Diagnosis and Genome Analysis of Lettuce Necrotic Yellows Virus Subgroups

Priyadarshana Ajithkumar

A thesis submitted to
Auckland University of Technology
in partial fulfilment of the requirements for the degree of
Master of Science (MSc)

2018

School of Science

Supervisor: Dr. Colleen Higgins

Dedicating this thesis to my

beloved grandparents

Abstract

Lettuce necrotic yellows virus is the type species of *Cytorhabdovirus*. This plant virus causes a disease that is most frequently reported in lettuce in Australia and New Zealand. The lettuce necrotic yellows virus (LNYV) population comprises two subgroups; subgroup I and subgroup II. The subgroups were previously identified by phylogenetic analysis of LNYV, and a diagnostic method distinguishing these subgroups has not yet been developed. In the current study, a diagnostic test for the LNYV subgroups based on reverse transcription polymerase chain reaction (RT-PCR) and RT-PCR- restriction fragment length polymorphism (RFLP) was developed and used for subgroup diagnosis. Subgroup specific primers were designed and tested on known infected samples. RT-PCR diagnosis of LNYV subgroups with these primers requires the subgroup specific primers are used in separate reactions, requiring two reactions for each sample being tested. The RT-PCR-RFLP diagnostic test allows amplification of an LNYV sequence using all the subgroup primers combined, followed by a restriction digest to generate a diagnostic pattern of DNA fragments that can be identified by gel electrophoresis.. The previously designed primers, BCNG1/BCNG2, and LNYV_440F/LNYV_1185R primer pairs were used for LNYV diagnosis. The conditions of these primers were re-optimised for the use in AUT laboratory. The as above mentioned primers were used to test for LNYV and its subgroup on potentially LNYV infected plants collected from Auckland, Waikato and Canterbury. A total of ten samples were tested positive for LNYV; three were subgroup I, six were subgroup II and one sample was LNYV subgroup unknown. LNYV subgroups can now be diagnosed more rapidly than by the previously used sequencing and phylogenetic analysis. The results also showed that LNYV_440F/LNYV_1185R primer pair was more efficient than BCNG1/BCNG2 primers to detect LNYV. The false negative results caused by BCNG1/BCNG2 primer pair could be due to RNA degradation.

Only one complete genome of LNYV (subgroup I) has been reported, which was obtained from an Australian isolate. In this study, the complete genomes of LNYV subgroups I and II from New Zealand isolates were sequenced by Illumina HiSeq. Phylogenetic analyses of LNYV genomes and all the available cytorhabdovirus and nucleorhabdovirus genomes were carried out. The results showed that LNYV subgroup I genomes are most closely related to each other than to subgroup II. Lettuce yellow mottle virus was the most closely related to LNYV. Phylogenetic analyses of the LNYV nucleocapsid gene

sequences were also performed. The amino acid phylogenetic analysis shows that the AU9 isolate (subgroup II from Australia) appears to be closely related to the common ancestor, which indicates the origin of subgroup II. Since the complete genome or other gene sequences from the AU9 isolate are not available, the origin of LNYV cannot be confirmed. More samples from both Australia and New Zealand are necessary to understand these relationships more clearly.

LNYV subgroup I isolate has not been detected in Australia since 1993 and subgroup II may have outcompeted subgroup I in Australia, while this has not occurred in New Zealand. It was hypothesised by previously that subgroup II may have a more efficient relationship with the insect vectors and hosts. The glycoprotein was specifically analysed in the current study because rhabdoviruses use glycoprotein to attach and penetrate to the insect vectors/plant hosts. It was hypothesised that analysis of the glycoprotein may help to determine if subgroup II has a higher efficient relationship with insect/plant hosts than subgroup I. Six characteristics of glycoprotein sequence and 2D structure were analysed. It showed there were differences between the subgroups. However, a 3D structure and mutational analysis are needed to determine if the differences affect its association with the insect/plant hosts.

Table of contents

Chapter 1 General introduction

Chapter 2 Development of diagnostic tests for LNYV subgroups

Chapter 3 LNYV genome sequence and glycoprotein analysis

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of Tables

Attestation of Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgements), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning.

Priyadarshana Ajithkumar

Acknowledgement

It is my pleasure to acknowledge and extend my heartfelt gratitude to those individuals who supported and helped me throughout my Master's research.

First and foremost, I offer my sincerest gratitude to my supervisor Dr. Colleen Higgins, for her support, encouragement, patience and motivation during this research period. It was a great privilege and honour working with an amazing and dedicated supervisor like her. Her advice and confidence in me helped bring out my potentials, improved my skills and I was able to overcome various challenging situations. Her exceptional scientific knowledge and critical thinking helped me a lot on this journey. I want to thank her for opening the world of plant viral research and I was greatly inspired by her. I have learned not only about research but also self-discipline, taking responsibility and perseverance.

I am grateful to Chris Puli'uvea who is a part of our research team, for his great efforts to support and motivate me from the beginning of this project until the very end. I am thanking him for patiently training me, always keeping a watchful eye on my progression and making this journey enjoyable. His positive attitude, advice and optimism helped me a lot.

I would like to thank everyone in our research team; Dr. Gardette Valmonte-Cortes, Lee Rabbidge, Elizabeth Buckley, Tony Hull and Timothy Lawrence for your valuable suggestions, discussions, ideas and enjoyable moments. Many thanks to all the technicians and AUT staff who helped me along the way. I would also like to sincerely thank John Fletcher from Plant and Food Research for kindly providing me the Canterbury samples.

Finally, my dear Amma (mummy), Achan (daddy) and Abhi (brother) deserve special thanks for always being there for me and giving me confidence. Their endless support and love helped me focus completely on my research. I also want to thank my two dear friends Prexya Pradhan and Manjot Benipal for the endless encouragement and motivation during all these years. They have always stood by me every step of the way and encouraged me to pursue my dreams. My family and friends have always believed in me more than I believe in myself which enable me to fulfil my aspirations.

Abbreviations

bp	base pair
°C	degrees Celsius
M	molar
ml	millilitre
mM	milimolar
nt	nucleotide
µg	microgram
µl	microliter
kDa	kilodalton
4b	Movement protein
ADV	*Alfalfa dwarf virus*
AGO	Argonaute
ATP	Adenosine triphosphate
BEAST	Bayesian evolutionary analysis by sampling trees
BLAST	Basic Local Alignment Search Tool
BNYV	*Broccoli necrotic yellows virus*
BYSMV	*Barley yellow striate mosaic virus*
CBDaV	*Colocasia bobone disease-associated virus*
Ccd	Charge coupled device
cDNA	Complementary DNA
CMV	*Cucumber mosaic virus*
DAS-ELISA	Double antibody sandwich enzyme-linked immunosorbent assay
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotide triphosphates
ds	Double strand
DYVV	*Datura yellow vein virus*
ELISA	Enzyme-linked immunosorbent assay
EM	Electron microscopy
EMDV	*Eggplant mottled dwarf virus*
ER	Endoplasmic reticulum
G	Glycoprotein
GLRaV-3	*Grapevine leafroll-associated virus 3*

ICTV	International Committee on Taxonomy of Viruses
L	RNA dependent RNA polymerase
LMV	*Lettuce mosaic virus*
LNYV	*Lettuce necrotic yellows virus*
LYMoV	*Lettuce yellow mottle virus*
M	Matrix protein
MFSV	*Maize fine streak virus*
MIMV	*Maize Iranian mosaic virus*
ML	Maximum likelihood
MMV	*Maize mosaic virus*
MPI	Ministry of Primary Industries
mRNA	Messenger RNA
MW	Molecular weight
N	Nucleocapsid
NCBI	National centre for biotechnology information
NCMV	*Northern cereal mosaic virus*
NJ	Neighbour joining
NSS	Non subgroup specific
ORF	Open reading frame
P	Phosphoprotein
PCR	Polymerase chain reaction
PeVA	*Persimmon virus A*
PLRV	*Potato leafroll virus*
PYDV	*Potato yellow dwarf virus*
q-PCR	Real-time polymerase chain reaction
RDRP6	RNA dependent RNA polymerase 6
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
RSMV	*Rice stripe mosaic virus*
RSS	RNA silencing suppressor
RSV	*Respiratory syncytial virus*
RT-PCR	Reverse transcription-polymerase chain reaction
RYSV	*Rice yellow stunt virus*
SBYV	*Sweet beet yellow virus*

SCV	*Strawberry crinkle virus*
SGS3	Suppressor of gene silencing 3
SonV	*Sonchus virus*
Spp.	Species
ss	Single strand
SYNV	*Sonchus yellow net virus*
SYVV	*Sowthistle yellow vein virus*
TaVCV	*Taro vein chlorosis virus*
TBE	Tris/Borate/EDTA
TEM	Transmission electron microscope
Tm	Primer melting temperature
TMV	*Tobacco mosaic virus*
TSWV	*Tomato spotted wilt virus*
TVCV	*Turnip vein-clearing tobamovirus*
TuMV	*Turnip mosaic virus*
UV	Ultra violet
vRNP	Viral ribonucleoprotein
WstMV	*Wheat American striate mosaic virus*

Chapter 1

General Introduction

Chapter 1 General introduction

1.1 Introduction

Viruses are ubiquitous and intracellular parasites that are comprised of nucleic acid (DNA or RNA) and capsid protein (Breitbart and Rohwer 2005; Gelderblom 1996; Lodish et al. 2000). The majority of infectious viruses cause significant destruction to agriculture, human health and most living organisms. According to Zhao et al. (2017), 950 plant virus varieties had been published by the International Committee on Taxonomy of Viruses (ICTV) in the 2011 report. These viruses cause up to $60 billion loss to the world economy per year and $20 billion loss was caused by plant pathogenic viruses that infect food crops (Zhao et al. 2017). The frequency of such occurrence dramatically increases the need for high quality and accurate diagnostic methods to detect plant viruses, especially those impacting the economy. Rapid diagnosis could reduce further dispersion and increase the containment of the disease. Therefore, this aspect along with genome and glycoprotein analysis were studied with regards to Lettuce necrotic yellows virus (LNYV) subgroups, which infect lettuce. This viruses causes a negative impact on the economy, mainly in Australia and New Zealand; the lettuce industry is estimated to be worth $43 million annually to New Zealand (Nixon 2015). LNYV is taxonomically classified in the order Mononegavirales and family *Rhabdoviridae* (Dietzgen et al. 2007).

1.2 Family *Rhabdoviridae*

The family *Rhabdoviridae* is one of three viral families that infect a broad spectrum of hosts that include vertebrates (terrestrial and aquatic), invertebrates and plants (Dietzgen et al. 2017; Walker et al. 2011). The *Rhabdoviridae* are classified as negative sense, single-stranded RNA viruses with 18 currently recognised genera and 131 species (ICTV 2016). Four plant rhabdovirus genera are recognised; *Cytorhabdovirus*, *Nucleorhabdovirus*, *Dichorhavirus* and *Varicosavirus* (Dietzgen et al. 2017). Dichorhaviruses and Varicosaviruses consist of bipartite genomes while cytorhabdoviruses and nucleorhabdoviruses have monopartite genomes. Currently, among the ten assigned *Nucleorhabdovirus* spp., the complete genome of nine nucleorhabdoviruses have been sequenced (Dietzgen et al. 2017; ICTV 2016). Among the 11 assigned *Cytorhabdovirus* spp., the National Centre for Biotechnology Information

(NCBI) genome database (https://www.ncbi.nlm.nih.gov/genome/) contains the complete genome of ten (Table 3.1) (ICTV 2016). However, genomes for cytorhabdovirus isolates are generally unavailable, which will be described in the following section. Analysis of all subgroup genomes within a cytorhabdovirus would help to determine the difference in replication, transmission, protein structure (primary to tertiary), detection, pathogenicity, evolution and infection at the molecular level. This analysis may help to develop disease detection and control methods or strategies for the subgroup that is the most infectious.

Cytorhabdoviruses and nucleorhabdoviruses are differentiated by the site of replication within the host cells. Maturation of cytorhabdoviruses and nucleorhabdoviruses occurs in the cytoplasm and nucleus of infected plant cells, respectively, (Dietzgen et al. 2017). These genera are also distinguishable from each other at the molecular level (Higgins et al. 2016b; Mann and Dietzgen 2014). This demonstrates that development of molecular diagnosis methods will be able to differentiate these viruses.

1.3   Genus *Cytorhabdovirus*

The genus *Cytorhabdovirus* has the largest number of species within a plant rhabdovirus genus, consisting of 11 assigned species (ICTV 2016). Strawberry crinkle virus (SCV) was the first cytorhabdovirus to be discovered in 1932 (Dietzgen 2011). Barley yellow striate mosaic virus (BYSMV), lettuce necrotic yellows virus (LNYV) and SCV are the most economically important cytorhabdoviruses (Dietzgen et al. 2007; Sastry 2013a; Yan et al. 2015). Therefore, it is critical to research these three viruses, especially the viral transmission and diagnosis to reduce the impact of these viruses on the economy. This thesis focuses on LNYV.

1.4   *Lettuce necrotic yellows virus*

LNYV is the type species of the *Cytorhabdovirus* genus (Dietzgen et al. 2017; ICTV 2016). Compared to other species of this genus, LNYV is the most extensively researched. In 1954, *Lactuca sativa* (lettuce) was the first species to be identified as a host (Stubbs and Grogan 1963b). There was frequent reporting of LNYV infected plants from Australia and New Zealand with sporadic reports from other locations, namely Italy,

Spain and Great Britain (Higgins et al. 2016b). The current worldwide distribution of LNYV is unknown, although the frequency of reports from New Zealand and Australia suggests an antipodean origin. The virus could be present but unreported in other countries as a catastrophic outbreak or severe disease has not yet occurred. Therefore, it is essential to evaluate the LNYV transmission process and its optimum growth conditions to predict the likelihood of disease incidence.

1.4.1   Host plants

LNYV infects a broad range of flora, which includes both monocots and dicots. BYSMV, northern cereal mosaic virus (NCMV) and Wheat American striate mosaic virus (WstMV) exclusively infect monocots whereas broccoli necrotic yellows virus (BNYV), sonchus virus (SonV) and SCV infect dicots only (Dietzgen 2011). LNYV infection has been predominantly identified from Australia in many economically significant plants. These plants are lettuce, *L. serriola* (pricky lettuce), *Sonchus oleraceus* (sowthistle), *Reichardia tingitana*, *S. hydrophilus*, *Eucalyptus leucoxylon* ssp. *Megalocarpa, Cicer arietinum* (chickpea), *Calendula officinalis* (pot marigold calendula), *Lupinus albus* (white lupin), *L. angustifolius* (blue lupin), *Carthamus tinctorius* (safflower) and *Medicago polymorpha* (burr medic) (Behncken 1983; Edwardson and Christie 1991; Francki et al. 1989; Fry et al. 1973). *Allium sativum* L. (white Italian garlic) was the first monocotyledon species to be identified as a host (Sward 1990). In New Zealand, it has been detected in lettuce, *S. oleraceus* and the native plant *S. kirkii* (Fletcher et al. 2017; Fry et al. 1973; Higgins et al. 2016b). The genetic and environmental determinants of the LNYV host range is unknown. Identification of these determinants could explain the susceptibility to some hosts and not others.

Mechanical inoculation has been used to transmit LNYV into *Chenopodium quinoa, Gomphrena globosa* (globe amaranth), *Datura metel, D. stramonium, Petunia x hybrid, Macroptilium lathyroides* (Phasey bean)*, Nicotiana glutinosa, N. clevelandii, Spinacia oleracea* (spinach) and *Lycopersicon esculentum* (tomato, now called *Solanum lycopersicum*) (Behncken 1983; Dietzgen et al. 2007). *N. glutinosa* is recognised as a differential host for LNYV (Dietzgen et al. 2007). Differential hosts are known as indicator hosts that are easily susceptible and transmittable by the target viruses (Smith 1974). The hosts should have a short life-cycle, produce evident symptoms, and be able

to grow under lab conditions when compared to the naturally infected hosts (Reddy and Reddy 2012; Smith 1974). In contrast, mechanical inoculation of LNYV into *S. oleraceus* and lettuce was difficult (Dietzgen et al. 2007; Francki et al. 1989). There was no subsequent reporting of new LNYV hosts. The number of currently identified natural and mechanically inoculated hosts could be small and more hosts are likely to be identified in the future.

1.4.2   Transmission of LNYV by aphids

LNYV is persistently transmitted by aphids (*Hyperomyzus lactucae*) and non-transmissible by grafting, contact or pollen (Dietzgen et al. 2007; Muthaiyan 2009). Viral particles were identified in infected *H. lactucae* aphids within salivary glands, trachea, mycetome, muscles, brain, adipose tissue and epidermis, but not embryo cells, using light and electron microscopic analysis (O'Loughlin and Chambers 1967). It was concluded that LNYV replication occurs within aphids, but the virus is not transmitted to the progeny. Subsequent experiments of two generations confirmed that aphids indeed transmitted LNYV transovarially but at low frequency (Boakye and Randles 1974). Further work is required to confirm the transovarial transmission to more than two aphid generations. A detailed study of the transovarial transmission of plant viruses could help to identify the genetic or protein constituents involved in this mode of transmission. The constituents could be used to develop methods to reduce viral dispersion. LNYV glycoprotein could be one of the constituents, but there is no detailed study on its involvement in the transmission.

S. oleraceus is the natural host reservoir for aphids and LNYV but lettuce is a non-host for the aphids (Dietzgen et al. 2007). Experiments conducted by Boakye & Randles, 1974 showed constant exposure to light and low humidity (65% – 70%) for 24 hours of starvation increased the aphid settling on lettuce. Settling of aphids on *S. oleraceus* was comparatively higher in light/dark and dry/humid conditions with increasing duration of starvation. These results indicate there was no association between these abiotic factors and aphid settlement on *S. oleraceus* (Boakye and Randles 1974). These experiments were conducted in lab conditions rather than natural conditions; the results could be different in a natural environment because unlimited and uncontrolled factors are involved. These factors include altitude, temperature, UV light intensity, wind speed and

direction, humidity and latitude. Observational research on LNYV, hosts and aphid activity in nature has not been conducted, which could help to evaluate aphid transmission in a more natural setting.

LNYV transmission can occur from infected plants to uninfected aphids and conversely during feeding (Boakye and Randles 1974). The initiation of feeding occurs when an aphid insert its stylet into the leaves, injects saliva and consumes the phloem sap, simultaneously (Figure 1.6). This results in virus acquisition and transmission (Hull 2014). Boakye and Randles (1974), observed that prolonged feeding by aphids increases the likelihood of a successful LNYV transmission. In a study conducted by Fry et al. (1973), aphids were trapped in Mangere market garden, Auckland (1967 – 1970) and Pukekohe, Auckland (1969 – 1970) to analyse the abundance of aphids throughout the year. The highest number of aphids were between September to April (spring-summer) and was associated with an increase in LNYV incidence in lettuce during that period (Fry et al. 1973). Coalescing the results from both of the previously mentioned studies, it can be concluded that prolonged feeding and abundance of aphids during the spring-summer could be the main factors involved in high LNYV incidence in lettuce.

Climate is also one of the essential factors for aphid reproduction and is associated with the prevalence of LNYV disease incidence. Aphids undergo holocyclic (sexual) and anholocyclic (production of asexual females) reproduction. These cycles are determined by the niche and seasonally oriented (Bale 1991; Carver and Woolcock 1986). Both cycles occur simultaneously and consecutively. Aphids in the northern hemisphere undergo holocyclic reproduction. The host plants are *Sonchus* and *Ribes* during the summer and winter, respectively, consuming and depositing overwintering eggs on *Ribes rubrum* during the autumn in the northern hemisphere (Kalaisekar et al. 2017; Woiwod et al. 2000). In contrast, aphids could undergo holocyclic or anholocyclic in Australia and New Zealand, and thrive on *S. oleraceus* (Carver and Woolcock 1986). LNYV disease incidence in both hemispheres could be different due to the differences in aphid reproduction cycles. More research of LNYV in the northern hemisphere is required, which can be compared with the data available in the southern hemisphere. This information could be used to predict the LNYV disease incidence in both hemispheres.

LNYV is also transmitted by the aphid *H. carduellinus* and was found on *S. oleraceus*, *E. leucoxylon* ssp. *Megalocarpa* and *R. tingitana* (Randles and Carver 1971). The spread of LNYV by infected *H. carduellinus* was low in comparison to *H. lactucae* due to the limited geographical distribution of *H. carduellinus* (Fry et al. 1973; Randles and Carver 1971). A recent report by Fletcher et al. (2017) found evidence for the first time that *Nasonovia ribisnigri* (blackcurrant-lettuce aphid) could transmit LNYV in Marshland, Canterbury, New Zealand and Southbridge, Canterbury. A detailed study of infected *N. ribisnigri* has not been carried out yet. Even though LNYV was predominately transmitted by *H. lactucae*, there could be other unidentified insect vectors that can transmit LNYV to other hosts.

1.4.3 Symptoms

LNYV causes a broad spectrum of visual symptoms on infected host plants including lettuce that could result in consumer rejection (Fry et al. 1973). The primary symptoms of lettuce are discolouration, necrosis and mosaic in premature leaves, chlorosis in mature leaves, growth inhibition, leaf curling and gradual death (Figure 1.1) (Dietzgen et al. 2007; Fry et al. 1973). Symptoms on infected chickpea are mainly exhibited on the growth tips and shoots. This includes necrosis on petioles, leaves and axillary nodes, growth inhibition and discolouration. Severe infection causes plants to become feeble and mortal (Behncken 1983). In contrast, LNYV infected *Sonchus* sp. are asymptomatic (Martinez et al. 2013). Symptoms exhibited by the natural or mechanically inoculated hosts is the preliminary indication of LNYV infection; however, this is not a reliable diagnosis method, which will be discussed in section 1.4.10.1.
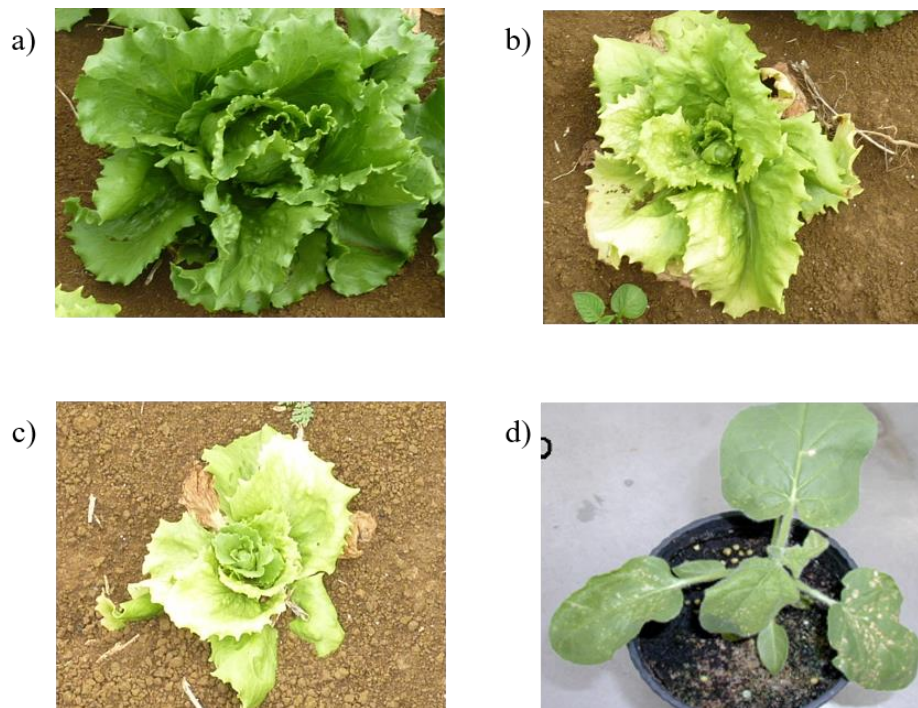
Figure 1.1 Comparison of LNYV infected lettuce and *N. glutinosa*. a) Healthy lettuce plant. b) Late infection of LNYV symptoms in lettuce. c) Early infection of LNYV in lettuce. d) Severe LNYV infected *N*. glutinosa. (Photos a, b, and c provided by Colleen Higgins, Auckland University of Technology; Photo d from (Dietzgen et al. 2007).

1.4.4   Epidemiology

LNYV was first discovered during 1954 as a 'destructive virus' infecting lettuce in Victoria, Australia and caused up to 100% crop loss (Stubbs and Grogan 1963a). Initially, it was misidentified as tomato spotted wilt virus (TSWV) - TSWV infects lettuce and induces similar symptoms to LNYV (Dietzgen et al. 2007). Subsequent research conducted by Stubbs and Grogan (1963b) identified that TSWV was not the infectious agent but rather a novel plant virus, naming it lettuce necrotic yellows virus. By 1961, the infected lettuce crops were found in Victoria, New South Wales, Queensland and South Australia (Stubbs and Grogan 1963b). Furthermore, nine infected samples were identified and sequenced from various geographical regions in Australia between 1985 – 2000 (Callaghan and Dietzgen 2005). These results were the initial reporting of LNYV in Australia. Lettuce cultivation in Australia is mainly in Perth, southwest Western Australia, Victoria, New South Wales, South Australia, Tasmania and southeast Queensland (Deuter et al. n.d.). Since there is low lettuce cultivation in Northern Territory, northern Queensland and northern Western Australia, LNYV infected lettuce

was not reported from these regions. However, it could have been present in plants that are economically not significant or because no outbreaks had occurred. There have been no reports of LNYV in Tasmania and more research is required to confirm the presence or absence in this State.

The first report of LNYV in New Zealand was from Blenheim, Marlborough in 1965 (Fry et al. 1973). The most severe incident occurred during 1969 in Waimauku, Auckland, where almost half of the lettuce crops in the area were infected (Fry et al. 1973). The New Zealand Plant Protection Society conducted a virus survey between 2003 – 2004 and identified LNYV infected lettuce from Pukekohe, Auckland, Gisborne and Hawke's Bay (Fletcher et al. 2005). A study conducted by Higgins et al. (2016b), identified six infected lettuce samples from Harrisville, Auckland, in 2011. A recent survey of lettuce virus disease undertaken by Plant and Food Research identified LNYV infected lettuce from various locations in Canterbury, New Zealand for the first-time (Fletcher et al. 2017). LNYV infection reports have been mainly confined to the North Island rather the South Island, possibly due to the greater lettuce cultivation in the North Island.

1.4.5   LNYV disease control

Management of plant viral diseases is challenging due to the complexity of viruses and transmission through various insect vectors (Sastry 2013b). The primary cause of LNYV infection on lettuce is due to the abundance of infected *S. oleraceus* near the lettuce crops (Teakle 1997). During the 1950s, an exotic species of rabbit was abundant in Australia that became invasive and was declared as a pest. Myxomatosis virus was used as a biological control agent to eradicate the rabbit population (Dietzgen et al. 2007; Fenner 1983). *S. oleraceus* is the primary source of food for rabbits; the eradication of rabbit resulted in exponential growth of *S. oleraceus* and a sudden increase of LNYV infection outbreaks was observed (Dietzgen et al., 2007; Jackson, Francki, & Zuidema, 1987).

Herbicide has been used to eradicate weeds, particularly *S, oleraceus*, resulting in up to 70% disease on lettuce reduction (Dietzgen et al. 2007; Fry et al. 1973). It was hypothesised that LNYV infection could also be diminished by the introduction of the aphid parasites *Aphidius sonchi* and *Praon volucre* to eliminate *H. lactuace* in Australia (Carver and Woolcock 1986). However, both *A. sonchi* and *P. volucre* failed to establish

because of unsuitable environmental conditions for *A. sonchi* and unknown reason for *P. volucre* (Carver and Woolcock 1986). The effectiveness of such eradication methods was contradicted by Teakle (1997) who proposed that aphids travel a long distance. Therefore, aphids are unlikely ever be eliminated. These results indicate that there is still a need for an appropriate control method to be developed.

1.4.6   LNYV virion structure

LNYV virions are bacilliform shaped, enveloped with dimensions of 227 nm in length and 68 nm (Figure 1.2a) (Dietzgen et al. 2007). The virion is surrounded by a lipid envelope with outwardly projecting glycoprotein spikes (G, MW approximately ~78 kDa) (Figure 1.2b). The matrix protein (M, approximately ~19 kDa) connects the G protein and nucleocapsid. The nucleocapsid is comprised of the viral RNS genome complexed with the viral nucleoprotein (N, approximately ~57 kDa), the phosphoprotein (P) and the RNA dependent RNA polymerase (L). The latter two proteins have MW of approximately ~38 kDa and ~170 kDa, respectively (Figure 1.2b). The 4b protein is a movement protein. The location within the virion structure and the molecular weight is unknown (Dietzgen et al. 2007). The function of these proteins is discussed in more detail in section 1.4.7.
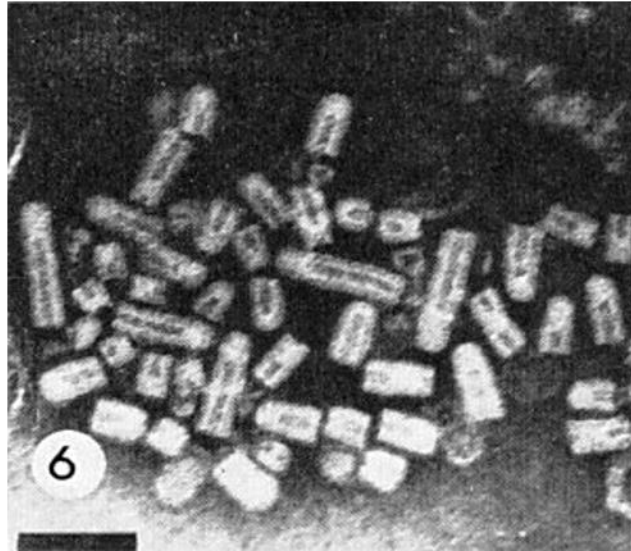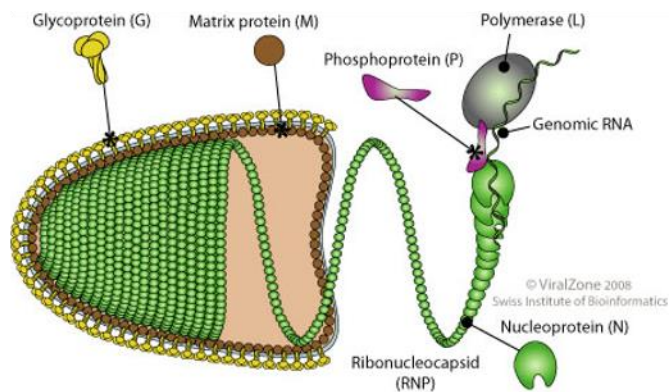
Figure 1.2: Electron micrograph and schematic diagram of LNYV virion. a) Electron micrograph of LNYV particle in an infected *S. oleraceus* leaf, scale bar: 200 nm. b) LNYV virion structure and proteins. Photo a from (Francki et al. 1989); Photo b from http://viralzone.expasy.org/all_by_species/77.html.

1.4.7 LNYV genome structure and protein functions

The LNYV RNA genome is negative sense single stranded of approximately 13,000 nucleotides (nt) in length (Wetzel et al. 1994a). The order of the genes in the genome is 3' leader (~ 84 nt) – N gene, P gene, 4b gene, M gene, G gene and L gene – 5' trailer (187 nt) (Figure 1.3) (Dietzgen et al. 2007). Each gene has an open reading frame (ORF), flanked by 3' untranslated regions (UTRs) and 5'UTRs. Between each gene are conserved polyadenylation signal, intergenic region and transcription start sequences of AUUCUUUU, GNU(C/U)(N)$_n$ACU, CUU, respectively (Bejerman et al. 2015; Dietzgen

et al. 2007; Higgins et al. 2016b). The (N)n is variable number of nucleotides (Dietzgen et al. 2007). The lettuce yellow mottle virus (LYMoV) polyadenylation signal sequence and the intergenic sequence are identical to that of LNYV whereas the LNYV transcription start sequence is identical to that of alfalfa dwarf virus (ADV) (Yang et al. 2016). The cytorhabdoviruses persimmon virus A (PeVA), NCMV, BYSMV, and rice stripe mosaic virus (RSMV) and the nucleorhabdoviruses rice yellow stunt virus (RYSV), potato yellow dwarf virus (PYDV), and sonchus yellow net virus (SYNV) have dissimilar polyadenylation signal, intergenic and transcription start sequences to LNYV (Bejerman et al. 2015; Yang et al. 2016). When all the viruses mentioned above are compared, LNYV is evolutionarily most closely related to LYMoV (Yang et al. 2016). Hence, the polyadenylation signal and intergenic sequences could be evolutionarily conserved among the closely related species. The intergenic region is likely to have multiple functions, primarily to regulate transcription and translation (Dorak 2017). No detailed study has been conducted on the LNYV intergenic sequence to determine how this region might impact the regulation of transcription and translation.
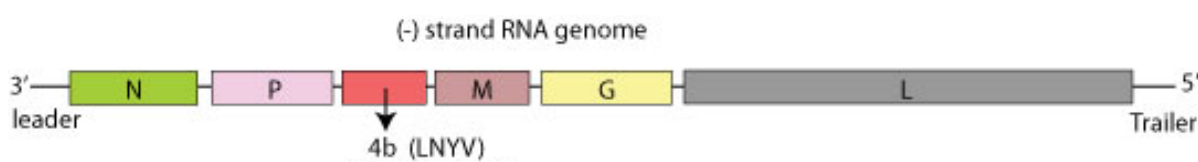
Figure 1.3: LNYV genome organisation with six genes.
Image from http://viralzone.expasy.org/all_by_species/77.html

N gene and protein

The viral nucleoprotein or nucleocapsid protein is encoded by the N gene and is the most studied gene of LNYV (Jackson et al. 2005). The functions of the N protein are genome encapsulation, regulation of transcription and replication, and association with the viral transcriptase enzyme. The first complete N gene sequence of LNYV was obtained from an isolate from garlic (Dietzgen et al. 2007; Wetzel et al. 1994b) and was shown to give rise to an mRNA of 1,377 nt encoding 459 amino acids (Figure 1.4). The 3' UTR has an end sequence of 'AAGAAAA' that may function as a polyadenylation signal (Wetzel et al. 1994b). Within the protein sequence, 21 locations were identified as potential phosphorylation sites and some of those regions were conserved among LNYV isolates

(Dietzgen et al. 2007). The functional significance of the multiple phosphorylation sites in LNYV and other plant rhabdoviruses is unknown. Tuffereau et al. (1985) conducted phosphorylation studies on rabies virus protein *in vitro* and showed that the hyperphosphorylation of the N protein resulted in a decrease of transcriptase activity. Similar studies could be conducted in LNYV infected plants/aphids to determine the impact of phosphorylation.

Sequencing and phylogenetic analysis of the N gene from 25 LNYV isolates from Australia and New Zealand showed that the LNYV population is made up of two subgroups. This will be discussed further in section 1.4.9 (Callaghan and Dietzgen 2005; Higgins et al. 2016b).

3'	5'UTR (78 nt)	ORF (1,377 nt)	3'UTR (75 – 81 nt)	5'
		Nucleocapsid gene		
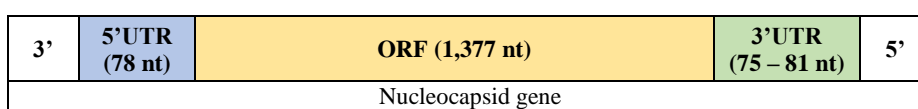
Figure 1.4: Nucleocapsid gene of LNYV with UTRs and ORF and the number of nucleotides.

P gene and protein

The P gene, also known as the 4a gene, encode a phosphoprotein. The P protein acts as the viral RNA silencing suppressor (RSS) (Dietzgen et al. 2007; Mann et al. 2016b). Mann et al. (2015) demonstrated that the LNYV P protein exhibited weak local RSS activity and hinders systemic silencing in *N. benthamiana.* In contrast, there was no LNYV P protein RSS activity within *Drosophila melanogaster* cells indicating the LNYV P protein specifically target the plant RNA silencing proteins rather than those in the insect vector (Mann et al. 2015). A follow-up study conducted by Mann et al. (2016b) showed that the LNYV P protein inhibits proteins involved in the RNA silencing pathways within *N. benthamiana.* These proteins include AGO1, AGO2, AGO4, RDR6 and SGS3, which require the C-terminal domain of the P protein for inhibition. The LNYV P protein interacts with the AGO1 protein and stabilise the protein, resulting in inhibition of microRNA-guided translation repression and cleavage of the target RNA. The P protein also attaches to the RDR6 and SGS3 proteins, which results in the RNA silencing amplification inhibition. The P protein is considered to be a new form of AGO

binding protein since the P sequence does not have GW/WG or F-box motifs (Mann et al. 2016b). The function of the LNYV P protein within the insect vector is unknown. Identification of the insect proteins that interact with the LNYV P protein would help understanding of LNYV replication within and transmission by the insect vectors.

4b gene and protein

The 4b gene encodes a protein now recognised as a movement protein of LNYV (Mann et al. 2016a; Martin et al. 2012; Wetzel et al. 1994a). The function was initially unknown until the first complete LNYV 4b gene sequence was obtained and analysed (Dietzgen et al. 2006; Wetzel et al. 1994a). The sequence resembled the movement proteins of viruses belonging to the *Trichovirus* and *Capillovirus* genera, which lead to the suggestion that 4b is a movement protein (Dietzgen et al. 2006). Other cytorhabdoviruses and nucleorhabdoviruses also contain similar genes as shown in Figure 1.5. The genes encoding the movement proteins are situated between the P gene and M gene within the genomes and are named as either P3, Sc4, Y or 4b. There is one movement protein gene for colocasia bobone disease-associated virus (CBDaV) (P3), ADV (P3), LYMoV (P3), PeVA (P3), LNYV (4b), rice yellow stunt virus (RYSV) (P3), Maize Iranian mosaic virus (MIMV) (P3), maize mosaic virus (MMV) (P3), taro vein chlorosis virus (TaVCV) (P3), eggplant mottled dwarf virus (EMDV) (Y), PYDV (Y), datura yellow vein virus (DYVV) (P3) and SYNV (SC4). In contrast, in this genome region other viruses have multiple ancillary genes could be involved in intercellular movement: BYSMV (P3 P4 P5 P6), NCMV (P3 P4 P5 P6) and maize fine streak virus (MFSV) (P3 P4) (Jackson et al. 2005; Mann et al. 2016a; Mann and Dietzgen 2014; Martin et al. 2012; Walker et al. 2011; Yan et al. 2015; Yang et al. 2016). The secondary structure of the movement protein has been predicted for a few of these viruses and indicated that there was structure similarity to that of the 30K superfamily of viral movement proteins that are involved in the cell-cell movement through plasmodesmata (Walker et al. 2011).

Using fluorescent protein localisation studies, Martin et al. (2012) found evidence that LNYV 4b and M proteins localise in the nucleus of *N. benthamiana*. This was unexpected since it has been accepted that the LNYV replication cycle occurs in the cytoplasm. The authors suggested that 4b may localise to the nucleus in order to recruit host transcription factors as has been found for the nucleorhabdovirus SYNV (Martin et al. 2012). This aspect has not been investigated further and is crucial to elucidate the LNYV infection

process in host plants. The M protein localisation will be discussed further in the following section. A recent experiment conducted by Mann et al. (2016a) showed that the LNYV 4b protein could trans-complement the movement protein (P30) of a turnip vein-clearing tobamovirus (TVCV) isolate containing a defective movement protein. The $LxD/N_{50-70}G$ motif is specific to the 30K superfamily; mutation in the LNYV 4b $LxD/N_{50-70}G$ motif resulted in movement dysfunction, indicating that this motif is essential for its function. This further supported 4b as a movement protein and was classified as a member of the 30K superfamily (Mann et al. 2016a). The association of LNYV 4b with insect vector proteins is unknown and this would be essential for understanding more deeply LNYV transmission.
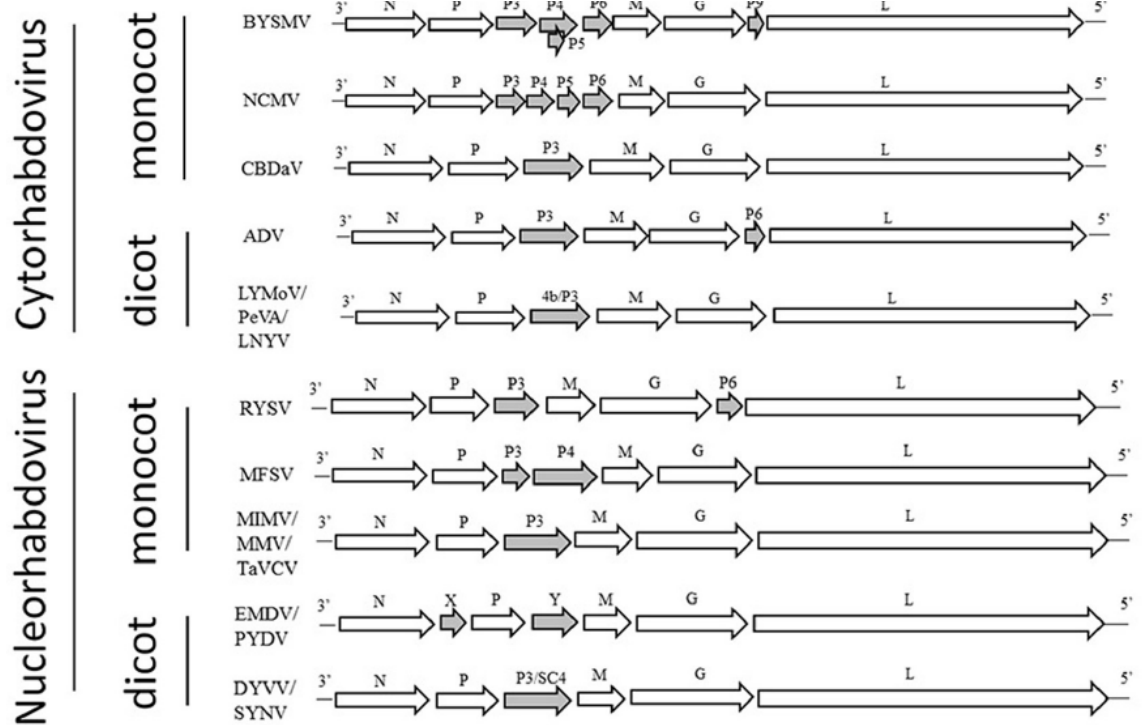
Figure 1.5: Genome organisation of cytorhabdoviruses and nucleorhabdoviruses. The ancillary genes are highlighted in grey (Yang et al. 2016).

M gene and protein

The LNYV M protein is encoded by the M gene (Dietzgen et al. 2007). The function of the M protein in LNYV is unknown but evidence from other rhabdoviruses suggest that it facilitates nucleocapsid condensation post viral maturation, regulates viral genome transcription and replication, increases budding of viruses into hosts and regulates host-cell transcription (Graham et al. 2008). As mentioned above Martin et al. (2012) showed the M protein was present in the nucleus of *N. benthamiana*. The authors hypothesised that the LNYV M protein inhibits export of host mRNA from the nucleus to reduce the resource competition by the virus, a process found to occur during infection by the animal vesicular stomatitis virus (VSV) (Martin et al. 2012). Protein-protein interactions were also examined in that study; the LNYV M protein was found to interact with itself and with the P protein. The M-M interaction has been observed within other plant rhabdoviruses and VSV, while the P-M interaction was unique to LNYV (Martin et al. 2012). The functional significance of these interactions in infected cells is unknown. LNYV protein localisation and protein-protein interaction studies should be carried out in aphids to determine the LNYV replication process in aphids.

