; 2-6 aug 1982. proceedings. eds. Technical report.

Gowda, M., Das, B., Makumbi, D., Babu, R., Semagn, K., Mahuku, G., Olsen, M. S., Bright, J. M., Beyene, Y., and Prasanna, B. M. (2015). Genome-wide association and genomic prediction of resistance to maize lethal necrosis disease in tropical maize germplasm. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., and Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology*, 29(7):644–52.

Guo, H., Song, X., Xie, C., Huo, Y., Zhang, F., Chen, X., Geng, Y., and Fang, R. (2013). Rice yellow stunt rhabdovirus protein 6 suppresses systemic RNA silencing by blocking RDR6-mediated secondary siRNA synthesis. *Molecular plant-microbe interactions : MPMI*, 26(8):927–36.

Guo, H. S. and Ding, S. W. (2002). A viral protein inhibits the long range signaling activity of the gene silencing signal. 21(3).

Haag, J. R., Brower-Toland, B., Krieger, E. K., Sidorenko, L., Nicora, C. D., Norbeck, A. D., Irsigler, A., LaRue, H., Brzeski, J., McGinnis, K., Ivashuta, S., Pasa-Tolic, L., Chandler, V. L., and Pikaard, C. S. (2014). Functional Diversification of Maize RNA Polymerase IV and V Subtypes via Alternative Catalytic Subunits. *Cell reports*, 9(1):378–90.

Hamilton, A., Voinnet, O., Chappell, L., and Baulcombe, D. (2002). Two classes of short interfering RNA in RNA silencing. 21(17):4671–4679.

Handley, J. a., Smith, G. R., Dale, J. L., and Harding, R. M. (1998). Sequence diversity in the CP coding region of eight sugarcane mosaic potyvirus isolates infecting sugarcane in Australia. *Archives of virology*, 143:1145–1153.

Hasiów-Jaroszewska, B., Fares, M. A., and Elena, S. F. (2014). Molecular Evolution of Viral Multifunctional Proteins: The Case of Potyvirus HC-Pro. *Journal of Molecular Evolution*, 78(1):75–86.

Heckman, K. L. and Pease, L. R. (2007). Gene splicing and mutagenesis by PCR-driven overlap extension. *Nature protocols*, 2(4):924–32.

Hilker, F., Allen, L., Bokil, V., Briggs, C., Feng, Z., Garrett, K. A., Gross, L., Hamelin, F., Jeger, M. J., Manore, C., Power, A., Redinbaugh, M., R?a, M., and Cunniffe, N. J. (2017). Modelling virus coinfection to inform management of maize lethal necrosis in Kenya. *Phytopathology*, (Accepted)(0):PHYTO–03–17–0080–FI.

Hillung, J., Elena, S. F., and Cuevas, J. M. (2013). Intra-specific variability and biological relevance of P3N-PIPO protein length in potyviruses. *BMC evolutionary biology*, 13(1):249.

Honaas, L. A., Wafula, E. K., Wickett, N. J., Der, J. P., Zhang, Y., Edger, P. P., Altman, N. S., Chris Pires, J., Leebens-Mack, J. H., and DePamphilis, C. W. (2016). Selecting superior de novo transcriptome assemblies: Lessons learned by leveraging the best plant genome. *PLoS ONE*, 11(1):1–42.

Hubert, L. and Arabie, P. (1985). Comparing partitions. *Journal of Classification*, 2(1):193–218.

Hudson, R. R. (2000). A new statistic for detecting genetic differentiation. *Genetics*, 155(4):2011–2014.

Hudson, R. R., Boos, D. D., and Kaplan, N. L. (1992a). A statistical test for detecting geographic subdivision. *Molecular Biology and Evolution*, 9(1):138–151.

Hudson, R. R., Slatkin, M., and Maddison, W. P. (1992b). Estimation of levels of gene flow from DNA sequence data. *Genetics*, 132(2):583–589.

Huson, D., Mitra, S., and Ruscheweyh, H. (2011). Integrative analysis of environmental sequences using MEGAN4. *Genome Research*, 21(9):1552–1560.

Huson, D. H. and Bryant, D. (2006). User Manual for SplitsTree4 V4.6. pages 1–57.

Iki, T., Tschopp, M.-A., and Voinnet, O. (2017). Biochemical and genetic functional dissection of the P38 viral suppressor of RNA silencing. *RNA (New York, N.Y.)*, 23(5):639–654.

Incarbone, M. and Dunoyer, P. (2013). RNA silencing and its suppression: novel insights from in planta analyses. *Trends in plant science*, 18(7):382–92.

Isabirye, B. E. and Rwomushana, I. (2016). Current and future potential distribution of maize chlorotic mottle virus and risk of maize lethal necrosis disease in Africa. 5(2):215–228.

Ishida, Y., Hiei, Y., and Komari, T. (2007). Agrobacterium-mediated transformation of maize. *Nature protocols*, 2(7):1614–21.

Ishida, Y., Saito, H., Ohta, S., Hiei, Y., Komari, T., and Kumashiro, T. (1996). High efficiency transformation of maize (Zea mays L.) mediated by Agrobacterium tumefaciens. *Nature Biotechnology*, 14:745–750.

Jaubert, M., Bhattacharjee, S., Mello, A. F. S., Perry, K. L., and Moffett, P. (2011). ARGONAUTE2 Mediates RNA-Silencing Antiviral Defenses against Potato virus X in Arabidopsis. *Plant Physiology*, 156(3):1556–1564.

Jensen, S. G. (1991). Seed Transmission of Maize Chlorotic Mottle Virus. *Plant Disease*, 75(5):497.

Jiang, X. Q., Meinke, L. J., Wright, R. J., Wilkinson, D. R., and Campbell, J. E. (1992). Maize chlorotic mottle virus in Hawaiian-grown maize: vector relations, host range and associated viruses. *Crop Protection*, 11(3):248–254.

Jiwan, S. D. and White, K. A. (2011). Subgenomic mRNA transcription in Tombusviridae. *RNA Biology*.