G gene and protein

The glycoprotein gene (G) encodes a transmembrane membrane protein that creates homotrimer spikes on the exterior envelope of the virion. In rhabdoviruses, G protein spikes are assumed to be essential for viral assembly, budding and penetration into the host cells (Dietzgen et al. 2017). These spikes could bind to surface receptors of the insect vector midgut epithelial cells (Ammar et al. 2009; Coll 1995a). The recognition is highly specific binding of G protein and was observed for PYDV (Ammar et al. 2009). The role for the G protein for entry into a plant cell is less clear. Wang et al. (2015) had inoculated an isolate of SYNV with a G gene mutation (deletion) on *N. benthamiana*, the virus caused systemic infection with the expected symptoms, however, morphogenesis of SYNV was arrested (Wang et al. 2015). This indicated the G protein is not required for systemic infection, but is required for correct virion assembly. Such experiments are required to understand the G protein involvement in LNYV morphogenesis within its plant hosts and insect vector.

The G protein sequence predicted from the published LNYV subgroup I genome contains a 25 amino acid signal sequence that could be used to direct polypeptides to the

endoplasmic reticulum (ER) (Da Poian et al. 2005; Dietzgen et al. 2007; Dietzgen et al. 2006). Within the signal sequence, there is a peptidase recognition sequence "VQG↓V", (↓ is the predicted cleavage site) (Dietzgen et al. 2007; Dietzgen et al. 2006). The LNYV G protein sequence also consists of three potential glycosylation sites and amino acid sequence that encodes the glycosylation sites is Asn-X-Ser/Thr (Dietzgen et al. 2007). According to Dutta et al. (2017), glycosylation is generally required for cell generation, association with hosts and signal transduction. The type of glycosylation appears may be host specific (Dietzgen & Francki, 1988). Until this thesis, the G gene and proteins sequences from a subgroup II isolate of LNYV were unavailable. Comparisons can now be made to determine if predicted functional regions are conserved between the LNYV subgroups.

L gene and protein

The L protein is encoded by the largest gene in the LNYV genome. This protein is required for the LNYV RNA replication and transcription (Dietzgen et al. 2007). Analysis of the L gene sequence of other rhabdoviruses identified six conserved domains, or blocks, each with unique functions (Redinbaugh and Hogenhout 2005). The functions of block I, block II, block IV, block V and block VI include multiple polymerase function, RNA recognition or nucleotide binding, nucleotide binding, catalysis and polyadenylation, respectively. Block III contains four conserved motifs and was located in all polymerases (Redinbaugh and Hogenhout 2005). LNYV has a conserved motif called the "GDN" motif that has a polymerase catalytic centre (Dietzgen et al. 2006). The GDN motif was identified in all negative sense RNA viruses. While the positive sense RNA virus contain the "(G)DD" motif (Klerks et al. 2004). In rhabdovirus infected cells, it appears the L gene is transcribed to the least extent so that the L mRNA is the least abundant of all the virus mRNAs, and gives rise to low level of the L protein (Regenmortel and Mahy 2010). This suggests detection of the L gene or protein would not be the best choice for diagnostic purposes.

1.4.8 Replication

1.4.8.1 Replication of LNYV in aphids

Replication of plant rhabdoviruses occurs in insect vectors (Mann and Dietzgen 2014). While there has been little research on LNYV and related cytorhabdoviruses, most of what is understood about rhabdovirus replication in insect vectors comes from studies of leafhoppers (Ammar et al. 2009). The initial aphid vector feeding behaviour on plants infected with rhabdovirus is described in section 1.4.2. Consumption of phloem sap from infected plants by the insect could lead to ingestion of plant rhabdovirus particles into the midgut of the intestine (Ammar et al. 2009; Hull 2014). The midgut is a significant barrier because it contains digestive enzymes, limited membrane receptors and initiation of innate immune response could eradicate and inhibit translocation of the viruses (Ammar et al. 2009). Hence, plant rhabdoviruses use glycoprotein spikes, which bind to the midgut epithelial cell receptors to escape from the midgut by receptor-mediated endocytosis (Mann and Dietzgen 2014). Virus translocates to the salivary glands for proliferation, replication and transmission to uninfected plants (Ammar et al. 2009; Hull 2014). Figure 1.6 shows the pathways of plant rhabdovirus within the leafhopper to reach the salivary glands, and includes hemolymph, nervous system and muscle tissues.

Leafhoppers (Cicadellidae) and aphids (Aphididae) belong to the taxonomic order Hemiptera. These organisms have similar way of transmitting plant viruses (Ammar et al. 2009; Fereres and Moreno 2009). Hence, LNYV could follow similar replication and movement processes in aphids. O'Loughlin and Chambers (1967) detected LNYV particles in the brain, fat cells, muscles, salivary glands, alimentary canals, trachea and mycetome of infected aphids. However, the definitive pathways for LNYV through the aphid have not been discovered; these could provide greater understanding of the transmissibility of the virus to its hosts.
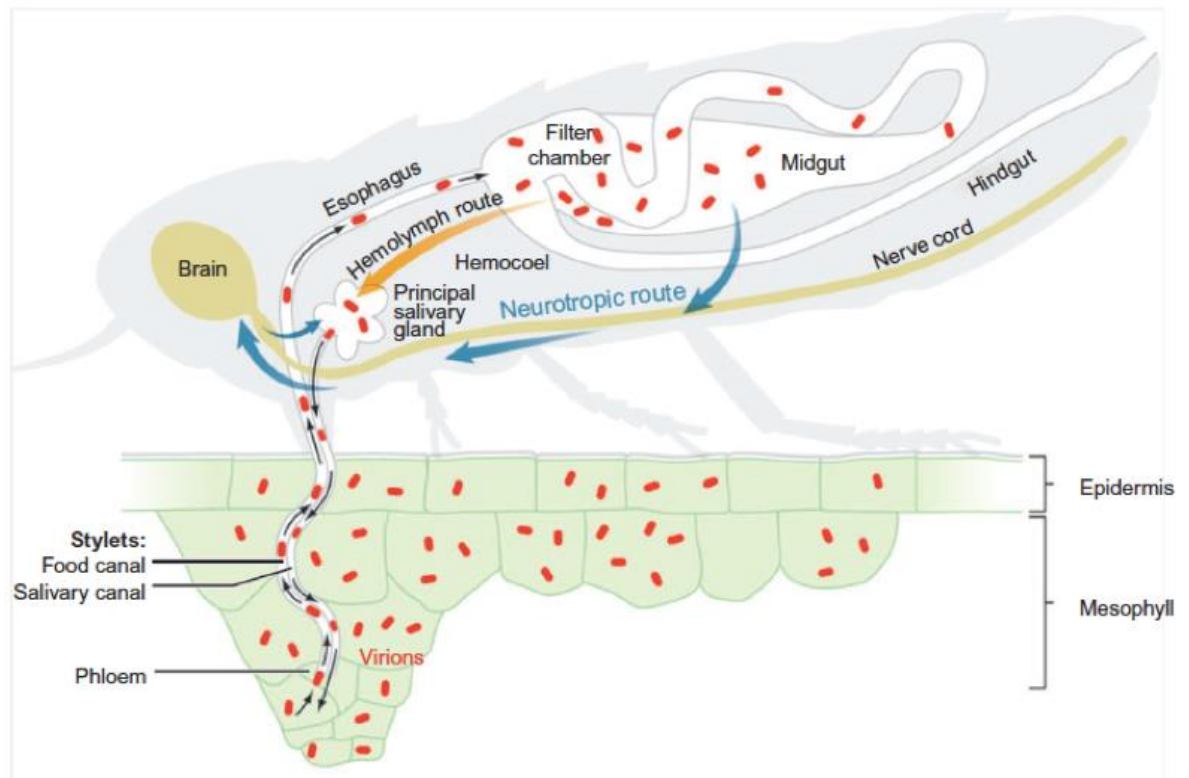
Figure 1.6: Diagram of a leafhopper feeding in a plant leaf along and the two pathways for plant rhabdovirus movement. Neurotropic route (blue arrow), hemolymph route (brown arrow) and ▬ = virion (Hull 2014).

1.4.8.2   Virus transcription and replication in plant cells

The cytorhabdovirus and nucleorhabdovirus infection cycles are distinctive by their replication sites, cytoplasm and nucleus, respectively, in the infected plant cells, as shown in Figure 1.7 (Dietzgen et al. 2017). Insect vector penetration into the plant cell during feeding, or mechanical abrasion, allow entry of cytorhabdoviruses and nucleorhabdoviruses when the glycoprotein binds to the plasma membrane receptors. Viruses uncoat the G protein and M protein on the cell's ER membrane after which the viral ribonucleocapsid (vRNP), which contains the negative sense RNA genome, is released into the cytoplasm. Then the nucleorhabdoviruses and cytorhabdoviruses replicate in the appropriate locations mentioned above (Figure 1.7) (Dietzgen et al. 2017; Gerlier and Lyles 2011; Hull 2014).
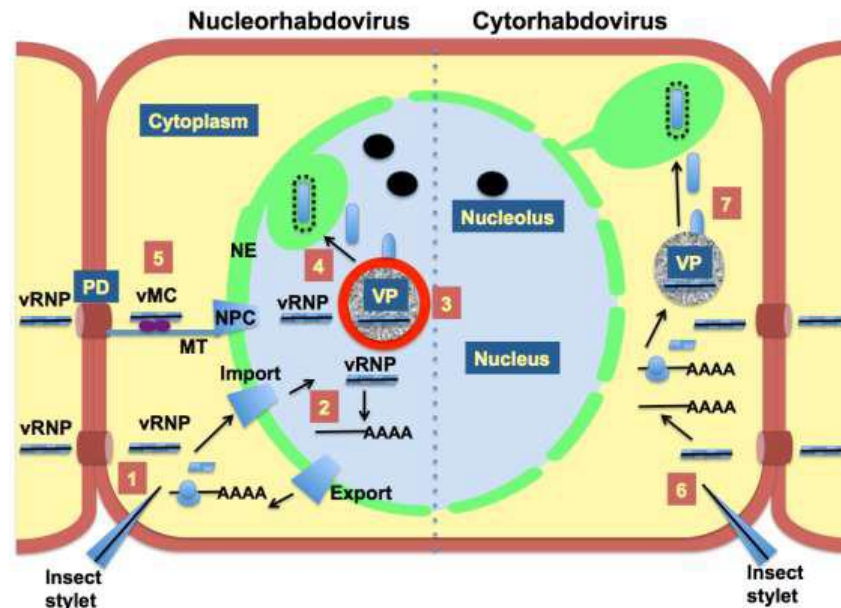
Figure 1.7: Replication process of *Cytorhabdovirus* and *Nucleorhabdovirus* comparison in plant cells. 1) Penetration of nucleorhabdoviruses into the plant cell wall by an insect vector and viral ribonucleoprotein (vRNP) move into the nucleus. 2) vRNP transcribes to mRNA, exports out to the cytoplasm and mRNA translates to make the proteins. Proteins are imported back into the nucleus to make the viroplasm (VP). 3) Replication of the RNA genome occur within VP. 4) Mature virions exports from the VP and binds to the perinuclear space. 5) Mature virions get transmitted by insect vectors, vRNP uses nuclear pore complex (NPC) to leave the nucleus and vRNP spread within the plants by viral movement complex through plasmodesmata. 6) Cytorhabdoviruses has a similar process to nucleorhabdoviruses but the replication process occurs in the cytoplasm. 7) The mature virions propagate in the ER and moves by using the movement proteins (Dietzgen et al., 2017).

The cytorhabdovirus vRNP associates with the ER, and the L protein activates viral gene transcription to generate the positive sense viral mRNA for each gene (Conzelmann 1998; Hull 2014). The L protein transcribes the 5' leader RNA of each gene and the L protein halts. The transcription resumes once the L protein reaches the transcription start signal. The 5' terminal signal is capped. When the L protein reaches the transcription stop terminal site, transcription stops and is polyadenylated (Figure 1.8) (Conzelmann 1998). Every viral gene is transcribed independently and the amount of transcription decreases from the N gene to the L gene (Hull 2014). The transcription results in the production of viral mRNAs for each viral protein, which are translated to make the viral proteins.
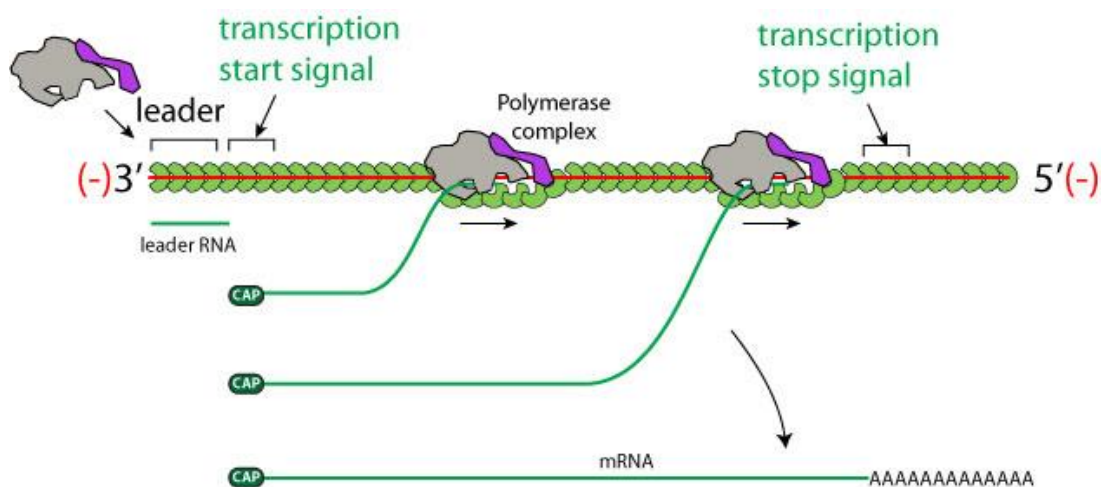
Figure 1.8: Translation of negative sense RNA genome.
Image from https://viralzone.expasy.org/1917

The viral proteins are required for the morphogenesis to form the viroplasms, a 'thread-like structure'. Replication begins when adequate amount of N protein is produced. Replication of the viral genome and additional viral mRNAs are translated within the viroplasms (Dietzgen et al. 2017; Hull 2014; Mann and Dietzgen 2014; Redinbaugh and Hogenhout 2005). The L protein replicates the negative sense RNA genome to produce the positive sense RNA genome (antigenome) by binding to the 3' leader. Replication of the antigenome is initiated when the L proteins bind to the 5' trailer to generate the complete negative sense RNA genome (Figure 1.9) (Conzelmann 1998; Dietzgen et al. 2007). Maturation occurs via matrix protein mediated condensation of nuclecapsid where the G protein is accumulated in the ER. Matured virions propagate in the ER and translocate through plasmodesmata using the movement proteins (Figure 1.7) (Dietzgen et al. 2017; Mann and Dietzgen 2014).
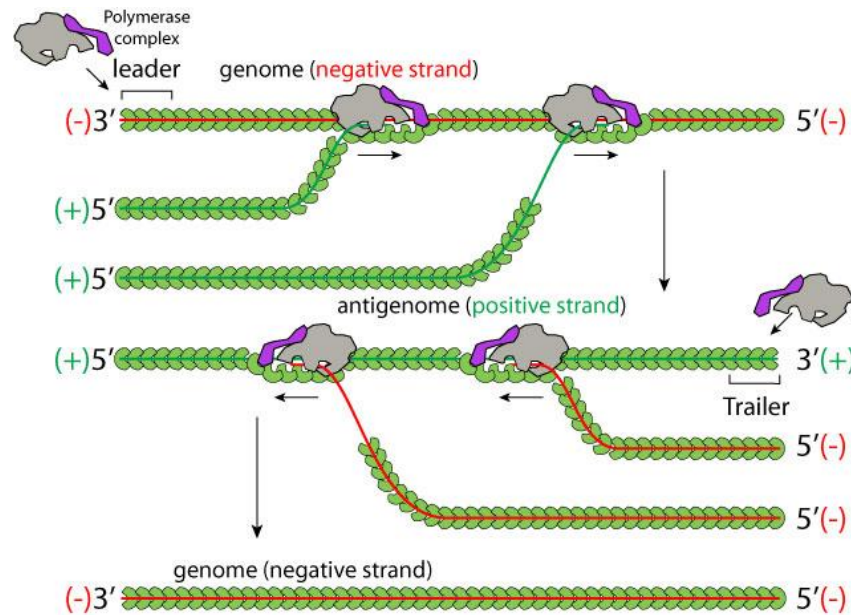
Figure 1.9: Replication of negative sense RNA genome.
Image from https://viralzone.expasy.org/1096

1.4.9 Phylogenetic analysis and subgroups

Phylogenetic analysis is a fundamental method to determine the evolutionary relationships of organisms (Brown 2002). The L, G, and N genes have been used to construct phylogenetic trees of LNYV and other rhabdoviruses. Phylogenies of L gene sequences showed that cytorhabdoviruses (including LNYV) and nucleorhabdoviruses were separated into two clades with a monophyletic origin (Figure 1.10) (Ito et al. 2013; Li et al. 2015; Yang et al. 2016). Even though phylogenetic analyses support the cytopathological findings of two distinguishing groups, the genes or sequences involved in the replication differences is unknown.

The L gene phylogenetic analysis showed that LNYV is closely related to LYMoV, followed by PeVA (Ito et al. 2013; Yang et al. 2016). This was also observed in the G gene and N gene phylogenetic analysis of cytorhabdoviruses (Yang et al. 2016). The most closely related cytorhabdovirus of LNYV could change as more cytorhabdoviruses are identified and included in the analyses.
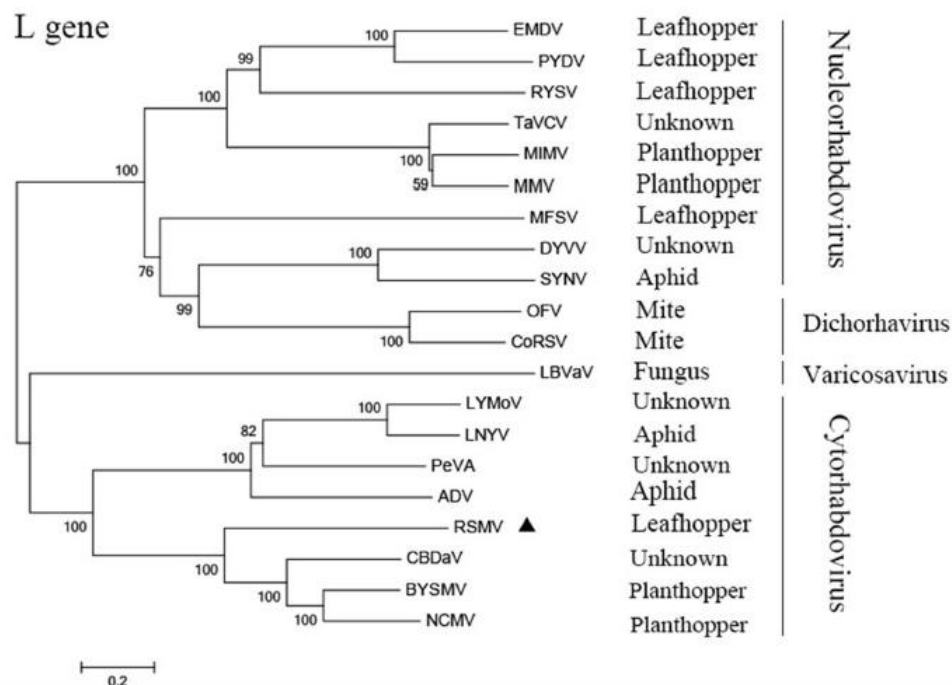
Figure 1.10: L gene phylogenetic tree of cytorhabdoviruses, nucleorhabdoviruese, dichorhaviruses and varicosavirus using neighbour-joining method. ▲ = Rice stripe mosaic virus (Yang et al. 2016).

There has been little study of the variability within cytorhabdovirus species, only LNYV and SCV have been studied. The phylogenetic analysis of N gene sequences from LNYV isolates in Australia revealed that LNYV was further separated into two subgroups: subgroup I and subgroup II, which had two and seven isolates, respectively (Callaghan and Dietzgen 2005) (Figure 1.11). This was also observed for SCV (Klerks et al. 2004). The significance of LNYV subgroups in the infected plant hosts and insect vectors is unknown. Higgins et al (2016) suggested that subgroup II may be outcompeting subgroup I, since it appears to have dispersed more rapidly. It is critical to evaluate the relationship between the subgroups to determine if such competition is actually occurring as this will impact the development of a LNYV control mechanism.
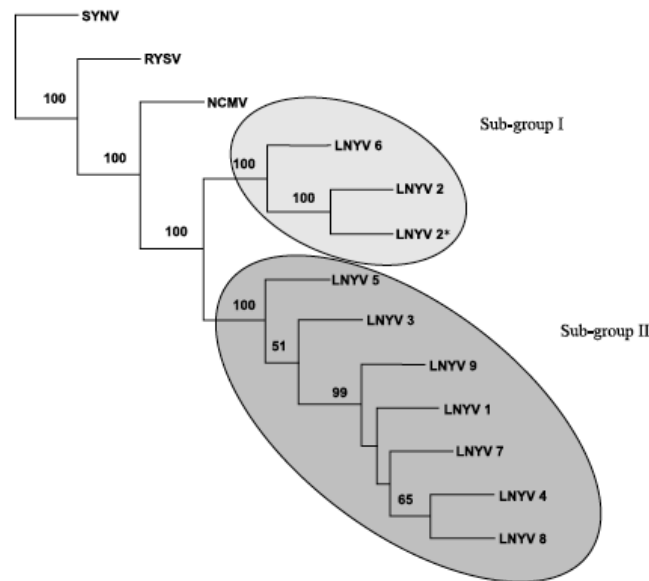
Figure 1.11: Phylogenetic tree of Australian LNYV isolates N gene nucleotide sequences using maximum parsimony method (Callaghan & Dietzgen, 2005).

Initial LNYV research in New Zealand was executed in 1965 (Fry et al. 1973). Recently, the first detailed genetic analysis was conducted by Higgins et al. (2016b) in New Zealand with six LNYV infected lettuce samples detected from Auckland. The N genes of those isolates were sequenced and phylogenetic analysis was performed. The evolutionary tree showed that five isolates were in subgroup I and one isolate was in subgroup II (Figure 1.12) (Higgins et al. 2016b). The molecular clock analysis indicated that LNYV appeared in Australia and New Zealand around 500 years ago with subgroup I and subgroup II emerging around 150 and 75 years ago, respectively (Higgins et al. 2016b). Subgroup I LNYV isolates have not been detected in Australia since 1993 and is now considered to be extinct there (Callaghan and Dietzgen 2005). A stated above, it has been proposed that subgroup II may have out competed subgroup I in Australia, which could have led to the extinction of subgroup I isolates and this has not yet happened in New Zealand (Higgins et al. 2016b). No follow up study has been conducted to determine if subgroup II does indeed out compete subgroup I, but if it does, it is critical to identify the genes or sequences involved that could increase the transmission efficiency.
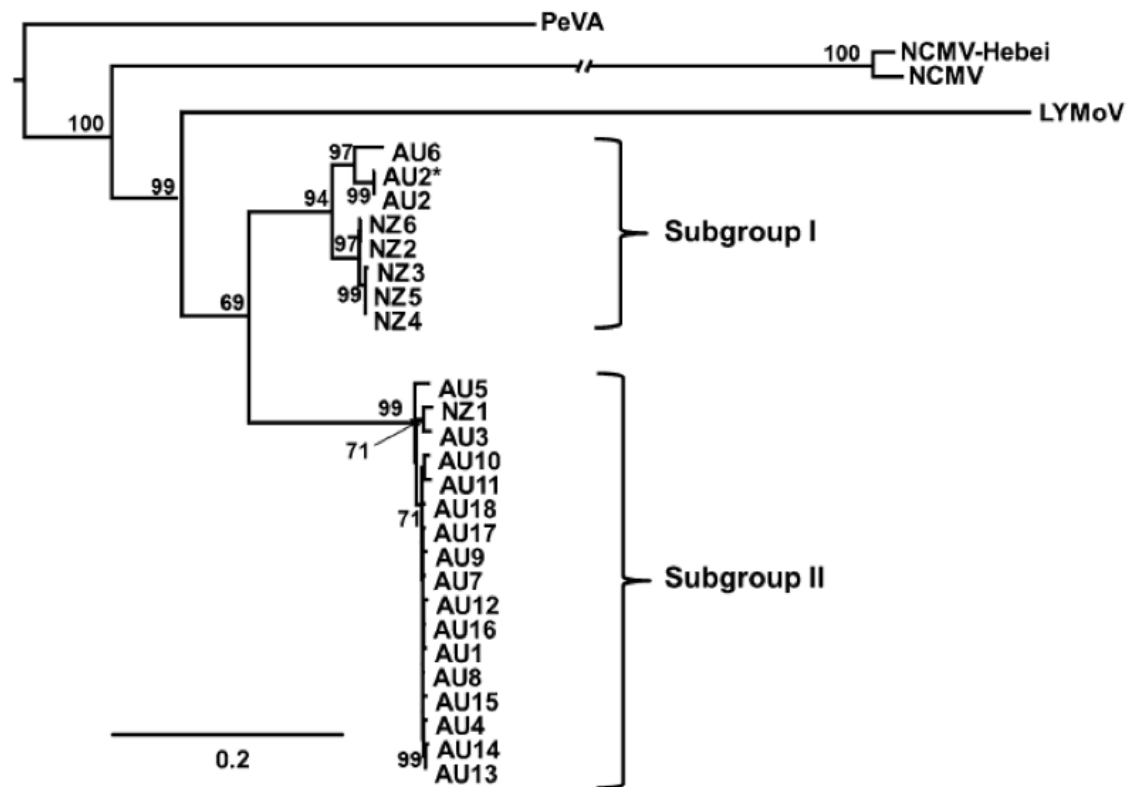
Figure 1.12: Phylogenetic tree of Australian and New Zealand LNYV isolates N gene nucleotide sequences using maximum likelihood (Higgins et al., 2016).

1.4.10   Diagnosis of LNYV and subgroups

1.4.10.1   Diagnosis of LNYV by conventional methods

Symptom-based diagnosis

Diagnosis of LNYV has progressed significantly over the past few decades. LNYV was initially detected by mechanical transmission or inoculation of infected leaves to the susceptible indicator host *N. glutinosa*. Development of symptoms on *N. glutinosa* was used to diagnose the virus (Chu and Francki 1982; Dietzgen et al. 2007). Exclusive use of symptom-based diagnosis can increase the likelihood of misdiagnosis, for example, 'virus-like' symptoms could be caused by other biotic or abiotic factors. Furthermore, different viruses may cause similar symptoms (Naidu and Hughes 2003); LNYV is commonly misdiagnosed as TSWV due to the similarity of symptoms (Dietzgen et al. 2007). Therefore, highly experienced individuals are required for accurately diagnose LNYV and follow up tests are mandatory for confirmation (Dietzgen et al. 2007; Naidu

and Hughes 2003; Narayanasamy 2011b; Sastry 2013a). Thus, symptom-based diagnosis of LNYV is not accurate, reliable or rapid and another method is required.

Electron microscopy

Electron microscopy (EM) can be used to observe plant viral morphology and is used to detect viruses (Naidu and Hughes 2003). EM utilises high-speed electrons to illuminate an image from a specimen at a shorter wavelength and higher magnification than light microscope for high resolution (Bozzola and Russell 1999; Egerton 2005). Purified LNYV particles were first visualised under a EM (Harrison and Crowley 1965). EM was used to identify unique physical characteristics of LNYV for identification and is capable of distinguishing LNYV from TSWV (Chu and Francki 1982). The physical characteristics that can be identified include viral architecture, size, particle characterisation, morphogenesis, replication and cytological modification of a specimen (Baker et al. 1985; Stussi-Garaud et al. 1994). However, it is expensive to purchase and maintain an electron microscope, and requires expert skills to operate. Furthermore, highly concentrated virus samples are required for the analysis (Fields et al. 2007). It is unknown if LNYV subgroup I and subgroup II can be distinguished by TEM but it is highly unlikely that the subgroups have significant differences in morphology to distinguish them. Therefore, it would be appropriate to use other methods such as serological or molecular tests to detect the LNYV subgroups.

Gel double-diffusion and immunodiffusion tests

Gel double-diffusion test, also known as the Ouchterlony test, is the use of an antibody to detect a specific virus using an agar gel (Dijkstra and Jager 1998). An infected sample and diagnosing antibody are placed into separate wells in an agar plate and then allowed to diffuse towards each other. If the virus contains the antigen specific to the antibody, a precipitate can be observed in the agar gel where they meet (Dijkstra and Jager 1998). Such a test was developed to detect LNYV in the study conducted by Harrison and Crowley (1965). Two types of antisera (antiserum G and antiserum L) were obtained by injecting partially purified LNYV into rabbits. A faint precipitate developed from LNYV infected lettuce and *N. glutinosa* samples using both antisera. However, results also showed that the precipitation occurred from healthy plants. Hence, it was considered to be an inefficient method for diagnosis of LNYV (Harrison and Crowley 1965). Follow-up research was carried out by McLean et al. (1971) to develop an antiserum with a high

degree of specificity to LNYV for use in immunodiffusion tests. This study also demonstrated that immunodiffusion tests are inefficient for detection of LNYV because the virus particles were non-diffusive through the agar and highly concentrated samples were required (McLean et al. 1971).

Enzyme-linked immunosorbent assay

Enzyme-linked immunosorbent assay (ELISA) is commonly used for plant virus detection (Hull 2014). Chu and Francki (1982) developed a double antibody sandwich enzyme-linked immunosorbent assay (DAS-ELISA) to detect LNYV in infected *N. glutinosa* and *H. lactucae* using an enzyme conjugated anti-LNYV γ-globulin antibody as the secondary antibody and a non-enzyme conjugated antibody as the capture antibody. The results showed that *N. glutinosa* samples infected with LNYV were detectable by ELISA, which was 600x more sensitive than the mechanical transmission method. *H. lactucae* collected from the field that had LNYV infected plants also tested positive by ELISA. ELISA is faster (within 24hr) to obtain results than symptom-based diagnosis and, with the right antibody, highly specific to LNYV (Chu and Francki 1982). One of the limitations of ELISA is that it can be difficult to determine if the sample is positive or negative by ELISA; cross-reaction by more than one pathogen in a sample can occur and the total number of infective viruses cannot measured (Fox et al. 2006; Hull 2014). ELISA was used recently to identify eight LNYV infected lettuce samples from Canterbury in a report by Fletcher et al. (2017) and these samples were sent to AUT to identify the subgroups in the current study. ELISA is capable of identifying virus species, but viral isolates or strains are often indistinguishable as isolate/strain-specific antibodies may not be produced (Boonham et al. 2014). For LNYV, ELISA can distinguish it from TSWV, which symptom studies often can't, but LNYV subgroups may not be discernible.

1.4.10.2 Molecular methods to detect LNYV

Molecular technologies allow detection of viral isolates or strains more accurately and efficiently than serological detection methods without the requirement to produce an antibody (Narayanasamy 2008). The invention of the polymerase chain reaction (PCR) was a breakthrough in molecular biology; the concept is simple but crucial for detecting plant viruses and other organisms. In each PCR cycle, DNA fragments are rapidly amplified in number by 2^n where n is the number of cycles (Hull 2014). PCR consists of

three primary steps; denaturation, annealing and extension (Brunstein 2013). The required components are template DNA, primers (forward and reverse), deoxynucleotide triphosphates (dNTPs), *Taq* DNA polymerase, buffer, magnesium as a cation and deionised water (Pelt-Verkuil et al. 2008). The PCR analysis commences with an increase of temperature to denature the dsDNA template and the primers to ssDNA. This is followed by a rapid decrease in temperature that allows the annealing of the primers to the specific DNA sequence – the target region to be amplified. The optimal annealing temperature is generally determined experimentally. The temperature is increased to allow *Taq* DNA polymerase to attach nucleotides to the primers in the order defined by the template strand (extension) to form a nascent DNA strand (Brunstein 2013). This process is repeated for a number of cycles, often 35, to obtain the final PCR product (Figure 1.13). The PCR products can be analysed using gel electrophoresis and viruses can be detected based on the expected band on the gel (Hull 2014). LNYV genes cannot be directly amplified by using PCR for detection because LNYV has an RNA genome. It requires the use of reverse transcription PCR (RT-PCR), which will be discussed in section 1.4.10.3.
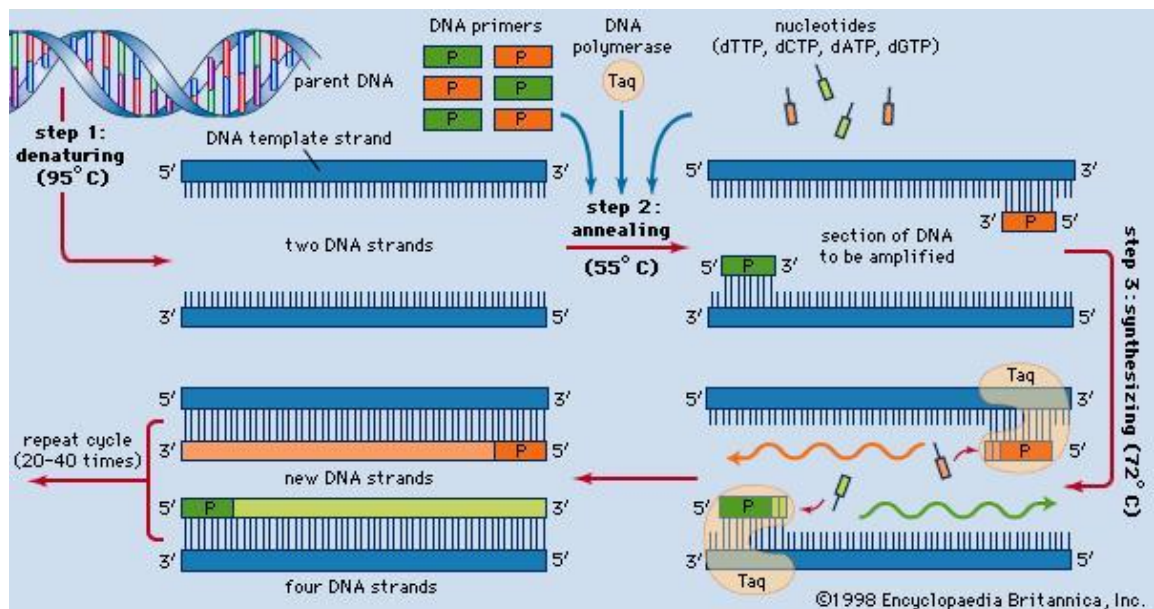
Figure 1.13: Polymerase chain reaction (PCR) process. The three main steps are: denaturation of dsDNA to ssDNA, annealing of primers to the template strand and synthesising (extension) of nascent DNA strand. The process is repeated 20-40 times.

Image from https://www.britannica.com/science/polymerase-chain-reaction

Real-time polymerase chain reaction (q-PCR), also called quantitative PCR, measures the amount of the nucleic acid from the sample in real time (Pelt-Verkuil et al. 2008). It uses fluorescent probes or fluorescent DNA-binding dyes that fluoresce during the amplification (Thermo Fisher Scientific 2016); fluorescence is detected and measured at the end of each cycle by the q-PCR machine. There is a directly proportional relationship between the fluorescent signal and the amount of PCR product (Thermo Fisher Scientific 2016). The amount of target template at the start of the reaction can be quantified by relative or absolute quantification. Absolute quantification requires a standard curve, obtained from known quantities of a series of standards. The accuracy of quantification is contingent on the standard curve. Relative quantification measures the change in nucleic acid amount compared to a reference sequence that is maintained in amount (Pfaffl 2006). q-PCR can be combined with RT-PCR to detect and quantify a target RNA sequence such as the genome of LNYV and LNYV subgroups. This aspect will be discussed in section 1.4.10.3.

There are many adaptions of PCR such as: multiplex PCR, nested PCR and RT-PCR. Multiplex PCR utilises multiple primers to amplify multiple target sequences simultaneously in a single reaction using a standard PCR (Hernandez-Rodriguez and Ramirez 2012; Markoulatos et al. 2002) (Figure 1.14). Nested PCR uses two consecutive runs of amplification with two pairs of primers one internal to the other. This decreases the likelihood of amplifying non-specific products, increasing the sensitivity and specificity (Haff 1994; Lorenz 2011; Yourno 1992). The initial amplification uses external primers to amplify a target region. Subsequently, the second pair of primers amplifies an internal region of the target sequence (Figure 1.15). This helps to verify the target sequences are the only sequences that were amplified (Boyanton and Rushton 2010; Hernandez-Rodriguez and Ramirez 2012; Pelt-Verkuil et al. 2008).
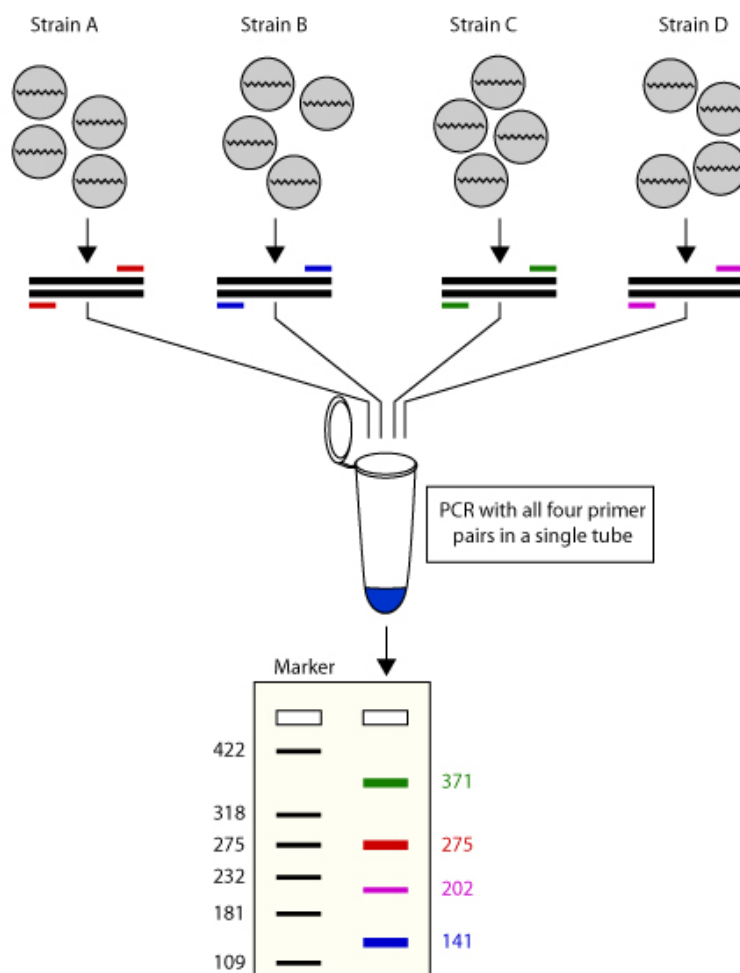
Figure 1.14: The principle of multiplex PCR to detect four different strains/sequences. Samples of four strains and four sets of primer are added to a single PCR. Four strains are identified by gel electrophoresis based on different sized bands.
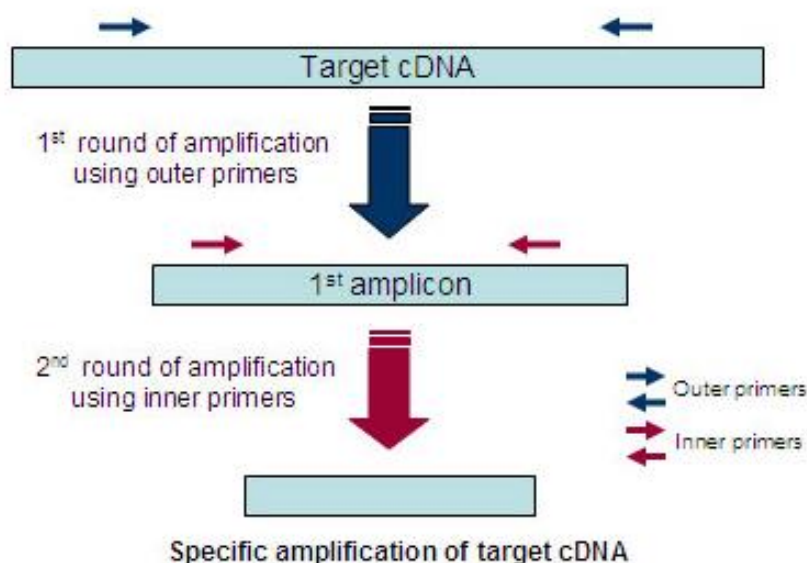
Figure 1.15: Principle of a nested PCR to amplify a specific target cDNA by using outer and inner primers.

Image from http://www.genofi.com/RefSpliceFAQ.asp

RT-PCR amplifies cDNA sequences from an RNA template strand. Components in RT-PCR are similar to a standard PCR but firstly requires reverse transcriptase. RT-PCR is initiated by the copying of the template RNA into a cDNA using the reverse transcriptase enzyme and a sequence specific primer. cDNA is then amplified as described above in a standard PCR (Yadav and Khurana 2016) (Figure 1.16). There are two methods of RT-PCR; one-step RT-PCR and two-step RT-PCR. In one-step RT-PCR, both cDNA synthesis and PCR amplification are performed in the same reaction tube, which is faster than two-step RT-PC and reduces contamination. In two-step RT-PCR, both of these processes are carried out in separate reactions (Cattoli and Monne 2009). LNYV has been detected using one-step RT-PCR (Callaghan and Dietzgen 2005; Higgins et al. 2016b). It was concluded by Higgins et al. (2016b) that the differences within N gene sequences could be used develop a subgroup-specific diagnostic RT-PCR. RT-PCR can be combined with qPCR for quantitative analysis.
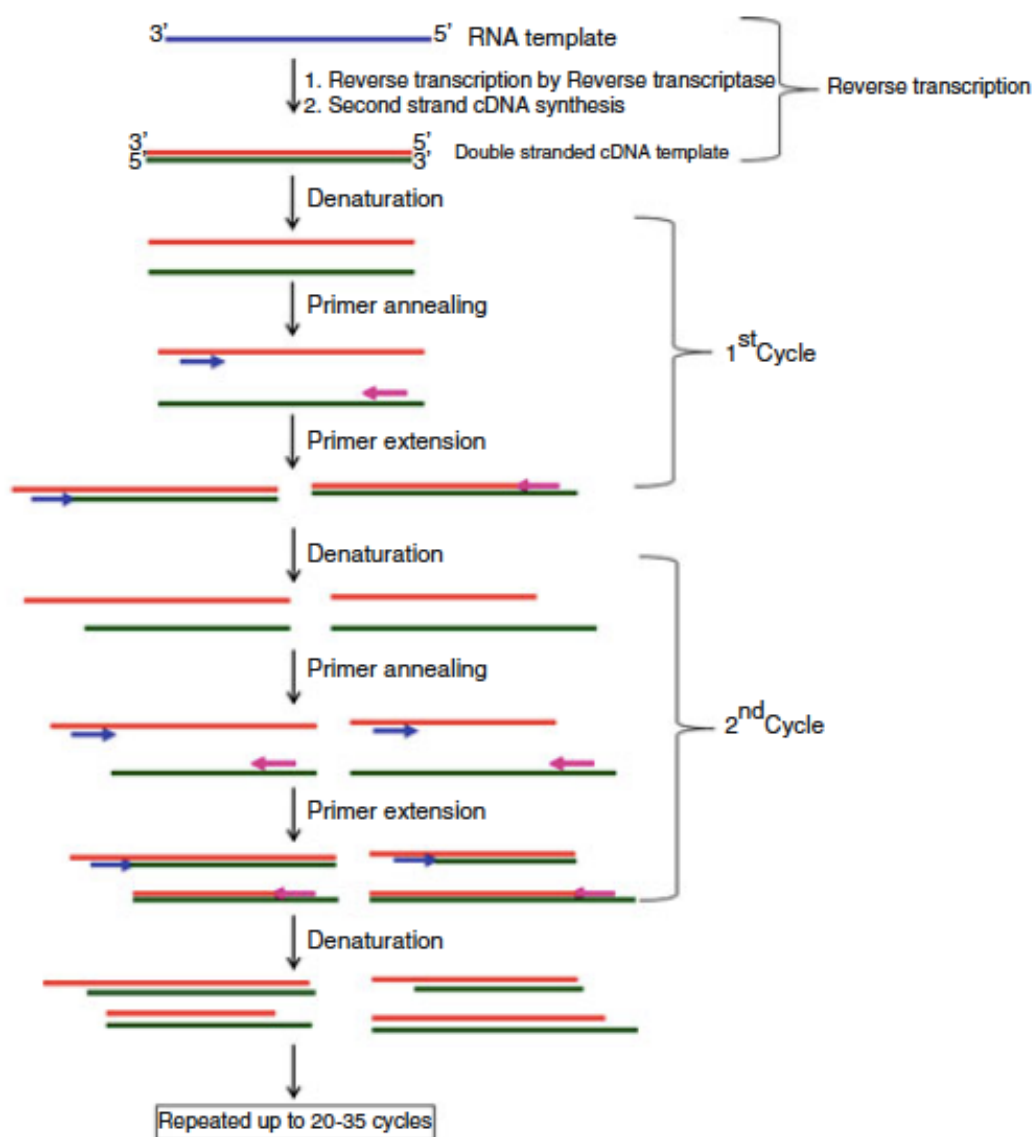
Figure 1.16: Reverse transcription polymerase chain reaction procedure (modified) (Yadav and Khurana 2016).