Jonczyk, M., Pathak, K. B., Sharma, M., and Nagy, P. D. (2007). Exploiting alternative subcellular location for replication: Tombusvirus replication switches to the endoplasmic reticulum in the absence of peroxisomes. *Virology*, 362(2):320–330.

Jones, J. D. G. and Dangl, J. L. (2006). The plant immune system. *Nature*, 444(7117):323–9.

Kalantidis, K., Tsagris, M., and Tabler, M. (2006). Spontaneous short-range silencing of a GFP transgene in Nicotiana benthamiana is possibly mediated by small quantities of siRNA that do not trigger systemic silencing. *Plant Journal*, 45(6):1006–1016.

Källberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., and Xu, J. (2012). Template-based protein structure modeling using the RaptorX web server. *Nat Protoc.*, 7(8):1511–1522.

Kang'ethe, E. (2011). Situation Analysis: Improving Food Safety in the Maize Value Chain in Kenya. Report prepared for FAO by Prof . Erastus Kang ' ethe College of Agriculture and Veterinary Science University of Nairobi. (September):1–89.

Karan, M., Noonea, D. F., Teakle, D. S., and Hacker, J. B. (1992). Susceptibility of pearl millet accessions and cultivars to Johnsongrass mosaic and sugarcane mosaic viruses in Queensland. *Australasian plant pathology*, 21(3):128–130.

Kendall, A., Mcdonald, M., Bian, W., Bowles, T., Baumgarten, S. C., Shi, J., Stewart, P. L., Bullitt, E., Gore, D., Irving, T. C., Havens, W. M., Ghabrial, S. A., Wall, J. S., and Stubbs, G. (2008). Structure of Flexible Filamentous Plant Viruses. 82(19):9546–9554.

Kibaki, J. J. and Francis, M. (2013). Controlling Maize Lethal Necrosis Disease via Vector Management. (March).

Kim, J. I., Kim, Y.-J., Lemey, P., Lee, I., Park, S., Bae, J.-Y., Kim, D., Kim, H., Jang, S.-I., Yang, J.-S., Kim, H., Kim, D.-W., Nam, J.-G., Kim, S. S., Kim, K., Myun Lee, J., Song, M. K., Song, D., Chang, J., Hong, K.-J., Bae, Y.-S., Song, J.-W., Lee, J.-S., and Park, M.-S. (2016). The recent ancestry of Middle East respiratory syndrome coronavirus in Korea has been shaped by recombination. *Scientific Reports*, 6(November 2015):18825.

Klinkong, T. and Sutabutra, T. (1983). A new virus disease of maize in thailand. In *International Maize Virus Disease Colloquium and Workshop, Wooster, Ohio (USA), 2-6 Aug 1982*. Ohio Agricultural Research and Development Center.

Kontra, L., Csorba, T., Tavazza, M., Lucioli, A., Tavazza, R., Moxon, S., Tisza, V., Medzihradszky, A., Turina, M., and Burgyán, J. (2016). Distinct Effects of p19 RNA Silencing Suppressor on Small RNA Mediated Pathways in Plants. *PLoS Pathogens*, 12(10):1–26.

Kotakis, C., Vrettos, N., Daskalaki, M. G., Kotzabasis, K., and Kalantidis, K. (2011). DCL3 and DCL4 are likely involved in the light intensity-RNA silencing cross talk in Nicotiana benthamiana. *Plant Signal Behav*, 6(8):1180–1182.

Kotakis, C., Vrettos, N., Kotsis, D., Tsagris, M., Kotzabasis, K., and Kalantidis, K. (2010). Light intensity affects RNA silencing of a transgene in Nicotiana benthamiana plants. *BMC plant biology*, 10:220.

Kovalev, N., Martín, I. F. d. C., Pogany, J., Barajas, D., Pathak, K., Risco, C., and Nagy, P. D. (2016). The role of viral RNA and co-opted cellular ESCRT-I and ESCRT-III factors in formation of tombusvirus spherules harboring the tombusvirus replicase. *Journal of Virology*, 90(January):JVI.02775–15.

Krueger, F. (2015). Trim Galore!: A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files.

Kusia, E. S., Subramanian, S., Nyasani, J. O., Khamis, F., Villinger, J., Ateka, E., and Pappu, H. R. (2015). First report of lethal necrosis disease associated with co-infection of finger millet with Maize chlorotic mottle virus and Sugarcane mosaic virus in Kenya. *Plant Disease*, 99(6):899.

Lafforgue, G., Martínez, F., Niu, Q.-W., Chua, N.-H., Daròs, J.-A., and Elena, S. F. (2013). Improving the effectiveness of artificial microRNA (amiR)-mediated resistance against Turnip mosaic virus by combining two amiRs or by targeting highly conserved viral genomic regions. *Journal of virology*, 87(14):8254–6.

Lafforgue, G., Martínez, F., Sardanyés, J., de la Iglesia, F., Niu, Q.-W., Lin, S.-S., Solé, R. V., Chua, N.-H., Daròs, J.-A., and Elena, S. F. (2011). Tempo and Mode of Plant RNA Virus Escape from RNA Interference-Mediated Resistance. *Journal of Virology*, 85(19):9686–9695.

Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods*, 9(4):357–359.

Latham, J. R. and Wilson, A. K. (2008). Transcomplementation and synergism in plants: implications for viral transgenes? *Molecular plant pathology*, 9(1):85–103.

Lauring, A. S. and Andino, R. (2010). Quasispecies theory and the behavior of RNA viruses. *PLoS pathogens*, 6(7):e1001005.

Leng, P., Ji, Q., Tao, Y., Ibrahim, R., and Pan, G. (2015). Characterization of Sugarcane Mosaic Virus Scmv1 and Scmv2 Resistance Regions by Regional Association Analysis in Maize. *PLoS ONE*, 10:1–18.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics (Oxford, England)*, 27(21):2987–93.