1.4.10.3 Diagnosis of LNYV using RT-PCR

RT-PCR analysis is one method used to detect LNYV (Callaghan and Dietzgen 2005; Higgins et al. 2016b). Different primer pairs targeting the LNYV N gene have been designed for this purpose. The LN1 and LN2 primers amplify 140 bp fragment from the N gene (Thomson and Dietzgen 1995). The BCNG1 and BCNG2 primers were designed to amplify the complete N gene (1,500 bp) of LNYV (Callaghan and Dietzgen 2005) while the LNYV_440F and LNYV_1185R primers amplify a 750 bp region in the N gene (Higgins et al. 2016b). The BCN3 and BCN4 primers amplify a 748 bp region in the N

gene (Higgins et al. 2016b). These primers are currently used to detect LNYV and were used in the current study for diagnosis.

1.4.11 Purification of RNA by crude and total RNA extraction

High RNA quality is necessary for RT-PCR analysis to detect plant viruses accurately because polyphenolic and polysaccharides from the leaves could interfere with the analysis (Yang et al. 2017). Total RNA was purified from samples using commercial RNA extraction kits and used in RT-PCR analyses to test for LNYV (Callaghan and Dietzgen 2005; Higgins et al. 2016b). The plant RNA extraction kits remove the plant components that could inhibit the RT-PCR and toxic solvents are not used in the procedure (Yang et al. 2017). RNA extraction kits can be expensive than conventional extraction methods (Yaffe et al. 2012). Thomson and Dietzgen (1995) developed a crude and rapid method for preparing crude template for detecting various plant viruses, including LNYV. The authors reported detecting LNYV by RT-PCR using the LN1 and LN2 primers. Since this procedure is rapid and cheaper, it was tested in the current study for LNYV diagnosis.

1.4.12 Sequencing plant virus genomes

1.4.12.1 Sanger sequencing

Plant virus genomes can be sequenced using several sequencing methods. The first complete plant virus genome was that of cauliflower mosaic virus (CaMV), a DNA virus that was sequenced using shotgun Sanger sequencing (Gardner et al. 1981; Houldcroft et al. 2017). Sanger sequencing, also referred as the 'chain termination method', was invented by Frederick Sanger who won the 1980 Nobel prize for this method, which is still the gold standard method for sequencing (Hartl and Ruvolo 2012; Roe 2014). To perform Sanger sequencing (sequencing by synthesis), first DNA is isolated from a sample of interest and a target region of a sequence is amplified or cloned for sequencing. A sequencing primer, deoxynucleotide (dNTPs), DNA polymerase and chain terminating dideoxynucleotides (ddNTPs) are added (Figure 1.17). The ddNTPs have a hydrogen at

the 3' carbon position, which cause termination of DNA synthesis. Each ddNTP is with a different fluorescent dye. and. A standard PCR is performed as mentioned in section 1.4.10.2, except that only one primer is included in the reaction. Thus, only synthesis of one strand of the target is synthesised. Once DNA polymerase incorporates one of the ddNTPs no new nucleotides can be added to the sequence (Figure 1.17). This process occurs every cycle and by the end of the last cycle, ddNTPs would have been added to all positions in the target sequence, creating a collection of synthesised molecules of different lengths, each terminating in a ddNTP labelled a specific colour. The fragments are then passed through capillary gel electrophoresis. As the fragments migrate through the gel, a laser is used illuminate the fragments as they pass causing them to fluoresce. The specific colour fluorescence is detected and stored on a computer. In this way a sequence is built up from the 5' to the 3' end, which is complementary to the template (Figure 1.17) (Heather and Chain 2016; Pelt-Verkuil et al. 2008). The genome of viruses can be obtained by using Sanger sequencing (Deng et al. 2015). However, for genome sequencing, when compared to other next-generation sequencing technologies, Sanger sequencing produces low amounts of data (250 kb). Furthermore, templates may require cloning, as well as primers to close any gaps where there is no sequence. It is also relatively time consuming, expensive and laborious (Ari and Arikan 2016; Franca et al. 2002; Husemann and Stoye 2009). These limitations increased the demand to develop more highly efficient sequencing technologies.

PCR in presence of fluorescent, chain-terminating nucleotides

Fragments run through gel electrophoresis

Laser beam

Photomultiplier

Fluorescent fragments detected by laser and represented on a chromatogram
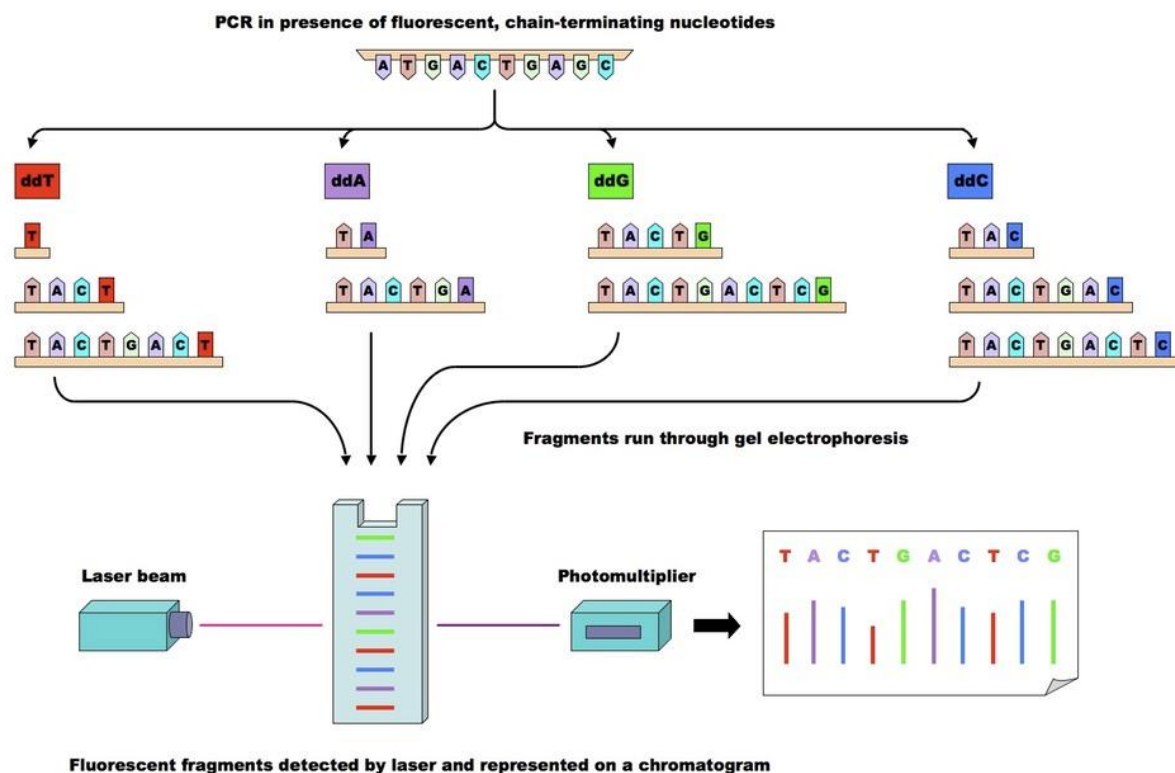
Figure 1.17: Sanger sequencing procedure overview.
Image: http://www.vce.bioninja.com.au/_Media/sequencing_med.jpeg

1.4.12.2   Next generation sequencing methods

Massively parallel sequencing technologies (also known as next-generation sequencing (NGS) have been developed to ameliorate the limitations of the previous sequencing technologies (Ari and Arikan 2016). Massively parallel sequencing means many templates can be sequenced simultaneously, generating large amounts of data. Furthermore, the cost per base is much reduced when compared with the first generation technologies such as Sanger sequencing (Liu et al. 2012). The uses of NGS in viral research include polymorphism analysis, transcriptome analysis, virus detection, identification of novel viruses and genome sequencing (Khalifa et al. 2016). The NGS technologies include pyrosequencing, ion torrent, Illumina and SOLiD (Valencia et al. 2013).

Pyrosequencing

Pyrosequencing is a sequencing by synthesis and bioluminescent technology that utilises detection of pyrophosphate to determine the sequence (Almeida et al. 2014). The process is initiated by the construction of a DNA library where each DNA is fragmented and ligated with adaptors at each end of the sequence. The library is then amplified using emulsion PCR where the fragments are hybridised to beads that have primers attached that are complementary to the adaptors on the fragments. Each bead has one DNA molecule attached, and the beads, in water, are then placed into oil to create an emulsion. PCR amplification is then carried out whereby each bead becomes saturated with one DNA sequence. The beads are then transferred to a flow cell so that each bead in placed into individual wells. Then the sequencing can begin (Ari and Arikan 2016).

During DNA synthesis, when a nucleotide is incorporated by DNA polymerase, a pyrophosphate (PPi) is released. This fact is taken advantage of; with the addition of adenosine 5´-phosphosulfate (APS) the enzyme ATP sulfurylase converts it to adenosine triphosphate (ATP) (Figure 1.18). The enzyme luciferase is also present, and uses ATP to luciferin to oxyluciferin, resulting in the production of light. The light is detected by a charge coupled device (CCD) and the peaks obtained from the light intensity can be used to determine the order of the sequence (Figure 1.18) (Ari and Arikan 2016; Méndez-García et al. 2018; Metzker 2010). Each nucleotide is washed sequentially over the flow cell, together with the enzymes and substrates, and the incorporation of a nucleotide is recorded as light is emitted. While pyrosequencing generates long reads (1 kb) and is faster than Sanger sequencing, it is no longer be available as Roche, the company that purchased and produced the machine, has shut down the technology (Méndez-García et al. 2018). Plant virus genome sequences from pyrosequencing data are still being published, for example, two new viruses genomes obtained from vanilla have been published using the pyrosequencing data (Grisoni et al. 2017).

The Principle of Pyrosequencing™ Technology

$(DNA)_n + dNTP \xrightarrow{\text{Polymerase}} (DNA)_{n+1} + PPi$

Sulfurylase

APS+PPi → ATP

luciferin → oxyluciferin

Luciferase

ATP → Light

light

time

nucleotide incorporation generates light seen as a peak in the Pyrogram

dNTP + Apyrase → dNDP + dNMP + phosphate

ATP + Apyrase → ADP + AMP + phosphate

nucleotide sequence

G C – A GG CC T

G C T A G C T

nucleotide added

Modified and published with permission from Biotage AB.
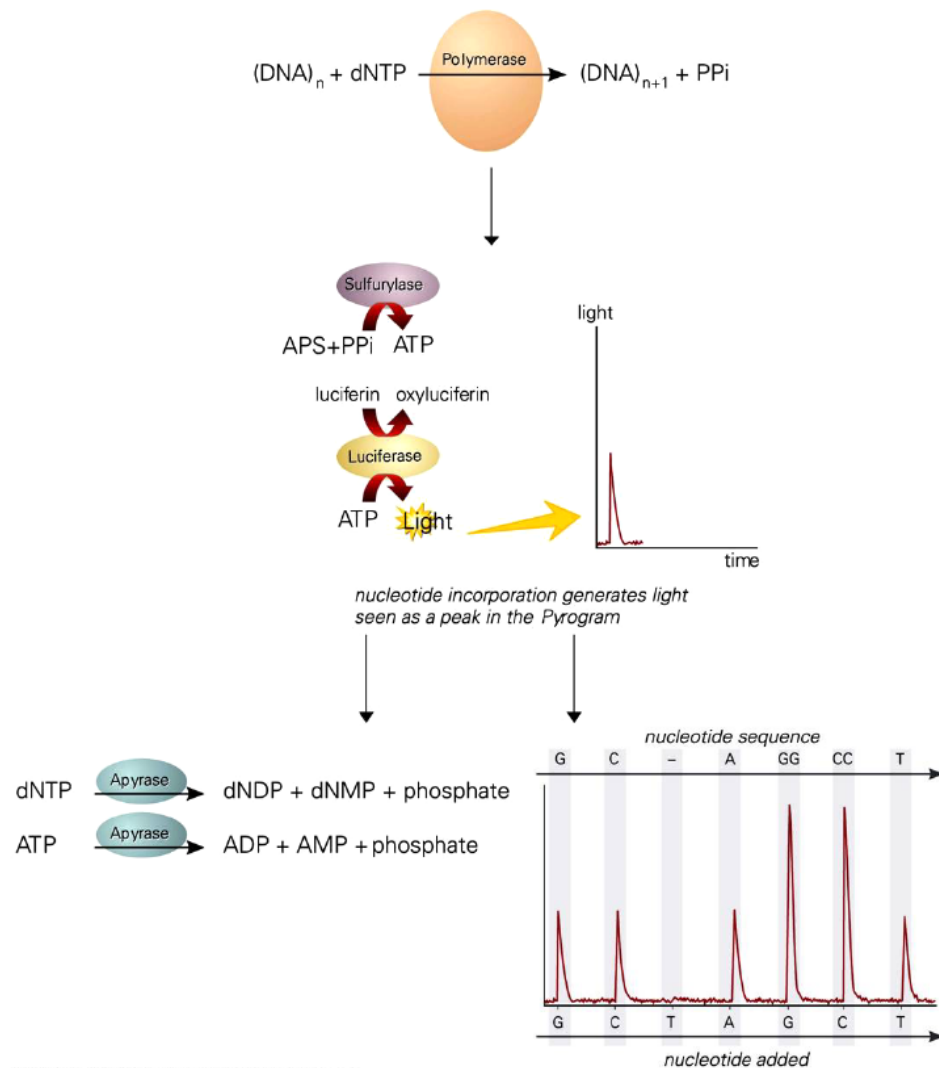
Figure 1.18: Pyrosequencing overview.
Image: https://www.qiagen.com/mx/resources/technologies/pyrosequencing-resource-center/technology-overview/?akamai-feo=off

Illumina sequencing

Illumina sequencing was developed by Shanker Balasubramanian and David Klenerman in 1998 (Balasubramanian 2015). Like pyrosequencing, the process of Illumina sequencing is initiated by the construction of libraries. The genomic DNA is fragmented and ligated with adaptors at each end of the sequence (Figure 1.19 a and b). In contrast to pyrosequencing, these fragments are added to the Illumina flow cell (Figure 1.19 c), which contains up to eight lanes, immobilised with oligonucleotides that are complementary to the adaptor sequences. The adapter-ligated DNA fragments anneal to

the complementary oligonucleotides and the adapter in the opposite end of DNA fragment anneals the complementary oligonucleotide by bending (Figure 1.19 d). The oligonucleotides perform like a primer and DNA polymerase synthesises the complementary sequence of the target DNA by inserting the dNTPs to form a dsDNA which is known as bridge amplification (Figure 1.19 d). The dsDNA is denatured to form a ssDNA and millions of sequences are amplified in the flow cell (Figure 1.19 d). Sequencing by synthesis is performed; all four nucleotides labelled with different coloured fluorophores are added. The incorporation of a nucleotide is recorded as fluorescence of a specific colour, and a sequence is then generated based on the order of colour emitted (Figure 1.19 e and f) (Illumina 2010; Metzker 2010; Raiol et al. 2014; Voelkerding et al. 2009). Khalifa et al. (2016) had compared the Illumina and Sanger sequencing by obtaining RNA virus sequences from five *Sclerotinia sclerotiorum* isolates. The results showed nine viral genomes were obtained from Sanger sequencing while the same genomes and additional genome were detected by Illumina sequencing. Illumina sequencing gave more than 99.3% identity with sequences obtained by Sanger sequencing. This result demonstrated that there is high accuracy in the Illumina sequencing, which can generate up to 1000 Gb data per run with high coverage, even though it produces short sequence reads (125 bp) (Ari and Arikan 2016). It is currently the predominantly used sequencing method for plant viral genomes when compared to other NGS technologies (Blawid et al. 2017).
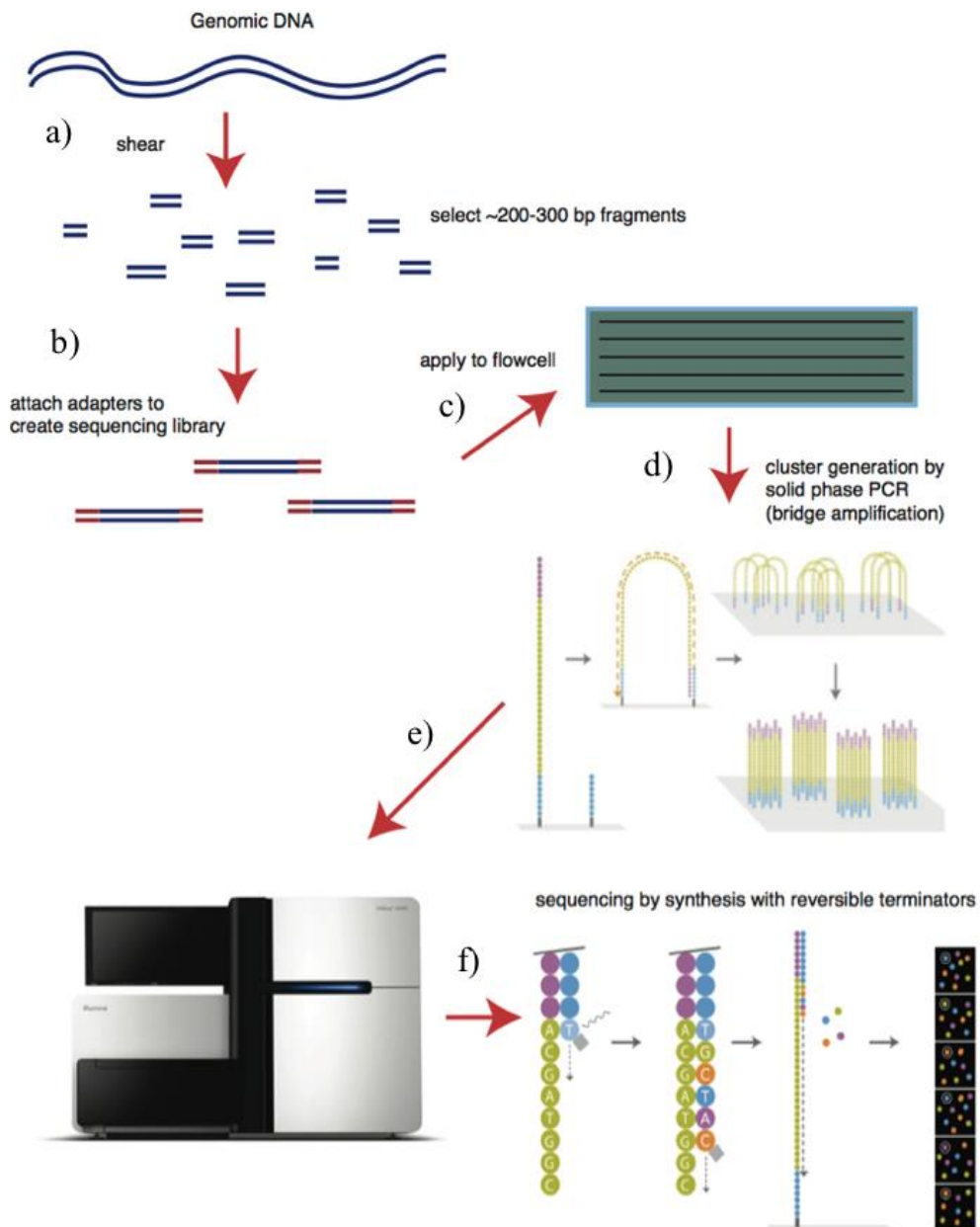
Figure 1.19: Illumina sequencing procedure overview. a) The genomic DNA is fragmented. b) Adapters are ligated to create the library. c) The adapter-ligated DNA fragments are applied to the flow cell. d) DNA fragments anneal to the oligonucleotide on the surface of the flow cell and bridge amplification. Millions of amplifications occur. e) Amplified sequenced are placed into a sequencing machine. f) Genome sequenced by sequencing by synthesis.

Image: https://bitesizebio.com/13546/sequencing-by-synthesis-explaining-the-illumina-sequencing-technology/ (modified)

SOLiD sequencing

Sequencing by oligonucleotide ligation detection (SOLiD) is a sequencing by ligation method which uses two-base-encoded probes that anneal to the template (Méndez-García et al. 2018; Metzker 2010). SOLiD had been used to identify sequences of new plant virus. For example, Sela et al. (2013) had identified a new virus that was infected in

watermelon using SOLiD sequencing and was named citrullus lanatus cryptic virus. When compared to other NGS sequencing methods, SOLiD produces a short read length of 50 bp, longer run times of up to 14 days and high sequencing cost (Méndez-García et al. 2018; Metzker 2010). As a result, this method has not become widely used.

Ion torrent

Ion torrent sequencing technology was officially released by Life Technologies during 2010 (Ari and Arikan 2016). This technology uses emulsion PCR and sequencing by synthesis in the same manner as pyrosequencing. In addition to the release of a Pi, a hydrogen ion is released when a nucleotide is incorporated. Ion torrent takes advantage of this. The order of the nucleotides incorporated is determined as the order of nucleotides applied to the flow cell is known. The release of a hydrogen ion causes a change in pH, which is detected by a complementary metal-oxide-semiconductor (CMOS) in the base of the flow cell well (Heather and Chain 2016; Méndez-García et al. 2018). To date there have been no plant virus genomes reported using ion torrent sequencing. Similar to pyrosequencing, this sequencing method is highly prone to errors in homopolymeric regions of the sequence. Further, it is currently more expensive than Illumina sequencing (Dervan and Shendure 2017; Kavak et al. 2016). Further developments and improvements in ion torrent sequencing should decrease the cost, enabling more plant virus genomes to be sequenced using this technology.

Pacific Biosciences single molecular real-time (SMRT) sequencing

Pacific Biosciences (PacBio) RS is the first third-generation sequencing released in 2011 (Bayés et al. 2012). PacBio sequencing does not require PCR amplification and is real-time sequencing using single molecular real-time (SMRT) sequencing technology (Rhoads and Au 2015). A single DNA polymerase molecule is bound to one molecule of DNA and immobilised on the bottom of a well with a zero mode waveguide detectors (ZMW detectors) that allow the light through the bottom of the well only. Each nucleotide is labelled with a different fluorophore at the alpha position rather than the gamma position as used for other technologies. When a nucleotide is incorporated, the fluorophore is released and fluoresces. The detector detects the light and sequences are determined (Heather and Chain 2016; Méndez-García et al. 2018; Nakano et al. 2017). SMRT technology can generate long read lengths of more than 60 kb, high accuracy of more than 99.999% and produce low bias since there is no amplification or cloning of the

templates (Nakano et al. 2017). This is contradicted by Dervan and Shendure (2017), who state that PacBio sequencing generates high errors of up to 20% per base. It is also more expensive when compared other sequencing technologies; however, continual development of this technology will likely decrease the cost in future (Dervan and Shendure 2017). Currently, there are no reports of plant virus genome sequence obtained from this technology, which could likely be due to the high error rate and expense.

Nanopore sequencing

Nanopore sequencing is one of the fourth-generation sequencing techniques and first released by Oxford Nanopore (Dervan and Shendure 2017; Feng et al. 2015). The sequencing is initiated when a DNA or RNA strand is translocated across an ion channel (alpha-hemolysin) using electrophoresis, resulting in a decrease of current across the channel (Feng et al. 2015; Heather and Chain 2016). The sequence can be determined by sensors which measure the physical changes that are specific to each nucleotide type (Ari and Arikan 2016; Feng et al. 2015). No plant virus genome has been reported using this technology yet. While there is high potential for this sequencing technology with regards to nucleic acid or protein analysis and application to various fields (Ari and Arikan 2016). The technology produces higher errors than PacBio (Dervan and Shendure 2017). This further supports the current preference to sequence plant virus genomes by Illumina sequencing over any other NGS technologies.

1.4.13   LNYV genome and gene sequence

The first and only complete published genome of LNYV is from an Australian subgroup I isolate that was sequenced by Sanger sequencing. The isolate was obtained from an infected garlic plant that was mechanically inoculated onto *N. glutinosa* (Dietzgen et al. 2006). The genome was 12,807 bp in length. No genomes from a subgroup II or New Zealand LNYV isolate have been sequenced yet. Comparison of a subgroup I with a subgroup II genome would help with determining any differences in transmission, replication and its survival. This may help understanding of why subgroup II appears to have dispersed more rapidly than subgroup I (Higgins et al, 2016). Comparing a New Zealand LNYV genome with that of an Australian isolate could also determine if there is a difference within the isolates or subgroups between the countries. This may help understand why subgroup I has become extinct in Australia but not in New Zealand.

There is a total of 25 LNYV N gene sequences available in the NCBI database. Among these sequences, three subgroup I sequences and 16 subgroup II sequences were obtained from Australian LNYV isolates (Callaghan and Dietzgen 2005; Higgins et al. 2016b). Five subgroup I sequences and one subgroup II sequence were obtained from New Zealand LNYV isolates (Higgins et al. 2016b). While the Australian sequences were from different locations collected at different times, the New Zealand isolates came from one location collected on one day. Thus, more New Zealand sequences need to be studied to understand the New Zealand LNYV population more fully and determine how wide spread each subgroup is.

1.5 Aims

The structure and aims of this thesis is as follows.

Chapter 2: Development of diagnostic tests for LNYV subgroups
LNYV related research had been carried out for the last 53 years, yet there is still much to learn. LNYV can be detected using RT-PCR analysis (Callaghan and Dietzgen 2005; Higgins et al. 2016b); however, there is no rapid method to diagnose the LNYV subgroup causing disease. The nucleotide differences in the available N gene sequences could be used to develop subgroup-specific diagnosis (Higgins et al. 2016b). Regarding the LNYV subgroup distribution in New Zealand, more samples are required to be collected and tested for LNYV subgroups.

Aims:
  ➤ Develop diagnostic methods to detect LNYV subgroups by RT-PCR and RT-PCR-RFLP
      o Design subgroup-specific primers and test on known LNYV infected plants
  ➤ Test BCNG1/2 and LNYV_440F/LNYV_1185R primers to detect LNYV
  ➤ Test symptomatic samples from Harrisville, Tuakau and Canterbury using BCNG1/2, LNYV_440F/LNYV_1185R and subgroup-specific primers to detect LNYV and the subgroups

> Test the TPS procedure for rapid template preparation described by Thomson and Dietzgen (1995) with RT-PCR on known LNYV infected samples

Chapter 3: LNYV genome sequence and glycoprotein sequence analysis

LNYV subgroup I seem to be extinct in Australia and Higgins et al (2016) suggested that it has been out-competed by subgroup II, while this has not occurred in New Zealand. BEAST analysis indicated that subgroup II has emerged more recently (75 years ago) than subgroup I (150 years ago). It was hypothesised by (Higgins et al. 2016b) that subgroup II was able to spread more rapidly than subgroup I because subgroup II has a more optimal association with the insect vectors and/or plant hosts than subgroup I. The G protein likely has an important role in LNYV transmission by the aphid. Therefore, a fundamental question arises as to whether or not subgroup II transmits more efficiently than subgroup I due to the specific differences between the LNYV subgroup G gene/protein sequences. These differences could have resulted in the extinction of subgroup I in Australia and may yet to occur in New Zealand. Comparison of the G sequences between subgroups and between Australian and New Zealand isolates requires a genome from a subgroup II isolate, as well as a New Zealand subgroup I isolate.

Aims:

> Sequence the complete genomes of subgroup I and subgroup II New Zealand isolates using Illumina sequencing.

> Analyse the G gene and protein sequences obtained from the genomes of New Zealand isolates and compare with the published LNYV subgroup I G gene.

o Identify differences within the sequences and 2D structures between the subgroups.

Chapter 2

Development of Diagnostic tests for LNYV subgroups

Chapter 2 Development of diagnostic tests for LNYV subgroups

2.1 Introduction

LNYV subgroups were first identified in Australia by Callaghan and Dietzgen (2005) from infected lettuce, garlic, bristly oxtongue and sowthistle samples. The nucleocapsid (N) gene was sequenced and phylogenetic analysis was conducted on ten isolates. Two distinct groups of isolates were identified and were designated as subgroup I and subgroup II, with three and seven isolates in each clade, respectively (Callaghan and Dietzgen 2005). Subsequently, a study conducted by Higgins et al. (2016b) identified LNYV in six lettuce samples, collected from Harrisville, Auckland, New Zealand as well as nine other Australian samples. Phylogenetic analysis of the N gene from those isolates showed that five of the New Zealand isolates belonged to subgroup I while one belonged to subgroup II and all new Australian isolates belonged to subgroup II.

2.1.1 Mechanical transmission test to detect LNYV subgroups

LNYV was mechanically transmitted into *N. glutinosa* and infected plants were diagnosed by evaluation of visual symptoms (Chu and Francki 1982). However, transmission tests do not diagnose LNYV subgroups because, according to Higgins et al. (2016b), there are no distinctive symptoms for each subgroup. Therefore, it can be concluded that mechanical transmission of LNYV to *N. glutinosa* is not suitable to determine which subgroup an isolate belongs to.

2.1.2 LNYV subgroup detection by electron microscopy

Plant viruses can be diagnosed using EM of infected plants based on viral morphology (Sastry 2013b). Harrison and Crowley (1965) used EM for LNYV detection; however, Bawden and Nixon (1951) reported that virus strains were not discriminated due to the high similarity in morphology. No EM studies have been done, but it is unlikely LNYV subgroups could be distinguished using this method.

2.1.3 Serological detection of LNYV subgroups

LNYV has been detected using DAS-ELISA from infected *S. oleraceus*, *N. glutinosa* and *H. lactucae* (Chu and Francki 1982). Fletcher et al. (2017) used indirect ELISA to diagnose LNYV infected lettuce samples. These latter lettuce samples were sent AUT to identify the subgroups using a molecular approach (reported in sections 2.3.4.2.2 and 2.3.4.2.3) because there is no diagnostic method for LNYV subgroups using ELISA. Such a serological assay would require each subgroup to have a specific epitope and as well as developing a high quality antibody for each subgroup. According to Naidu and Hughes (2003), it is difficult to obtain highly specific antibodies to viruses, thus it would be an even greater challenge to develop subgroup specific antibodies for detection.

2.1.4 Detection of LNYV subgroup using RT-PCR analysis.

RT-PCR analysis has been used to diagnose LNYV previously. Two primer pairs were designed for this purpose. These primers are BCNG1/BCNG2 and LNYV_440F/LNYV_1185R (Callaghan and Dietzgen 2005; Higgins et al. 2016b) but subgroup specific primers have not been designed for LNYV subgroup diagnosis. Studies carried out on nepovirus by Digiaro et al. (2007) showed development of strain specific RT-PCR is feasible. Based on this study, it can be presumed that the RT-PCR can be used to detect LNYV subgroups.

RT-PCR can be combined with restriction fragment length polymorphism (RFLP) for sequence specific diagnosis. For this method, the PCR product generated by RT-PCR is digested with a specific restriction endonuclease and fragments separated by gel electrophoresis (Sastry 2013a). The virus strains can be distinguished as the RT-PCR products from each have a different number of sites for the restriction enzyme, leading to the generation of fragments of different sizes. The fragment pattern is specific to each strain. Such a method has been used to identify strains of eggplant mottled dwarf virus (Parrella and Greco 2016) and has the potential to distinguish the LNYV subgroups.

2.1.5   Rapid extraction method

RNA extraction is a time consuming process. Having a rapid method for template preparation for RT-PCR-based diagnostics would give a more rapid diagnosis and allow higher throughput of samples. The template preparation solution (TPS) procedure described by Thomson & Dietzgen (1995) enabled the release of LNYV more rapidly than other plant or viral nucleic acid extraction procedures. LNYV was released from fresh and frozen lettuce and *N. glutinosa* samples and LNYV detected using RT-PCR. The procedure appeared straightforward, with one step preparation, and did not include the use of organic solvents (Thomson and Dietzgen 1995). Many organic solvents are hazardous, extraction requires multiple steps and increases the likelihood of cross contamination (Xiao et al. 2015). Therefore, the TPS procedure was assessed in combination with RT-PCR for rapid diagnosis of LNYV subgroups.

2.1.6   Aims

In this chapter, the aim was to develop diagnostic tests for LNYV subgroups using RT-PCR and RT-PCR RFLP and to expand the knowledge of the LNYV population in New Zealand. Subgroup specific primers were designed and tested on known LNYV infected samples obtained from the study by Higgins et al. (2016b) to optimise the reaction conditions. The BCNG1/BCNG2 and LNYV_440F/LNYV_1185R primer pairs were used to initially diagnose LNYV following optimisation in the AUT laboratory. These primers were then used with previously untested samples, following which the LNYV subgroup was identified. The remaining samples that were collected from Harrisville in 2011 but not tested in the study by Higgins et al. (2016b) were tested using the optimised primers. As mentioned previously, the Canterbury samples were already identified as infected by LNYV by ELISA. These samples were sent to AUT for subgroup identification by PCR. Additional samples that were collected from Tuakau, Waikato were also tested for LNYV and its subgroups.

2.2 Materials and methods

2.2.1 Plant materials

Plant leaf samples showing symptoms of LNYV infection were collected from three locations of New Zealand to test for LNYV (Table 2.1). A total of 26 samples were collected by Colleen Higgins (Auckland University of Technology) from a single lettuce farm in Harrisville, Auckland during December, 2011 and stored at -80 °C until use. Other samples were collected from Tuakau, Waikato in January, 2015 by Colleen Higgins (Auckland University of Technology) and stored at -80 °C. Canterbury samples were collected by John Fletcher, Plant and Food Research, during January, 2017 and stored at -80 °C (Fletcher et al. 2017). All samples were tested for LNYV and its subgroups using molecular diagnosis as described below.

Table 2.1: Plant samples obtained from various locations during 2011 and 2015 summer.

Sample name	Plant species	Location	Sampled date
H2	*Lactuca sativa*	Harrisville, Auckland	December, 2011
H3			
H5			
H6			
H7			
H8			
H9			
H10			
H11			
H12			
H14			
H16			
H17			
H18			
H21	*Senecio vulgaris*		
H22			
H25			
H26			
H35	*Lactuca sativa*		
SC28	*Sonchus asper*	Tuakau, Waikato	January, 2015
SC29			
SC30			
SC31			
SC32			
SC33	Sowthistle		
SC34	*Lactuca sativa*		
SC35			
SC36			
SC37			
SC38	Sowthistle		
SC40			
SC41			
Scott 2 (S2)	*Lactuca sativa*	Southbridge, Canterbury	February, 2017
Scott11 (S11)			
Scott13 (S13)			
Scott14 (S14)			
Kiesanowski9 (K9)		Marshlands, Canterbury	
Watson4 (W4)			
Leaderband2 (L2)		Chertsey, Canterbury	

2.2.2 RNA extraction

2.2.2.1 RNA extraction using Spectrum™ plant total RNA extraction kit

Total RNA was extracted from leaf samples using the Spectrum™ Plant Total RNA Extraction Kit (Sigma Aldrich, Missouri, USA). The manufacturer's instructions were modified to obtain a higher concentration of RNA (R. Dietzgen, personal communication, November 17, 2016). Approximately 100 – 200 mg of the leaf samples was used for each extraction. Samples were ground in liquid nitrogen to form a powder. Cells were lysed with 500 µl of the lysis solution/2-mercaptoethanol mixture, according to the manufacturer's instructions. The sample was vortexed vigorously for 30 seconds and then centrifuged at 15,000 x g (Eppendorf 5430R Centrifuge, New South Wales, Australia) for 3 minutes to separate the pellet and supernatant. All centrifugation steps were performed at room temperature. The lysate supernatant was pipetted into a filtration column and centrifuged at 15,000 x g for 1 minute to remove the large debris. Following protocol A, 500 µl of binding solution was added to the flow through and the mixed solution was added onto a binding column. After centrifugation at 15,000 x g for 1 minute, the residue was discarded. DNA was digested by using the One-Column DNase digestion kit (Sigma Aldrich, Missouri, USA) according to the manufacturer's instructions. Wash solution I (300 µl) was added to the binding column and centrifuged at 15,000 x g for 1 minute. A master mix was prepared by combining 10 µl of DNase I and 70 µl of DNase digestion buffer for each sample. For each sample, 80 µl of this master mix was added to the column and incubated for 15 minutes at room temperature. After the incubation, 500 µl of wash solution I was pipetted onto the binding column, which was then centrifuged at 15,000 x g for 1 minute to complete the DNA digestion procedure. The RNA extraction procedure was continued by adding 500 µl of wash solution II to the binding column, centrifuging at 15,000 x g for 30 seconds, this was repeated once. The binding column was centrifuged at 15,000 x g for 1 minute to dry the column and transferred into a new 2 ml collection tube. Elution solution of 25 µl was added and incubated at room temperature for 1 minute. After centrifugation at 15,000 x g for 1 minute to elute the RNA. A further 25 µl of elution solution was added to the same binding column and the elution process was repeated to increase the RNA concentration. The final RNA extract was stored in -80 °C freezer.

2.2.2.2   RNA extraction using the Quick RNA™ MiniPrep kit

RNA samples were also extracted using the Quick-RNA™ MiniPrep kit (Zymo Research, California, USA) with the manufacturer's instructions being slightly modified to reduce RNA degradation. Around 100 – 200 mg of frozen plant tissue was ground in liquid nitrogen to form a powder. RNA lysis buffer of 600 µl was added to the leaf powder and incubated at room temperature for 5 – 10 minutes to form a liquid solution. The solution was centrifuged at 10,000 x g (Eppendorf 5430 R Centrifuge, New South Wales, Australia) for 1 minute to separate the pellet and supernatant. All centrifugation steps were performed at room temperature. The supernatant was pipetted into a spin-away filter column and centrifuged at 10,000 x g for 1 minute. Hydrous ethanol (96% ethanol) of 600 µl was added to the flow-through. This was transferred to a Zymo-spin IIICG column and centrifuged at 12,000 x g for 30 seconds. DNA was digested using the In-column DNase I treatment; 400 µl of RNA wash buffer was added and centrifuged at 12,000 x g for 30 seconds. DNase I reaction master mix was prepared by combining 5 µl of lyophilized DNase I^2 and 75 µl of DNA digestion buffer for each sample. All 80 µl of the master mix was used per sample. After incubation for 15 minutes at room temperature, samples were centrifuged at 12,000 x g for 30 seconds. RNA prep buffer of 400 µl was added and centrifuged at 12,000 x g for 30 seconds. RNA wash buffer of 700 µl was added and centrifuged at 12,000 x g for 30 seconds, followed by another wash with 400 µl of RNA wash buffer, then centrifuging at 12,000 x g for 2 minutes. DNase/RNase-free water of 50 µl was added to the column, which was finally centrifuged at 12,000 x g for 30 seconds to elute the RNA. The purified RNA extract was stored in – 80 °C.

2.2.3   RNA concentration and integrity

The concentration of the extracted RNA samples was quantified by measuring the absorbance of 2 µl using a GE Nanovue spectrophotometer (GE Biosciences, New Jersey, USA) using the default settings. The elution solution of 2 µl was used to calibrate the spectrophotometer. The purity of the RNA was measured as the A_{260}/A_{280} and A_{260}/A_{230} absorbance ratios.

Agarose gel electrophoresis was used to evaluate the integrity of RNA using a Mini-sub® Cell GT Cell Gel electrophoresis chamber (Bio-Rad, Auckland, New Zealand). RNA

extract of 3 – 5 µl and 3 µl of 0.1 µg/µl 100 bp DNA ladder (Solis BioDyne, Estonia) were used. Ethidium bromide at a final concentration of 10 µg/ml was used as an intercalating agent and agarose gel was immersed in 1 X TBE buffer. Gels were electrophoresed at 75 Volts for 50 minutes and viewed under UV light using the Alpha Imager (Protein Simple, California, USA).

2.2.4   Primer design

High-quality primers are essential for an efficient and specific RT-PCR analysis to detect the specific organism for diagnosis (Jeong et al. 2014; Narayanasamy 2011a). LNYV subgroup specific primers were designed to distinguish between the subgroups by RT-PCR. All available sequences for the LNYV N gene were downloaded from NCBI Genbank (https://www.ncbi.nlm.nih.gov/) (Table 2.2) and were aligned using MUSCLE alignment software (http://www.drive5.com/muscle). The alignment was used to identify multiple potential subgroup specific and non-subgroup specific primers manually. A total of three primers were selected: one non-specific reverse primer (binds to both subgroups) and two subgroup specific forward primers were tested on known LNYV infected samples. Melting temperature (Tm), GC content, specificity, primer length and hairpin are the criteria for a standard PCR primer design (Lorenz 2011). Primers were analysed using OligoAnalyser 3.1 program (https://sg.idtdna.com/calc/analyzer) with default settings to identify these parameters. The parameters were 39.4% – 57.9% GC content, ~57 °C Tm, 19 nt to 28 nt length and -2.41 – 0.041 kcal.mole^{-1} hairpin. All of these criteria were used to test the subgroup specific and non-specific subgroup primers for high specificity, reduce non-specific PCR products and primer dimer (Table 2.3). The species specificity was tested using the NCBI BLASTn search (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

Table 2.2: LNYV N gene sequences from the NCBI Genbank database.

Published isolate code	Referred to as	Subgroup	Plant host	Accession number	Size (bp)
AU2	Not applicable	Subgroup I	*Allium sativum*	L30103	1,533
AU2*				AJ746191	1,533
AU6				AJ746195	1,533
AU1		Subgroup II	*Lactuca sativa*	AJ746190	1,539
AU3				AJ746192	1,539
AU4				AJ746193	1,538
AU5				AJ746194	1,539
AU7			*Picris echioides*	AJ746196	1,539
AU8			*Sonchus oleraceus*	AJ746197	1,538
AU9				AJ746198	1,539
AU10			*Lactuca sativa*	KP109940	1,539
AU11				KP109941	1,539
AU12				KP109942	1,539
AU13				KP109943	1,539
AU14			*Sonchus oleraceus*	KP109944	1,504
AU15			*Lactuca sativa*	KP109945	748
AU16				KP109946	748
AU17				KP109947	1,537
AU18			*Actites megalocarpus*	KP109948	1,539
NZ2	H27[+]	Subgroup I	*Lactuca sativa*	KP109950	1,516
NZ3	H28[+]			KP109951	1,516
NZ4	H29[+]			KP109952	1.516
NZ5	H30[+]			KP109953	1,516
NZ6	H33[+]			KP109954	1,516
NZ1	H19[+]	Subgroup II		KP109949	1,540

AU2 isolate was mechanically transferred to *N. glutinosa* to obtain AU2* sequence (Callaghan and Dietzgen 2005)

+ LNYV infected samples used in Higgins et al (2016)

AU = Australian isolate   NZ = New Zealand isolate

2.2.5   One-step RT-PCR

One-step RT-PCR was performed using the InVitrogen SuperScript[TM] III One-Step RT-PCR with Platinum[TM] *Taq* DNA polymerase kit (Thermofisher, Massachusetts, USA). Optimisation of the RT-PCR conditions is described below. The samples were amplified on a Bibby Scientific[TM] Techne[TM] TC-512 Gradient Thermal Cycler (Fisher Scientific, England, UK). Primers used for the analysis are shown in Table 2.3.

Table 2.3: LNYV N gene primers for RT-PCR detection.

Primer name	Nucleotide sequence 5' → 3'	Expected product size (bp)	Reference
BCNG1	TCT GGG TAT TGG TTC GGG AAA AGA GTG	1500	(Callaghan and Dietzgen 2005)
BCNG2	AGT ATT CAT AAA CTG ATG TGG TTT CTC		
LNYV_440F	TGA CAC AGA TTC AGA ACA ACT C	746	(Higgins et al. 2016b)
LNYV_1185R	CGG ACA ATC CAT CTC CAC TA		
LNYVNS1F889 (Subgroup I)	ACC TGA AGT CTT ATC CAC ATG GAC TTT A	212	Subgroup specific primers used in this study
LNYVNS1S2R110 (Non-subgroup specific)	TGG CCG GAC AAT CCA TCT C		
LNYVNS2F892 (Subgroup II)	CGA AGT GTT ATC GAC ATG GAC ACT G	209	

2.2.5.1 RT-PCR analysis using subgroup specific primers

a) Initial RT-PCR conditions

PCR product was amplified from RNA extracted samples as follows: 1 µl of RNA template, 2 µl each of 10 µM non-subgroup specific primer and subgroup I or subgroup II specific primer (Table 2.3), 12.5 µl of 2X Reaction mix, 1 µl SuperScript™ III RT/ Platinum *Taq* Mix and autoclaved distilled water to a total volume of 25 µl. The RT-PCR conditions were 1 cycle of cDNA synthesis for 30 minutes at 50 °C, 1 cycle of pre-denaturation for 2 minutes at 94 °C, 40 cycles of PCR amplification with denaturation for 15 seconds at 94 °C, annealing for 30 seconds at 55 °C, extension for 1 minute 50 seconds at 68 °C and final extension for 5 minutes at 68 °C. The amplified PCR product was stored at -20 °C.

b) Gradient PCR

Gradient PCR is used to determine the optimum annealing temperature for primers. It was performed on known LNYV infected samples and negative controls using the non-subgroup specific and subgroup I primer. The annealing temperature range was 55 °C to 65 °C. The 25 µl volume were set up as described in section 2.2.5.1a. The RT-PCR conditions were 1 cycle of cDNA synthesis for 30 minutes at 50 °C, 1 cycle of pre-denaturation for 2 minutes at 94 °C, 40 cycles of PCR amplification with denaturation for 15 seconds at 94 °C, annealing for 30 seconds at 60 °C Δ 5 °C (55 °C to 65 °C), extension for 1 minute 50 seconds at 68 °C and final extension for 5 minutes at 68 °C. The amplified

PCR product was stored at -20 °C. Similar reactions were performed using subgroup II and non-subgroup specific primers with annealing temperature ranging from 45 °C to 55 °C.

c) Final optimised RT-PCR conditions

The amount of template RNA, primer concentration, cDNA synthesis temperature, PCR cycle number, annealing temperature and extension time conditions were required to be optimised. RNA samples were amplified using 100 – 300 ng of RNA sample extract, 0.5 µl each of 10 µM non-subgroup specific and subgroup I or subgroup II specific primers, 12.5 µl of 2X Reaction mix, 1 µl SuperScript^{TM} III RT/ Platinum *Taq* Mix and autoclaved distilled water to a total volume of 25 µl. The RT-PCR conditions were 1 cycle of cDNA synthesis for 30 minutes at 55 °C, 1 cycle of pre-denaturation for 2 minutes at 94 °C, 30 cycles of PCR amplification with denaturation for 15 seconds at 94 °C, annealing for 30 seconds at 65°C, extension for 30 seconds at 68 °C and final extension for 5 minutes at 68 °C. The amplified PCR product was stored at -20 °C.

For RT-PCRs where the total volume was 12.5 µl, the optimised conditions were 100 – 300 ng RNA sample extract, 0.25 µl each of 10 µM non-subgroup specific and subgroup I or subgroup II primers, 6.25 µl of 2X Reaction mix, 0.5 µl SuperScript^{TM} III RT/ Platinum *Taq* Mix and autoclaved distilled water to 12.5 µl. The RT-PCR conditions were as described in the previous paragraph. These conditions were used where the non-subgroup specific primer was used with either subgroup specific primers as well as in a multiplex RT-PCR with all three primers.

2.2.5.2 RT-PCR analysis using the BCNG1/BCNG2 primers

a) Initial RT-PCR conditions

Initial RT-PCR conditions using BCNG1 and BCNG2 primers (Table 2.3) for detection of LNYV were as described by Callaghan and Dietzgen (2005). The SuperScript^{TM} III One-Step RT-PCR with Platinum^{TM} *Taq* DNA polymerase kit (Thermofisher, Massachusetts, USA) were used. For 25 µl volume, 1 µl template RNA, 1 µl each of 10 µM BCNG1 and BCNG2 primers, 12.5 µl of 2X Reaction mix, 1 µl SuperScript^{TM} III RT/ Platinum *Taq* Mix and autoclaved distilled water were used. The RT-PCR conditions

were as described by Callaghan and Dietzgen (2005), 1 cycle of cDNA synthesis for 30 minutes at 50 °C, 1 cycle of pre-denaturation for 2 minutes at 94 °C, 35 cycles of PCR amplification with denaturation for 30 seconds at 94 °C, annealing for 30 seconds at 50 °C, extension for 2 minutes at 72 °C, final extension for 10 minutes at 72 °C. The amplified PCR product was stored at -20 °C.

b) Final optimised RT-PCR conditions

The amount of RNA template, PCR cycles, extension and final extension temperature conditions were required to be optimised. The RT-PCR conditions for a 25 µl volume were 250 ng – 500 ng template RNA, 1 µl each of 10 µM BCNG1 and BCNG2 primers, 12.5 µl of 2X Reaction mix, 1 µl SuperScript^{TM} III RT/ Platinum *Taq* Mix and autoclaved distilled water. The RT-PCR conditions were 1 cycle of cDNA synthesis for 30 minutes at 50 °C, 1 cycle of pre-denaturation for 2 minutes at 94 °C, 40 cycles of PCR amplification with denaturation for 30 seconds at 94 °C, annealing for 30 seconds at 50 °C, extension for 2 minutes at 68 °C and final extension for 10 minutes at 68 °C. The amplified PCR product was stored at -20 °C.

2.2.5.3 RT-PCR analysis using the LNYV_440F/LNYV_1185R primers

a) Initial RT-PCR conditions

The RT-PCR conditions using LNYV_440F and LNYV_1185R primers (Table 2.3) for LNYV detection were as described by Higgins et al. (2016b) with conditions modified for use in the AUT laboratory. For 25 µl volume, 250 ng template RNA, 1 µl each of 10 µM LNYV_440F and LNYV_1185R primers, 12.5 µl of 2X Reaction mix, 1 µl SuperScript^{TM} III RT/ Platinum *Taq* Mix and autoclaved distilled water. The RT-PCR conditions were 1 cycle of cDNA synthesis for 30 minutes at 50 °C, 1 cycle of pre-denaturation for 2 minutes at 94 °C, 40 cycles of PCR amplification with denaturation for 30 seconds at 94 °C, annealing for 30 seconds at 50 °C, extension for 1 minutes at 68 °C and final extension for 5 minutes at 68 °C. The amplified PCR product was stored at -20 °C.

b)   Final optimised RT-PCR conditions

The number of PCR cycles condition was required to be optimised for the LNYV_440F and LNYV_1185R primers. The optimised RT-PCR conditions for a 25 µl volume were 250 ng template RNA, 1 µl each of 10 µM LNYV_440F and LNYV_1185R primers, 12.5 µl of 2X Reaction mix, 1 µl SuperScript^{TM} III RT/ Platinum *Taq* Mix and autoclaved distilled water. The amplification conditions were 1 cycle of cDNA synthesis for 30 minutes at 50 °C, 1 cycle of pre-denaturation for 2 minutes at 94 °C, 30 cycles of PCR amplification with denaturation for 30 seconds at 94 °C, annealing for 30 seconds at 50 °C, extension for 1 minutes at 68 °C and final extension for 5 minutes at 68 °C. The amplified PCR product was stored at -20 °C. The optimised conditions for the reactions with a total volume of 12.5 µl were all reaction components in the mentioned above was scaled down by half and the RT-PCR conditions were the same.

2.2.6   Testing for LNYV in plant samples

Samples listed in Table 2.1 were tested for LNYV using the primer pairs BCNG1/2 and LNYV_440F/LNYV_1185R primers by RT-PCR as described in sections 2.2.5.2b and 2.2.5.3b, respectively. The LNYV subgroups were detected using the subgroup specific primers developed in the current study, which was described in section 2.2.5.1c.

2.2.6.1   Testing for LNYV using the BCNG1/BCNG2 primers

The BCNG1 and BCNG2 primers amplify the complete LNYV N gene resulting in a PCR product of approximately 1,500 bp (Callaghan and Dietzgen 2005). It was intended that all previously untested samples would be analysed using the BCNG1/BCNG2 primers, but since one Harrisville sample (H3) that was tested negative with these primers was found to be positive with other primers (sections 2.2.6.2 and 2.2.6.3). BCNG1/BCNG2 were considered unreliable and no further samples were tested. The BCNG1/BCNG2 primers were concluded to be inefficient to detect LNYV and more false negative results are likely to occur if this primer pair was used to test the remaining samples. Therefore, the remaining samples was tested using the LNYV_440F/LNYV_1185R primers.

The optimised conditions described in 2.2.5.2b were used to test the 11 samples using the BCNG1/BCNG2 primers in a 25 µl volume. The positive control for the RT-PCR was the LNYV infected H19 sample that was previously identified by Higgins et al. (2016b) and two negative controls - healthy uninfected lettuce RNA extract and a no template control (NTC). The negative controls were used to identify contamination and non-specific primer binding in the RT-PCR.

2.2.6.2   Testing for LNYV using the LNYV_440F/LNYV_1185R primers

The LNYV_440F and LNYV_1185R primers amplify a PCR product of approximately 750 bp from the LNYV N gene (Higgins et al. 2016b). The 11 plant samples that were previously tested (section 2.2.6.1) were retested with these primers. The remaining eight samples from Harrisville, 13 Tuakau and four Canterbury samples were also tested in a 12.5 µl volume. The optimised conditions described in 2.2.5.3b were used for the analysis. The positive controls were LNYV infected H29 (subgroup I) and H19 (subgroup II) samples (Higgins et al. 2016b), and negative controls were the same as described in 2.2.6.1.

2.2.6.3   Testing for LNYV subgroups using the subgroup specific primers

The subgroup specific primers were used to detect LNYV subgroups, amplifying products of approximately 200 bp. All the plant samples obtained from Harrisville, Tuakau and Canterbury were retested using these primers in a 12.5 µl one-step RT-PCR, as described in section 2.2.5.1c. The positive and negative controls described in 2.2.6.1 were used for the analysis.

2.2.7   Template Preparation Solution (TPS) procedure

2.2.7.1   Extraction using TPS

Detection of LNYV using template preparation solution (TPS) procedure with RT-PCR has the potential to provide a rapid template preparation method for virus detection. Thomson and Dietzgen (1995) demonstrated that TPS1 solution with incubations at 95 °C or 60 °C were suitable to detect LNYV in lettuce and *N. glutinosa*. This procedure was assessed on known LNYV infected and uninfected samples. TPS1 solution contained

100 mM Tris-HCl, 1.0 M KCl and 10 mM EDTA, pH of 8.42. Around 1 mg of infected or uninfected leaves were ground in 100 µl of the TPS1 solution. Samples were centrifuged for a few seconds, incubated at 95 °C for 10 minutes using a heating block (Model: MiniT-100, Allsheng, Hangzhou city, China) and the supernatant diluted 10 fold in Millipore water. Undiluted supernatant and diluted supernatant of 1 µl were tested in the SuperScript™ III One-step RT-PCR system with Platinum™ *Taq* DNA polymerase using the various primer pairs as described below.

2.2.7.2   Testing the TPS extracted samples using the subgroup specific primers

Templates obtained from the TPS extraction procedure described in section 2.2.7.1 were tested for LNYV subgroups using one-step RT-PCR with the non-subgroup specific and subgroup specific primers. The optimised conditions described in the section 2.2.5.1c were used for the analysis, except that 1 µl and 0.5 µl of each sample was used for the RT-PCR analysis in a 12.5 µl volume. The negative controls were the same as described in 2.2.6.1.

2.2.7.3   Testing the TPS extracted samples using the LNYV_440F / LNYV_1185R primers

LNYV infected samples that were extracted using the TPS as described in section 2.2.7.1 were tested for LNYV using the LNYV_440F and LNYV_1185R primers. The optimised conditions described in section 2.2.5.3b were used for the analysis, except that 1 µl of each sample was used in a 12.5 µl volume. The negative controls were the same as described in 2.2.6.1.

2.2.8   Agarose gel electrophoresis

Products obtained from the RT-PCR analyses were visualised using agarose gel electrophoresis in a Mini-sub® Cell GT Cell Gel electrophoresis chamber (Bio-Rad, Auckland, New Zealand). Samples were electrophoresed at 75 Volts for 50 minutes in 1X TBE buffer with ethidium bromide at a final concentration of 10 µg/ml. For the subgroup specific RT-PCR detection, 1.5% agarose was used, while for all other primers, 1% was used. The DNA ladders used were 3 µl of 100 ng/µl Kapa universal ladder (Kapa

Biosystem, Massachusetts, USA) or 3 µl of 100 ng/µl 100 bp DNA ladder (Solis BioDyne, Estonia). The gels were viewed under UV light using an Alpha Imager (Protein Simple, California, USA).

2.2.9 RT-PCR-Restriction Fragment Length Polymorphism (RFLP) to detect subgroups

A RT-PCR-RFLP method was developed to distinguish the LNYV subgroups. The LNYV genome sequence region amplified by the subgroup specific primers was analysed for restriction enzyme sites that would distinguish each subgroup. An enzyme was chosen that gave different restriction fragments patterns for each subgroup.

LNYV was amplified from known positive samples using the non-subgroup specific, subgroup I and subgroup II primers in a multiplex RT-PCR. The optimised conditions described in the section 2.2.5.1c were used for the analysis in a 25 µl volume. The restriction enzyme used for RFLP was *Mae*III. A digest was set up with 10 µl of amplified PCR product, 12.5 µl of 2X *Mae*III incubation buffer, 1 unit of *Mae*III restriction enzyme (Sigma-Aldrich, New Zealand) and autoclaved distilled water to a final volume of 25 µl. The digest was incubated at 55 °C for 1 hr in a heating block (Model: MiniT-100, Allsheng, Hangzhou city, China). Digested and non-digested products were separated by gel electrophoresis as described in section 2.2.8, using 4% agarose electrophoresed at 75 Volts for 90 minutes.

2.3 Results

2.3.1 Subgroup specific primer design

Nucleotide primers are short DNA sequences that initiate DNA synthesis during replication and is a critical component used for PCR amplification. The specificity of the primers determines the amplification of a particular target sequence (Pelt-Verkuil et al. 2008). Published LNYV N gene sequences from various isolates were aligned to identify regions that might be suitable for designing primers that would distinguish the LNV subgroups (Figure 2.1). It was difficult to find positions within the target region that would give rise to different sized PCR products that could be easily distinguished by electrophoresis. Thus, one primer was designed that was not subgroup specific that would act as the reverse primer for each of two subgroup specific forward primers. Identification of these primers was as described below.

In total, six pairs of subgroup specific primers and three non-subgroup specific primers were identified as candidates for RT-PCR analysis (Table 2.4 and Figure 2.1). Among these, four pairs of subgroup specific primers (Table 2.4 referred as, 1 – 8) had fewer than six nucleotide differences between subgroup I and subgroup II primers. This low number of nucleotide differences between the subgroups suggested these primers may not be subgroup specific. Subgroup specific primer pairs 9 to 12 had greater than six nucleotide differences between the subgroups and these primer pairs may be suitable for LNYV subgroup diagnosis.

Closer inspection of the subgroup I specific primer 9 showed it did not match five out of the eight subgroup I sequences sufficiently. This primer binding position had two number of nucleotide differences between the different sequences reducing confidence that this primer would amplify its target sequence. The subgroup II specific primer 10 did not match four out of the 17 subgroup II sequences. The subgroup specific primer 11* matched all subgroup I isolates while subgroup II specific primer 12* was mismatched with only one out of the 17 subgroup II sequences. One mismatched sequence was acceptable for the current analysis. Therefore, the subgroup specific primers 11* and 12* and non-subgroup specific (NSS) 3* primers were chosen for further analysis. The non-subgroup specific primer 3* (reverse), subgroup I primer 11* (forward) and subgroup II

primer 12* (forward) were named LNYVNS1S2R110, LNYVNS1F889 and LNYVNS2F892, respectively (Figure 2.2). A summary for these primers is provided in Table 2.5. The non-subgroup specific and subgroup I primers were expected to amplify a region of 212 bp, while non-subgroup specific and subgroup II primers were expected to amplify a region of 209 bp. For both sets of primers, the Tm was 57.3°C – 57.5°C and GC content 39.4% – 57.9% (Table 2.5). The BLASTn search shows that these primers are not similar to human, bacteria, fungal, plants, insects or viruses sequences, indicating that these primers are specific.
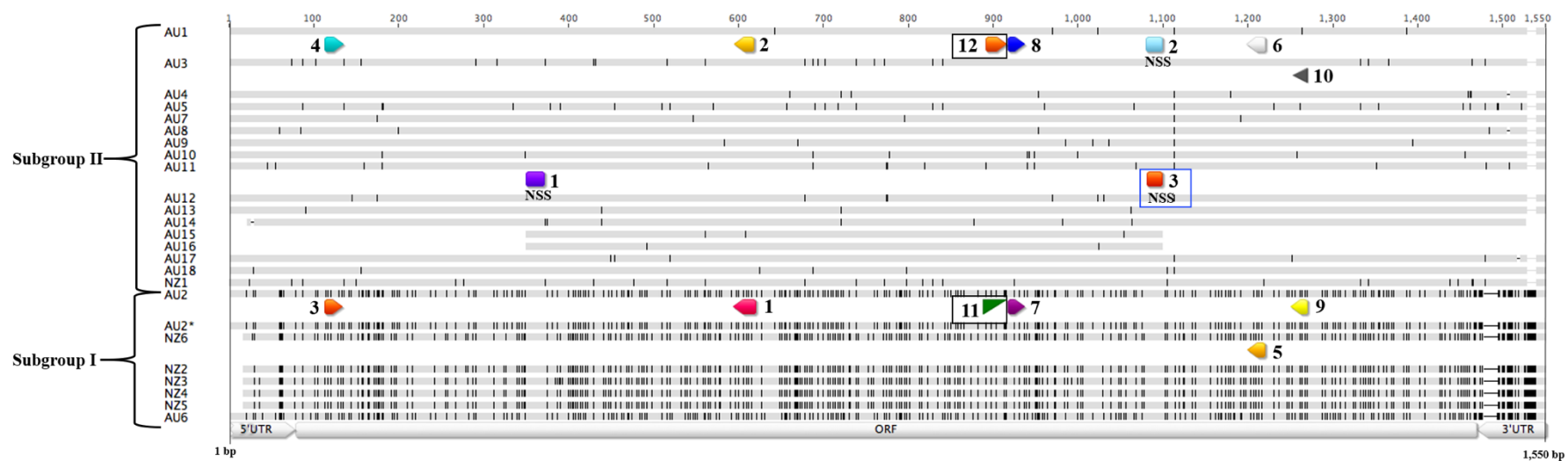
Figure 2.1: The relative locations of the candidate non-subgroup specific, subgroup I and subgroup II primers within the LNYV N gene. The numbers refer to primers in Table 2.4. Black vertical lines within the alignment indicate where the nucleotides differ from the consensus sequence. The black coloured boxes indicate the finalised subgroup specific primers and blue box indicates the finalised non-subgroup specific (NSS) primer. Numbers at top the figure indicate nucleotide position.
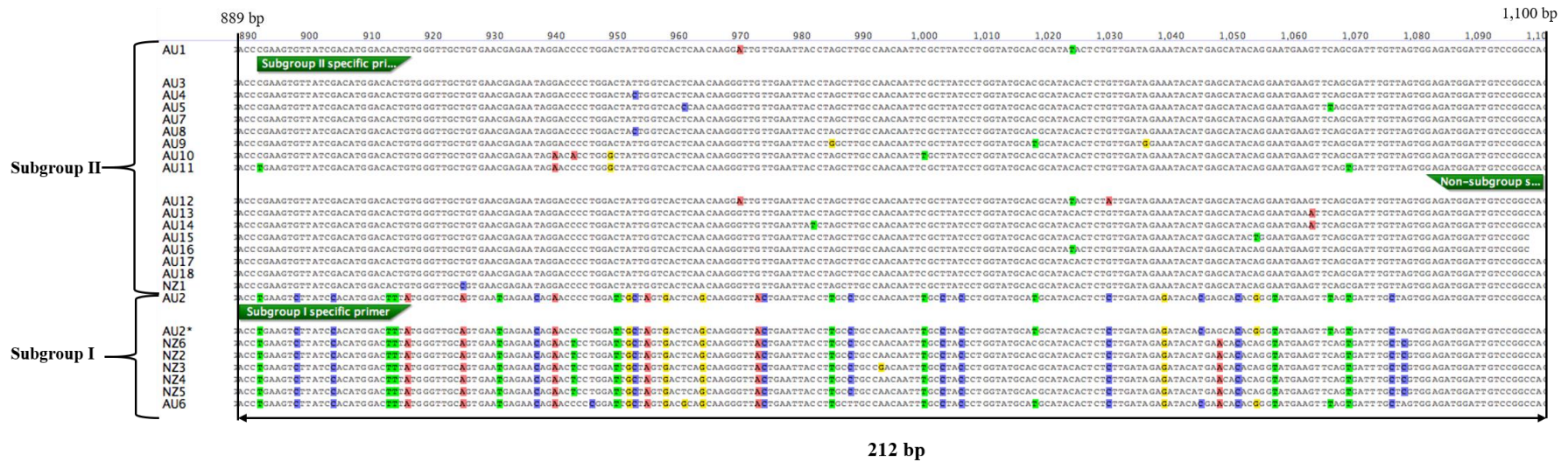
Figure 2.2: The location of subgroup specific and non-subgroup specific primers in the LNYV N gene. Colours within the alignment indicate where the nucleotides differ from the consensus sequence. Numbers at top of the figure indicate nucleotide position. * Sequence of AU2 after 10 years of mechanical transfer as described in Callaghan & Dietzgen (2005).

Table 2.4: Summary of candidate non-subgroup specific, subgroup I and subgroup II primers.

Primers					No. differences in nt between subgroups (nt)	Approximate product size (bp)	No. of sequences matched/ not matched with sub I sequences
Referred as	Non-subgroup specific primers	Subgroup	Referred as	Subgroups specific primers (5' → 3')			No. of sequences matched/ not matched with sub II sequences
Non-subgroup specific (NSS) 1	GAG TTG TTC TGA ATC TGT GTC A	Subgroup I	1	GCT TCA TAT CTG GCT CTC TTC TCT TCA	5	268	Matched all sub I sequences
		Subgroup II	2	TCG TAC CTA GCC CTC TTT TCC TC			Did not match 1 out of 17 sub II sequences
		Subgroup I	3	AAC AAT TGA GGA AAG AGA GGG	5	236	Matched all sub I sequences
		Subgroup II	4	AGC AGT TAA GGA AAG AAA GGT C			Matched all sub II sequences
		Subgroup I	5	GCC CCA AAA TAT TTT GGG TCC C	5	870	Did not match 2 out of 8 sub I sequences
		Subgroup II	6	CC CCG AAG TAC TTA GGA TCC CA			Matched all sub II sequences
NSS 2	GGA GAT GGA TTG TCC GGC CA	Subgroup I	7	GGG TTG CAG TGA ATG AGA AC	3	183	Matched all sub I sequences
		Subgroup II	8	GGG TTG CTG TGA ACG AGA AT			Did not match 1 out of 17 sub II sequences
		Subgroup I	9	GAC AGA CGC AGC AAC GTA TA	7	189	Did not match 5 out of 8 sub I sequences
		Subgroup II	10	AAC GGA GGC GGC AAC ATA			Did not match 4 out of 17 sub II sequences
NSS 3*	TGG CCG GAC AAT CCA TCT C	Subgroup I	11*	ACC TGA AGT CTT ATC CAC ATG GAC TTT A	6	200	Matched all sub I sequences
		Subgroup II	12*	CGA AGT GTT ATC GAC ATG GAC ACT G			Did not match 1 out of 17 sub II sequences

Red coloured is the difference in nucleotides between the LNYV subgroup I and subgroup II primers. nt = nucleotides. bp = base pair. Sub = subgroup
* Finalised primers

Table 2.5: Summary of the finalised non-subgroup specific, subgroup I and subgroup II primers with the parameters.

Name	Sequence (5'→3')	Length (nt)	Product size (bp)	Tm (°C)	GC content (%)	Hair pin ΔG (kcal.mole^{-1})
LNYVNS1F889 (Subgroup I)	ACC TGA AGT CTT ATC CAC ATG GAC TTT A	28	212	57.4	39.3	-1.97 – -1.11
LNYVNS1S2R110 (Non-subgroup specific)	TGG CCG GAC AAT CCA TCT C	19	NA	57.3	57.9	0.04 – -0.38
LNYVNS2F892 (Subgroup II)	CGA AGT GTT ATC GAC ATG GAC ACT G	25	209	57.5	48	-1.67 – -2.41

NA = not applicable

2.3.2 Template RNA quality and integrity

An efficient RT-PCR analysis requires high quality and non-degraded RNA (Reiter and Ptatfl 2011). Non-denaturing agarose gel electrophoresis was carried out to determine if the RNA was degraded. Figure 2.3 shows an example gel with two extractions of a known LNYV positive sample, H33 (subgroup I). Intact 28S and 18S rRNA bands can be seen for both samples, as well 5S rRNA. The DNA markers indicate how far the gel had run. RNA concentration was determined by measuring absorbance at 260 nm (Table 2.6). Sample purity was determined by measuring the A_{260}/A_{280} and A_{260}/A_{230} ratios. Good quality RNA should have ratios of 1.9 – 2.1 and >2 for A_{260}/A_{280} and A_{260}/A_{230}, respectively (Berthiaume and Morgan 2010). The ratios of H33 samples were >2, which indicated that RNA was good quality and these samples could be efficiently used for further experiments (Table 2.6).
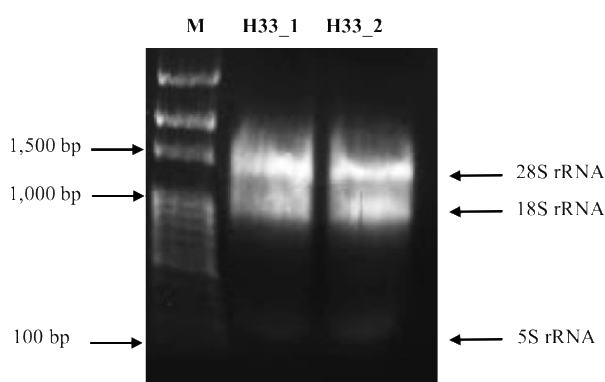
Figure 2.3: Non-denaturing agarose gel electrophoresis of total RNA from LNYV infected H33 samples (subgroup I). Lane M: 100bp ladder (300ng). Arrows indicated the 28S rRNA, 18S rRNA, 5S rRNA and sizes of the ladder.

Table 2.6: The A_{260}/A_{280} and A_{260}/A_{230} ratios of LNYV infected H33 samples.

Samples	A_{260}/A_{280} ratio	A_{260}/A_{230} ratio
H33_1	2.205	2.179
H33_2	2.216	2.192

2.3.3 Optimisation of amplification using subgroup specific, BCNG1/BCNG2 and LNYV_440F/LNYV_1185R primers

One step RT-PCR synthesises cDNA by reverse transcriptase and amplifies the DNA in a single tube, and has been used previously for detection of LNYV (Callaghan and Dietzgen 2005; Higgins et al. 2016b). Previously identified LNYV infected samples were used to optimise RT-PCR amplification using the subgroup specific, BCNG1/BCNG2, LNYV_440F/ LNYV_1185R primers to detect LNYV and subgroups. Figure 2.4 shows the relative location of these primers within the N gene. The negative controls were used to identify the occurrence of contamination and non-specific primer binding in the RT-PCR.
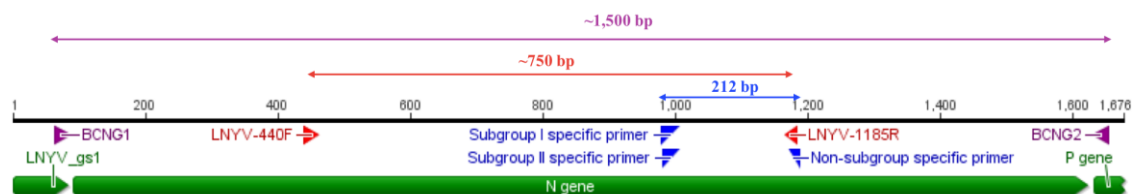
Figure 2.4: The locations of subgroup specific/non-subgroup specific, LNYV_440F/LNYV_1185R and BCNG1/BCNG2 to the LNYV N gene. LNYV_gs1 = 3' leader.

2.3.3.1 Subgroup specific primers

2.3.3.1.1 Optimisation RT-PCR conditions for subgroup I detection

The subgroup I primer combination (that is, non-subgroup specific primer paired with the subgroup I specific primer) was tested for its specificity and ability to amplify the product of the expected size of 212 bp. Known LNYV infected samples were used as templates in one-step RT-PCR with an annealing temperature of 55 °C. Figure 2.5 shows a PCR product of approximately 200 bp was amplified in all LNYV subgroup I, subgroup II and uninfected samples, but not in the no template control (NTC). Amplification with these primers was only expected in the subgroup I samples, namely H27, H28, H29, H30 and H33. This primer combination appears to bind in a manner that is not subgroup specific nor virus specific under these reaction conditions, possibility due to an inappropriate

annealing temperature. To test this, temperature gradient PCR was performed to determine the optimum annealing temperature.
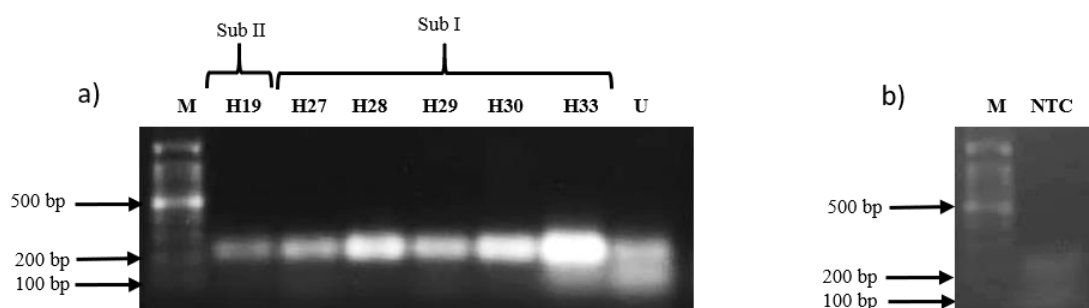
Figure 2.5: Analysis of RT-PCR by agarose gel electrophoresis using subgroup I primer combination with an annealing temperature of 55°C on known infected samples. a) Lane M: 100bp ladder, H19, H27, H28, H29, H30, H33 and Lane U: U (uninfected lettuce). b) Lane M: 100bp ladder and Lane NTC: no template control. Sub: subgroup. Samples were analysed on separate gels on the same day and all RT-PCRs were carried out at the same time.

2.3.3.1.2 Gradient PCR using subgroup I specific primer to optimise primer the annealing temperature

Gradient PCR is used to determine the optimum annealing temperature of primers by using a gradient of temperatures (Kennedy and Oswald 2011). This was carried out for the subgroup I specific primer pair using the subgroup I sample H33, as the template. The annealing temperatures were programmed to be between 55 °C – 65 °C in increments of 2 °C. The primer pair efficiently amplified the expected sized product from the infected H33 sample at all annealing temperatures (Figure 2.6a). The uninfected samples had a very faint DNA band at about 200 bp at all temperatures while the NTC reactions had no product (Figure 2.6b and c). This indicates that this range of temperatures is suitable for this primer pair for subgroup I detection. However, other PCR conditions need to be optimised to reduce unexpected PCR products in uninfected lettuce.
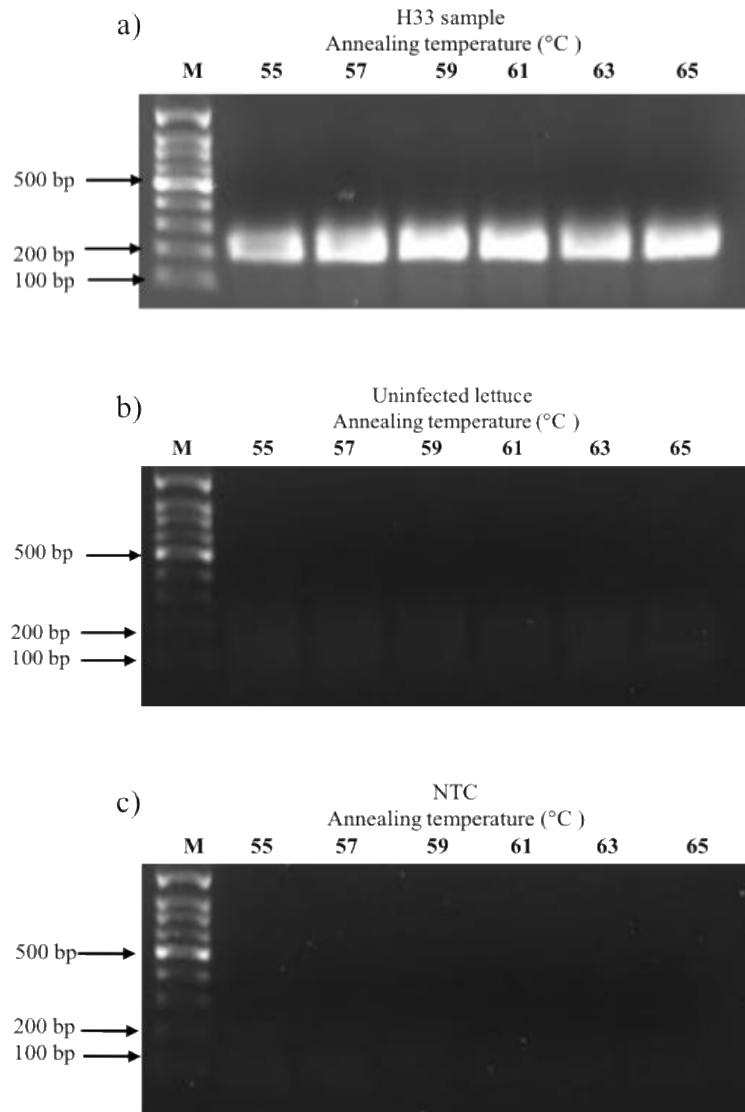
Figure 2.6: Agarose gel electrophoresis of PCR gradient at annealing temperature of 55°C - 65°C using the subgroup I primer combination in H33 sample and negative controls. a) H33 subgroup I sample b) uninfected lettuce and c) NTC. Lane M: 100bp ladder. Samples were analysed on separate gels on the same day and all RT-PCRs were carried out at the same time.

2.3.3.1.3 Optimisation of annealing temperature for subgroup I specific primers

The gradient PCR results from the previous section suggested that 65 °C could be the optimum temperature for the subgroup I primers. Other known LNYV infected samples were tested using RT-PCR with the annealing temperature of 65 °C using this primer combination. Figure 2.7 shows that all LNYV infected samples, including the subgroup II sample, gave a PCR product of ~200 bp. The false positive result could have been due to inappropriate PCR conditions and reagent concentrations. Negative controls had no DNA, indicating there was no contamination in the reaction (Figure 2.7b)
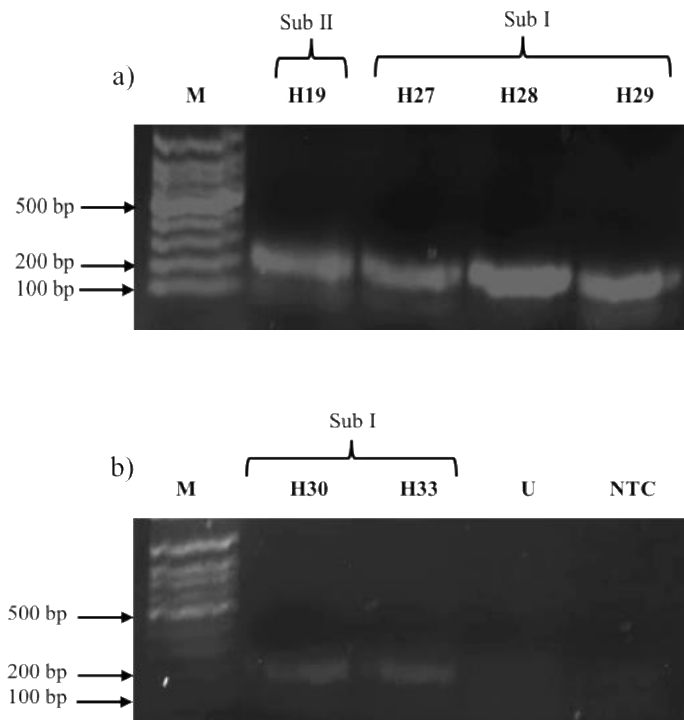
Figure 2.7: Analysis of RT-PCR by agarose gel electrophoresis using the subgroup I primer combination with an annealing temperature of 65°C on known LNYV infected subgroup I and subgroup II samples. a) subgroup I and subgroup II samples. b) subgroup I samples, Lane U: uninfected lettuce and Lane NTC: no template control. Sub: subgroup. Lane M: 100bp ladder. Samples were analysed on separate gels on the same day and all RT-PCRs were carried out at the same time.

2.3.3.1.4 Gradient PCR of subgroup II specific primer to optimise the primer annealing temperature

Gradient PCR was conducted on the subgroup II primer combination (non-subgroup specific primer paired with the subgroup II specific primer) to determine the optimum annealing temperature using the subgroup II sample, H19, as the template. The annealing temperature was between 45 °C and 55 °C in increments of 2 °C because both subgroup I specific primer combination pair and subgroup II specific primer combination pair were carried out at the same time. Figure 2.8a shows that the expected ~200 bp PCR product size was amplified at all annealing temperatures for H19, with most product produced between 51 °C and 55 °C, suggesting any temperature within this range would be suitable. Unexpected faint DNA bands were present in the NTC samples below 51 °C (Figure 2.8b). These bands were absent when the temperature was raised above 51 °C, indicating

that increasing the annealing temperature reduced the non-specific bands. Therefore, the subgroup II primers were then tested with an annealing temperature of 65 °C.
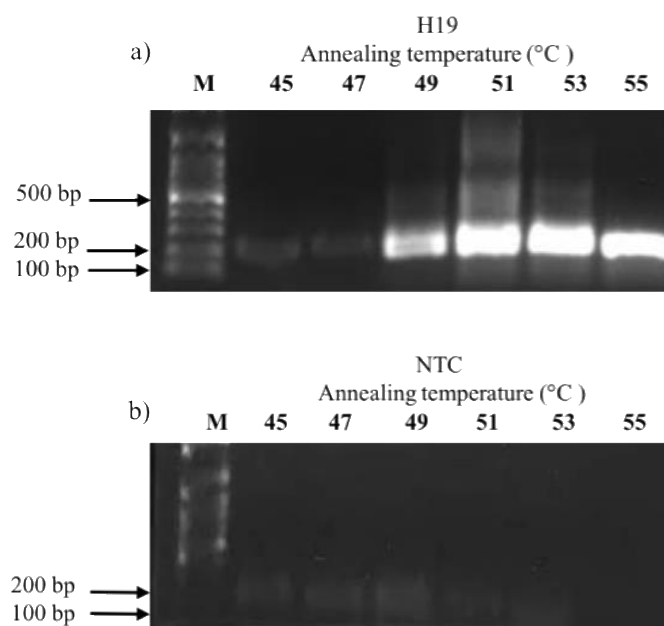
Figure 2.8: PCR gradient of H19 sample and NTC at annealing temperature of 50°C Δ 5°C using the subgroup II primer combination. a) H19 sample b) NTC. Lane M: 100 bp ladder. Samples were analysed on separate gels on the same day and all RT-PCRs were carried out at the same time.

2.3.3.1.5 Optimisation of annealing temperature for subgroup II specific primer

Gradient PCR indicated that the subgroup II primer pair efficiently amplified the ~200 bp product at an annealing temperature of up to 55 °C and high annealing temperatures could reduce the non-specific products in NTC. Figure 2.9a shows amplification of the expected ~200 bp product in the subgroup I H19 sample. There was also faint smearing at approximately ~200 bp and 100 bp in subgroup I samples and the negative controls (Figure 2.9a and b). Bands less than 100 bp were possibly primer dimer. The smearing could have been caused by an inappropriate reaction mixture. The subgroup I primer pair works well at 65 °C, so this temperature was tested for the subgroup II primer pair using both subgroup I and II templates.
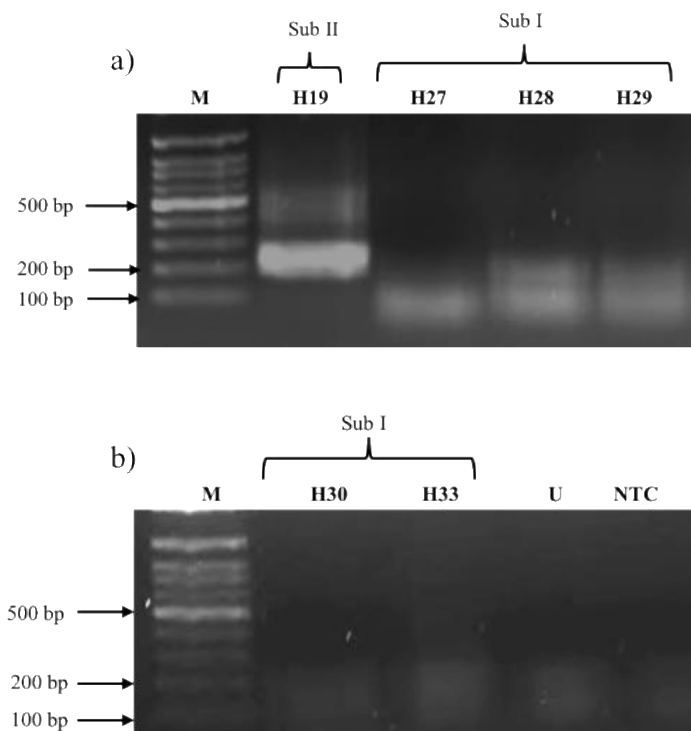
Figure 2.9: Analysis of RT-PCR by agarose gel electrophoresis using the subgroup II primer combination with an annealing temperature of 65°C on known LNYV subgroup I and subgroup II infected samples. a) subgroup I and subgroup II samples. b) subgroup I samples, Lane U: uninfected lettuce and Lane NTC: no template control. Sub: subgroup. Lane M: 100bp ladder. Samples were analysed on separate gels on the same day and all RT-PCRs were carried out at the same time.