Li, L., Wang, X., and Zhou, G. (2007). Analyses of maize embryo invasion by Sugarcane mosaic virus. *Plant Science*, 172(1):131–138.

Li, M., Li, Y., Xia, Z., Di, D., Zhang, A., Miao, H., Zhou, T., and Fan, Z. (2017). Characterization of small interfering RNAs derived from Rice black streaked dwarf virus in infected maize plants by deep sequencing. *Virus Research*, 228:66–74.

Li, Y., Liu, R., Zhou, T., and Fan, Z. (2013). Genetic diversity and population structure of Sugarcane mosaic virus. *Virus Research*, 171(1):242–246.

Li, Z., Chen, J., Han, L., Wen, J., Chen, G., and Li, H. (2016). Association mapping resolving the major loci Scmv2 conferring resistance to sugarcane mosaic virus in maize. *European Journal of Plant Pathology*, pages 385–391.

Librado, P. and Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25(11):1451–1452.

Lin, S. S., Wu, H. W., Elena, S. F., Chen, K. C., Niu, Q. W., Yeh, S. D., Chen, C. C., and Chua, N. H. (2009). Molecular evolution of a viral non-coding sequence under the selective pressure of amiRNA-mediated silencing. *PLoS Pathogens*, 5(2).

Lindbo, J. A. and Falk, B. W. (2017). The Impact of Coat Protein-Mediated Virus Resistance in Applied Plant Pathology and Basic Research. *Phytopathology*, 107(6):624–634.

Liu, Q., Liu, H., Gong, Y., Tao, Y., Jiang, L., Zuo, W., and Xu, M. (2017). An Atypical Thioredoxin Imparts Early Resistance to Sugarcane Mosaic Virus in Maize. *Molecular Plant*, (March):483–497.

López-Moya, J. J., Valli, A., and García, J. A. (2009). Potyviridae. In *Encyclopedia of Life Sciences*. John Wiley & Sons, Ltd, Chichester.

López-Moya, J. J., Wang, R. Y., and Pirone, T. P. (1999). Context of the coat protein DAG motif affects potyvirus transmissibility by aphids. *Journal of General Virology*, 80(12):3281–3288.

Lorenz, R., Bernhart, S. H., zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., and Hofacker, I. L. (2011). {ViennaRNA} Package 2.0. *Algorithms for Molecular Biology*, 6(1):26.

Louie, R. (1980). Sugarcane Mosaic Virus in Kenya. *Plant Disease*, 64(10):944.

Lukanda, M., Owati, A., Ogunsanya, P., Valimunzigha, K., Katsongo, K., Ndemere, H., and Kumar, P. L. (2014). First Report of Maize chlorotic mottle virus Infecting Maize in the Democratic Republic of the Congo. *Plant Disease*, 98(10):1448–1448.

Lynch, M. and Crease, T. (1990). The analysis of population survey data on DNA sequence variation. *Molecular Biology and Evolution*, 7(4):377–394.

Mahuku, G., Lockhart, B. E., Wanjala, B., Jones, M. W., Kimunye, J. N., Stewart, L. R., Cassone, B. J., Subramanian, S., Nyasani, J., Kusia, E., Kumar, L., Niblett, C. L., Kiggundu, A., Asea, G., Pappu, H., Wangai, A., Prasanna, B. M., and Redinbaugh, M. (2015). Maize lethal necrosis (MLN), an emerging threat to maize-based food security in sub-Saharan Africa. *Phytopathology*.

Martin, D. P., Murrell, B., Golden, M., Khoosal, a., and Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution*, 1(1):1–5.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads.

Martin-Hernandez, A. M. and Baulcombe, D. C. (2008). Tobacco Rattle Virus 16-Kilodalton Protein Encodes a Suppressor of RNA Silencing That Allows Transient Viral Entry in Meristems. *Journal of Virology*, 82(8):4064–4071.

Martínez, F., Lafforgue, G., Morelli, M. J., González-Candelas, F., Chua, N.-H., Daròs, J.-A., and Elena, S. F. (2012). Ultradeep Sequencing Analysis of Population Dynamics of Virus Escape Mutants in RNAi-Mediated Resistant Plants. *Molecular Biology and Evolution*, 29(11):3297–3307.

Martínez-Turiño, S. and Hernández, C. (2009). Inhibition of RNA silencing by the coat protein of Pelargonium flower break virus: distinctions from closely related suppressors. *The Journal of general virology*, 90(Pt 2):519–25.

Mascia, T. and Gallitelli, D. (2016). Synergies and antagonisms in virus interactions. *Plant Science*, 252:176–192.

Mehle, N. and Ravnikar, M. (2012). Plant viruses in aqueous environment – Survival, water mediated transmission and detection. *Water Research*, 46(16):4902–4917.

Meister, G. (2013). Argonaute proteins: functional insights and emerging roles. *Nature reviews. Genetics*, 14(7):447–59.

Melnyk, C., Molnar, A., Bassett, A., and Baulcombe, D. (2011). Mobile 24 nt Small RNAs Direct Transcriptional Gene Silencing in the Root Meristems of Arabidopsis thaliana. *Current Biology*, 21(19):1678–1683.

Meng, C., Chen, J., Peng, J., and Wong, S. M. (2006). Host-induced avirulence of hibiscus chlorotic ringspot virus mutants correlates with reduced gene-silencing suppression activity. *Journal of General Virology*, 87(2):451–459.

Mérai, Z., Kerényi, Z., and Kertész, S. (2006). Double-Stranded RNA Binding May Be a General Plant RNA Viral Strategy To Suppress RNA Silencing Double-Stranded RNA Binding May Be a General Plant RNA Viral Strategy To Suppress RNA Silencing. *Journal of Virology*, 80(12):5747–5756.

Mérai, Z., Kerényi, Z., Molnár, A., Barta, E., Válóczi, A., Bisztray, G., Havelda, Z., Burgyán, J., and Silhavy, D. (2005). Aureusvirus P14 is an efficient RNA silencing suppressor that binds double-stranded RNAs without size specificity. *Journal of virology*, 79(11):7217–26.