2.3.3.1.6 Optimisation of subgroup specific primers using the appropriate kit and conditions

It was recognised that a kit designed for use in RT-qPCR was accidentally used for the PCR gradient experiments with incorrect RT-PCR conditions. Once the error was identified, a comparison of the recommended conditions was done to determine if any differences in reaction conditions might be the causes of the non-specific amplification (Table 2.7). Based on the information provided with the one-step qRT-PCR kit, the 2X reaction mix contained a $MgSO_4$ concentration of 3 mM, while the one-step RT-PCR kit had a $MgSO_4$ concentration of 1.6 mM. The primer concentrations were also too high. Further, reactions with a standard volume of RNA was used as the template rather than a standard amount, so that the amount of RNA between reactions varied. This was standardised to 300 ng per reaction. The temperature and time for the RT-PCR was also found to be suboptimal.

Table 2.7: Summarises conditions modified to optimise the subgroup specific primers.

Parameter	Modifications in RT-PCR	
	Initial incorrect conditions	Final conditions
One step RT-PCR kit	SuperScript™ III Platinum™ One-Step RT-qPCR	SuperScript™ III Platinum™ One-Step RT-PCR
$MgSO_4$ concentration	3 mM	1.6 mM
Primer concentration	2 µl of 10 µM	0.5 µl of 10 µM
Template RNA	1 µl	300 ng
cDNA synthesis	50 °C	55°C
Extension time	1 minute 50 seconds	30 seconds

These correct conditions were tested on LNYV infected subgroup I and subgroup II samples. Amplification with the subgroup I primer combination resulted in the expected PCR product of ~200 bp in all subgroup I samples except the H27 sample (Figure 2.10a). Negative controls had no PCR products (Figure 2.10a). Non-amplification of the product from the H27 sample could have been due to degradation of the template RNA. The H19 subgroup II sample had non-specific PCR products indicating the PCR conditions were still inappropriate (Figure 2.10a). When the modified PCR conditions were tested with the subgroup II primer combination, strong amplification was observed for the H19 subgroup II sample (Figure 2.10b). However, there were also faint products in subgroup I samples indicating further optimisation was also required for these primers (Figure 2.10b).
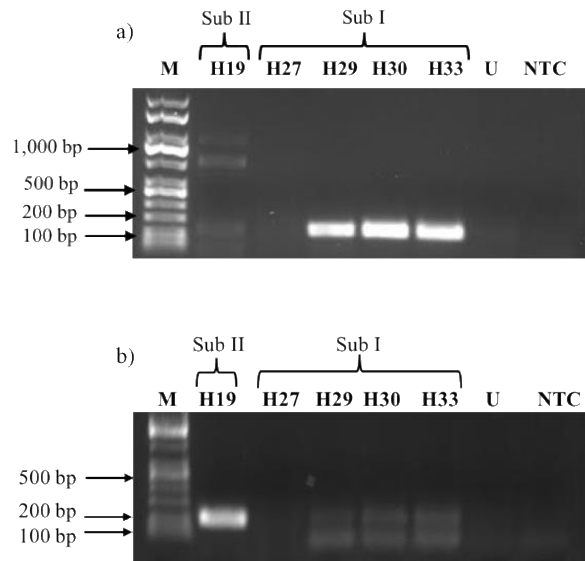
Figure 2.10: Analysis of RT-PCR by agarose gel electrophoresis using the subgroup specific primers with an annealing temperature of 65°C and modified conditions. a) subgroup I primer combination b) subgroup II primer combination. Lane M: 100bp ladder, Lane U: uninfected lettuce and Lane NTC: no template control (negative control). Samples were analysed on separate gels on the same day and all RT-PCRs were carried out at the same time.

2.3.3.1.7 Final optimised RT-PCR conditions for subgroup specific primers

Previous experiments showed the subgroup primers were not subgroup specific under the conditions used. This could have been due to over-amplification of sequences by excessive PCR cycles. Therefore, the number of cycles was reduced from 40 to 30. The H19 (subgroup II) and H29 (subgroup I) samples were tested using the subgroup II primer pair with the modified PCR conditions at an annealing temperature of 65 °C. Figure 2.11a shows that the H19 subgroup II sample had the ~200 bp product as expected, while the subgroup I sample H29 and negative control samples had no product. This demonstrated that there was no contamination in the PCR and reducing the number of cycles with an annealing temperature of 65 °C allowed subgroup II specific amplification. The specificity was further tested with more subgroup I samples (H27, H28 and H33); only the H19 sample produced of the expected product (Figure 2.11b).
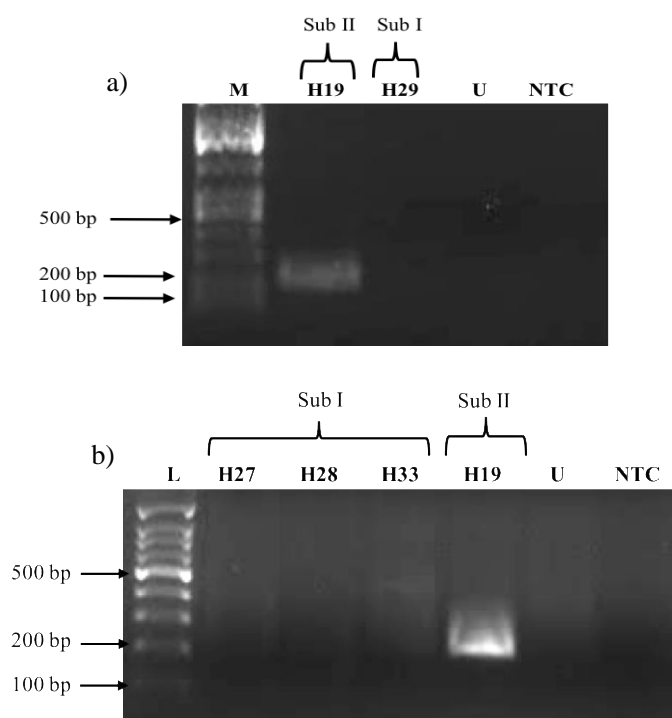
Figure 2.11: Analysis of RT-PCR by agarose gel electrophoresis using subgroup II primer pair combination with annealing temperature of 65°C and 30 number of cycles on known LNYV infected samples. a) H29 subgroup I sample b) H27, H28 and H33 subgroup I samples. Lane M: 100bp ladder, Lane U: uninfected lettuce and Lane NTC: no template control (negative control).

LNYV subgroup I samples were tested using the subgroup I primer pair using the optimised PCR conditions described above. Figure 2.12 shows these conditions allow amplification of the expected product in subgroup I samples with negative controls showing no amplification.

A final test of specificity was carried out on subgroup I and subgroup II primers were finally tested for subgroup specificity. Figure 2.13 shows the subgroup II primer pair amplified only the H19 subgroup II sample, while the subgroup I primers amplified only the H29 subgroup I sample. This shows that the conditions used are optimal for subgroup specific amplification of the LNYV N gene.
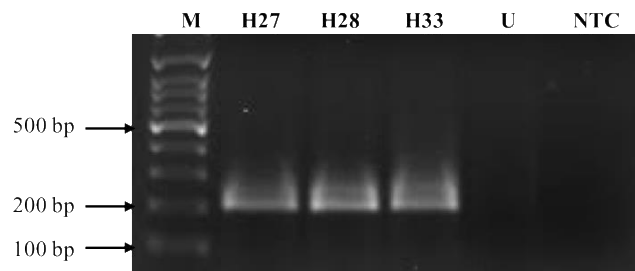
Figure 2.12: Analysis of RT-PCR by agarose gel electrophoresis subgroup I primers with the optimised conditions on known LNYV infected samples. Lane M: 100bp ladder, Lane U: uninfected lettuce and Lane NTC: no template control (negative control).
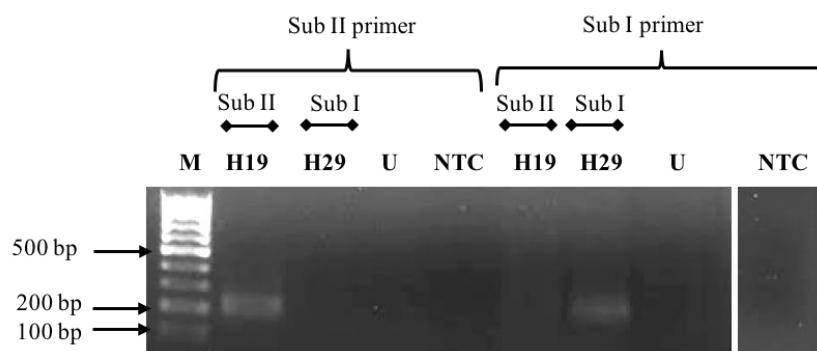
Figure 2.13: Analysis of RT-PCR by agarose gel electrophoresis using the subgroup specific primer pair combinations with the optimised conditions on known LNYV infected samples. Lane M: 100bp ladder.

2.3.3.2   BCNG1 and BCNG2 primers

2.3.3.2.1 Optimisation of RT-PCR conditions for the BCNG1/BCNG2 primers

The BCNG1 and BCNG2 primers amplifies the complete N gene sequence of LNYV from both subgroups giving rise to a PCR product of 1,500 bp. These primers were tested on known LNYV infected samples and negative controls using the conditions described by Callaghan and Dietzgen (2005). All samples had no PCR product, only primer dimer (Figure 2.14). This indicated that these conditions were unsuitable for using these primers within the AUT laboratory. The false negative results could have been due to an inadequate quantity of RNA template because, initially, a standard volume was used for the analysis rather than an appropriate amount. A standard amount of RNA template (250 ng to 500 ng) was used in the subsequent analysis. Furthermore, the kit used by Callaghan

and Dietzgen (2005) was the SuperScript II™ One-step™ RT-PCR, whereas the kit used in the current study was the newer version of the kit SuperScript III™ One-step™ RT-PCR. The extension temperature for the primers was modified from 72 °C to 68 °C according to the instructions provided in the SuperScript III kit. With these modifications, no products were obtained from the infected samples and negative controls (Figure 2.15). Although there was no contamination in the reaction and the PCR had failed to detect LNYV, even after conditions were modified.
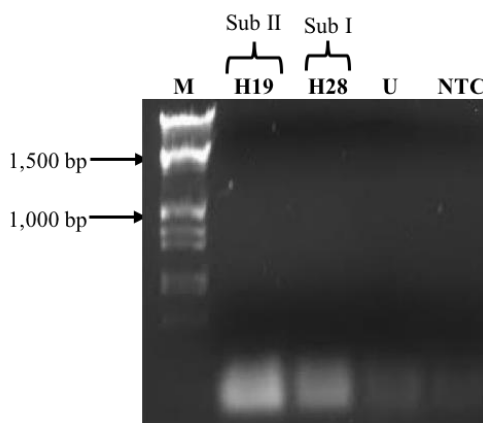
Figure 2.14: Agarose gel electrophoresis of the amplified products from the RT-PCR using the BCNG1/BCNG2 primers on LNYV infected samples. Lane M: Kapa universal ladder, H19, H28, Lane U: uninfected lettuce and Lane NTC: no template control (negative control). Sub: subgroup
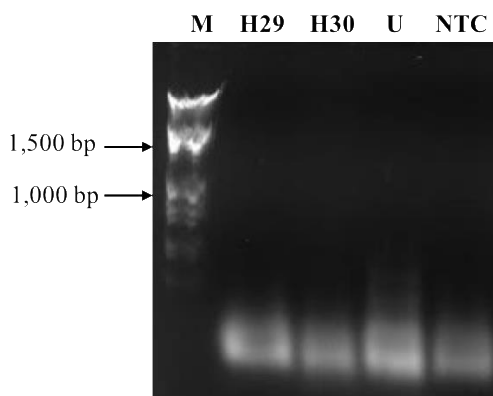
Figure 2.15: Agarose gel electrophoresis of the amplified products from the RT-PCR using the BCNG1/BCNG2 primers with the modified conditions to test LNYV infected samples. Lane M: Kapa universal ladder, Lane U: uninfected lettuce and Lane NTC: no template control (negative control).

2.3.3.2.2 Final optimisation of PCR conditions for the BCNG1/BCNG2 primers

The false negative results from the infected samples could have been due to insufficient PCR cycles, resulting in inconsistent amplification. This may be because the PCR product is reasonably large at 1,500 bp and therefore, may require more cycles to produce detectable levels of product, especially if the viral RNA is present at a low level. Thus, the number of PCR cycles was increased from 35 to 40. This change was tested on an infected LNYV sample (H30) with negative controls and resulted in consistent amplification of the expected 1,500 bp product (Figure 2.16). These conditions allow detection of LNYV using the BCNG1 and BCNG2 primers.
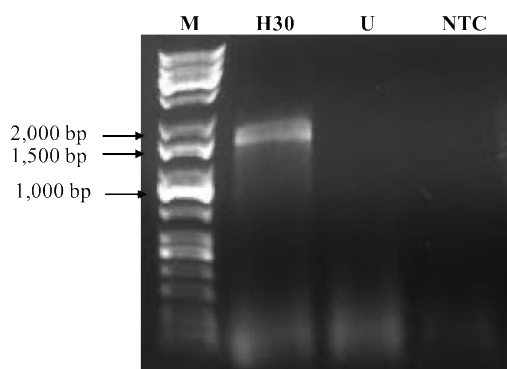
Figure 2.16: Agarose gel electrophoresis of the amplified products from the RT-PCR using the BCNG1/BCNG2 primers with PCR cycles of 40 on LNYV infected samples. Lane M: Kapa universal ladder, Lane U: uninfected lettuce and Lane NTC: no template control (negative control).

2.3.3.3   LNYV_440F/LNYV_1185R primers

2.3.3.3.1 Optimisation of annealing temperature and number of PCR cycles

The LNYV_440F and LNYV_1185R primers amplify a 750 bp PCR product from the LNYV N gene. These primers had been designed by Joe Tang from Ministry of Primary Industries (MPI), based on the primers BCNG3 and BCNG4 designed by Ben Callaghan, University of Queensland (Callaghan 2005). These new primers were of a better design for routine virus diagnosis according to parameters used by MPI. The LNYV_440F and LNYV_1185R primers were tested on previously identified LNYV subgroup I and subgroup II infected samples using an annealing temperature of 50 °C as used previously

(J. Tang, personal communication). Uninfected lettuce and no template control (NTC) were the negative controls. Figure 2.17 shows a product of around 750 bp was amplified from LNYV infected samples and the negative controls. The products in uninfected and NTC are not the same size, likely not from LNYV. They are probably the result of non-specific binding to an unknown template. The false positive result for the negative controls could have been due to the high number of PCR cycles, or contamination. The number of cycles was decreased from 40 to 35 and the experiment was repeated. The infected samples gave a PCR product of approximately 750 bp, the uninfected lettuce sample had a faint band the same size as seen in Figure 2.17, Lane U, while no product was present in the NTC sample (Figure 2.18). The presence of a faint product in the uninfected sample indicated the PCR conditions remain unsuitable for LNYV detection and require further optimisation.
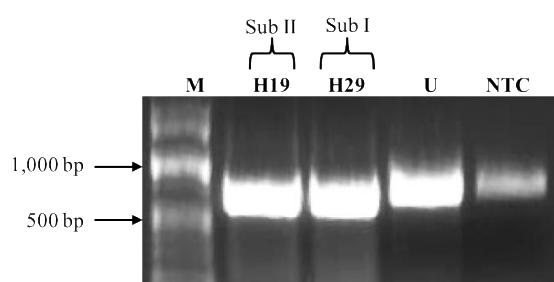
Figure 2.17: Agarose gel electrophoresis analysis of amplified products from LNYV infected and uninfected samples using LNYV_440F and LNYV_1185R primers. Lane M: Kapa universal ladder, Lane U: uninfected lettuce and Lane NTC: no template control (negative control).
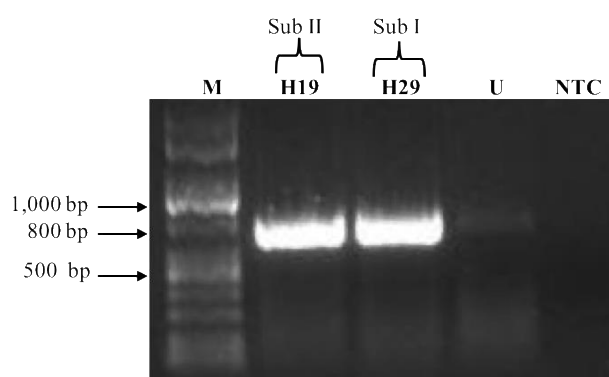
Figure 2.18: Agarose gel electrophoresis analysis of amplified products from LNYV infected and uninfected samples using LNYV_440F and LNYV_1185R primers with a PCR cycle number of 35. Lane M: Kapa universal ladder, Lane U: uninfected lettuce and Lane NTC: no template control (negative control).

2.3.3.3.2 Final optimised PCR conditions for LNYV_440F and LNYV_1185R primers

Amplification of the LNYV N gene using the LNYV_440F and LNYV_1185R primers in a reaction of 40 cycles resulted in false positive results. Reducing the number of PCR cycles to 30, the experiment was repeated. Figure 2.19 shows the expected PCR product in LNYV infected samples H19 and H29 only; false positives were eliminated. Thus, these PCR conditions were suitable for molecular diagnosis of LNYV infection using these primers.
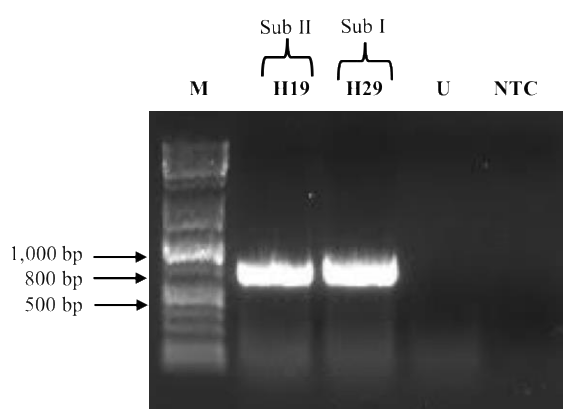
Figure 2.19: Agarose gel electrophoresis analysis of amplified products from LNYV infected and uninfected samples using LNYV_440F and LNYV_1185R primers with PCR cycles of 30 number. Lane M: Kapa universal ladder, Lane U: uninfected lettuce and Lane NTC: no template control (negative control).

2.3.4 Diagnosing LNYV in field collected plants

2.3.4.1 Visual symptoms of plants

Lettuce samples showing symptoms of LNYV were collected from Harrisville, Auckland and Canterbury to confirm the presence of LNYV (Fletcher et al. 2017; Higgins et al. 2016b). Example samples H3, H10, H18 and H35 are shown in Figure 2.20. H3 shows the expected symptoms for a late infection, while H10, H18 and H35 show symptoms of early infection. These samples exhibited chlorosis, bronze and faded green colour on the primary leaves. When compared to the H3 sample, the H10, H18 and H35 samples had

stunted growth. These symptomatic lettuces could indicate LNYV infection and were tested using the different primers conditions described in section 2.3.3.
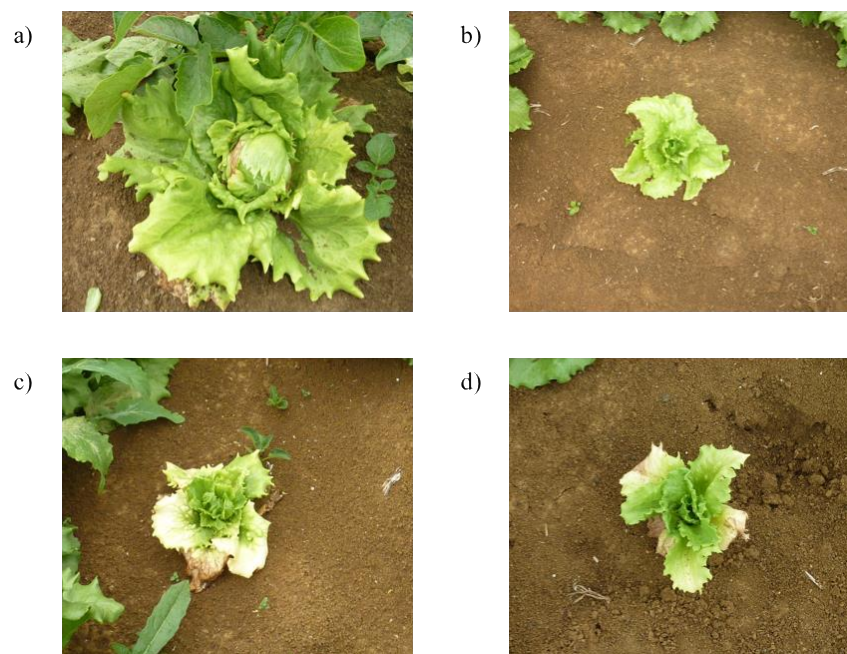
Figure 2.20: Examples of lettuce plants showing symptoms of LNYV infection collected from Harrisville, Auckland. a) H3 b) H10 c) H18 and d) H35 samples. (Photos were provided by Colleen Higgins, Auckland University of Technology).

2.3.4.2 Molecular testing of Harrisville, Tuakau and Canterbury samples for LNYV and subgroups

The LNYV primers described in section 2.3.3 were used to diagnose LNYV infection in plant samples from three locations of New Zealand. These locations were Harrisville (Auckland), Tuakau (Waikato) and Canterbury. Samples were tested using each primer pair to increase the confidence of the results and to confirm the reliability of the primers.

2.3.4.2.1 Molecular testing of 11 Harrisville samples for LNYV using the BCNG1/BCNG2 primers

BCNG1 and BCNG2 primers have been used by others to detect LNYV in plant samples from Australia and New Zealand in one step RT-PCR analysis (Callaghan and Dietzgen 2005; Higgins et al. 2016b). Higgins et al. (2016b) described the identification of six LNYV samples from Harrisville, Auckland; namely NZ1-6, which are the published names of the samples H19, H27, H28, H29, H30 and H33, respectively. Several samples (H2, H3, H5, H6, H8, H9, H10, H11, H12, H14 and H18) from this location had not been tested. These were tested in the study presented here using the BCNG1 and BCNG2 primers with H19 and H29 as the positive control. Uninfected lettuce and NTC were negative controls. H5, H14 and H18 samples tested positive for LNYV with a PCR product of the expected 1,500 bp (Figure 2.21). All other samples and the negative controls contained no PCR products (Figure 2.21). It was concluded that the H5, H14 and H18 were infected by LNYV and other samples were likely to be uninfected. This was surprising since they showed symptoms of LNYV infection, therefore, these samples were retested using LNYV_440F/LNYV_1185R and subgroup specific primers for confirmation. The samples from Tuakau (Waikato) and Canterbury were not tested using the BCNG1 and BCNG2 primers.
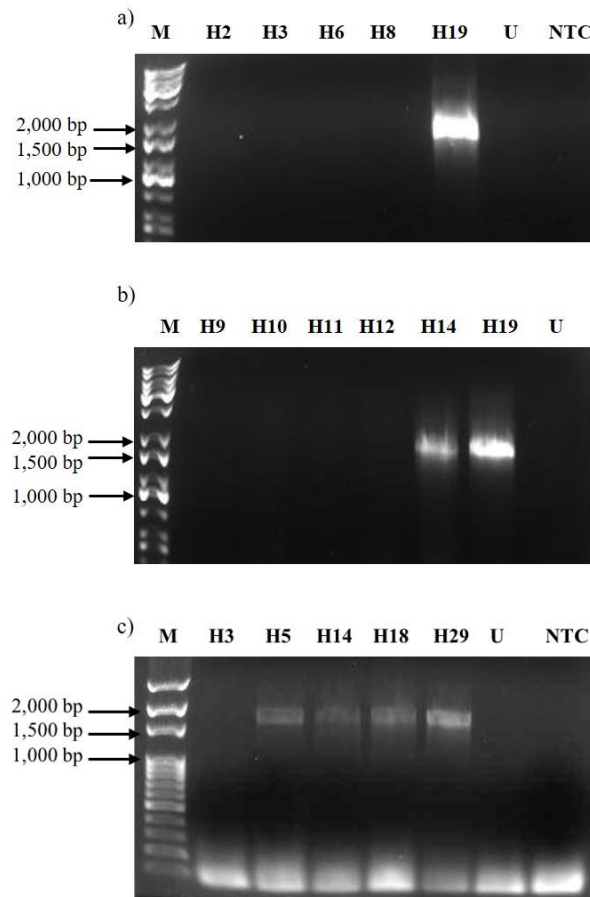
Figure 2.21: The RT-PCR amplification of LNYV with the BCNG1 and BCNG2 primers. a) Lane M: Kapa universal ladder, H2, H3, H6 and H8 samples b) Lane M: Kapa universal ladder, H9, H10, H11, H12 and H14 samples. c) Lane M: 100 bp ladder, H3, H5, H14 and H18. Lane H19: H19 (positive control) and Lane H29: H29 (positive control).

2.3.4.2.2 Testing of Harrisville, Tuakau and Canterbury samples for LNYV using LNYV_440F/LNYV_1185R primers

a) Testing of Harrisville samples

The LNYV_440F and LNYV_1185R primers were used to detect LNYV in the lettuce plants that had tested negative using the BCNG1/BCNGs primers. All samples tested negative with these primers, except for H3. The H29 sample was the positive control. The H3 sample contained a very faint band and the positive samples had a bright band at the expected size of ~750 bp (Figure 2.22). This confirmed that seven of the samples were uninfected by LNYV and the H3 sample was an infected sample. For a final confirmation, some of these samples were retested using the subgroup specific primers.
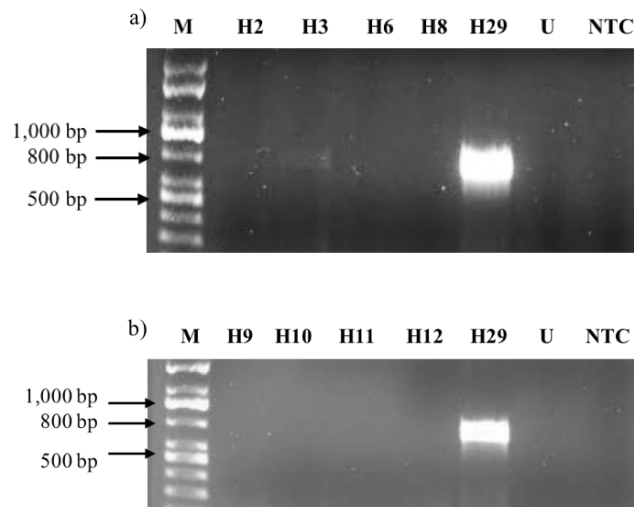
Figure 2.22: Re-testing of negative lettuce samples using the LNYV_440F and LNYV_1185R primers. a) Lane M: Kapa ladder, H2, H3, H6 and H8 samples b) Lane M: Kapa ladder, H9, H10, H11 and H12 samples. Lane H29: H29 (positive control).

The remaining samples from Harrisville (H7, H16, H17, H21, H22, H25, H26 and H35) were tested using the same primers. No products were detected for these samples and negative controls with exception of the H18 sample (Figure 2.23). The H18 sample was shown to be infected with LNYV while the other eight samples appeared to be uninfected. These samples were also tested using the subgroup specific primers for final confirmation.

Based on symptoms in Figure 2.20, it was particularly surprising that H10 and H35 samples had tested negative for LNYV. These were retested along with H5, H14, H18 as well as the positive control H29. PCR products of 750 bp were detected in H5, H14, H18 and the positive control samples but not in H10, H35 and the negative control samples (Figure 2.23c and d). This confirmed that H10 and H35 were not infected with LNYV.
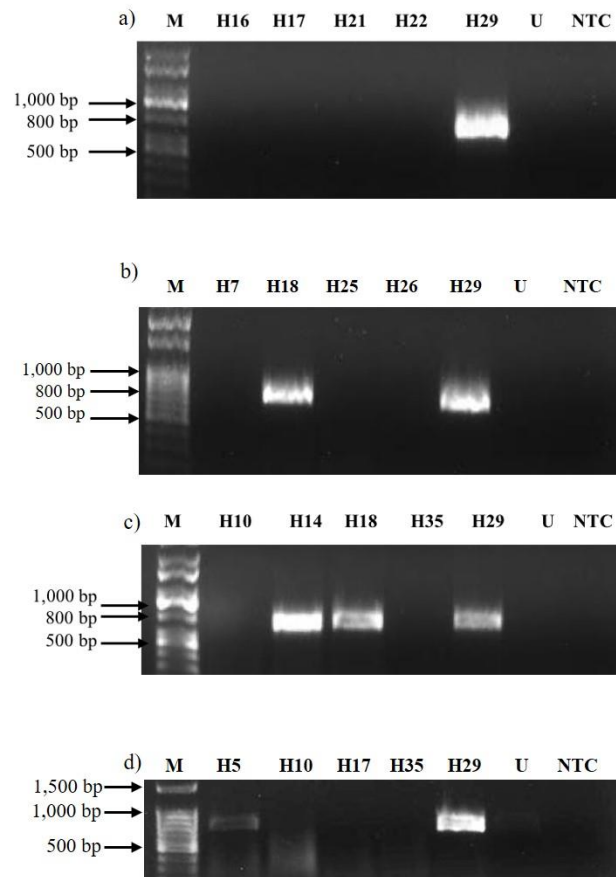
Figure 2.23: Re-testing of negative lettuce samples using the LNYV_440F and LNYV_1185R primers. a) Lane M: Kapa universal ladder, H16, H17, H22 and H22 samples. b) Lane M: Kapa universal ladder, H7, H18, H25 and H26 samples. c) Lane M: 100bp ladder, H10, H14, H18 and H35. d) Lane M: 100bp ladder, H5, H10, H17 and H35. H29: H29 (positive control).

b)   Testing of Tuakau samples

All the samples collected from Tuakau (S28, S29, S30, S31, S32, S33, S34, S35, S36, S37, S39, S40 and S41 samples) were also evaluated by RT-PCR using the LNYV_440F and LNYV_1185R primers. The expected PCR product was undetected in all samples and the negative controls (Figure 2.24). The positive control H19 gave the expected PCR product of 750 bp (Figure 2.24). This indicated that the Tuakau samples were not infected with LNYV. These samples were retested using the subgroup specific primers for the final confirmation.
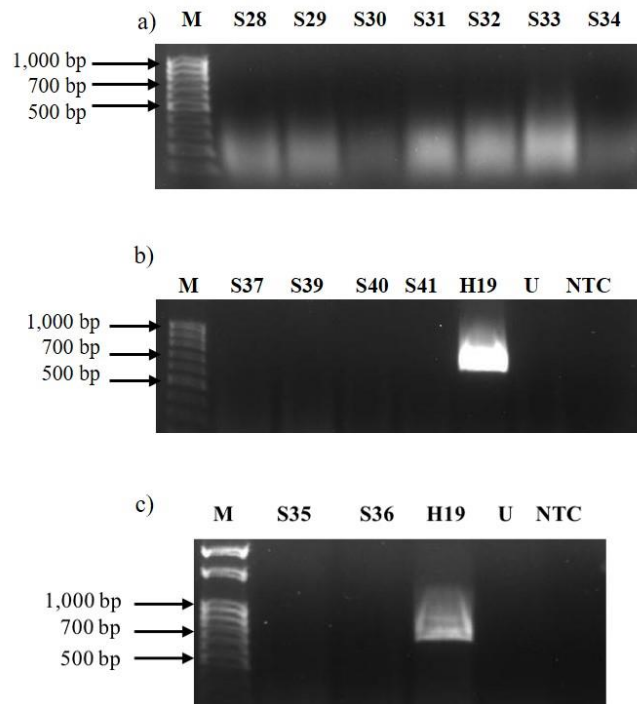
Figure 2.24: The detection of LNYV in Tuakau samples using the LNYV_440F and LNYV_1185R primers. a) S28, S29, S30, S31, S32, S33 and S34 samples b) S35 and S36 c) S37, S39 S40 and S41. Lane M: 100bp ladder and Lane H19: H19 (positive control). Samples were analysed on separate gels on the same day and all RT-PCRs were carried out at the same time.

c)   **Testing of Canterbury samples**

The NZ Institute of Plant and Food Research had conducted virus surveys in Canterbury, New Zealand and seven LNYV infected lettuce samples were identified by ELISA (Fletcher et al. 2017). These samples were sent to AUT for RT-PCR diagnosis and were initially tested with subgroup specific primers (section 2.3.4.2.3c), but four out of the seven samples (S2, S11 and L2) appeared to be LNYV negative. There was low PCR product in K9 sample. Therefore, RNA of these four samples was re-extracted and re-tested using the LNYV_440F and LNYV_1185R primers. Figure 2.25 shows that S11, K9, L2 and the positive control samples had a PCR product of 750 bp. No PCR products were detected in the S2 sample and S11 sample also had non-specific products (Figure 2.25a).
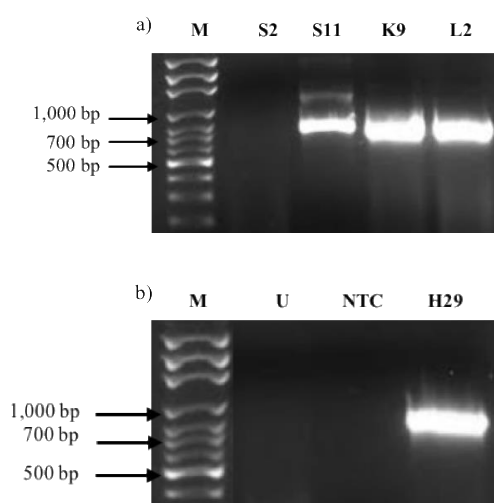
Figure 2.25: Detection of LNYV infected Canterbury samples using LNYV_440F and LNYV_1185R primers. a) S2, S11, K9 and L2 b) Lane U: uninfected lettuce, Lane NTC: NTC and Lane H29: H29 (positive control). Lane M: 100bp ladder. Samples were analysed on separate gels on the same day and all RT-PCRs were carried out at the same time.

2.3.4.2.3 Testing of Harrisville, Tuakau and Canterbury samples using subgroup specific primers to detect LNYV subgroups

a) Testing of Harrisville samples

All the Harrisville samples were tested for LNYV subgroups using the subgroup specific primers for the final validation. The results obtained from the RT-PCR with the subgroup I primers showed the expected product for H3 and H14 samples, these had tested positive previously (Figure 2.26). The other ten samples had no PCR products, as seen previously (sections 2.3.4.2.1 and 2.3.4.2.2a) indicating that these samples were not infected with LNYV subgroup I. Therefore, H3 and H14 were infected with LNYV subgroup I. H5 and H18 samples had tested positive for LNYV with the previously described primer pair, but negative for subgroup I, indicating it was not infected with LNYV of this subgroup.
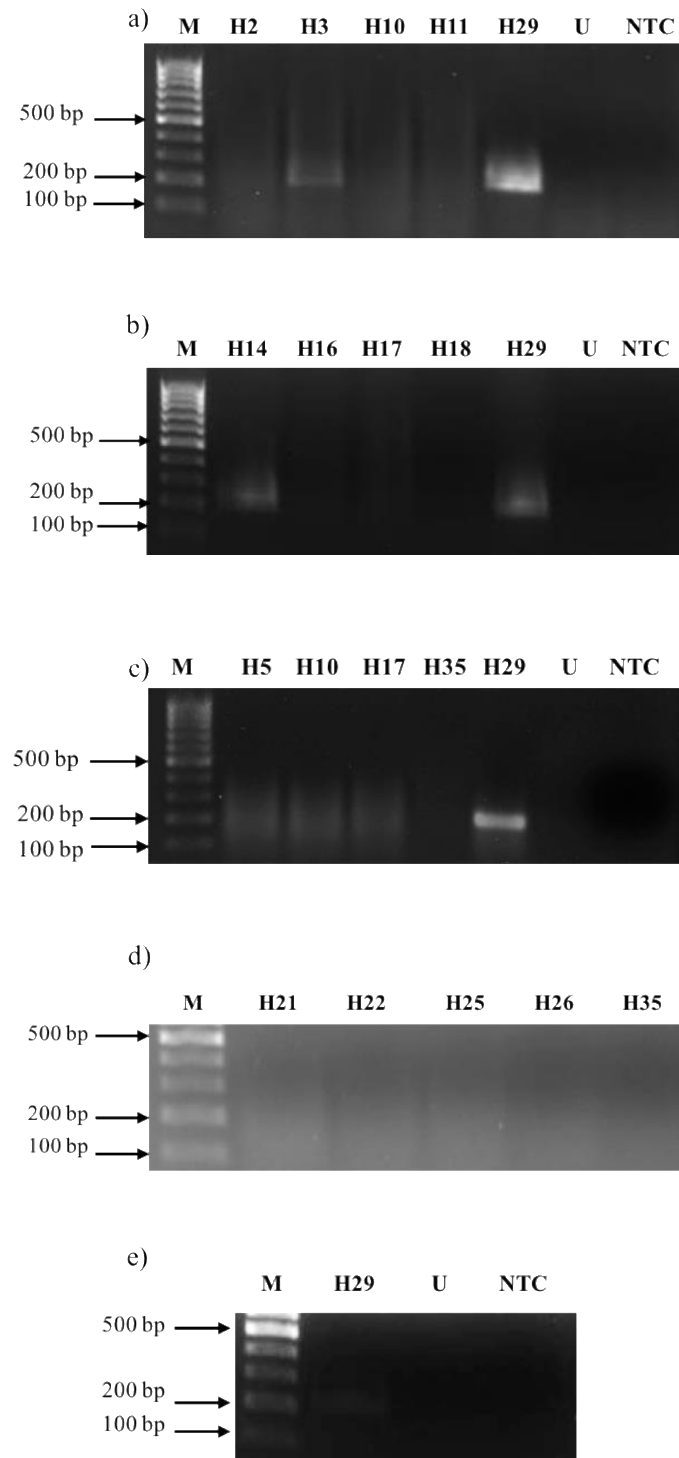
Figure 2.26: The retesting of Harrisville samples (H) using subgroup I primer combination. a) H2, H3, H6 and H8. b) H14, H16, H17 and H18. c) H5, H10, H17 and H35 d) H21, H22, H25, H26 and H35. Lane M: 100bp ladder and Lane H29: H29 (positive control).

The same samples mentioned above were retested for subgroup II using the subgroup II primers. The subgroup II N gene sequence was amplified in the H18 sample but not in any others, including H3 and H14 (Figure 2.27). H3 and H14 samples tested positive for LNYV subgroup I and appeared negative for subgroup II (Figure 2.27a). Thus, H18 was infected with LNYV subgroup II. However, H5 was tested negative for LNYV subgroup II (Figure 2.27c). As mentioned in the previous paragraph, H5 was also tested negative for subgroup I. This could be due to RNA degradation or the subgroup specific primers are not specific to amplify the N gene sequence of H5 sample.
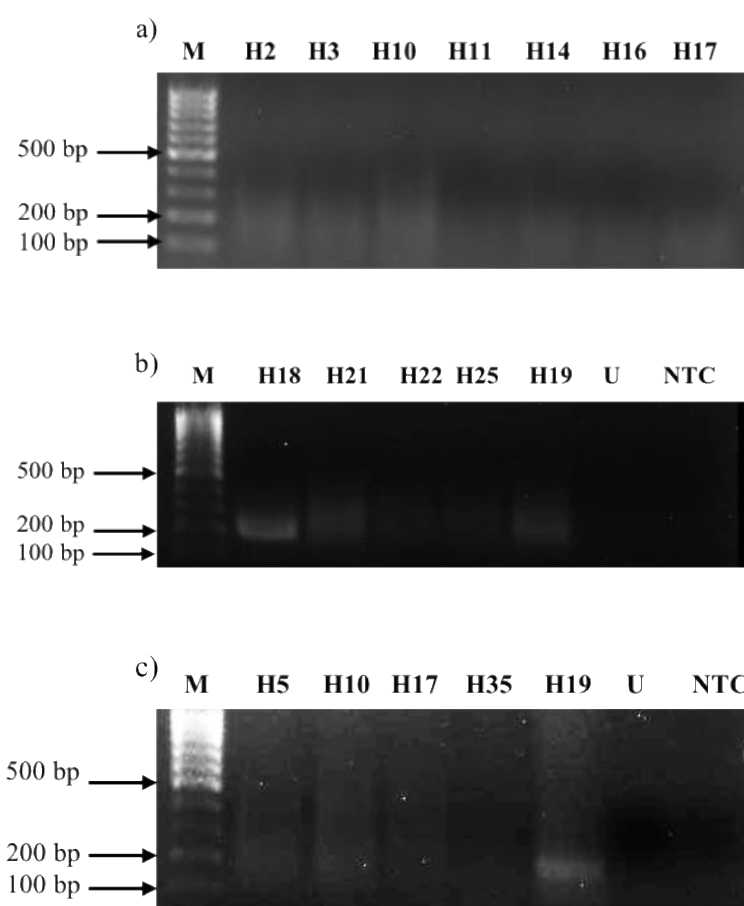
Figure 2.27 The retesting of Harrisville samples (H) using subgroup II primer combination. a) H2, H3, H10, H11, H14, H16 and H17 b) H18, H21, H22 and H25. c) H5, H10, H17 and H35 Lane M: 100bp ladder and Lane H19: H19 (positive control).

b)   Testing of Tuakau samples

The Tuakau samples were also retested for LNYV subgroups. Both subgroup analyses gave no PCR products at the expected size of ~200 bp, supporting the finding described in section 2.3.4.2.2b (Figures 2.28 and 2.29). Thus, these plants were not infected with LNYV.
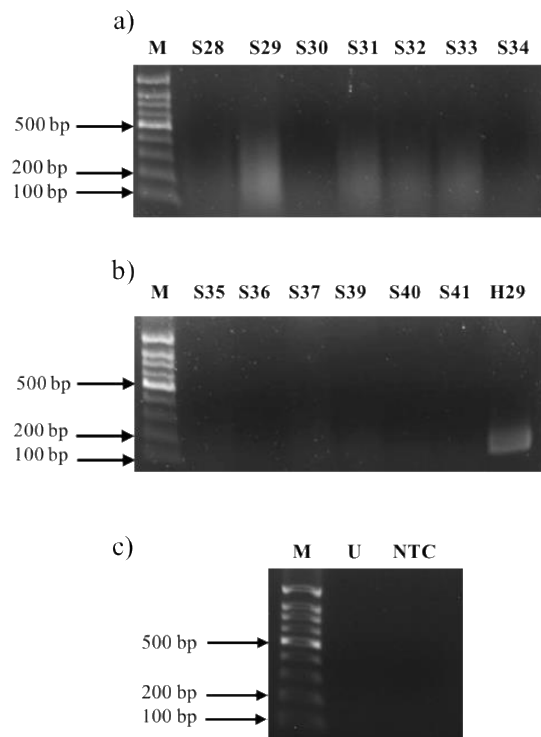
Figure 2.28: The retesting of Tuakau samples (S) using subgroup I primer combination. a) S28, S29, S30, S31, S32, S33 and S34. b) S35, S36, S37, S39, S40 and S41. Lane M: 100bp ladder and Lane H29: H29 (positive). Samples were analysed on separate gels on the same day and all RT-PCRs were carried out at the same time.
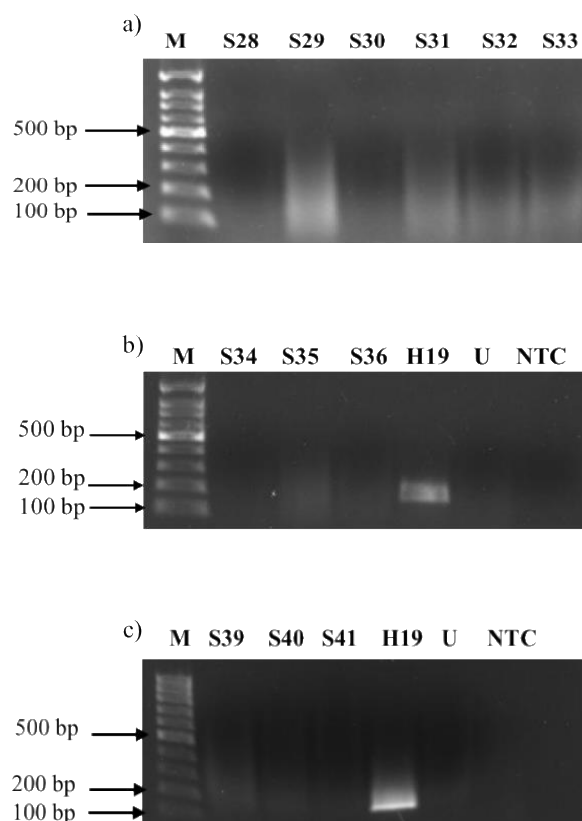
Figure 2.29: The retesting of Tuakau samples using subgroup II primer combination. a) S28, S29, S30, S31, S32 and S33. b) S34, S35 and S36. c) S39, S40 and S41. Lane M: 100bp ladder and Lane H19: H19 sample (positive control).

c) Testing of Canterbury samples

Samples from Canterbury, New Zealand that tested positive for LNYV by ELISA (Fletcher et al. 2017) and were also tested with subgroup specific primers in the present study. Figure 2.30 show that S13, S14, K9 and W4 samples were subgroup I or subgroup II positive using the subgroup I and subgroup II primers. The S2, S11, and L2 samples were subgroup I and subgroup II negative (Figure 2.30). The RNA of these three samples and K9 were re-extracted and tested with LNYV_440F/LNYV_1185R primers in section 2.3.4.2.2.c. S11, K9 and L2 samples were positive with those primers and were retested using each subgroup specific primer pairs to identify the subgroup in all of the samples. Figure 2.31 show that K9 sample was infected with subgroup I while the remaining samples were subgroup I negative using the subgroup I primer pair. The S11, S13, S14, W4 and L2 samples were subgroup II positive and K9 was subgroup II negative using the subgroup II primer pair (Figure 2.32).
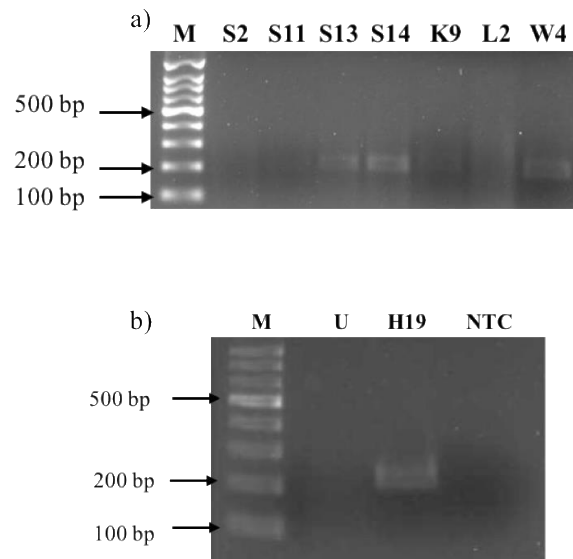
Figure 2.30: The retesting of Canterbury samples using non-subgroup specific subgroup I and subgroup II primers. a) Lane S2: Scott2, Lane S11: Scott11, Lane S13: Scott13, Lane S14: Scott14, Lane K9: Kiesanowski9, Lane W4: Watson4 and Lane L2: Leaderband2. Lane M: 100bp ladder and Lane H19: H19 sample (positive control). Samples were analysed on separate gels on the same day and all RT-PCRs were carried out at the same time.
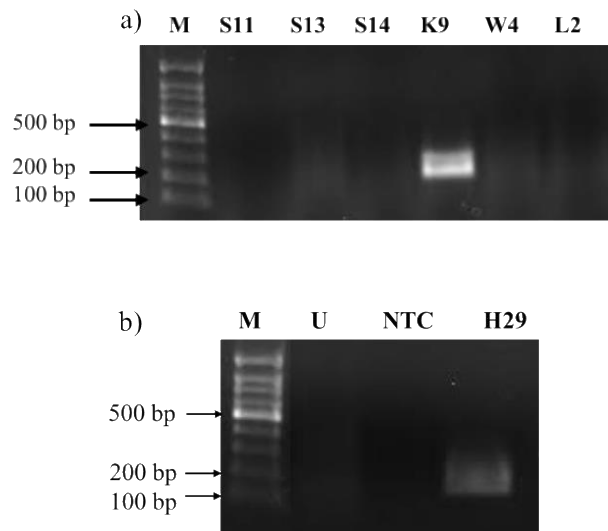
Figure 2.31: The retesting of Canterbury samples using subgroup I primer combination. a) Lane S11: Scott11, Lane S13: Scott13, Lane S14: Scott14, Lane K9: Kiesanowski9, Lane W4: Watson4 and Lane L2: Leaderband2. Lane M: 100bp ladder and Lane H29: H29 sample (positive control). Samples were analysed on separate gels on the same day and all RT-PCRs were carried out at the same time.
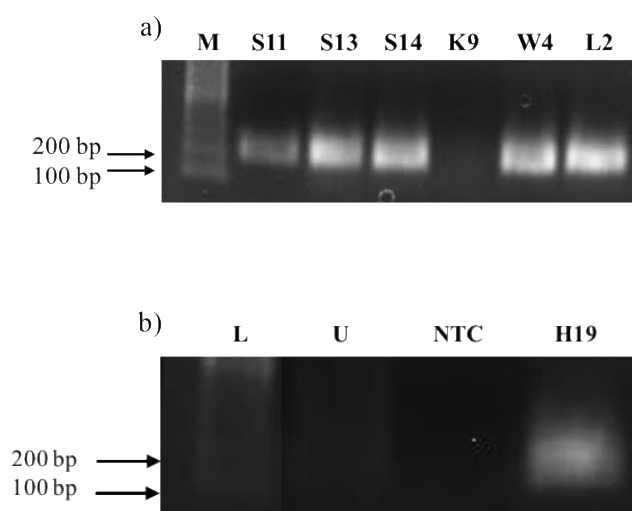
Figure 2.32: The retesting of Canterbury samples using subgroup II primer combination. a) Lane S11: Scott11, Lane S13: Scott13, Lane S14: Scott14, Lane K9: Kiesanowski9, Lane W4: Watson4 and Lane L2: Leaderband2. Lane M: 100bp ladder and Lane H19: H19 sample (positive control). Samples were analysed on separate gels on the same day and all RT-PCRs were carried out at the same time.

2.3.4.3 Summary of LNYV positive samples from Harrisville and Canterbury

The Harrisville and Canterbury samples that had tested positive for LNYV and its subgroups using the BCNG1/BCNG2, LNYV_440F/LNYV_1185R and subgroup specific primer pairs are summarised in Table 2.8. The results showed that H3 had tested LNYV negative using the BCNG1/BCNG2 primers, while the sample was LNYV positive using the LNYV_440F/LNYV_1185R and subgroup I primer pairs. This result indicated that a false negative result was obtained using the BCNG1/BCNG2 primers. Hence, some of Harrisville samples, all Tuakau and Canterbury samples were not tested using the BCNG1/BCNG2 primers. H14, H18 and all Canterbury samples were tested positive by the LNYV_440F/LNYV_1185R and subgroup specific primer pairs except S2. H5 was tested positive by the BCNG1/BCNG2 primers and LNYV_440F/LNYV_1185R primer while this sample was tested negative by the subgroup specific primer pairs. H3, H14 and K9 were infected with LNYV subgroup I, while H18, L2, W4, S11, S13 and S14 were infected with LNYV subgroup II. The subgroup of H5 is unknown.

Table 2.8: Summary of LNYV infected sample results using various primer pairs.

Location	LNYV infected samples	Primers			
		BCNG1 and BCNG2	LNYV_440F and LNYV_1185R	Subgroup I	Subgroup II
Harrisville#, Auckland	H2	-	-	-	-
	H3	-	+	+	-
	H5	+	+	-	-
	H6	-	-	NT	NT
	H7	NT	-	NT	NT
	H8	-	-	NT	NT
	H9	-	-	NT	NT
	H10	-	-	-	-
	H11	-	-	-	-
	H12	-	-	NT	NT
	H14	+	+	+	-
	H16	NT	-	-	-
	H17	NT	-	-	-
	H18	+	+	-	+
	H21	NT	-	-	-
	H22	NT	-	-	-
	H25	NT	-	-	-
	H26	NT	-	-	NT
	H35	NT	-	-	NT
Tuakau#, Waikato	S28	NT	-	-	-
	S29	NT	-	-	-
	S30	NT	-	-	-
	S31	NT	-	-	-
	S32	NT	-	-	-
	S33	NT	-	-	-
	S34	NT	-	-	-
	S35	NT	-	-	-
	S36	NT	-	-	-
	S37	NT	-	-	-
	S39	NT	-	-	-
	S40	NT	-	-	-
	S41	NT	-	-	-
Cherstey*, Canterbury	L2	NT	+	-	+
Marshland*, Canterbury	K9	NT	+	+	-
	W4	NT	NT	-	+
Southbridge*, Canterbury	S2	NT	-	-	-
	S11	NT	+	-	+
	S13	NT	NT	-	+
	S14	NT	NT	-	+

+ LNYV detected   - LNYV non-detected   NT: not tested
#Samples collected by (C. Higgins, AUT)
*Samples collected and published (Fletcher et al. 2017)

2.3.4.4 Distribution of LNYV subgroups in New Zealand

LNYV subgroup distribution in New Zealand using the available data and results obtained from the current study is shown in Figure 2.33. It shows that LNYV infected lettuce samples collected from Harrisville was mostly infected by subgroup I than subgroup II, while there was a higher number of subgroup II infected lettuce samples rather than subgroup I in Canterbury. Since the sample size of 16 and sampled locations of four are small, more sampling is required from North Island and South Island to obtain a better understanding of LNYV subgroup distribution in New Zealand.
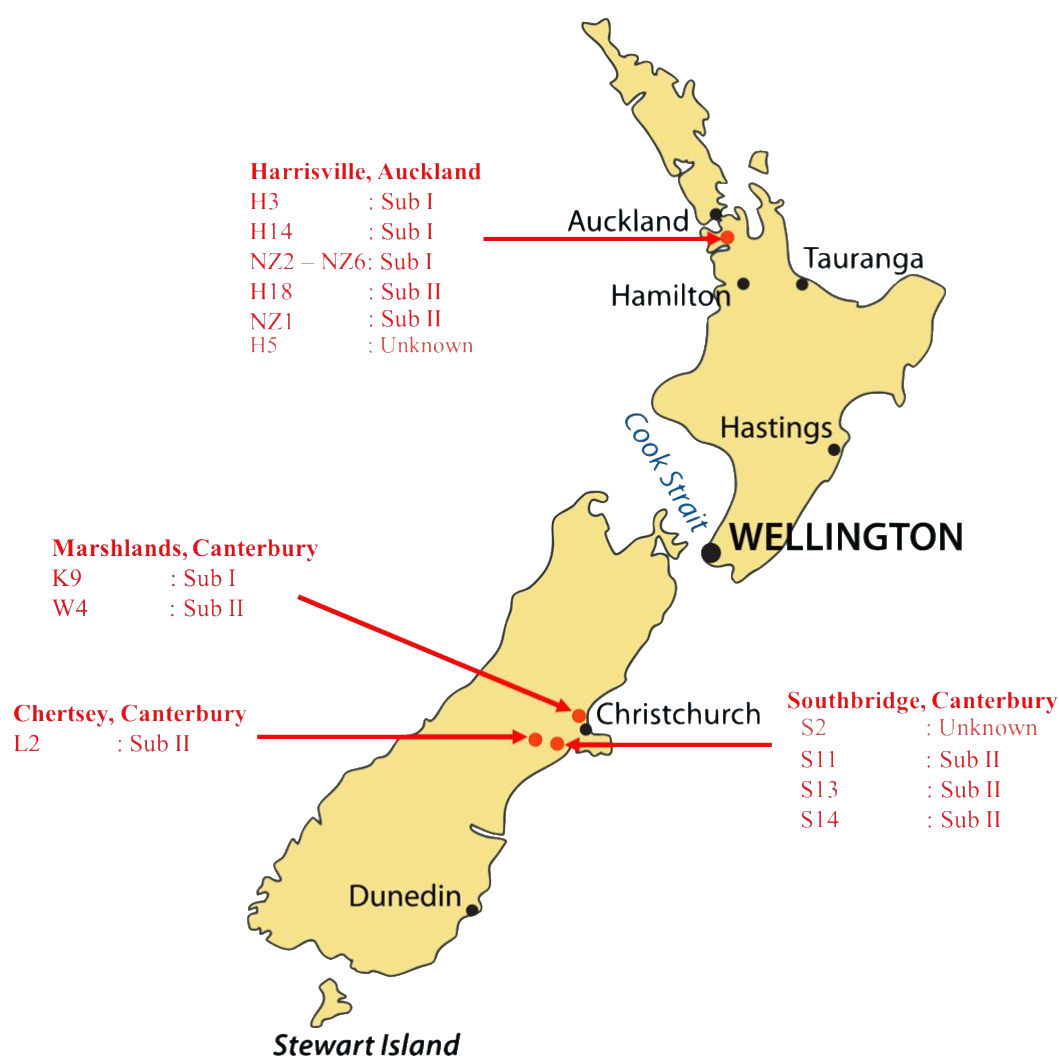
Figure 2.33: Distribution of LNYV infected lettuce samples collected from Harrisville, Auckland, Cherstey, Marshland and Southbridge. ●: New Zealand cities. ●: Sampled locations

Image: http://newzealandmap.facts.co/newzealandmapof/NewZealandMapwithCities.png (modified)

2.3.4.5 Detection of LNYV using the TPS extraction method

All the previous diagnosis was carried out on samples where purified RNA had been extracted. For high throughput diagnosis, this extraction procedure is laborious and expensive; a method for virus detection using crude plant extract would be valuable for diagnosis. A previous report described the use of a crude preparation for detection of various viruses, including LNYV (Thomson and Dietzgen 1995). LNYV was detected in a template prepared by incubation at 60 °C and 95 °C in TPS1 solution, which contained 100 mM Tris-HCl, 1.0 M KCl and 10 mM EDTA, pH of 8.42. There was no major difference in detection using either temperatures (Thomson and Dietzgen 1995). This method was evaluated for LNYV detection from frozen lettuce, fresh and frozen *N. glutinosa* using the different primer pairs.

2.3.4.5.1 Testing subgroup I sample extracted with TPS1 using the subgroup I primer combination

Using 100 µl of TPS1 solution, 100 mg LNYV infected H29 (subgroup I) leaf sample was ground and incubated at 95 ˚C. Undiluted samples and samples diluted 1/10 in Millipore water as described in Thomson and Dietzgen (1995) were used to carry out one step RT-PCR using either 1 µl or 0.5 µl of sample. The negative controls consisted of undiluted uninfected lettuce and no template control (NTC) samples. RT-PCR with subgroup I specific primer gave no product (Figure 2.34a). This false negative result may have been due to the high incubation temperature of 95 ˚C, which may have degraded the RNA or due to the high volume of TPS1 solution, which could have resulted in a too high KCl concentration. Consequently, LNYV infected *N. glutinosa* was retested using an incubation temperature of 60 °C to detect LNYV. Also, the TPS1 volume used for extraction was reduced to 30 µl and 50 µl, which would decrease the KCl amount. Ground templates were again tested as diluted and undiluted solutions. Once again, no PCR product was amplified using subgroup I specific primers (Figure 2.34b). Further testing was done using the LNYV_440F/LNYV_1185R primers.
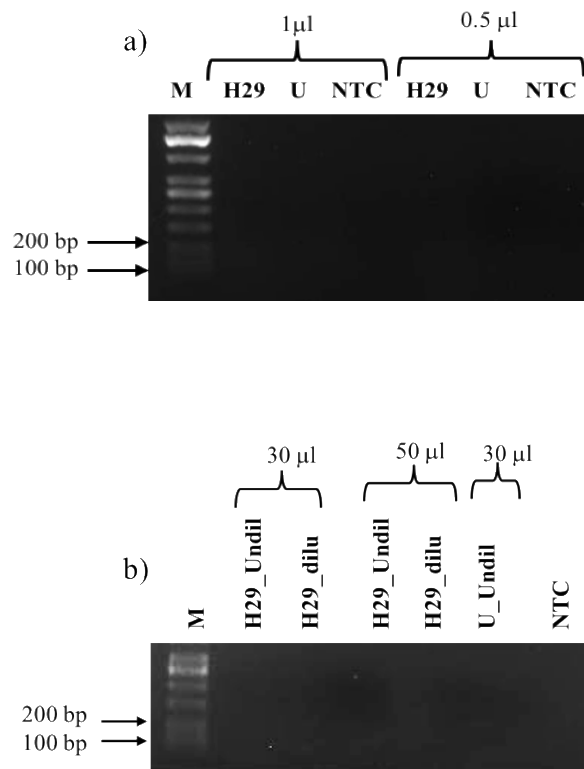
Figure 2.34: Detection of LNYV infected subgroup sample with TPS1 solution using the subgroup I primer combination. a) Lane H29_1μl: 1μl of H29 (subgroup I), Lane U_1μl: 1μl of uninfected, Lane H29_0.5μl: 0.5μl of H29 and Lane U_0.5μl: 0.5μl of uninfected lettuce sample. b) Lane H29_Undil: undiluted H29 (subgroup I), Lane H29_dilu: diluted H29 and Lane U_undil: undiluted uninfected lettuce. Lane M: Kapa universal ladder. Lane NTC: no template control (negative control)

2.3.4.5.2 Testing LNYY infected sample extracted with TPS using the LNYV_440F/ LNYV_1185R primers

The LNYV infected sample extracted with 30 μl of TPS1 at an incubation temperature of 60 °C was retested using the LNYV_440F and LNYV_1185R primers. The infected sample and negative controls gave no PCR product (Figure 2.35). This confirmed that the false negative seen with the subgroup I specific primers was not due to the primers. Due to lack of time, no further testing of this extraction procedure was carried out.
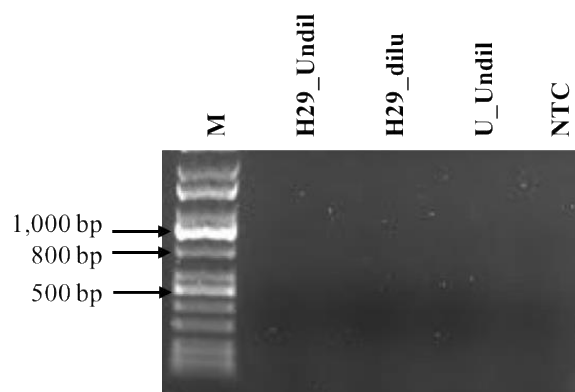
Figure 2.35: The retesting of LNYV infected sample in 30μl of TPS1 using the LNYV_440F and LNYV_1185R primers. Lane M: 100bp ladder, Lane H29_Undil: undiluted H29, Lane H29_dilu: diluted H29 and Lane U_undil: undiluted uninfected lettuce.

2.3.5   Detection of LNYV subgroups using the RT-PCR-RFLP

LNYV subgroups can be diagnosed by endpoint RT-PCR but both subgroups cannot be diagnosed simultaneously because subgroup I and subgroup II primer combinations amplify similar sized PCR product of ~200 bp. Therefore, RT-PCR-RFLP was developed to diagnose the subgroups simultaneously.

To use RT-PCR-RFLP to determine the LNYV subgroup infecting the plants, the LNYV N gene sequence had to be amplified in the presence of subgroup I, subgroup II and non-subgroup specific primers in a single reaction. The optimised conditions described in section 2.2.5.1c. were tested on infected samples to determine if they would give rise to a produce when all three primers are present. Both known positive samples of subgroup I (H29) and subgroup II (H19) had the expected PCR product of ~200 bp (Figure 2.36). The negative controls had no PCR products indicating the conditions were suitable for combining all subgroup primers, and did not require additional optimisation.
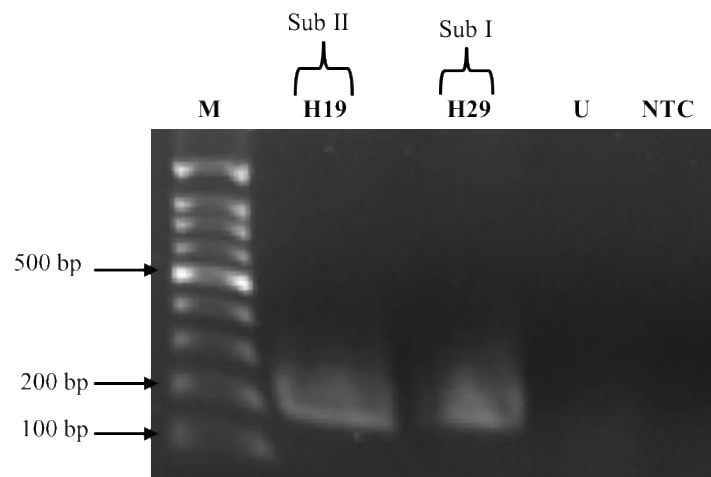
Figure 2.36: Testing of LNYV infected subgroup I and subgroup II samples in a multiplex RT-PCR using subgroup specific primers. Lane M = 100bp ladder. Sub: subgroup.

For diagnosis of the LNYV subgroup using the primers as designed, two RT-PCR would have to be run for each sample. An alternative approach is to use all three primers in a single PCR and identify the subgroup by a subgroup specific restriction fragment length polymorphism. *Mae*III restriction enzyme was identified as a suitable enzyme to cleave the 212 bp PCR product sequence, expected to cleave in a subgroup specific manner. Subgroup I isolates were expected to generate three different fragment sizes of 14 bp, 67 bp and 131 bp, while subgroup II isolates would produce two fragments of 64 bp and 145 bp (Figure 2.37).
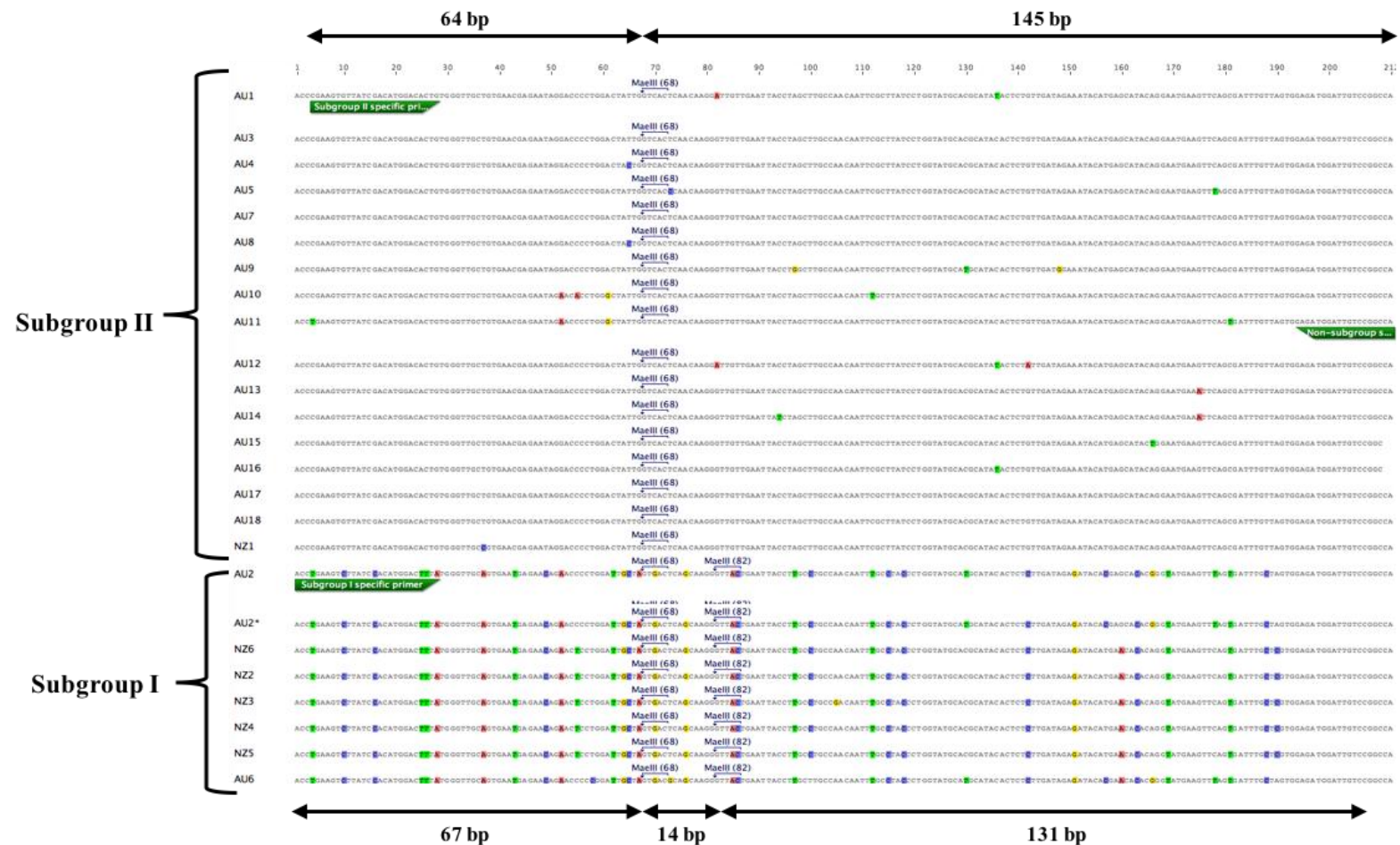
Figure 2.37: The restriction sites of *Mae*III enzyme within the subgroup I and subgroup II specific PCR products. Colours within the alignment indicate where the nucleotides differ between LNYV isolates. Numbers at top the figure indicate nucleotide position. Green: primers

Previously identified LNYV subgroup I and subgroup II samples were amplified using the subgroup I and subgroup II primers in a one step RT-PCR as described above. These PCR products were digested with *Mae*III restriction enzyme. The results show, as expected, that subgroup II & subgroup I gave two fragments, while samples without the enzyme had one DNA band at ~200 bp (Figure 2.38). For subgroup I, three fragments were expected, but in the gel system used, only the 131 bp and 67 bp fragments were visible. The 14 bp is likely too faint to see or has run off the gel. Regardless, the fragments sizes are expected and different from that observed for subgroup II. Thus, each subgroup gave the expected sized fragments based on the relative band locations in the gel. Therefore, using the three primers for RT-PCR, followed by RFLP analysis using *Mae*III can be used to diagnose LNYV and the subgroup.
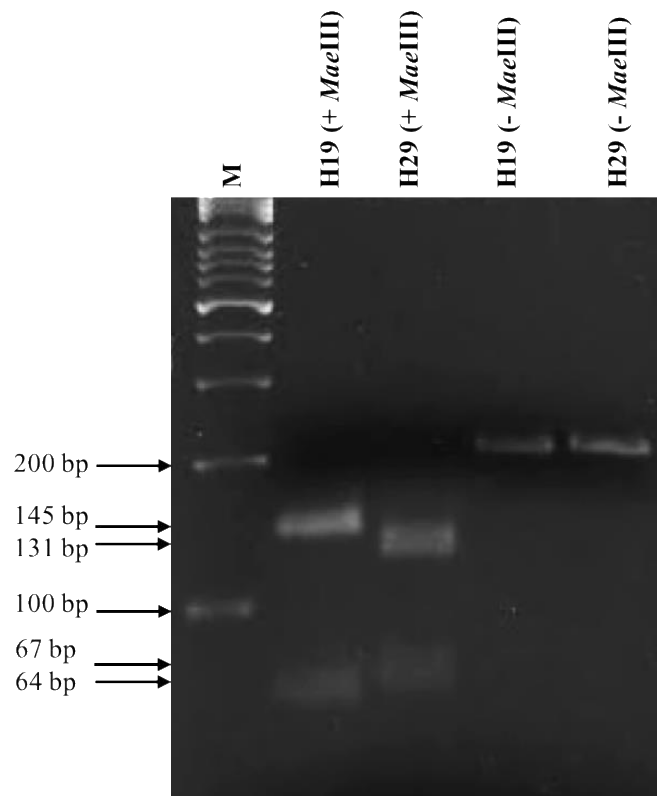
Figure 2.38: RT-PCR-RFLP analysis of known LNYV infected samples in a 4% agarose gel electrophoresis using the restriction enzyme *Mae*III. Lane M: 100bp ladder, Lane H19 (+ *Mae*III): H19 with *Mae*III, Lane H29 (+ *Mae*III): H29 with *Mae*III, Lane H19 (- *Mae*III): H19 without *Mae*III and Lane H29 (- *Mae*III): H29 without *Mae*III.

2.4 Discussion

Distinctive symptoms are the initial indication of LNYV in infected plants (Dietzgen et al. 2007). LNYV and TSWV disease symptoms on infected lettuce are often indistinguishable, while LNYV infection on *Sonchus* sp. is asymptomatic (Dietzgen et al. 2007). This hindrance increased the requirement to develop alternative diagnostic tests that included ELISA and RT-PCR for LNYV (Dietzgen et al. 2007). The N gene is primarily used as a target gene for RT-PCR diagnostics because infected tissues contained N gene transcript in high amounts compared to the other LNYV genes due to the expression gradient (Callaghan and Dietzgen 2005; Higgins et al. 2016b).

LNYV population is made up of two subgroups. Sequencing of the N gene and phylogenetic comparisons with the available LNYV N gene sequences has been required to identify the subgroup of any given isolate (Callaghan & Dietzgen, 2005; Higgins et al., 2016). Rapid and efficient molecular diagnosis of subgroups is undeveloped, hitherto. Hence, one of the aims of this study was to develop diagnostic tests for LNYV and its subgroups by RT-PCR and RT-PCR-RFLP analysis. This was accomplished by designing subgroup specific primers and optimising the conditions for their use. The conditions for the previously designed primers (BCNG1/BCNG2 and LNYV_440F/LNYV_1185R) used to detect LNYV were also re-optimised.

Primer design

Primer design is one of the fundamental requirements that govern the efficiency and specificity of PCR analysis (Dieffenbach et al. 1993). Subgroup specific primers had been designed by previous researchers to detect various plant virus subgroups for viruses such as grapevine-infecting nepoviruses and cucumber mosaic virus (CMV) (Bald-Blume et al. 2017; Digiaro et al. 2007). Various subgroup specific and non-subgroup specific primers were designed in the current study to detect LNYV subgroups by RT-PCR by aligning the N gene sequences using the MUSCLE alignment. The alignment clearly distinguished the two subgroups. The 5'untranslated region (UTR), open reading frame (ORF) and 3'UTR regions were identified in the alignment from previously published data (Dietzgen et al. 2007). As found previously, the alignment in Figure 2.1 shows the 5'UTR and ORF region are more conserved than the 3'UTR (Callaghan and Dietzgen 2005; Higgins et al. 2016b). The 5'UTR and ORFs were particularly examined for the current primer design. Ideally, subgroup specific primers would amplify target gene

sequences giving PCR products of different sizes for each subgroup, so they can be easily distinguishable by electrophoresis. For instance, the nepovirus subgroups (A, B and C) were detected using RT-PCR with three different primers that amplified subgroup A (255 bp), subgroup B (390 bp) and subgroup C (640 bp) (Digiaro et al. 2007). This analysis enabled the detection of subgroups simultaneously and separation of band sizes for each subgroup helped to discriminate between the subgroups (Digiaro et al. 2007). Similar PCR size separation for each LNYV subgroup was the aim when subgroup specific primers were being developed; however, this proved difficult.

Extensive analysis of LNYV N gene sequences indicated that there was low sequence heterogeneity between the subgroups, thus, it was not possible to design two unique primers with two different PCR product sizes for each subgroup. Therefore, six sets of subgroup I and subgroup II primers and three non-subgroup specific (binds to both subgroups) primers were designed. This phenomenon of low sequence heterogeneity was also observed in the L gene of the cytorhabdovirus SCV subgroups (group I and group II) (Klerks et al. 2004). Simultaneous detection of each LNYV subgroups by exclusive use of endpoint RT-PCR analysis was not possible using the approach described in the current study. If using endpoint PCR with electrophoresis, each subgroup must be tested for independently. An alternative approach, as shown here is RT-PCR combined with RFLP. Another alternative approach using the subgroup specific primers is to carry out a RT-qPCR following by melting temperature analysis.

Further testing was carried out to determine the optimum primer set. RT-PCR consists of various primer selection criteria variables that determine an efficient, sensitive and specific amplification (Singh et al. 2000). The variables are GC content, hairpin loop, primer-dimer, melting temperature (Tm), specificity and primer length (Singh et al. 2000). Specificity is one of the more critical variables that were considered when designing the LNYV subgroup specific primers. The specificity was determined by analysing the number of nucleotide differences between the subgroup specific primers, the total non-complementary LNYV N gene sequences and verification of the specificity by BLASTn search. The finalised LNYV subgroup specific primers contained six nucleotide differences between the subgroups (Figure 2.2). In the case of respiratory syncytial virus (RSV), subgroup A and subgroup B were detected by RT-PCR analysis using subgroup specific primers. The subgroup specific primers contained six nucleotide

differences between the subgroups (Hu et al. 2003). This result indicated that six differences in nucleotides between the subgroup primers were appropriate and would be expected to efficiently distinguish between LNYV subgroups for diagnosis. The findings from the current study indicate that after the optimisation of RT-PCR conditions, the subgroup I primer pair combination amplified only subgroup I isolates while the subgroup II primer pair combination only amplified subgroup II isolates. Amplification with subgroup I and subgroup II primer combinations could indicate the presence of both subgroup isolates in a sample. These subgroup specific primers can be used to diagnose the LNYV subgroups efficiently.

According to Stadhouders et al. (2010), the primer efficiency is depended upon the similarity between the primer and target sequences. A high-number of mismatches could reduce the primer binding to the target sequences, which may result in no amplification of the target gene sequences (Stadhouders et al. 2010). To verify the sequence complementarity within each subgroup to determine the specificity, the subgroup specific primers were aligned and compared with available LNYV N gene sequences (template sequences). Among the 25 LNYV N gene template sequences analysed, one mismatching template sequence to the primers was tolerated, within each subgroup. More than one non-complementary template sequence may not detect the entire LNYV subgroup population and could have resulted in a false negative. The subgroup I primer pair had sequence complementarity within all subgroup I template sequences and subgroup II primer pair had one mismatching template sequence within the subgroup II template sequences.

The final selection criterion to be considered was the specificity of the primer sequences. A BLASTn search suggested the primers were highly unlikely to amplify DNA of non-targeted organisms that included bacteria, fungi, other viruses, plants, animals such as insects, mollusc and human cells. The BLASTn search is commonly used to examine primer specificity (Whyte and Greer 2005). The subgroup specific primers are unlikely to amplify sequences from these organisms and occurrence of false positives should be low during LNYV subgroup diagnosis.

Among the six pairs of LNYV subgroup primers, one pair (the finalised pair) satisfied these three criteria that were mentioned above. Comprehensive analysis of the LNYV

primer specificity reduces the risk of deceptive results caused by inadequacy in primer design, increases the efficiency of primers and subgroup detection. The remaining criteria were also verified to determine the efficiency of the finalised primers. The finalised primers had lengths between 19 nt to 28 nt and these primer lengths are within the acceptable range. According to Lorsch (2013), primer sequence of >18 nt increases the specificity of the target sequence. However, increased length could decrease the efficiency of primer binding to the template sequence; the optimum Tm must be > 54°C to ensure high primer efficiency (Lorsch 2013). The Tm values for the subgroup specific and non-subgroup primers were between 57.3 °C to 57. 5°C as required. The hairpin ΔG were between -1.97 Kcal.mole to 2.41 Kcal.mole and the GC content was between 39.4% to 57.9%. The recommend hairpin ΔG should be > -9 Kcal.mole and GC content should be between 35% to 65% because these factors could affect the secondary structure of the primers (Integrated DNA Technologies 2017; Lorsch 2013). These results signified that both variables met the recommended criteria. Designing and verification of the LNYV primers was the primary aim of LNYV subgroup diagnosis. The results substantiated that all the primer designing criteria were indeed satisfied by the finalised primers.

Optimisation of LNYV and subgroup specific primers

The use of the subgroup specific primers was optimised by using the LNYV positive samples and RT-PCR analysis for subgroup diagnosis. RT-PCR analysis is considered to be highly sensitive and specific to detect for plant RNA viruses (Babu et al. 2017; Yang et al. 2017). The accuracy, specificity, stringency and reliability of RT-PCR are determined by the optimisation of primers to reduce the production of primer dimers, non-specific products, and no PCR products (Reiter and Ptatfl 2011). It is necessary to optimise the use of novel primers or previously designed primers since different kits are used for the analysis in different laboratories. Earlier studies had detected LNYV using two pairs of primers in the RT-PCR analysis: BCNG1/BCNG2 and LNYV_440F/LNYV_1185R (Callaghan and Dietzgen 2005; Higgins et al. 2016b). Conditions for these primers for LNYV detection were optimised since a different RT-PCR kit was used in the current study. Use of the finalised subgroup specific primers was also optimised.

Primer annealing temperature is one of the important conditions to be optimised (Pelt-Verkuil et al. 2008). The current study showed that when the annealing temperature was

less than 65 °C, the LNYV subgroup specific primers amplified non-specific and unexpected PCR products. A similar finding by Rychlik et al. (1990) affirmed that low annealing temperatures could cause the amplification of nonspecific products. While high temperature increases the specificity between the primer and the target sequence, it could decrease the yield, or no PCR product maybe generated (Lee et al. 1997; Life Science 2017; Rychlik et al. 1990). Hence, gradient PCR were performed to identify the optimum annealing temperature; 65°C was determined as the optimum temperature for the LNYV subgroup specific primers.

The predominant issue during the optimisation of LNYV subgroup specific primers and BCNG1/BCNG2 primers was the use of the wrong RT-PCR kit (SuperScript™ III Platinum™ One-Step qRT-PCR Kit). According to the manufacturer's protocol, the 2X reaction mix of this kit contained a final magnesium concentration of 3 mM, while the SuperScript™ III Platinum™ One-Step RT-PCR Kit had a magnesium concentration of 1.6 mM. High magnesium concentration may have decreased the specificity and resulted in non-specific products in the analysis. It is known that was an inverse relationship exists between the magnesium concentration and primer specificity in a standard PCR analysis (Life Science 2017; Lorenz 2011). The presence of non-specific products could also be due to the presence of contaminants in the samples or reagents that may have occurred during the PCR preparation. PCR is considered highly susceptible to contamination by aerosolised DNA, materials or improper handling of samples during their preparation (O'Connell 2002). Due to the high susceptibility to contamination, stringent procedures were taken in the current research to reduce contamination. By using new materials, reagents, the practice of effective handling techniques and laminar flow hood, amplification of nonspecific products and contamination were reduced, with increased reliability for detection LNYV and subgroups.

High primer concentration and number of PCR cycles could also lead to amplification of nonspecific products and primer dimers in PCR analysis (Innis and Gelfand 1990; Life Science 2017). According to the procedure described by Callaghan and Dietzgen (2005), the primer concentration for BCNG1/BCNG2 primers was reduced from 0.8 µM to 0.4 µM. The subgroup specific primer concentrations were reduced from 0.8 µM to 0.2 µM because according to the SuperScript™ III Platinum™ One-Step RT-PCR Kit the recommended primer concentration was between 0.15 µM to 0.5 µM for a standard RT-

PCR. The final optimisation to reduce the nonspecific products in subgroup specific and LNYV_440F/LNYV_1185R primers was to decrease the number of PCR cycles from 40 to 30. PCR of 40 cycles was recommended in the SuperScript™ III Platinum™ One-Step RT-PCR Kit protocol. High PCR cycles could cause a plateau effect (Innis and Gelfand 1999). Initially, the amplification of PCR products is increased exponentially then, as the amount of components are consumed within the reaction, the amplification decreases and eventually reaches a plateau stage (Kochanowski and Udo 1999). The plateau effect is caused by high PCR cycles that increase the likelihood of mispriming and can result in non-specific products when the reaction has reached the plateau phase (Pestana et al. 2010). Reduction of primer concentration and the number of PCR cycles eliminated the production of nonspecific products in the RT-PCR analysis of the current LNYV study.

RT-PCR is the predominantly used plant RNA virus detection method (Lima et al. 2012). LNYV has substantially diversified hosts that include monocots and dicots (Dietzgen et al. 2007). Therefore, it is essential to develop a detection method that has the capability to detect in multitudinous hosts. RT-PCR analysis is a more sensitive method than serological detection; for example, SCV is a cytorhabdovirus that was detected in herbaceous hosts by polyclonal antisera (Klerks et al. 2004; Posthuma et al. 2002). However, it was undetectable in an infected strawberry due to poor sensitivity of the assay, inhibitors or low viral distribution within the leaves. As a result, an RT-PCR detection system was developed to detect SCV in all infected plants including strawberries and aphids (Klerks et al. 2004; Posthuma et al. 2002). Furthermore, ELISA is known to be less sensitive when compared to molecular methods for the detection of plant virus strains (Jeong et al. 2014). Serological methods maybe unreliable for detecting viruses in all plant hosts and could result in a false negative. Hence, development of a LNYV subgroup detection system using RT-PCR analysis would be more efficient for detecting LNYV in various hosts and insect vectors than serological methods.

While RT-PCR is generally more sensitive than other conventional methods such as ELISA (Jeong et al. 2014) it can give false results when the reaction conditions are not optimal. A few of the RT-PCR analyses of LNYV positive samples using the BCNG1/BCNG2 primers gave false negative results, most likely due to RNA degradation of those samples. If the RNA was degraded to some degree, a large product such as the one amplified by the BCNG1/BCNG2 primers at 1,500 bp is more difficult to amplify.

According to the study by Pelt-Verkuil et al. (2008), the reliability of RT-PCR depends upon RNA integrity and degradation within the samples. RNase enzymes are depend on the presence of metallic ions, water molecules, oxidisation by reactive oxygen species (ROS), and frequent freeze-thawing instigates the RNA degradation process (Fabre et al. 2014). Although RNA is stored at -20°C to -80°C to reduce the degradation, it does not inhibit the process as RNases are functional even in cryogenic conditions, resulting in some degradation (Arrigo 1994; Fabre et al. 2014). A reliable method to reduce RNA degradation was developed by dehydration of RNA, which was placed in air and moisture tight containers in an anhydrous and anoxic conditions (Fabre et al. 2014). Such a method could be used to store LNYV RNA samples to reduce the degradation and false negative diagnosis with these primers in the future.

Testing for LNYV and subgroups

The first LNYV incidence on lettuce in New Zealand was reported in 1965 from Blenheim, Marlborough, followed by significant lettuce crop collapse in 1969 in Waimauku, Auckland (Fry et al. 1973). Since then, LNYV has been identified in various locations of New Zealand, including Auckland, Gisborne and Hawke's Bay (Fletcher et al. 2005; Higgins et al. 2016b). The latest LNYV infection of lettuce was reported by Plant and Food Research in Canterbury (Fletcher et al. 2017). A total of 25 samples were collected from Harrisville, Auckland, in 2011; six samples were already tested and identified as LNYV positive prior to the current study (Higgins et al. 2016b). Several samples remained untested. In the current study, BCNG1/BCNG2 primers were used to test 11 Harrisville samples that showed symptoms of LNYV, collected from the above study; the H5, H14 and H18 samples were detected as LNYV positive. The remaining eight samples were tested using LNYV_440F/LNYV_1185R or subgroup specific primers by RT-PCR. Both sets of primers had detected one sample that was infected by LNYV. In total, two samples were infected by LNYV subgroup I and one sample was subgroup II. However, the subgroup specific primers had failed to detect the H5 sample which will be discussed subsequently. The BCNG1/BCNG2 primers had failed to detect the H3 infected sample but this sample was determined to be infected with LNYV by the other two primers pairs (LNYV_440F/LNYV_1185R and subgroup specific primers). When compared with the other primer pairs, the BCNG1/BCNG2 primers amplified the largest PCR fragment of ~1,500 bp while LNYV_440F/LNYV_1185R and subgroup specific primers amplified ~750 bp and ~200 bp fragments, respectively.