Miller, S. and Krijnse-Locker, J. (2008). Modification of intracellular membrane structures for virus replication. *Nature Reviews Microbiology*, 6(5):363–374.

Mitter, N., Worrall, E. A., Robinson, K. E., Li, P., Jain, R. G., Taochy, C., Fletcher, S. J., Carroll, B. J., Lu, G. Q. M., and Xu, Z. P. (2017). Clay nanosheets for topical delivery of RNAi for sustained protection against plant viruses. *Nature Plants*, 16207(January).

Mlotshwa, S., Pruss, G. J., and Vance, V. (2008). Small RNAs in viral infection and host defense. *Trends in Plant Science*, 13(7):375–382.

Molnar, A., Melnyk, C., and Baulcombe, D. C. (2011). Silencing signals in plants: a long journey for small RNAs. *Genome Biology*, 12(1):215.

Monaghan, J. and Zipfel, C. (2012). Plant pattern recognition receptor complexes at the plasma membrane. *Current Opinion in Plant Biology*, 15(4):349–357.

Moradi, Z., Mehrvar, M., Nazifi, E., and Zakiaghl, M. (2016). The complete genome sequences of two naturally occurring recombinant isolates of Sugarcane mosaic virus from Iran. *Virus Genes*, 52(2):270–280.

Moreno, I. M., Malpica, J. M., Díaz-Pendón, J. A., Moriones, E., Fraile, A., and García-Arenal, F. (2004). Variability and genetic structure of the population of watermelon mosaic virus infecting melon in Spain. *Virology*, 318(1):451–460.

Moreno, I. M., Malpica, J. M., Rodríguez-Cerezo, E., and García-Arenal, F. (1997). A mutation in tomato aspermy cucumovirus that abolishes cell-to-cell movement is maintained to high levels in the viral RNA population by complementation. *Journal of virology*, 71(12):9157–9162.

Moya, A., Holmes, E. C., and González-Candelas, F. (2004). The population genetics and evolutionary epidemiology of RNA viruses. *Nature reviews. Microbiology*, 2(4):279–288.

Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., and Kosakovsky Pond, S. L. (2012). Detecting individual sites subject to episodic diversifying selection. *PLoS Genetics*, 8(7).

Nagy, P. D. (2016). Tombusvirus-Host Interactions : Co-Opted Evolutionarily Conserved Host Factors Take Center Court. *Annual Review of Virology*, 3(September):1–25.

Nagy, P. D. and Bujarski, J. J. (1998). Silencing Homologous RNA Recombination Hot Spots with GC-Rich Sequences in Brome Mosaic Virus. 72(2):1122–1130.

Nagy, P. D., Strating, J. R., and van Kuppeveld, F. J. (2016). Building Viral Replication Organelles: Close Encounters of the Membrane Types. *PLoS Pathogens*, 12(10):6–11.

Nakagawa, T., Suzuki, T., Murata, S., Nakamura, S., Hino, T., Maeo, K., Tabata, R., Kawai, T., Tanaka, K., Niwa, Y., Watanabe, Y., Nakamura, K., Kimura, T., and Ishiguro, S. (2007). Improved Gateway binary vectors: high-performance vectors for creation of fusion constructs in transgenic analysis of plants. *Bioscience, biotechnology, and biochemistry*, 71(8):2095–100.

Nault, L. R., Styer, W. P., Coffey, M. E., Gordon, D. T., Negi, L. S., and Niblett, C. L. (1978). Transmission of maize chlorotic mottle virus by chrysomelid beetles. *Phytopathology*, 68(7):1071–1074.

Newburn, L. R. and White, K. A. (2015). Cis-acting RNA elements in positive-strand RNA plant virus genomes. *Virology*, 479-480:434–43.

Niblett, C. L. and Claflin, L. E. (1978). Corn lethal necrosis-a new virus disease of corn in Kansas. *Plant disease reporter*, 62(1):15–19.

Nicholson, A. W. (1999). Function, mechanism and regulation of bacterial ribonucleases. *FEMS Microbiology Reviews*, 23(3):371–390.

Niu, Q.-W., Lin, S.-S., Reyes, J. L., Chen, K.-C., Wu, H.-W., Yeh, S.-D., and Chua, N.-H. (2006a). Expression of artificial microRNAs in transgenic Arabidopsis thaliana confers virus resistance. *Nature Biotechnology*, 24(11):1420–1428.

Niu, Q.-W., Lin, S.-S., Reyes, J. L., Chen, K.-C., Wu, H.-W., Yeh, S.-D., and Chua, N.-H. (2006b). Expression of artificial microRNAs in transgenic Arabidopsis thaliana confers virus resistance. *Nature biotechnology*, 24(11):1420–8.

Ohshima, K., Tomitaka, Y., Wood, J. T., Minematsu, Y., Kajiyama, H., Tomimura, K., and Gibbs, A. J. (2007). Patterns of recombination in turnip mosaic virus genomic sequences indicate hotspots of recombination. *Journal of General Virology*, 88(1):298–315.

Olspert, A., Chung, B. Y.-W., Atkins, J. F., Carr, J. P., and Firth, A. E. (2015). Transcriptional slippage in the positive-sense RNA virus family Potyviridae. *EMBO reports*, page e201540509.

Ossowski, S., Schwab, R., and Weigel, D. (2008). Gene silencing in plants using artificial microRNAs and other small RNAs. *The Plant journal*, 53(4):674–90.

Padhi, A. and Ramu, K. (2011). Genomic evidence of intraspecific recombination in sugarcane mosaic virus. *Virus genes*, 42(2):282–5.

Pasin, F., Simón-Mateo, C., and García, J. A. (2014). The Hypervariable Amino-Terminus of P1 Protease Modulates Potyviral Replication and Host Defense Responses. *PLoS Pathogens*, 10(3).