Two conditions that could affect the amplification of a long PCR product: RNA degradation and less than optimal PCR conditions (Rosner et al. 1997). The failure in detection of LNYV by the BCNG1/BCNG2 primers was unlikely to be due to the PCR conditions as above-mentioned because these primers were optimised to detect LNYV infected samples. Amplification of a large PCR product can be more difficult, especially if RNA degradation has started to occur as may have happened since the sample had been stored for a long period of time. This problem may have also been exacerbated if the virus was at low titre, which may have been the case since a lower amount of PCR product was consistently amplified from the H3 sample compared to H14, H18 and positive controls. This result was supported by the comparative study of long and short RT-PCR products to detect prunus necrotic ringspot virus (PNRV) (Rosner et al. 1997). It was concluded that a low concentration of PNRV had failed to be detected by primers that amplified a large PCR product. It was efficiently detected by the primers that amplified a short PCR product because there was higher DNA polymerase efficiency to synthesise short DNA fragments (Rosner et al. 1997). Therefore, the H3 RNA sample may have been partially degraded and this may have inhibited the amplification of the complete N gene sequence by the BCNG1/BCNG2 primers. Nevertheless, the subgroup specific and LNYV_440F/ LNYV_1185R primers had tolerated partial degradation of RNA samples and were more reliable for diagnosis.

The H5 subgroup was not detected by the subgroup specific primers and the subgroup was undetermined. This indicate that the subgroup specific primer may not detect all LNYV population and could be due to the genetic diversity within the N gene sequence. In the future, the complete N gene should be sequenced from all LNYV positive samples from this study including the H5 sample and the subgroup specific primers should be re-designed or further optimisation of primers could enable the detection of larger LNYV subgroup population. The subgroup of H5 can be determined by the phylogenetic analysis of the H5 N gene sequence.

Pukekohe, Auckland is one of the highest lettuce producing regions in New Zealand (Walker 2005). LNYV infected samples were previously identified in Pukekohe and the nearby region, Harrisville (Fletcher et al. 2005; Higgins et al. 2016b). It was hypothesised in the current research that there was a high likelihood of LNYV infected lettuce or weeds

in Tuakau, Waikato, which is also geographically located near to Pukekohe. Potentially LNYV infected samples were collected from Tuakau and were tested using the multiple primer pairs for LNYV detection. All samples were LNYV negative by the RT-PCR analyses. These results indicated that specific sampling sites might not be infected. More sampling and testing are required to determine the presence or absence of LNYV in Tuakau region.

Canterbury is the second highest lettuce producing region in New Zealand and has been known to grow more than 133ha of lettuce (Fagan et al. 2010). A recently published report by Fletcher et al. (2017) confirmed LNYV infection in Canterbury between 2014 to 2016. A total of seven LNYV infected lettuce samples were identified by ELISA. These samples were collected from the Marshland, Southbridge and Chertsey regions of Canterbury (Fletcher et al. 2017). Subsequently, these seven samples were tested using the subgroup specific primers to identify the LNYV subgroups infecting these plants. The results showed that all Southbridge samples were infected by subgroup II, Marshland samples were infected by both subgroups and Chertsey sample was infected by subgroup I (Figure 2.33 and Table 2.8). Sample S2 from Southbridge was shown to be uninfected by LNYV using LNYV_440F/LNYV_1185R and subgroup specific primers. This could have been attributed to a very low viral titre and / or unequal distribution of the virus within the sample. This result was supported by the ELISA results which had a low reaction for this sample (J. Fletcher, personal communication, March 14, 2017).

Plant viruses are non-uniformly distributed within infected hosts. The distribution is determined by the gene expression of the viruses and hosts, the plant immune system and abiotic factors (Hull 2014). As observed and mentioned above, the false negative results for LNYV detection may not have been due to the incapability of the RT-PCR based diagnosis but rather a sampling error where leaf tissue with little virus was tested. This type of sampling error was observed and substantiated by the study conducted by Komínek et al. (2009) who had tested for grapevine virus A (GVA) using RT-PCR analysis. The results from their RT-PCR analysis was inconsistent due to the irregular distribution of the virus. Similar to the inconsistent results for H3 as described above. This signified that multiple stratified random sampling within the various locations of the LNYV infected leaves should be taken to obtain an accurate diagnosis.

Epidemiology of plant viruses is essential for disease surveillance and control (Chen and Siede 2007). The results from the Canterbury samples showed that there was a higher number of subgroup II infected lettuce samples than subgroup I. The Harrisville results from the current and previous study by Higgins et al. (2016b) had the opposite to the Canterbury results. It was anticipated that H18 would be infected with subgroup II because it was located near a previously identified subgroup II sample (H19). Hence, there was a high possibility that the same aphid could disseminate the same subgroup. The sample size from both locations was too small (<10) to conclude the prevalence, abundance and the distribution of LNYV subgroups in New Zealand. More sampling from Canterbury and other locations around New Zealand is required to get a complete picture of the distribution of each subgroup. However, this study showed that both subgroups are present in both the North and South islands of New Zealand, expanding the findings of Higgins et al. (2016b).

LNYV molecular diagnosis and symptoms

Symptoms of LNYV are similar to TSWV (Chu and Francki 1982; Dietzgen et al. 2007). Hence, there is a high likelihood of misdiagnosis. The H3, H10, H18 and H35 samples had symptoms associated with LNYV in the present study. After testing with multiple primers, it was confirmed that H3 and H18 were infected with LNYV; however, H10 and H35 were not. Both of these plants could have been infected by TSWV or lettuce mosaic virus. Further testing is required to test this hypothesis.

TPS test

The template preparation solution (TPS) procedure was used as a crude virus release method to detect various plant viruses, including LNYV, by RT-PCR method (Thomson and Dietzgen 1995). The procedure was more rapid than other RNA extraction methods because it does not require organic solvents or ethanol precipitation that can be difficult to extraction (Narayanasamy 2011a; Thomson and Dietzgen 1995). LNYV was detected using the TPS1 solution with an incubation at 60 °C and 95 °C undiluted (Thomson and Dietzgen 1995). In this present research, LNYV infected *N. glutinosa* samples were tested using the published procedure to determine if this could be used for rapid and higher throughput diagnosis of LNYV subgroups. It was carried out multiple times by modifying various conditions. The modified conditions were incubation temperature for extraction, volume of the TPS solution used for RT-PCR, diluted and undiluted samples. The LNYV

was undetected in all experiments. The false negative results could have been due to PCR inhibitors within the *N. glutinosa* TPS extracts. An earlier study confirmed that all plants contain pectin, polysaccharide, polyphenols and these components are considered to be PCR inhibitors (Schrader et al. 2012; Stagnati et al. 2017). This conclusion was supported by the study conducted by Dietzgen (2003) that used the TPS procedure to detect TSWV in anemone, capsicum, lettuce and tomato leaves with false negative results. The results indicated that detection of TSWV in 1/20 dilution of TPS solution were affected by PCR inhibitors. The ideal dilution required for detection was ≥ 1:100 dilution to reduce the inhibitors (Dietzgen 2003; Dietzgen et al. 2005). It was concluded that this procedure was inappropriate because there was inadequate information about PCR inhibitors in each plant (Dietzgen et al. 2005). This method could be retested for LNYV detection by increasing the template dilution to 1:100; however, according to Albrechtsen (2006) ≥ 100 fold dilution could reduce the sensitivity and detection, especially for a low virus titre. It is not guaranteed if the increase of dilution would enable the detection of LNYV or subgroup using this procedure, but should be tested. Regardless, a rapid template preparation method for detection of RNA viruses would be very useful for large scale diagnostics.

LNYV infected lettuce and *N. glutinosa* leaves were detected previously by using the TPS extraction procedure combined with RT-PCR using the SuperScript II RNase H- kit with LN1 and LN2 primers (Thomson and Dietzgen 1995). In the present study, the extraction procedure was tested on LNYV infected *N. glutinosa* using Superscript III One-Step RT-PCR kit with LNYV_440F/LNYV_1185R and subgroup specific primers. The main two distinctions between the procedures were the use of different kits and primers. According to the SuperScript III One-Step RT-PCR protocol, the RNase H activity was reduced to obtain higher cDNA yield and to increase the efficiency. RNase H hydrolyses the RNA during the hybridisation of DNA and RNA (Cerritelli and Crouch 2009). RNase H is used in RT-PCR during cDNA synthesis to degrade the mRNA (Mullan et al. 2002). A study conducted by Levesque-Sergerie et al. (2007) investigated the limitations of various commercial RT enzymes including SuperScript II and SuperScript III in RT-qPCR. It was identified that SuperScript III had a higher inhibition percentage (79.3 ± 2.9%.) than SuperScript II (59.7 ± 6.8%). Based on these results, it was concluded that the reduction of RNase H activity in the Superscript III kit might have reduced the sensitivity for detection of LNYV in the TPS1 solution by making it more

sensitive to PCR inhibitors such as high KCl and cell components in the TPS extraction, causing the false negative results. The use of different primers could not reduce the failure of LNYV detection in infected samples. To conclude, the TPS procedure was an inefficient method to combine with RT-PCR analysis to detect LNYV and subgroups.

RT-PCR RFLP

Restriction fragment length polymorphism (RFLP) utilises restriction enzymes that cleave dsDNA at specific restriction sites and forms variable lengths of DNA fragments (Mittal et al. 2013). The fragments are separated and visualised by gel electrophoresis. It is occasionally combined with PCR analysis in a method cell PCR-RFLP, which has been used to detect plant viruses (Tennant and Fermin 2015). For example, the EMDV subgroups were detected and differentiated by RT-PCR-RFLP (Parrella and Greco 2016). Even though LNYV subgroups can be detected using endpoint RT-PCR, two reactions were required, one for each subgroup. The gel electrophoresis result from subgroup RT-PCR analysis was unable to differentiate between the two subgroups because both subgroup specific primers amplify the PCR products of almost the same size. Therefore, RT-PCR-RFLP was developed to diagnose the LNYV subgroups in a single reaction. *Mae*III enzyme was predicted to cleave the subgroup I PCR product into three fragments (14 bp, 67 bp and 131 bp) and subgroup II into two fragments (64 bp and 145 bp) (Figure 2.39). Gel electrophoresis showed two fragments for both subgroups with the 14 bp fragment of subgroup I was undetected; according to Lee (2012) agarose is unable to detect very small sized DNA fragments less than 50 bp. Polyacrylamide, which is capable of detecting DNA fragments between 5 bp to 500 bp (Lee 2012), may have been used to detect the 14bp subgroup I fragment. The expected fragment sizes were observed by gel electrophoresis, which indicated that *Mae*III had cleaved both subgroups efficiently and this procedure can be used to diagnose LNYV and identify the LNYV subgroups.

Future research should focus on developing an RT-qPCR assay with high resolution melting analysis. It is faster, quantifiable and more sensitive to detect low titre plant RNA viruses. Further, there is no post-PCR procedures such as gel electrophoresis (Mackay et al. 2002). According to Simko (2016), high resolution melting analysis is capable of detecting small variations in the target sequence, post-PCR. The amalgamation of RT-qPCR with high resolution melting analysis has been used to identify plant virus subgroups, for example, the grapevine leafroll-associated virus 3 (GLRaV-3) subgroups

were detected and differentiated by RT-qPCR with high resolution melting analysis (Bester et al. 2012). Similar research could be conducted for LNYV subgroups using the subgroup specific primers. As demonstrated, all three primers can be used in a multiplex reaction for subgroup detection. The subgroup specific primers give rise to products that are at the upper end of the desired size for real time PCR products, and the primers are capable of amplifying their target sequences at the higher annealing temperature of 65 $^{\circ}$C. Thus, it is expected that these primers would be useful in a RT-qPCR diagnostic assay.

Summary

Figure 2.39 shows a summary of process that could be used to confirm LNYV infection and identification of subgroup. LNYV can be detected using BCNG1/BCNG2 and LNYV_440F/LNYV_1185R primers by RT-PCR analysis (Figure 2.39). Of the two primer pairs, based on the present research study, it is recommended to use LNYV_440F/LNYV_1185R primers because these were demonstrated to be more reliable and efficient than the BCNG1/BCNG2 primers. If a negative result was obtained from either of the primer pairs, it should be retested using alternative primers for reconfirmation. LNYV positive samples from these primers should be tested using either RT-PCR or RT-PCR-RFLP (Figure 2.39) with subgroup specific primers. A positive result from the subgroup I primer combination reflects infection with LNYV subgroup I while negative results were considered as likely uninfected by subgroup I (Figure 2.39). The same result was applied when subgroup II primers were tested on the samples (Figure 2.39). Of course, LNYV infection can be diagnosed using all three subgroup primers. If this test is found to be positive, the subgroup can identified by RFLP analysis of the RT-PCR product by digesting it with the restriction enzyme *Mae*III to identify LNYV subgroups (Figure 2.39). If all the three fragments of the expected sizes or the larger two were obtained, the sample was infected by subgroup I (Figure 2.39). LNYV subgroup II was identified if two fragments with expected sizes were obtained (Figure 2.39). This study is considered to be the first LNYV subgroup detection by RT-PCR analysis.
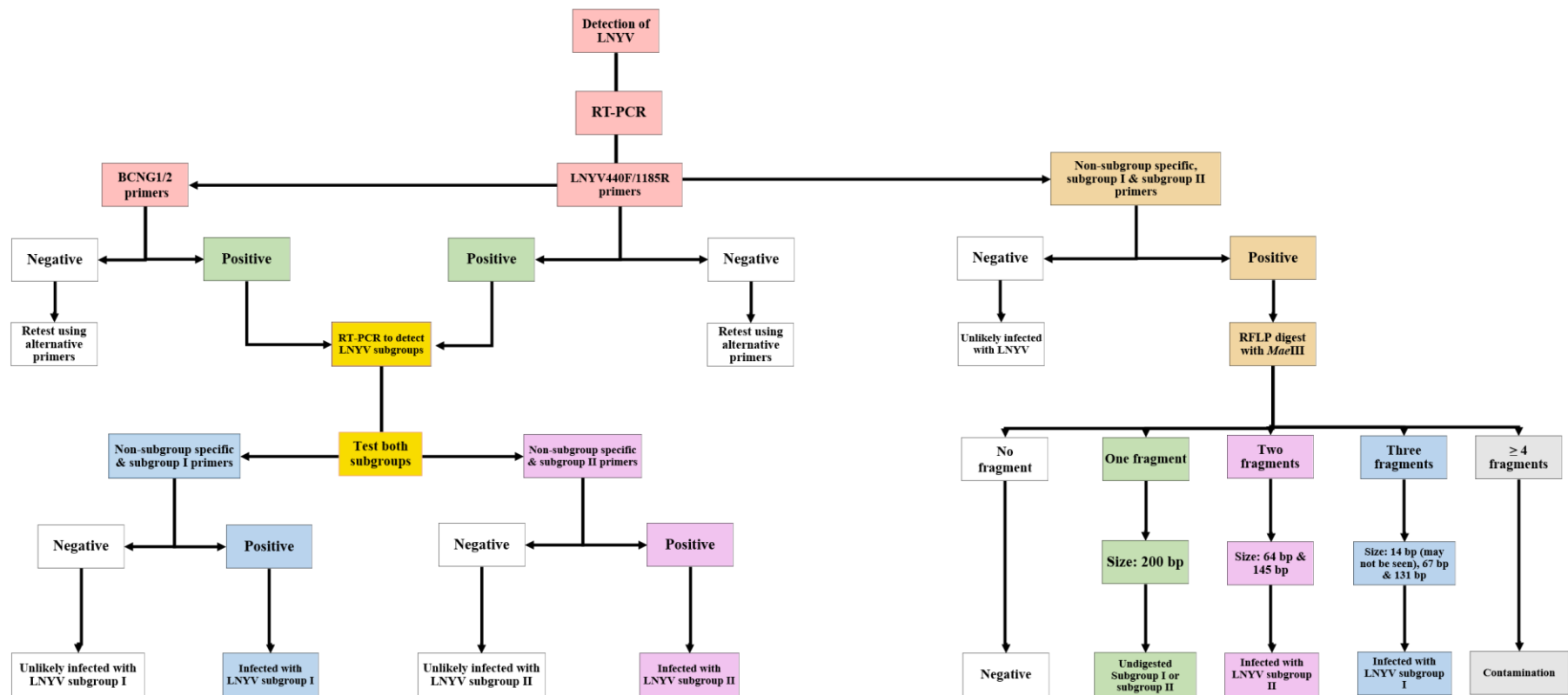
Figure 2.39: Summary of LNYV and subgroup detection methods by RT-PCR and RT-PCR RFLP.

Red = Detection of LNYV   Brown = Detection of LNYV and subgroup by RT-PCR RFLP   Yellow = Subgroup detection by RT-PCR

Green = LNYV positive detection   White = LNYV/subgroup negative   Blue = Subgroup I positive

Pink = Subgroup II positive   Grey = Contamination

Chapter 3

LNYV Genome Sequence and Glycoprotein Sequence Analysis

Chapter 3 LNYV genome sequence and glycoprotein analysis

3.1 Introduction

The genomes of plant viruses comprise information that encodes for the proteins that determine the viral anatomy, function and interaction with hosts/vectors (Hull 2014; Mandahar 2006). The viral genome also consists of non-coding regions that regulate the genome multiplication and expression (Hull 2014). Analysis of a complete viral genome provides an overview of the genetic composition of the virus and helps to elucidate the host/vector interactions with the plant virus (Kaur et al. 2016). The genetic pathways associated with each gene and expression could be determined to explain the interactions (Kaur et al. 2016). The first and only available complete genome of LNYV, which is from a subgroup I isolate, was sequenced using Sanger sequencing. The sequence came from an isolate originally from an infected garlic plant that had been mechanically inoculated onto *N. glutinosa* in Australia (Dietzgen et al. 2006). The genome consists of 12,807 bp and six genes in total. The six genes are nucleocapsid (N), phosphoprotein (P), 4b, matrix (M), glycoprotein (G) and RNA dependent RNA polymerase (L). The LNYV genome was used to identify the total number of nucleotides, 5' leader nt, 3' trailer nt and the total number of amino acids encoded by each ORF. The L gene was used to perform phylogenetic analysis to understand the evolutionary relationship of LNYV with other plant rhabdoviruses (Dietzgen et al. 2006).

Until this study, a genome from a LNYV subgroup II isolate had not been sequenced. Furthermore, the genome of a New Zealand isolate had not been sequenced. The genome of a subgroup II isolate could help to understand the evolutionary relationship between the subgroups and other plant rhabdoviruses. Differences in nucleotides, amino acids, 2D and 3D structures of the proteins or genes between the subgroups may assist to identify differences that are responsible for association with the hosts or vectors. Further, these differences may account for the relatively more rapid dispersal of subgroup II.

3.1.1 Glycoprotein analysis

The glycoprotein (G protein) is located on the surface of cytorhabdoviruses and is required for the virus interaction with the insect vectors (Mann and Dietzgen 2014). The G protein may be involved in the transmission of LNYV by aphids (*H. lactucae*) to hosts. Dietzgen et al. (2006) examined the LNYV G gene in detail; signal peptide, glycosylation sites and putative transmembrane domain were identified in the G protein (Dietzgen et al. 2006). In aphids, LNYV is likely to translocate from the insect's midgut to various organs of the body, including the salivary glands, for transmission (Redinbaugh and Hogenhout 2005). Analysis of plant and animal rhabdovirus infection of insect vectors indicates that the virus penetrates through the midgut by receptor-mediated endocytosis using viral glycoprotein receptors (Ammar et al. 2009; Redinbaugh and Hogenhout 2005). No studies have been carried out looking at this process for LNYV specifically, however, it is assumed that this virus would use a similar mechanism. Subgroup II seems to be more efficiently dispersed and may have out competed subgroup I in Australia. Higgins et al. (2016b) hypothesised that this may have happened via more efficient attachment to the insect or via more efficient replication in either or both the insect or plant hosts. In the current study, the G gene and protein were specifically analysed since this is the part of the virus most likely to have a role in attachment to the plant hosts and insect vectors. Analysis of the G sequences may help to determine if subgroup II could more efficiently associate with insect vectors than subgroup I. Since the G gene of the subgroup II is not available, the genetic and protein differences between the subgroups in this genome region is unknown.

3.1.2 Aims

One of the aims of this study was to obtain the complete genome of LNYV subgroup I and subgroup II from New Zealand isolates by Illumina RNA sequencing. The evolution of LNYV subgroups was analysed using phylogenetic analysis of the genome and N gene. The G gene was obtained from both genomes and differences in nucleotide, amino acid sequence and 2D structure of the G proteins were identified to determine differences between the subgroups.

3.2 Materials and methods

3.2.1 Genome and gene analysis summary

The summary of the procedure used in the current study to analyse the LNYV genome and gene is given in Figure 3.1. The RNA was extracted from LNYV subgroup I and subgroup II infected plants. RNA purity, concentration and integrity were measured and the presence of LNYV confirmed by RT-PCR analysis. Samples were sent for Illumina sequencing. The quality of the sequences was analysed by FastQC, and *de novo* assembly was performed to obtain contigs; LNYV contigs were identified by BLASTn search. The analyses diverged into two pathways; pathway A and pathway B. Pathway A was a *de novo* assembly of the LNYV contigs, while pathway B analysis was a reference assembly of the LNYV contigs using the published LNYV genome as the reference. The consensus genome sequences obtained from both pathways were compared and used to determine which pathway would be used for the subsequent analysis. The pathway B consensus genome sequences were used for the bioinformatics analyses that included genome annotation, genome, phylogenetic and glycoprotein analyses. The specific methods used are detailed below.
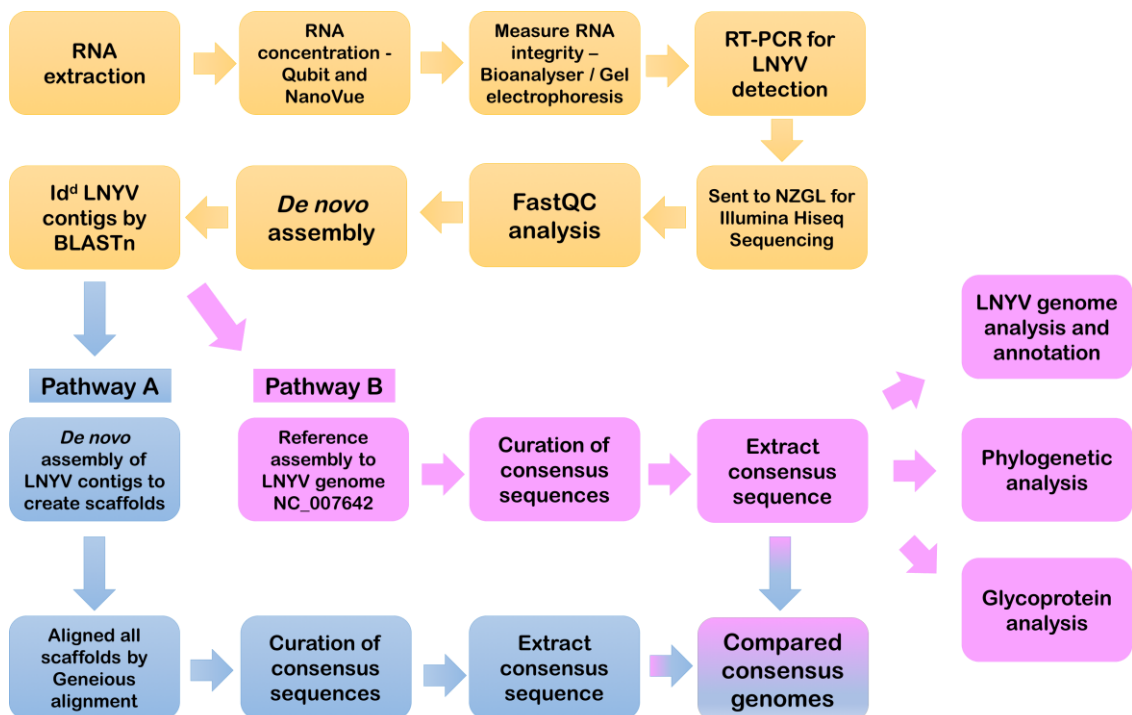
Figure 3.1: LNYV genome and gene analysis procedure. NZGL: New Zealand Genomics Limited

3.2.2 RNA extraction, concentration, purity and integrity analysis for sequencing

Plant leaf samples used for this study were previously identified isolates obtained from LNYV infected lettuce, namely H29 (subgroup I), H33 (subgroup I) samples and H19 (subgroup II). The H19 sample was mechanically inoculated onto *N. glutinosa* to obtain sample for sequencing as the lettuce sample was almost depleted (Higgins et al. 2016b). Different LNYV isolates from each subgroup were extracted to produce RNA of sufficient quality for sequencing for subgroup I, leaf samples from plants infected with isolates H29 or H33 were extracted. For subgroup II, lettuce infected with the LNYV isolate H19 (subgroup II) needed to be inoculated onto *N. glutinosa* to increase the amount of infected material. RNA was also extracted from lettuce infected with the subgroup II isolate H18 identified in section 2.3.4.2.3a. These samples were used for the subsequent experiments. For each LNYV infected sample, approximately ten leaf samples were selected for RNA extraction. The RNA was extracted using the Spectrum™ Plant Total RNA extraction kit (Sigma Aldrich, Saint Louis, Missouri, USA) as described in section 2.2.2.1. The RNA integrity of the 28S rRNA, 18S rRNA and 5S rRNA was evaluated using the non-denaturing agarose gel electrophoresis, the concentration and purity of the RNA extracts were determined using a NanoVue spectrophotometer as described in section 2.2.3.

3.2.3 RNA concentration quantification by Qubit

RNA concentration is more accurately quantified using a fluorometric assay (Illumina 2011b). The Qubit® 2.0 fluorometer (Thermo Fisher Scientific, Massachusetts, USA) was used to determine the RNA concentration according to the manufacturer's instructions. The Qubit® working solution was prepared by diluting the Qubit® RNA HS Reagent to a 1:200 dilution in Qubit® RNA HS Buffer. Qubit® working solution (190 µl) was added to each standard (10 µl) and vortexed for three seconds. For each sample, 199 µl of Qubit® working solution was combined with 1 µl of the sample and vortexed for three seconds. All samples including the standards, were incubated at room temperature for two minutes to allow the binding of the dye. Subsequently, standard 1 with a final concentration of 0 ng/µl in TE buffer and standard 2 with a final concentration of 10 ng/µl in TE buffer were inserted into the sample chamber to calibrate the fluorometer. Then samples were inserted into the sample chamber sequentially and the concentration of the RNA was determined from the standard curve.

3.2.4 Evaluation of template RNA integrity and RNA Integrity Number (RIN) using a bioanalyzer

RNA integrity is evaluated more precisely using a bioanalyzer (Illumina 2011b). The Agilent RNA 6000 Pico kit (Agilent Technologies, Waldbronn, Germany) was used according to the manufacturer's instructions. The provided RNA ladder was denatured for two minutes at 70 °C using a heating block (Model: MiniT-100, Allsheng, Hangzhou city, China) and 90 µl of RNase free water was added to the ladder. The samples were diluted to a final concentration of approximately 2,500 pg/µl and denatured at 70 °C for two minutes using a heating block. To prepare the gel, 550 µl of RNA 6000 Pico gel matrix was added into the spin filter and centrifuged at 1,500 x g for ten minutes at room temperature (Eppendorf 5424R Centrifuge, New South Wales, Australia) and 65 µl of the filtered gel was used in each well for the analysis. The gel-dye mix was prepared by using the RNA 6000 Pico dye concentrate, which was vortexed for ten seconds and 1 µl of the dye added to the filtered gel. The mixture was vortexed for a few seconds and centrifuged at 13,000 x g for ten minutes at room temperature. Gel-dye of 9 µl was added into the **G** labelled well shown in Figure 3.2 and the plunger was positioned at 1 ml. The chip priming station was closed, the plunger was placed under the clip and released after 30 seconds. The plunger was pulled to the 1 ml position after 5 seconds and 9 µl of the gel dye was pipetted into each of the two wells labelled as G shown in Figure 3.2. The RNA 6000 Pico conditioning solution of 9 µl was pipetted into the CS shown in Figure 3.2, 5 µl of RNA 6000 Pico marker was added to the 11 sample wells and the ladder well (Figure 3.2). The denatured ladder of 1 µl was added to the well that was labelled as the ladder and 1 µl of the sample was added to each of the 11 sample wells. The chip was vortexed for one minute using the bioanalyzer Chip IKA vortexer (IKA MS 3, Agilent Technologies, Waldbronn, Germany) at 2,400 rpm. Finally, the chip was analysed in the Agilent 2100 (default settings) for plant RNA pico analysis. The samples that had a RNA integrity number (RIN) >4.5 with the greatest RNA integrity were selected to be tested for LNYV.

Figure 3.2: RNA pico chip for bioanalyser by Agilent technologies.
Image: https://web.uri.edu/gsc/files/RNAChip.jpg

3.2.5 RT-PCR analysis to detect LNYV

The RT-PCR was used to verify the presence of LNYV within the sample before sequencing. The H19 and H33 samples that contained the highest RIN values and RNA integrity were tested using the LNYV_440F and LNYV_1185R primers in RT-PCR. The optimised conditions described in section 2.2.5.3b were used for the analysis. The infected positive controls samples were H19 and H29 samples used by Higgins et al. (2016b). The negative controls were an uninfected lettuce sample and no template control (NTC) to identify contamination.

3.2.6 Illumina sequencing and post sequencing quality analysis

The H33 (subgroup I) and H19 (subgroup II) RNA extract was sent to New Zealand Genomics Limited (NZGL) (New Zealand Genomics Limited, Otago Genomics Facility, University of Otago, Dunedin, New Zealand) to sequence the complete genome of LNYV by Illumina Hiseq 2500 by paired-end/mate-paired RNA sequencing. The FastQC data analysis examines the quality of high throughput sequencing data including Illumina sequences (Andrews 2016). The raw data obtained from the Illumina sequencing were analysed by FastQC using the MutiQC report (http://www.bioinformatics.babraham.ac.uk/). The FastQC analysis contained seven main criteria; mean quality scores, per sequence quality scores, per sequence GC content,

per base N content and sequence duplication levels. After the examination of the FastQC files provided by NZGL, sequences were assembled by *de novo* and reference assembly.

3.2.7   *De novo* assembly

De novo sequence assembly was used to assemble sequence reads to obtain final LNYV genome sequences using the Geneious 6.0.6 software (http://www.geneious.com/). Both subgroups were paired by name, the relative orientation of forward/reverse (Illumina short read kit) and expected distance of 105 nt. The *de novo* assembly was performed according to the procedure by Puli'uvea et al. (2017); Wylie and Jones (2011) and was modified to obtain higher coverage. The 5' end and 3' end of each read were trimmed by 10 bp with low sensitivity / fastest, one million sequence reads were assembled. Such assemblies were done 30 times for H19 and 20 times for H33. Approximately 1000 contiguous sequences (contigs) were obtained from each assembly. The contigs were sorted by the number of sequences and length. A BLASTn search (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was carried out using the longest 10 (H33) and 20 (H19) contigs with the highest number of sequence reads to identify the LNYV contigs. At this point, the analyses diverge into two pathways; pathway A and pathway B.

3.2.7.1   Pathway A: *de novo* assembly

In pathway A, the LNYV contigs from each assembly were reassembled by *de novo* assembly to obtain the final scaffolds. The settings were medium sensitivity/ fast and using the existing trim regions. The scaffolds were aligned using the Geneious alignment in the Geneious 6.0.6 software. The settings for the alignment were 65% similarity, gap open penalty of 12, gap extension penalty of 3, alignment type was global alignment with free end gaps and refinement iterations of 2. The consensus genome sequence within the alignment was manually curated and then compared with the genome sequence obtained from pathway B.

3.2.7.2   Pathway B: reference assembly

In pathway B, the LNYV contigs that were obtained from the initial *de novo* assembly as described in section 3.2.7 were reassembled by reference assembly using the published

LNYV genome NC_007642 as the reference sequence to obtain the final scaffold. The settings used were medium sensitivity/ fast, iterative fine tuning up to five times and using the existing trim regions. The consensus sequence within the scaffold was curated manually and verified by BLASTn search. The genome was annotated using the reference genome (NC_007642). The genome sequence obtained from pathway B was used for the genome, phylogenetic and glycoprotein analyses.

3.2.8 Phylogenetic analysis

Phylogenetic analysis is used to examine the evolutionary relationship of organisms or genes in a phylogeny tree (Lemey et al. 2009). This type of analysis by comparing the assembled LNYV genomes to all currently available cytorhabdovirus, nucleorhabdovirus and LNYV genomes from the NCBI database (Table 3.1). The CLUSTAL alignment and neighbour joining (NJ) clustering methods with 1,000 bootstraps were used for the analysis using Geneious 6.0.6 software. Bovine ephemeral fever virus (BEFV) was selected as the outgroup based on the study conducted by Dietzgen et al. (2006). A CLUSTAL alignment and maximum likelihood (ML) tree was also constructed using the MEGA7 software with 1,000 bootstrap replicates on the same genome sequences as mentioned above (http://www.megasoftware.net/). Evolutionary model testing identified the General Time Reversible+G+I as being appropriate with a BIC score of 458,139.57. Pairwise distances were estimated using the Tamura-Nei model with a BIC score of 458,618.35.

Table 3.1: Viruses used for the genome phylogenetic analysis.

Genus	Virus	Accession
Cytorhabodvirus	Alfalfa dwarf virus	NC_028237
	Barley yellow striate mosaic virus	NC_028244
	Colocasia bobone disease-associated virus	NC_034551
	Lettuce necrotic yellows virus (published)	NC_007642
	LNYV H19	This study
	LNYV H33	This study
	Lettuce yellow mottle virus	NC_011532
	Northern cereal mosaic virus	NC_002251
	Persimmon virus A	NC_018381
	Wuhan Insect virus 4	NC_031225
	Wuhan insect virus 5	NC_031227
	Wuhan Insect virus 6	NC_031232
Nucleorhabodvirus	Datura yellow vein virus	NC_028231
	Eggplant mottled dwarf virus	NC_025389
	Maize fine streak virus	NC_005974
	Maize Iranian mosaic virus	NC_011542
	Maize mosaic virus	NC_005975
	Potato yellow dwarf virus	NC_016136
	Rice yellow stunt virus	NC_003746
	Sonchus yellow net virus	NC_001615
	Taro vein chlorosis virus	NC_006942
Ephemerovirus	Bovine ephemeral fever virus	NC_002526

Phylogenetic analysis of the LNYV N gene was also carried out. Based on the study conducted by Higgins et al. (2016b), the N gene ORF sequences were extracted from all published LNYV isolates (Table 2.2) and used for phylogenetic analysis. The N gene ORF sequences were also extracted from the H19, H33, LYMoV, NCMV and PeVA genomes (Table 3.1). The PeVA was used as the outgroup and N gene sequences were aligned by MUSCLE using Geneious 6.0.6 software. A maximum likelihood tree was calculated using the Tamura-3-parameter model+G (BIC score 8,265.47) with 1,000 bootstrap replicates using MEGA7 software. Another phylogenetic tree was determined using the amino acid sequences predicted from all the N gene sequences from the viruses mentioned above using the same procedure, but using the LG+G model (BIC score 4,676.85) with 1,000 bootstrap replicates.

3.2.9   Glycoprotein analysis

Plant viral gene sequence analysis is the preliminary study to understand the virus and vector interactions (Hull 2014). The glycoprotein sequences were extracted from the LNYV genome consensus sequences from H19, H33 and the published LNYV genome sequence. These three gene sequences were aligned using MUSCLE. The glycosylation site, polyA signal, repeat region, signal peptide and transcript regions were identified according to the study conducted by Callaghan (2005). Geneious 6.0.6 software (http://www.geneious.com/) was used to predict the 2D structure of the G gene sequences for the three isolates.

3.3 Results

3.3.1 Template RNA integrity analysis using the agarose gel electrophoresis

RNA quality can be evaluated by using agarose gel electrophoresis and high quality RNA samples are used for Illumina sequencing because degraded sample may decrease the read coverage (Davila et al. 2016; Sharma and Chaudhary 2016; Sheng et al. 2017). Non-denaturing agarose gel electrophoresis was used to determine the RNA integrity of the extracted LNYV samples of subgroup I and subgroup II. The total RNA was extracted from ten samples of H29 (subgroup I). All ten samples of H29 isolates had complete degradation of the ribosomal RNAs (data not shown). The bioanalyzer results in section 3.3.4 showed that the majority of H29 samples had degraded ribosomal RNAs and those samples were also unsuitable for sequencing. Therefore, ten LNYV lettuce infected with isolate H33 (subgroup I) RNA samples were extracted and tested for LNYV. Figure 3.3 shows that the H33 samples; 4, 7 and 8 had intact 28S rRNA and 18S rRNA bands, while the 5S rRNA band was faint. It also shows that H33 samples 1, 9 and 10 had partial degradation of the 28S rRNA and 18S rRNA since the amount of the 28S rRNA appeared less than that of the 18S rRNA. The remaining H33 samples were completely degraded (Figure 3.3). These samples were tested using the bioanalyzer for a final confirmation.
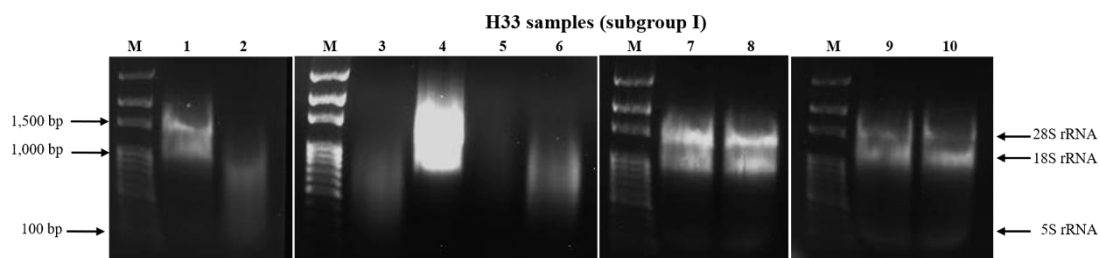
Figure 3.3: Non-denaturing agarose gel electrophoresis of total RNA from H33 (subgroup I) samples. Lane M: 100bp ladder (300ng). Lane number: sample number.

The total RNA was extracted from six samples of H18 (subgroup II). Figure 3.4a shows that four H18 samples (1, 2, 4 and 6) had intact 28S rRNA and 18S rRNA bands, but the 5S rRNA bands were too faint. It also shows that for the H18 samples 3 and 5 ribosomal RNAs were degraded. Hence, no intact ribosomal RNA bands were present in these samples. These results were confirmed using the bioanalyzer. The subsequent analysis in section 3.3.5 confirmed that the H18 samples were unsuitable for sequencing because

LNYV could not be detected by RT-PCR. Therefore, ten *N. glutinosa* infected with isolate H19 (subgroup II) RNA samples were extracted and tested for LNYV. The ribosomal RNAs of 2, 3, 5, 6, 7 and 8 H19 samples were partially degraded (Figure 3.4b) while samples 1 and 4 of H19 was degraded. The remaining H19 samples had complete degradation (data not shown) and all of these samples were tested using the bioanalyzer.
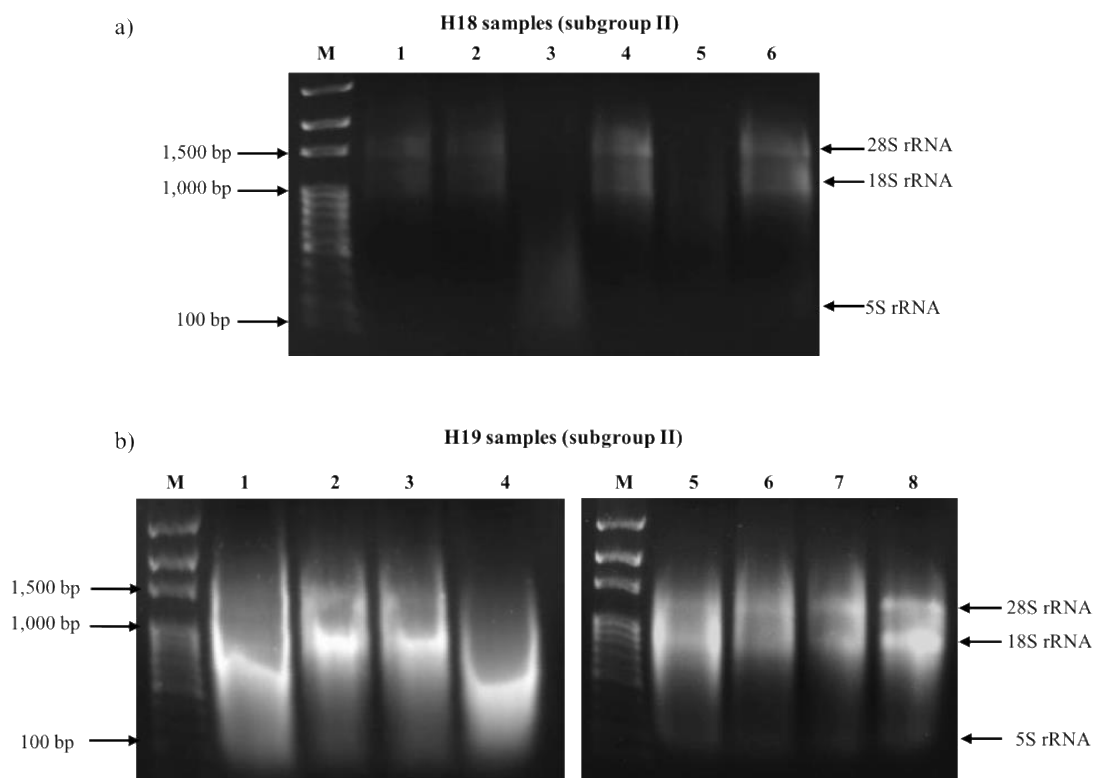
Figure 3.4: Non-denaturing agarose gel electrophoresis of total RNA from LNYV infected H18 (subgroup II) samples. Lane M: 100bp ladder (300ng). a) H18 samples. b) H19 samples. Lane number: sample number.

3.3.2   Template RNA purity and concentration analysis using the NanoVue

RNA purity can be measured by NanoVue spectrophotometer (Bernardo et al. 2013). It was determined by evaluating the A_{260}/A_{280} and A_{260}/A_{230} ratios. According to Otago Genomic & Bioinformatics Facility (2016), good quality RNA should have ratios of >1.8 and >1.5 for A_{260}/A_{280} and A_{260}/A_{230}, respectively. The RNA concentration for H33 samples was between 29.2 ng/µl – 674 ng/µl. The absorbance ratios of the H33, H18 and H19 samples are shown in Table 3.2. This table shows that RNA of all H33 samples

except H33_10 had A_{260}/A_{280} and A_{260}/A_{230} ratios were >2, indicating that these samples were of sufficient purity and suitable for genome sequencing. The H33_10 sample had A_{260}/A_{280} and A_{260}/A_{230} ratios 1.87 and 0.849, respectively, indicating that the sample was impure and unsuitable for genome sequencing. The RNA concentration of non-degraded H18 samples was between 41.6ng/µl – 149.6ng/µl. Almost all H18 samples had A_{260}/A_{280} ratios that were >2 and A_{260}/A_{230} ratios that were >1.5 except H18_1, which had an A_{260}/A_{230} ratio of 1.3. This result signified that H18_2, H18_4 and H18_6 RNA samples were suitable for sequencing and H18_1 was unsuitable because it contained impurities. The RNA concentrations for H19 samples were between 683.2 ng/µl – 1632 ng/µl. All degraded or partially degraded H19 samples had highly pure RNA with A_{260}/A_{280} and A_{260}/A_{230} ratios >2, which indicated that these H19 samples were acceptable for genome sequencing.