Patil, B. L. and Fauquet, C. M. (2015). Light intensity and temperature affect systemic spread of silencing signal in transient agroinfiltration studies. *Molecular Plant Pathology*, 16(5):484–494.

Paul, D. and Bartenschlager, R. (2013). Architecture and biogenesis of plus-strand RNA virus replication factories. *World J Virol May*, 12(22):32–48.

Peng, Y.-h., Kadoury, D., Huet, H., Wang, Y., and Raccah, B. (1998). Mutations in the HC-Pro gene of zucchini yellow mosaic potyvirus : effects on aphid transmission and binding to purified virions. *Journal of General Virology*, 79:897–904.

Perera, M. F., Filippone, M. P., Ramallo, C. J., Cuenya, M. I., García, M. L., Ploper, L. D., and Castagnaro, A. P. (2009). Genetic Diversity Among Viruses Associated with Sugarcane Mosaic Disease in Tucumán , Argentina. *Phytopathology*, 99(1):38–49.

Perez-Canamas, M. and Hernandez, C. (2015). Key importance of small RNA binding for the activity of a Glycine-Tryptophan (GW) motif-containing viral suppressor of RNA silencing. *Journal of Biological Chemistry*, 290(5):3106–3120.

Persley, D. M. and Greber, R. S. (1977). Additional Field Hosts of Sugarcane A Disease of Duboisia Caused by Tomato Spotted Wilt Virus (TSWV). *Australasian plant pathology*, 6(4):54.

Pinel, A., N'Guessan, P., Bousalem, M., and Fargette, D. (2000). Molecular variability of geographically distinct isolates of Rice yellow mottle virus in Africa. *Archives of Virology*, 145(8):1621–1638.

Posada, D. and Crandall, K. (2002). The Effect of Recombination on the Accuracy of Phylogeny Estimation. *Journal of molecular evolution*, 54(3):396–402.

Powers, J. G., Sit, T. L., Heinsohn, C., George, C. G., Kim, K. H., and Lommel, S. A. (2008). The Red clover necrotic mosaic virus RNA-2 encoded movement protein is a second suppressor of RNA silencing. *Virology*, 381(2):277–286.

Pratt, C. F., Constantine, K. L., and Murphy, S. T. (2017). Economic impacts of invasive alien species on African smallholder livelihoods. *Global Food Security*, (November 2016):1–7.

Pyott, D. E., Sheehan, E., and Molnar, A. (2016). Engineering of CRISPR / Cas9-mediated potyvirus resistance in transgene-free Arabidopsis plants. *Molecular plant pathology*, 7(20 16):1276–1288.

Qian, Y., Cheng, Y., Cheng, X., Jiang, H., Zhu, S., and Cheng, B. (2011). Identification and characterization of Dicer-like, Argonaute and RNA-dependent RNA polymerase gene families in maize. *Plant cell reports*, 30(7):1347–63.

Qu, J., Ye, J., and Fang, R. (2007). Artificial microRNA-mediated virus resistance in plants. *Journal of virology*, 81(12):6690–9.

Quito-Avila, D. F., Alvarez, R. A., and Mendoza, A. A. (2016). Occurrence of maize lethal necrosis in Ecuador : a disease without boundaries ? *European Journal of Plant Pathology*, 1914.

Rehmsmeier, M., Steffen, P., Höchsmann, M., Giegerich, R., and Ho, M. (2004). Fast and effective prediction of microRNA / target duplexes. *Spring*, (2003):1507–1517.

Revers, F., Le Gall, O., Candresse, T., Le Romancer, M., and Dunez, J. (1996). Frequent occurrence of recombinant potyvirus isolates. *Journal of General Virology*, 77(8):1953–1965.

Robinson, D. J., Ryabov, E. V., Raj, S. K., Roberts, I. M., and Taliansky, M. E. (1999). Satellite RNA is essential for encapsidation of groundnut rosette umbravirus RNA by groundnut rosette assistor luteovirus coat protein. *Virology*, 254(1):105–114.

Robinson, K. E., Worrall, E. a., and Mitter, N. (2014). Double stranded RNA expression and its topical application for non-transgenic resistance to plant viruses. *Journal of Plant Biochemistry and Biotechnology*, 23(3):231–237.

Ronde, D. D., Butterbach, P., and Kormelink, R. (2014). Dominant resistance against plant viruses. *Frontiers in plant science*, 5(June):1–17.

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., and Huelsenbeck, J. P. (2012). Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, 61(3):539–542.

Roth, B. M., Pruss, G. J., and Vance, V. B. (2004). Review-Plant viral suppressors of RNA silencing. *Virus Research*, 102(1):97–108.

Rubio, L., Abou-Jawdah, Y., Lin, H. X., and Falk, B. W. (2001a). Geographically distant isolates of the crinivirus Cucurbit yellow stunting disorder virus show very low genetic diversity in the coat protein gene. *Journal of General Virology*, 82(4):929–933.

Rubio, L., Ayllón, M. A., Kong, P., Fernández, A., Polek, M., Guerri, J., Moreno, P., and Falk, B. W. (2001b). Genetic variation of Citrus tristeza virus isolates from California and Spain: evidence for mixed infections and recombination. *Journal of virology*, 75(17):8054–62.

Rybicki, E. P. (2015). A Top Ten list for economically important plant viruses. *Archives of Virology*, 160(1):17–20.

Sanfaçon, H. (2015). Plant translation factors and virus resistance. *Viruses*, 7(7):3392–3419.

Sanford, J. C. and Johnston, S. A. (1985). The Concept of Parasite-Derived Resistance–Deriving Resistance Genes from the Parasite ' s Own Genome t Department of Horticultural Science , Cornell University ,. *Journal of Theoretical Biology*, 113:395–405.

Scheets, K. (1998). Maize Chlorotic Mottle Machlomovirus and Wheat Streak Mosaic Rymovirus Concentrations Increase in the Synergistic Disease Corn Lethal Necrosis. *Virology*, 242(1):28–38.

Scheets, K. (2000). Maize Chlorotic Mottle Machlomovirus Expresses Its Coat Protein from a 1.47-kb Subgenomic RNA and Makes a 0.34-kb Subgenomic RNA. *Virology*, 267(1):90–101.

Scheets, K. (2016). Analysis of gene functions in Maize chlorotic mottle virus. *Virus Research*.

Scheets, K., Khosravi-Far, R., and Nutter, R. C. (1993). Transcripts of a maize chlorotic mottle virus cDNA clone replicate in maize protoplasts and infect maize plants. *Virology*, 193(2):1006–9.

Scholthof, H. B. (2006). The Tombusvirus encoded P19: from irrelevance to elegance. 4(May):405–411.

Schwab, R., Maizel, A., Ruiz-ferrer, V., Garcia, D., Bayer, M., Crespi, M., Voinnet, O., and Martienssen, R. A. (2009). Endogenous TasiRNAs Mediate Non-Cell Autonomous Effects on Gene Regulation in Arabidopsis thaliana. *PLoS ONE*, 4(6):4–9.

Schwach, F., Vaistij, F. E., Jones, L., and Baulcombe, D. C. (2005). An RNA-dependent RNA-polymerase prevents meristem invasion by Potato virus X and is required for the activity but not the production of a systemic silencing signal. *Plant Physiology*, 138:1842–1852.

Seo, J.-K., Ohshima, K., Lee, H.-G., Son, M., Choi, H.-S., Lee, S.-H., Sohn, S.-H., and Kim, K.-H. (2009). Molecular variability and genetic structure of the population of soybean mosaic virus based on the analysis of complete genome sequences. *Virology*, 393(1):91–103.

Shi, X. M., Miller, H., Verchot, J., Carrington, J. C., and Vance, V. B. (1997). Mutations in the region encoding the central domain of helper component-proteinase (HC-Pro) eliminate potato virus X/potyviral synergism. *Virology*, 231(1):35–42.

Shukla, D. D., Strike, P. M., Tracy, S. L., Gough, K. H., and Ward, C. W. (1988). The N and C Termini of the Coat Proteins of Potyviruses Are Surface-located and the N Terminus Contains the Major Virus-specific Epitopes. *Journal of General Virology*, 69(7):1497–1508.

Shukla, D. D., Ward, C. W., and Brunt, A. A. (1994). The Potyviridae.

Shulla, A. and Randall, G. (2016). (+) RNA virus replication compartments: A safe home for (most) viral replication. *Current Opinion in Microbiology*, 32:82–88.

Singh, A., Taneja, J., Dasgupta, I., and Mukherjee, S. K. (2015). Development of plants resistant to tomato geminiviruses using artificial trans-acting small interfering RNA. *Molecular plant pathology*, 16:724–734.

Sit, T. L. and Lommel, S. A. (2010). *Encyclopedia of Life Sciences*. John Wiley & Sons, Ltd, Chichester, UK.

Sit, T. L. and Lommel, S. A. (2015). Tombusviridae. *eLS*, (November):1–9.

Smith, N. A., Singh, S. P., Wang, M.-B., Stoutjesdijk, P. A., Green, A. G., and Waterhouse, P. M. (2000). brief communications. *Nature*, 407(September):319–320.

Snipes, K. (2014). Kenya Maize Lethal Necrosis - The growing challenge in Eastern Africa. (September 2011).

Srisink, S., Noonea, D. F., Teakle, D. S., and Ryan, C. C. (1993). Brachiaria piligera and Sorghum verticilliflorum are natural hosts of two different strains of sugarcane mosaic virus in Australia. *Australasian plant pathology*, 22:94–97.

Stam, M., Mol, J., and Kooter, J. (1997). The Silence of Genes in Transgenic Plants. *Annals of Botany*, 79:3–12.

Stenger, D. C. and French, R. (2008). Complete nucleotide sequence of a maize chlorotic mottle virus isolate from Nebraska. *Archives of Virology*, 153(5):995–997.

Stenger, D. C., Young, B. A., Qu, F., Morris, T. J., and French, R. (2007). Wheat streak mosaic virus Lacking Helper Component-Proteinase Is Competent to Produce Disease Synergism in Double Infections with Maize chlorotic mottle virus. 97(10):1213–1221.

Stewart, L. R., Jarugula, S., Zhao, Y., Qu, F., and Marty, D. (2017). Identification of a maize chlorotic dwarf virus silencing suppressor protein. *Virology*, 504(November 2016):88–95.

Syller, J. (2012). Facilitative and antagonistic interactions between plant viruses in mixed infections. *Molecular plant pathology*, 13(2):204–16.

Szittya, G., Silhavy, D., Molnár, A., Havelda, Z., Lovas, Á., Lakatos, L., Bánfalvi, Z., and Burgyán, J. (2003). Low temperature inhibits RNA silencing-mediated defence by the control of siRNA generation. *EMBO Journal*, 22(3):633–640.

Sztuba-Solińska, J., Urbanowicz, A., Figlerowicz, M., and Bujarski, J. J. (2011). RNA-RNA recombination in plant virus replication and evolution. *Annual review of phytopathology*, 49:415–43.

Tajima, F. (1993). Statistical analysis of DNA polymorphism.

Takeda, A., Iwasaki, S., Watanabe, T., Utsumi, M., and Watanabe, Y. (2008). The mechanism selecting the guide strand from small RNA duplexes is different among Argonaute proteins. *Plant and Cell Physiology*, 49(4):493–500.

Takeda, A., Tsukuda, M., Mizumoto, H., Okamoto, K., Kaido, M., Mise, K., and Okuno, T. (2005). A plant RNA virus suppresses RNA silencing through viral RNA replication. *The EMBO journal*, 24(17):3147–3157.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular biology and evolution*, 30(12):2725–9.

Teakle, D. and Grylls, N. (1973). Four strains of sugarcane mosaic virus infecting cereals and other grasses in Australia. *Australian Journal of Agricultural Research*, 24(24):465–477.

Teyssandier, E.E., S. F. N. and Bo., E. (1983). Maize virus diseases in Argentina. In *Proceedings International Maize Virus Disease Colloquium and Workshop*, pages 87–92.

Tollenaere, C., Susi, H., and Laine, A. L. (2015). Evolutionary and Epidemiological Implications of Multiple Infection in Plants. *Trends in Plant Science*, 21(1):80–90.

Tournier, B., Tabler, M., and Kalantidis, K. (2006). Phloem flow strongly influences the systemic spread of silencing in GFP Nicotiana benthamiana plants. *Plant Journal*, 47(3):383–394.

Tschopp, M.-a., Iki, T., Brosnan, C. A., Jullien, P. E., and Pumplin, N. (2017). A complex of Arabidopsis DRB proteins can impair dsRNA processing. *RNA*, 23:782–797.

Tugume, A. K., Cuellar, W. J., Mukasa, S. B., and Valkonen, J. P. T. (2010). Molecular genetic analysis of virus isolates from wild and cultivated plants demonstrates that East Africa is a hotspot for the evolution and diversification of Sweet potato feathery mottle virus. *Molecular Ecology*, 19(15):3139–3156.

Uyemoto, J. K. (1983). Biology and control of maize chlorotic mottle virus. *Plant disease*, 67(1):7–10.

Valli, A., García, J. A., and López-Moya, J. J. (2015). Potyviridae. *eLS*, pages 1–10.

Valli, A., López-Moya, J. J., and García, J. A. (2007). Recombination and gene duplication in the evolutionary diversification of P1 proteins in the family Potyviridae. *The Journal of general virology*, 88(Pt 3):1016–28.

Valli, A., Martín-Hernández, A. M., López-Moya, J. J., and García, J. A. (2006). RNA silencing suppression by a second copy of the P1 serine protease of Cucumber vein yellowing ipomovirus, a member of the family Potyviridae that lacks the cysteine protease HCPro. *Journal of virology*, 80(20):10055–10063.

Varanda, C. M. R., Machado, M., Martel, P., Nolasco, G., Clara, M. I. E., and Félix, M. R. (2014a). Genetic diversity of the coat protein of olive mild mosaic virus (OMMV) and tobacco necrosis virus D (TNV-D) isolates and its structural implications. *PLoS ONE*, 9(10).

Varanda, C. M. R., Nolasco, G., Clara, M. I., and Félix, M. R. (2014b). Genetic diversity of the coat protein of olive latent virus 1 isolates. *Archives of Virology*, 159(6):1351–1357.

Vaucheret, H., Nussaume, L., Palauqui, J., Quillere, I., and Elmayan, T. (1997). A Transcriptionally Active State Is Required for Post-Transcriptional Silencing (Cosuppression) of Nitrate Reductase Host Genes and Transgenes. *The Plant cell*, 9(8):1495–1504.

Vazquez, F., Legrand, S., and Windels, D. (2010). The biosynthetic pathways and biological scopes of plant small RNAs. *Trends in plant science*, 15(6):337–45.

Vignuzzi, M., Stone, J. K., Arnold, J. J., Cameron, C. E., and Andino, R. (2006). Cooperative Interactions Within a Viral Population. *Biochemistry*, 439(7074):344–348.

Viswanathan, R. and Balamuralikrishnan, M. (2005). Impact of Mosaic Infection on Growth and Yield of Sugarcane. *Sugar Tech*, 7:61–65.

Wang, C.-y., Zhang, Q.-f., Gao, Y.-z., Xie, L., Li, H.-m., and Hong, J. (2015a). Uncoating Mechanism of Carnation Mottle Virus Revealed by Cryo-EM Single Particle Analysis. *Nature Publishing Group*, 5:3–9.

Wang, D., Yu, C., Wang, G., Shi, K., Li, F., and Yuan, X. (2015b). Phylogenetic and recombination analysis of Tobacco bushy top virus in China. *Virology Journal*, 12(1):111.

Wang, J. G., Zheng, H. Y., Chen, H. R., Adams, M. J., and Chen, J. P. (2010a). Molecular diversities of sugarcane mosaic virus and sorghum mosaic virus isolates from Yunnan Province, China. *Journal of Phytopathology*, 158(6):427–432.

Wang, Q., Zhang, C., Wang, C., Qian, Y., Li, Z., Hong, J., and Zhou, X. (2017). Further characterization of Maize chlorotic mottle virus and its synergistic interaction with Sugarcane mosaic virus in maize. *Scientific Reports*, 7(November 2016):39960.

Wang, X.-B., Wu, Q., Ito, T., Cillo, F., Li, W.-X., Chen, X., Yu, J.-L., and Ding, S.-W. (2010b). RNAi-mediated viral immunity requires amplification of virus-derived siRNAs in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences*, 107(1):484–489.

Wangai, A. W., Redinbaugh, M. G., Kinyua, Z. M., Miano, D. W., Leley, P. K., Kasina, M., Mahuku, G., Scheets, K., and Jeffers, D. (2012). First Report of Maize chlorotic mottle virus and Maize Lethal Necrosis in Kenya. *Plant Disease*, 96(10):1582–1582.

Waterhouse, P. M., Graham, M. W., and Wang, M.-B. (1998). Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. *Proceedings of the National Academy of Sciences*, 95(23):13959–13964.

Watson, S. J., Welkers, M. R. A., Depledge, D. P., Coulter, E., Breuer, J. M., de Jong, M. D., and Kellam, P. (2013). Viral population analysis and minority-variant detection using short read next-generation sequencing. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 368(1614):20120205.

Wheeler, T. and von Braun, J. (2013). Climate change impacts on global food security. *Science*, 341(August):479–485.

Wintermantel, W. M., Cortez, A. a., Anchieta, A. G., Gulati-Sakhuja, A., and Hladky, L. L. (2008). Co-infection by two criniviruses alters accumulation of each virus in a host-specific manner and influences efficiency of virus transmission. *Phytopathology*, 98(12):1340–1345.

Wright, S. (1951). The Genetical Structure of Populations. *Annals of Eugenics*, 15:322–354.

Wu, B., Pogany, J., Na, H., Nicholson, B. L., Nagy, P. D., and White, K. A. (2009). A discontinuous RNA platform mediates RNA virus replication: Building an integrated model for RNA-based regulation of viral processes. *PLoS Pathogens*, 5(3):1–13.

Wu, J., Yang, Z., Wang, Y., Zheng, L., Ye, R., Ji, Y., Zhao, S., Ji, S., Liu, R., Xu, L., Zheng, H., Zhou, Y., Zhang, X., Cao, X., Xie, L., Wu, Z., Qi, Y., and Li, Y. (2015). Viral-inducible Argonaute18 confers broad-spectrum virus resistance in rice by sequestering a host microRNA. *eLife*, 4:e05733.

Wu, L., Zu, X., Wang, S., and Chen, Y. (2012). Sugarcane mosaic virus – Long history but still a threat to industry. *Crop Protection*, 42:74–78.

Wu, Q., Wang, X., and Ding, S. W. (2010). Viral suppressors of RNA-Based viral immunity: Host targets. *Cell Host and Microbe*, 8(1):12–15.

Xia, Z., Peng, J., Li, Y., Chen, L., Li, S., Zhou, T., and Fan, Z. (2014). Characterization of small interfering rnas derived from sugarcane mosaic virus in infected maize plants by deep sequencing. *PLoS ONE*, 9(5):28–33.

Xia, Z., Zhao, Z., Chen, L., Li, M., Zhou, T., Deng, C., Zhou, Q., and Fan, Z. (2016). Synergistic infection of two viruses MCMV and SCMV increases the accumulations of both MCMV and MCMV-derived siRNAs in maize. *Scientific Reports*, 6(November 2015):20520.

Xiao, X. W., Frenkel, M. J., Teakle, D. S., Ward, C. W., and Shukla, D. D. (1993). Sequence diversity in the surface-exposed amino-terminal region of the coat proteins of seven strains of sugarcane mosaic virus correlates with their host range. *Archives of Virology*, pages 399–408.

Xie, L., Zhang, J., Wang, Q., Meng, C., Hong, J., and Zhou, X. (2011). Characterization of Maize Chlorotic Mottle Virus Associated with Maize Lethal Necrosis Disease in China. *Journal of Phytopathology*, 159(3):191–193.

Xie, Q. and Guo, H. S. (2006). Systemic antiviral silencing in plants. *Virus Research*, 118(1-2):1–6.

Xie, X., Chen, W., Fu, Q., Zhang, P., An, T., Cui, A., and An, D. (2016). Molecular variability and distribution of Sugarcane mosaic virus in Shanxi, China. *PLoS ONE*, 11(3):1–12.

Xu, G., Greene, G. H., Yoo, H., Liu, L., Marqués, J., Motley, J., and Dong, X. (2017). Global translational reprogramming is a fundamental layer of immune regulation in plants. *Nature*.

Xu, K. and Nagy, P. D. (2016). Enrichment of Phosphatidylethanolamine in Viral Replication Compartments via Co-opting the Endosomal Rab5 Small GTPase by a Positive-Strand RNA Virus. *PLoS Biology*, 14(10):1–24.

Yahaya, A., Al Rwahnih, M., Dangora, D. B., Gregg, L., Alegbejo, M. D., Lava Kumar, P., and Alabi, O. J. (2017). First Report of Maize yellow mosaic virus Infecting Sugarcane (Saccharum spp.) and Itch Grass (Rottboellia cochinchinensis) in Nigeria. *Plant Disease*, 101(7):1335–1335.

Yang, S.-J., Carter, S. A., Cole, A. B., Cheng, N.-H., and Nelson, R. S. (2004). A natural variant of a host RNA-dependent RNA polymerase is associated with increased susceptibility to viruses by Nicotiana benthamiana. *Proceedings of the National Academy of Sciences of the United States of America*, 101(16):6297–302.

Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R. W., Steward, R., and Chen, X. (2005). Methylation as a Crucial Step in Plant microRNA Biogenesis. *Science*, 307(February):932–936.

Zhan, B., Lang, F., Zhou, T., and Fan, Z. (2016). Nuclear import of Maize chlorotic mottle virus capsid protein is mediated by importin-α. *European Journal of Plant Pathology*.

Zhang, L., Chia, J.-M., Kumari, S., Stein, J. C., Liu, Z., Narechania, A., Maher, C. a., Guill, K., McMullen, M. D., and Ware, D. (2009). A genome-wide characterization of microRNA genes in maize. *PLoS genetics*, 5(11):e1000716.

Zhang, Y., Yin, X., Yang, A., Li, G., and Zhang, J. (2005). Stability of inheritance of transgenes in maize (Zea mays L.) lines produced using different transformation methods. *Euphytica*, 144(1-2):11–22.

Zhao, M., Ho, H., Wu, Y., He, Y., and Li, M. (2014). Western Flower Thrips (Frankliniella occidentalis) Transmits Maize Chlorotic Mottle Virus. *Journal of Phytopathology*, 162(7-8):532–536.

Zhong, Y., Guo, A., Li, C., Zhuang, B., Lai, M., Wei, C., Luo, J., and Li, Y. (2005). Identification of a naturally occurring recombinant isolate of Sugarcane mosaic virus causing maize dwarf mosaic disease. *Virus Genes*, 30(1):75–83.

Zhu, X. and Ye, K. (2014). Cmr4 is the slicer in the RNA-targeting Cmr CRISPR complex. *Nucleic acids research*, pages gku1355–.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, 31(13):3406–3415.

Zwart, M. P., Willemsen, A., Daròs, J. A., and Elena, S. F. (2014). Experimental evolution of pseudogenization and gene loss in a plant RNA virus. *Molecular Biology and Evolution*, 31(1):121–134.