Table 3.2: NanoVue spectrophotometer RNA concentration, A_{260}/A_{280} and A_{260}/A_{230} ratios of LNYV infected H33, H18 and H19 samples.

Subgroup	Samples	RNA integrity	RNA Concentration (ng/µl)	A_{260}/A_{280} ratio	A_{260}/A_{230} ratio
Subgroup I	H33_1	Partially degraded	409.6	2.216	2.231
	H33_4	Non-degraded	674	2.171	2.271
	H33_7		365.2	2.205	2.179
	H33_8		410.4	2.216	2.192
	H33_9	Partially degraded	399.2	2.218	2.198
	H33_10		29.2	1.872	0.849
Subgroup II	H18_1	Non-degraded	41.6	2.261	1.333
	H18_2		63.6	2.208	1.828
	H18_4		43.2	2.204	1.687
	H18_6		149.6	2.226	1.938
	H19_1	Partially degraded	1632	2.213	2.348
	H19_2		732	2.202	2.254
	H19_3	Degraded	683.2	2.170	2.236
	H19_4		1167	2.174	2.323
	H19_5	Partially degraded	1344	2.188	2.311
	H19_6		838.4	2.172	2.251
	H19_7		838.8	2.169	2.234
	H19_8		1080	2.195	2.348

3.3.3   RNA concentration using the Qubit

RNA concentration can be quantified accurately by a Qubit fluorometer (Korpelainen et al. 2014). The recommended RNA concentration measured from the Qubit fluorometer for Illumina sequencing is between 100 ng – ~500 ng/µl (2016). Table 3.3 shows that the RNA concentrations of the H33 samples was between 136 ng/µl – 170 ng/µl, the H18 samples were between 32.2 ng/µl – 67 ng/µl and the H19 sample concentrations were between 688 ng/µl – 1360 ng/µl. The results indicated that all H33 samples had the recommended concentrations while the H18 samples had lower RNA concentrations than recommended. All H19 samples had higher concentrations than recommended and required dilution in RNAase free water for sequencing. The subsequent bioanalyzer and RT-PCR results in section 3.3.4 and 3.3.5, respectively, showed that H33_1, H33_4 samples and H19_1, were suitable for sequencing. Therefore, the H33_1 and H33_4 samples were pooled, which increased the concentration to 192 ng/µl. The H19_1 sample was diluted to ~500 ng/µl to meet the recommended concentration for genome sequencing.

Table 3.3: RNA concentrations of H18, H29 and H33 infected samples as measured using the Qubit fluorometer.

Subgroup	Sample	RNA concentration (ng/µl)
Subgroup I	H33_1	160
	H33_4	168
	H33_7	138
	H33_8	136
	H33_9	136
	H33_10	170
Subgroup II	H18_1	32.2
	H18_2	43.1
	H18_4	67
	H18_6	57
	H19_1	1360
	H19_2	774
	H19_3	688
	H19_4	1000
	H19_5	986
	H19_6	824
	H19_7	694
	H19_8	874

3.3.4 Evaluation of RNA integrity using a bioanalyzer

A bioanalyzer is capable of determining RNA integrity by pseudo-gel imagery and quantifying by RNA integrity number (RIN) (DeLong 2013; Malmanger et al. 2012). All samples of H33, H18, H19 and H29 samples were analysed in this way to confirm the RNA integrity. Figure 3.5a shows that the subgroup I H33_1, H33_4, H33_7, H33_8, H33_9 and H33_10 samples had intact 28S rRNA and 18S rRNA bands, with RIN values >6.8 (Table 3.4). Subgroup I H29 samples were all too degraded for further consideration (Figure 3.5a and b). Figure 3.5b and c show the results for subgroup II samples H18 and H19. H18_1, H18_4 and H18_6 had intact 28S rRNA and 18S rRNA bands. Samples H18_2 was partially degraded, and H18_5 was very degraded. All samples had RIN values >5 except for H18_5. All H19 samples were partially degraded except for H19_4 and H19_10, which were more degraded. The partially degraded samples had RIN values ~5. Samples H33_1, H33_4, H18_4, H18_6, H19_1, H19_2, H19_6 and H19_7 were chosen for further processing.
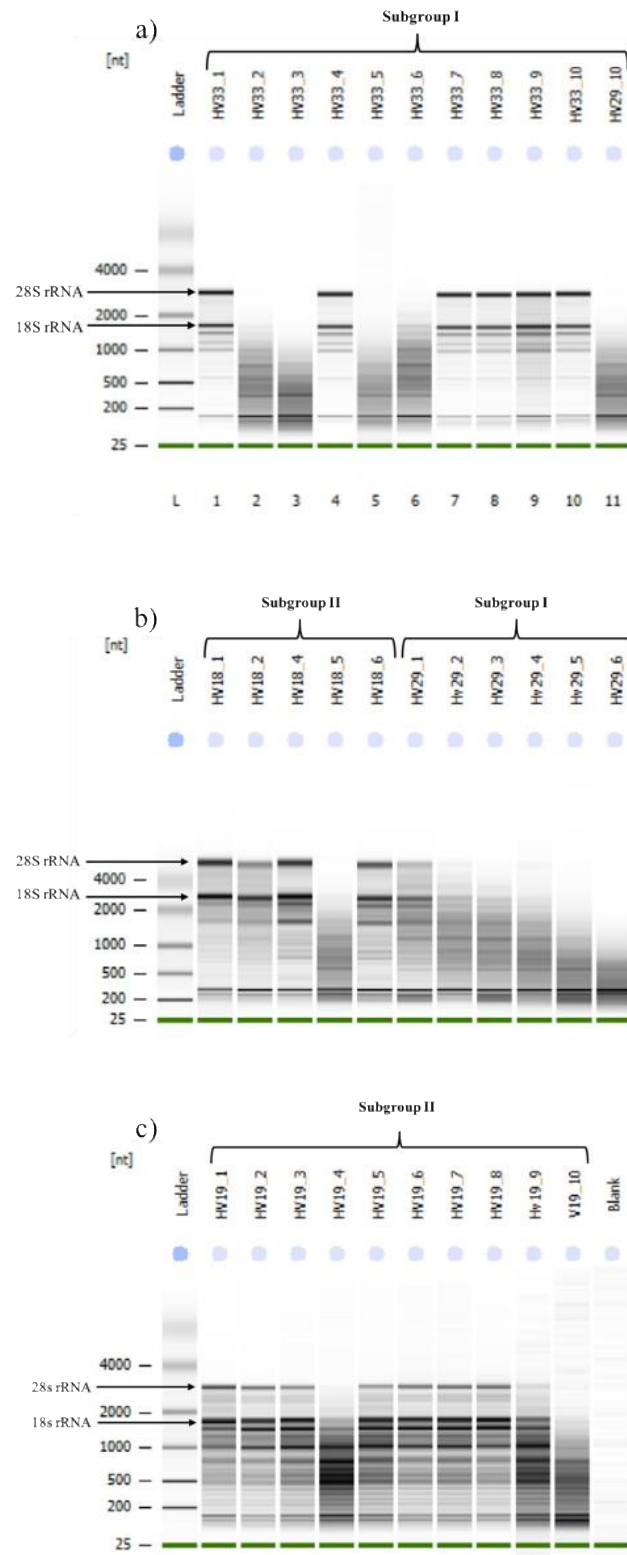
Figure 3.5: The bioanalyzer results for H33, H18, H29 and H19 RNA samples. a) H33 samples 1 – 10 and H29 sample 10. b) H18 samples 1 – 6 and H29 samples 1 – 6. c) H19 samples 1 – 10. The bands corresponding to 28Ss and 18s rRNA are indicated.

Table 3.4: RIN values and integrity for H29, H33, H18 and H19 using the bioanalyzer with RNA integrity.

Subgroup	Sample	RIN value	RNA integrity
Subgroup I	H29_1	3.4	Degraded
	H29_2	2	
	H29_3	1.9	
	H29_4	1.7	
	H29_5	1.9	
	H29_6	2	
	H29_10	1.6	
	H33_1	6.8	Non-degraded
	H33_2	1.7	Degraded
	H33_3	1.8	
	H33_4	7	Non-degraded
	H33_5	2.2	Degraded
	H33_6	1.9	
	H33_7	6.9	Non-degraded
	H33_8	7	
	H33_9	6	
	H33_10	6.9	
Subgroup II	H18_1	6.5	
	H18_2	5.1	Partially degraded
	H18_4	6.2	Non-degraded
	H18_5	1.5	Degraded
	H18_6	5.6	Non-degraded
	H19_1	5	Partially degraded
	H19_2	5	
	H19_3	4.7	
	H19_4	2.1	Degraded
	H19_5	3.9	Partially degraded
	H19_6	4.9	
	H19_7	4.8	
	H19_8	4.8	
	H19_9	2.7	
	H19_10	1.9	Degraded

3.3.5 Detection of LNYV using the LNYV_440F/LNYV_1185R primers

RT-PCR analysis was utilised to confirm the presence of LNYV in the extracted RNA, using the method described in section 2.2.5.3b. Since the RNA concentrations of subgroup I samples H33_1 and H33_4 were low, these samples were pooled to increase the amount of LNYV RNA. This was also done for subgroup II H18_4 and H18_6.

The H33 and H18 samples were tested using the LNYV_440F/LNYV_1185R primers. Figure 3.6a shows that the H33 RNA contained LNYV RNA since the expected PCR product of ~750 bp was obtained. The quality of the H18 RNA was generally better than that of H19, however, the H18 samples was negative for LNYV (Figure 3.6a). Therefore, the H19 samples with the best RIN values were tested for LNYV. These samples H19_1,

H19_2, H19_6 and H19_7 were positive for LNYV (Figure 3.6b). H19_1 was chosen for sequencing since it had a suitable RIN and RNA concentration.
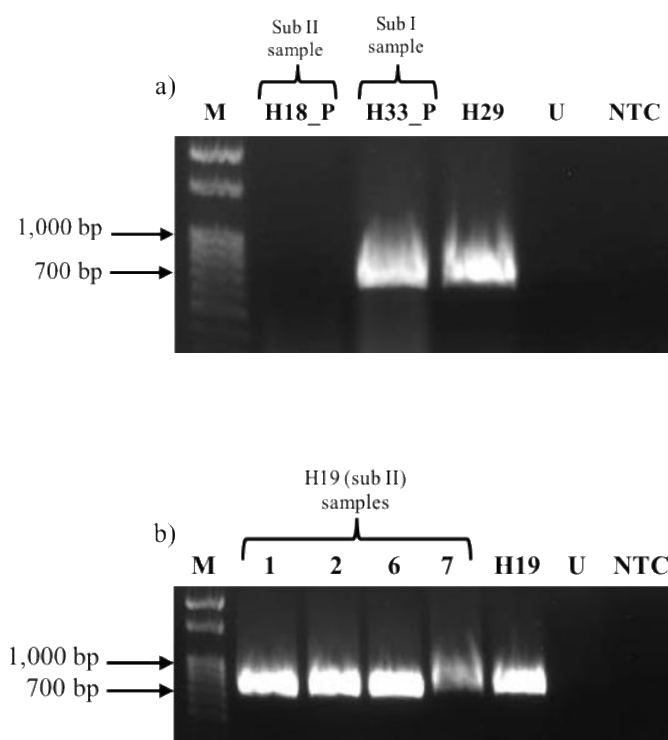
Figure 3.6: Testing of H33 (subgroup I), H18 (subgroup I) and H19 (subgroup II) samples for the presence of LNYV RNA using the LNYV_440F and LNYV_1140R primers for LNYV. a) Lane H18_P: H18 pooled (H18_4 and H18_6), Lane H33_P: H33 pooled (H33_1 and H33_4) and Lane H29: H29 (positive control). b) Lane number: H19 sample and Lane H19: H19 (positive control). Lane M: 100bp ladder (300ng), Lane U: uninfected lettuce sample and Lane NTC: no template control (negative control). Sub: subgroup

3.3.6 FastQC (Fast quality control) analysis

FastQC is one of the programmes that examines and provides read quality reports from millions of sequence reads (Korpelainen et al. 2014). It consists of five main criteria as mentioned in section 3.2.6. The sequence data obtained from NZGL contained two Fastq files (forward and reverse reads) for each subgroup sample. The background mean quality score graph is divided into three regions; red, orange and green, which represent poor-quality, reasonable quality and good quality, respectively (Figure 3.7). This figure shows that there was a sudden increase of Phred score from 32 to 36 and then a decrease of Phred score from 35 to 30 in all the reads. Nevertheless, all four reads were good quality and had a Phred score >30 in all base positions, indicating these reads were reliable for sequence analysis.
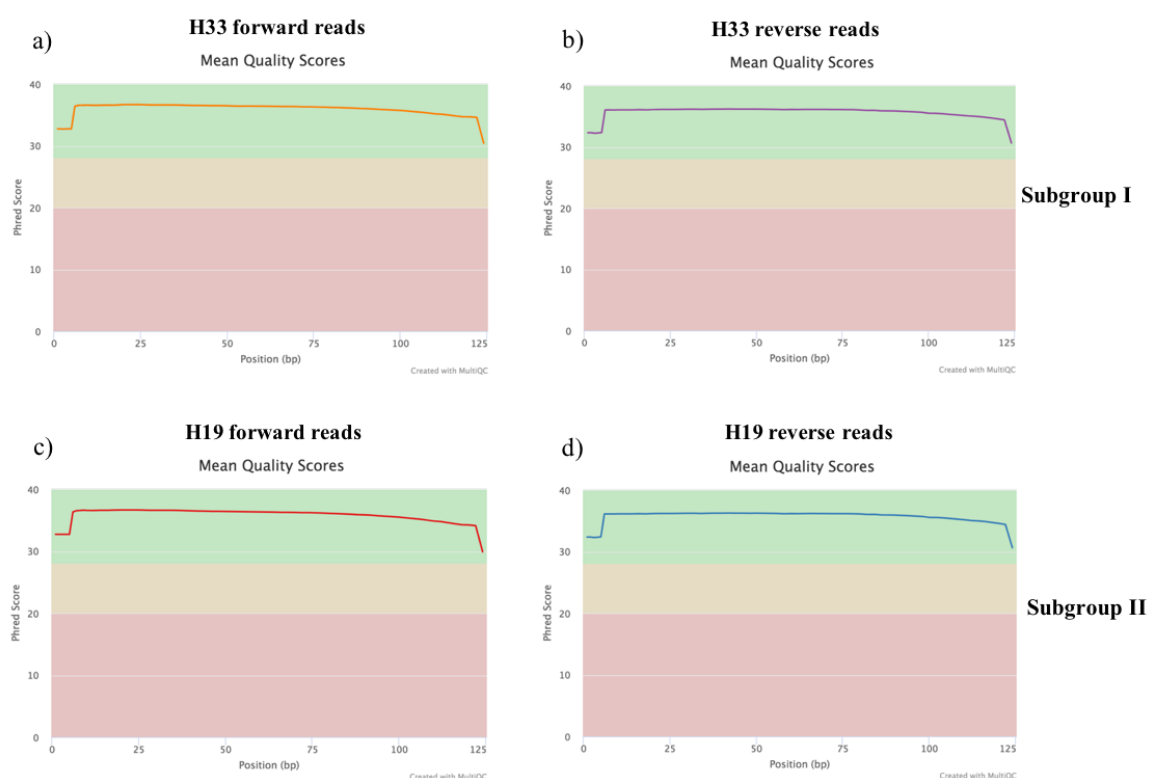
Figure 3.7: Per base sequence mean quality score plots of H33 and H19 reads by FastQC analysis. a) H33 forward reads. b) H33 reverse reads. c) H19 forward reads. d) H19 reverse reads. Green background: good quality sequences, orange background: reasonable quality sequences and red background: poor quality sequences.

The per sequence quality score helps to determine if a subset of the sequence contains poor quality sequences (Babraham Institute Enterprise n.d.-c). More than one million reads of H33 and H19 contain a Phred score of >30, signifying that there were no subsets within the reads that contained low-quality sequences (Figure 3.8). Hence, these sequence reads are appropriate for sequence analysis.
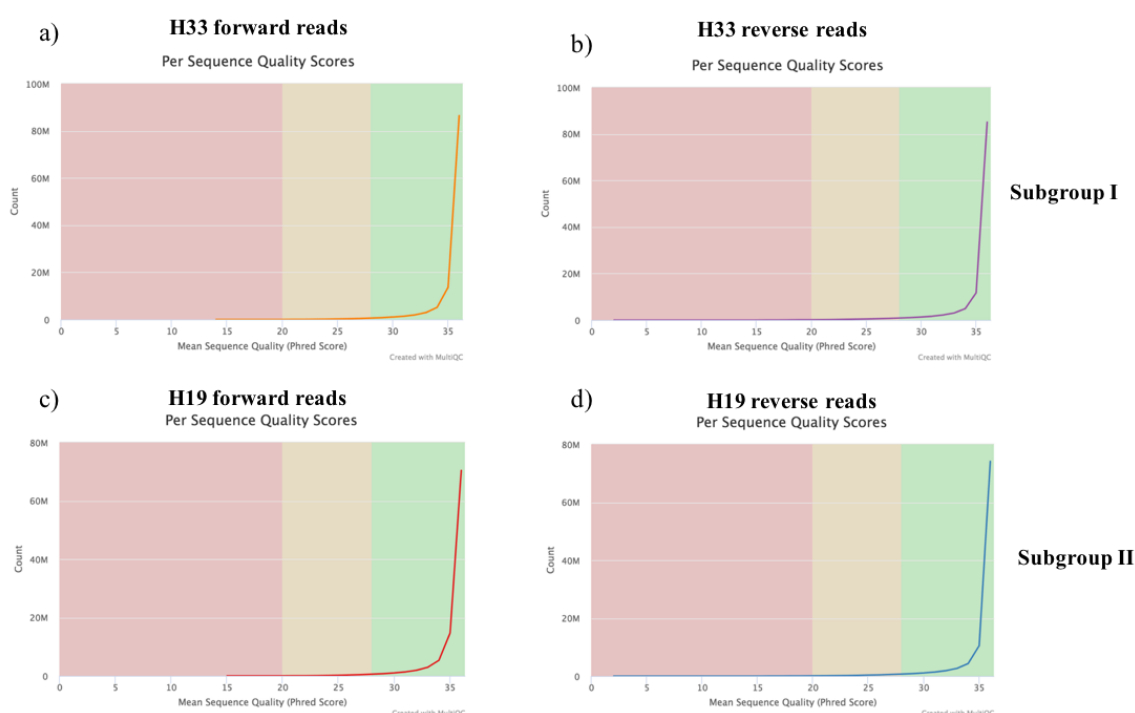
Figure 3.8: Per sequence mean quality score plots of H33 and H19 by FastQC analysis. a) H33 forward reads. b) H33 reverse reads. c) H19 forward reads. d) H19 reverse reads. Green background: good quality sequences, orange background: reasonable quality sequences and red background: poor quality sequences.

The per sequence GC content graph determines the average GC content in all the reads (Babraham Institute Enterprise n.d.-b). The forward and reverse reads of H33 and H19 are roughly normally distributed indicating that the samples were uncontaminated and reliable for analysis (Figure 3.9).
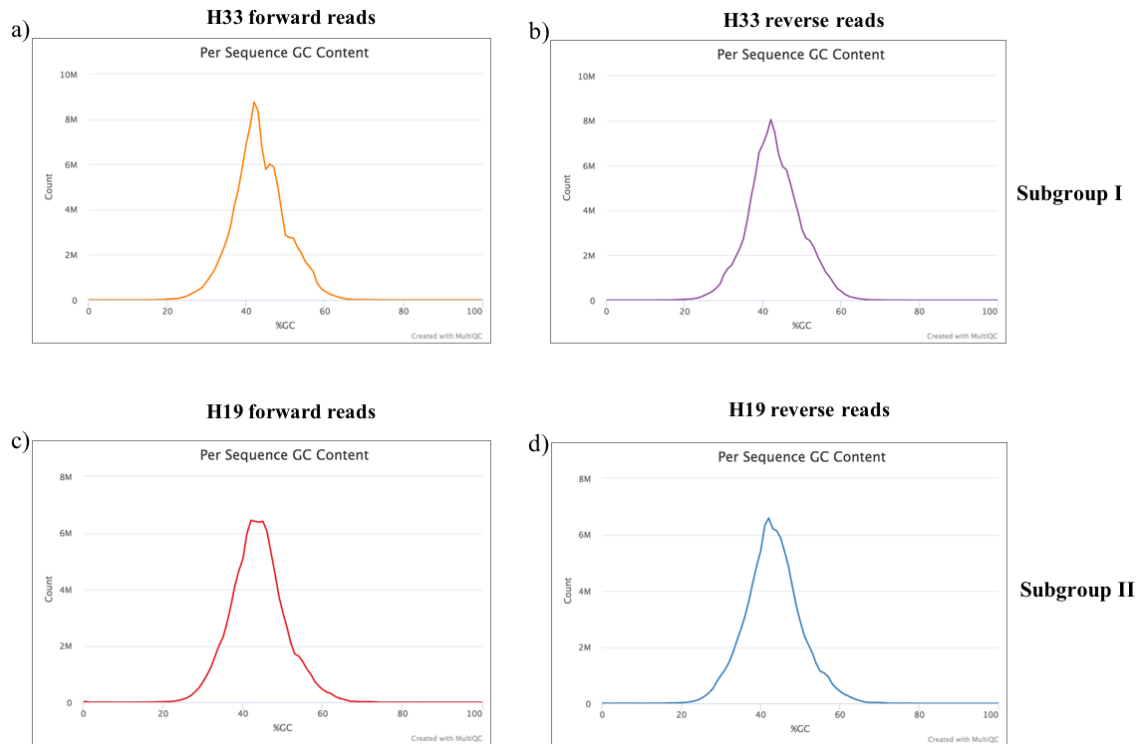
Figure 3.9: Per sequence GC distribution plots of H33 and H19 by FastQC analysis. a) H33 forward reads. b) H33 reverse reads. c) H19 forward reads. d) H19 reverse reads.

The Per base N content graph shows the use of N (any nucleotides) in a read position when there is insufficient confidence to determine the nucleotide (Babraham Institute Enterprise n.d.-a). Figure 3.10 shows that that the maximum percentage of N was 0.2% in the H33 reverse reads while the other reads had <0.2% of N count. The low percentage of N indicates that subgroup I and subgroup II had high quality reads.
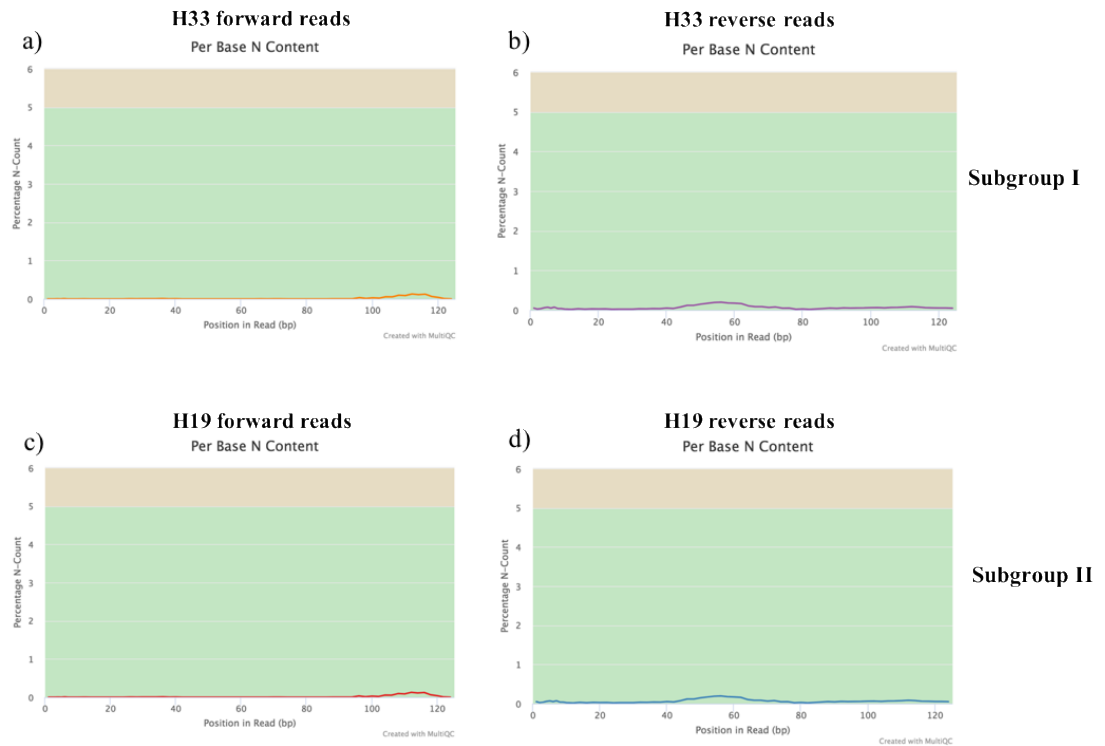
Figure 3.10: Per base N distribution plots of H33 and H19 reads by FastQC analysis. a) H33 forward reads. b) H33 reverse reads. c) H19 forward reads. d) H19 reverse reads.

Ideally, there would be only one of each sequence in a diverse DNA-seq library while this assumption is not applicable for RNA-seq libraries because RNA-seq libraries could contain highly expressed genes. Highly expressed genes would be indicated by high duplication level (Delhomme et al. 2014). Figure 3.11 shows relatively high duplication level in the H33 forward reads while others had low duplication. The identities of overrepresented sequences were not provided. These four reads satisfied all the criteria mentioned above and were reliable for the downstream analysis.
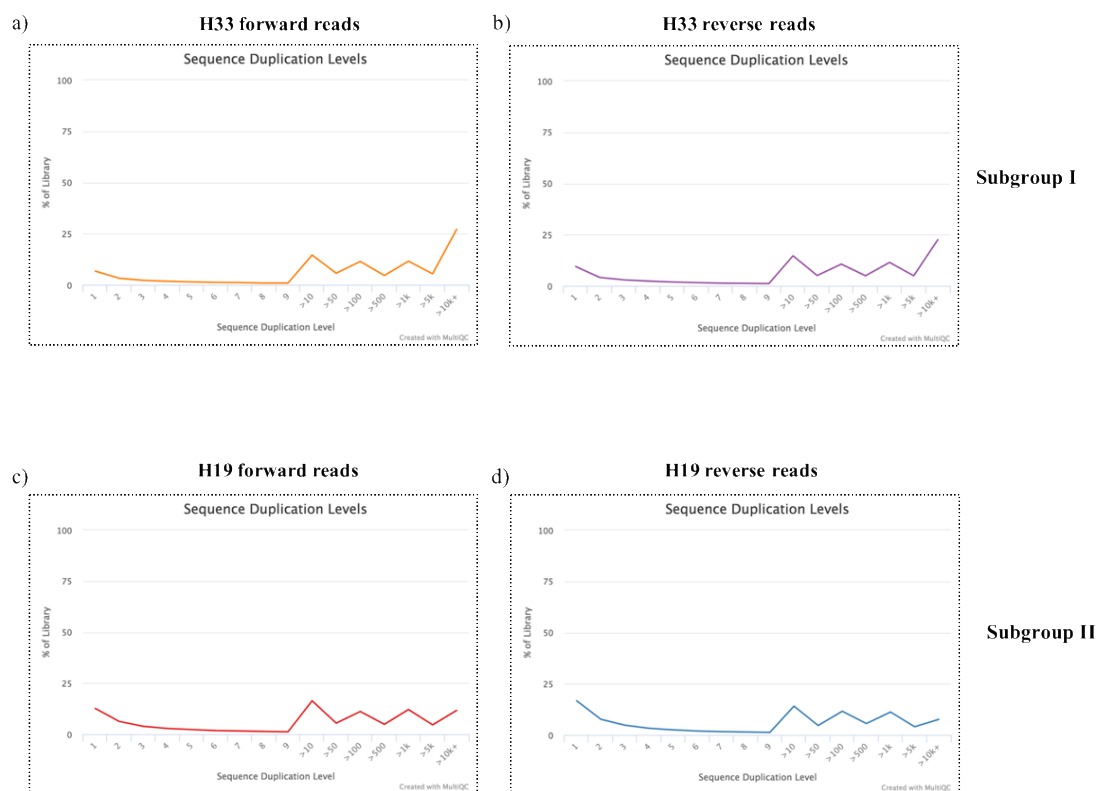
Figure 3.11: Per sequence duplication levels plots of H33 and H19 reads by FastQC analysis. a) H33 forward reads. b) H33 reverse reads. c) H19 forward reads. d) H19 reverse reads.

3.3.7 *De novo* assembly

De novo assembly is defined as the construction of sequences without the use of a reference sequence to guide the assembly (Dudley and Karczewski 2013). A summary of the assembly reports for each assembly of H33 and H19 is shown in Table 3.5. The H33 de novo assemblies generated on average; 890,303 reads, 77,272 contigs with 462 contigs ≥ 1,000 bp (Table 3.5). The H19 de novo assemblies generated on average; 963,889 reads, 126,776 contigs with 394 contigs ≥ 1000 bp (Table 3.5). The total number of reads and contigs is higher in H19 de novo assemblies is most likely due to the 10 additional assemblies compared to the H33 de novo assemblies. This was done because the LNYV subgroup II genome had not been sequenced before this and a higher number of assemblies would give greater confidence.

Table 3.5: Summary of the assembly report for each H33 and H19 *de novo* assembly.

Sample	Assembly no.	No. of Reads	Total no. of contigs	No. of contigs ≥1,000 bp
H33 (subgroup I)	1	807,766	68,713	534
	2	891,112	77,407	456
	3	894,956	77,854	438
	4	895,275	77,755	474
	5	893,925	77,482	454
	6	898,741	77,122	446
	7	890,782	77,409	441
	8	895,531	77,498	548
	9	891,685	77,698	442
	10	892,928	77,806	450
	11	890,772	77,953	457
	12	903,427	77,749	450
	13	902,846	77,944	439
	14	892,567	77,853	466
	15	901,657	77,795	444
	16	891,469	77,721	458
	17	892,957	78,001	434
	18	892,082	77,843	452
	19	892,036	77,936	441
	20	893,540	77,905	462
	Average	**890,303**	**77,272**	**459**
H19 (subgroup II)	1	963,341	125,947	401
	2	962,930	126,139	394
	3	963,222	126,527	393
	4	963,239	126,643	382
	5	963,558	126,326	409
	6	963,455	126,187	404
	7	964,158	126,443	397
	8	964,139	127,086	388
	9	964,483	126,847	410
	10	964,110	127,038	396
	11	965,155	127,393	391
	12	964,760	126,769	420
	13	965,026	127,153	376
	14	964,406	127,029	386
	15	964,732	127,026	399
	16	964,901	127,213	399
	17	964,426	127,605	404
	18	965,174	128,179	375
	19	964,738	127,850	363
	20	965,314	127,942	396
	21	964,185	126,946	399
	22	964,033	127,253	388
	23	962,189	126,513	389
	24	963,777	126,668	393
	25	962,057	126,006	388
	26	963,404	126,407	395
	27	962,177	126,012	387
	28	963,434	126,418	404
	29	962,430	126,677	399
	30	963,712	125,043	403
	Average	**963,889**	**126,776**	**394**

No. number

The LNYV contigs from the *de novo* assemblies were identified by BLASTn search. Table 3.6 shows the number of LNYV contigs were obtained for each H33 and H19 assembly. For H33 a total of 35 LNYV contigs were obtained from 20 million sequence reads. The length of the H33 contigs varied between 2,137 bp to 9,149 bp with an average of 4,508 bp (Table 3.6). The number of reads per contig for H33 varied between 3,243 and 15,039 with an average of 6,694. For H19 a total of 114 contigs were identified from the 30 million assemblies (Table 3.7). The H19 contig length varied between 702 bp and 8,791 bp with an average of 2,669 bp. The number of reads per contig for H19 varied between 3,473 and 26,890 number of reads with an average of 8,674.

Table 3.6: H33 LNYV contig number for each assembly and the number of reads per contig.

Assembly no.	LNYV contig no.	Length (bp)	No. of reads
1	Contig 7	5,993	8,009
2	Contig 8	6,843	11,319
	Contig 19	3,440	4,019
3	Contig 5	7,806	11,252
	Contig 14	2,137	4,345
	Contig 23	2,827	3,395
4	Contig 24	2,616	3,719
	Contig 26	2,641	3,306
5	Contig 7	6,171	10,218
	Contig 28	2,470	3,243
6	Contig 4	7,718	13,373
	Contig 19	3,322	3,937
7	Contig 6	9,149	15,039
	Contig 24	3,192	3,968
8	Contig 13	3,686	5,298
9	Contig 11	3,353	6,841
	Contig 12	4,906	6,491
10	Contig 7	6,029	11,185
	Contig 20	3,037	3,831
11	Contig 7	4,186	8,241
	Contig 14	4,310	5,296
12	Contig 14	3,876	4,901
13	Contig 8	6,271	7,601
14	Contig 11	4,231	8,425
	Contig 14	4,884	6,507
15	Contig 11	4,710	5,854
16	Contig 13	3,849	4,429
	Contig 15	3,226	4,185
17	Contig 5	5,577	10,105
	Contig 14	3,464	4,218
18	Contig 6	7,543	10,581
	Contig 14	3,029	6,168
19	Contig 9	4,185	7,315
	Contig 12	4,020	4,157
20	Contig 24	3,072	3,531
	Average	**4,508**	**6,694**

No. number

Table 3.7: H19 LNYV contig number for each assembly and the number of reads per contig.

Assembly no.	LNYV contig no.	Length (bp)	No. of reads
1	Contig 3	3,756	15,303
	Contig 5	2,674	9,682
	Contig 9	2,933	7,799
2	Contig 4	3,721	12,988
	Contig 13	1,444	5,101
3	Contig 4	7,657	23,688
	Contig 6	2,604	12,399
4	Contig 3	3,015	13,624
	Contig 6	3,650	9,744
	Contig 8	2,489	8,624
5	Contig 3	4,867	12,850
	Contig 6	2,105	9,871
	Contig 7	2,063	7,648
	Contig 8	2,274	6,503
	Contig 10	1,909	5,485
6	Contig 5	2,636	7,685
	Contig 6	2,086	7,141
	Contig 10	2,220	5,545
	Contig 11	1,370	5,280
	Contig 18	829	3,735
7	Contig 6	2,515	10,922
	Contig 7	3,446	8,855
	Contig 12	1,659	4,753
	Contig 18	1,659	3,743
	Contig 20	959	3,538
8	Contig 4	3,065	10,766
	Contig 5	2,060	7,747
	Contig 8	1,749	7,272
	Contig 16	1,563	3,921
	Contig 18	1,339	3,516
	Contig 19	702	3,473
9	Contig 4	3,437	11,471
	Contig 7	3,027	7,847
	Contig 10	2,362	6,811
	Contig 12	1,621	5,661
	Contig 14	1,110	5,342
10	Contig 6	2,962	9,429
	Contig 8	1,753	8,561

Assembly no.	LNYV contig no.	Length (bp)	No. of reads
10	Contig 9	3,066	8,437
	Contig 11	2,786	7,250
	Contig 13	1,399	5,070
11	Contig 6	2,091	9,729
	Contig 8	3,144	8,752
	Contig 10	2,382	6,972
	Contig 12	1,602	6,111
	Contig 14	1,942	5,642
12	Contig 3	4,861	18,714
	Contig 7	2,004	8,074
	Contig 11	2,409	6,200
	Contig 14	2,283	5,509
	Contig 17	1,562	4,356
13	Contig 5	3,692	10,877
	Contig 7	1,957	9,099
	Contig 8	2,849	6,984
	Contig 10	2,129	6,459
	Contig 13	1,353	5,313
14	Contig 5	2,508	11,659
	Contig 8	2,614	7,504
	Contig 9	2,781	7,424
	Contig 11	1,620	5,726
	Contig 13	2,067	5,310
15	Contig 4	4,630	14,866
	Contig 6	2,907	12,326
	Contig 10	3,326	9,157
16	Contig 4	4,304	11,345
	Contig 8	1,395	6,181
	Contig 11	1,996	5,610
	Contig 12	1,387	5,219
	Contig 14	2,006	4,856
	Contig 15	1,245	4,727
17	Contig 5	4,494	12,116
	Contig 6	1,805	8,735
	Contig 8	2,210	7,526
	Contig 13	1,374	5,033
	Contig 17	1,492	4,276
18	Contig 4	2,667	12,561

Assembly no.	LNYV contig no.	Length (bp)	No. of reads
18	Contig 5	3,960	10,541
	Contig 11	1,933	6,001
	Contig 13	1,412	5,411
	Contig 18	1,590	3,655
19	Contig 6	2,517	11,642
	Contig 9	3,178	8,428
	Contig 12	2,663	6,680
	Contig 13	1,990	6,252
	Contig 17	1,134	4,480
20	Contig 3	3,609	15,293
	Contig 4	5,227	13,545
	Contig 6	2,008	7,732
	Contig 16	1,547	4,454
21	Contig 3	3,990	13,321
	Contig 5	2,985	7,711
	Contig 9	1,565	5,582
22	Contig 4	4,263	16,986
	Contig 10	3,170	8,397
23	Contig 5	1,957	9,005
	Contig 6	2,726	8,848
	Contig 7	3,286	8,565
	Contig 10	2,011	6,128
24	Contig 3	3,384	13,326
	Contig 5	4,297	11,056
25	Contig 4	5,685	16,412
	Contig 10	1,554	6,402
26	Contig 7	2,423	7,499
	Contig 8	2,881	7,364
27	Contig 3	4,534	17,196
	Contig 4	3,300	12,558
	Contig 5	3,552	9,606
28	Contig 4	3,357	11,432
	Contig 5	2,257	10,458
	Contig 6	3,851	9,880
29	Contig 3	5,408	5,560
	Contig 5	3,405	12,010
30	Contig 3	8,791	26,890
	Contig 8	1,263	6,467

No. number

Average contig length: 2,669
Average number of reads: 8,674

3.3.8 Pathway A: *de novo* assembly of LNYV contigs and alignment of LNYV scaffolds

The 35 contigs obtained from the LNYV H33 *de novo* assemblies were reassembled by *de novo* assembly to generate three scaffolds. Scaffold 1, scaffold 2 and scaffold 3 consisted of 21 contigs, 8 contigs and 6 contigs, respectively (Figure 3.12). Medium sensitivity was used to obtain a fast *de novo* assembly. Higher sensitivity could increase the accuracy of the assembly. Fewer and longer scaffolds will be assembled in a higher sensitivity (Geneious 2013). The three scaffolds were aligned by Geneious alignment to obtain the complete H33 genome consensus sequence (Figure 3.12). This figure shows that there was low coverage and unidirectional reads in each of the contigs (only one directional reads) in the 1 bp to 472 bp region. The figure also shows the region from 2,409 bp to 3,796 bp also contained unidirectional reads in each of the contigs while the read coverage decreases between positions 12,192 bp to 12,766 bp (Figure 3.12). There were high coverage and bidirectional reads in each of the contigs (reads from both directions) in all other regions of the genome. Coverage has a proportional with number of aligned sequence reads and confidence in the base calls (Illumina 2018). More reads need to be analysed to obtain higher coverage could increase the confidence of the nucleotide positions in the sequence. The reads in each contigs were checked individually and all contigs contained bidirectional reads. It is unclear why certain regions within the contigs contain unidirectional reads after the assembly of all LNYV contigs.
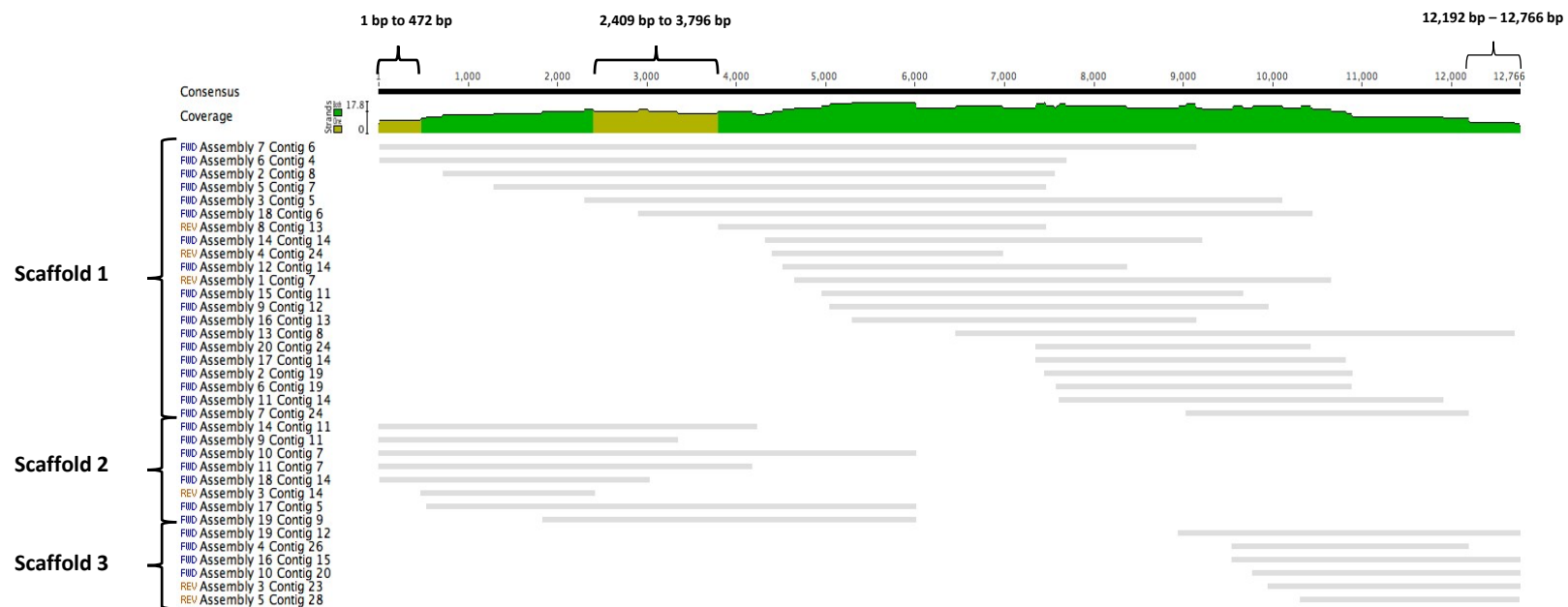
Figure 3.12: Geneious alignment of LNYV H33 subgroup I scaffolds and scaffolds contains the LNYV contigs. Gray: LNYV contigs. Green: bidirectional reads. Brown: unidirectional reads. FWD: forward direction. REV: reverse direction. Numbers at top the figure indicate nucleotide position.

A total of 114 LNYV H19 subgroup II contigs was reassembled by *de novo* assembly to generate four scaffolds. Scaffold 1, scaffold 2, scaffold 3 and scaffold 4 consist of 57 contigs, 51 contigs, 4 contigs and 2 contigs, respectively (Figure 3.13). The four scaffolds were aligned by Geneious alignment to obtain the complete H19 consensus sequence of the genome (Figure 3.13). Figure 3.13 shows the region from nucleotide position 1 to 116 contained unidirectional reads in each of the contigs, and the coverage decreases from 12,574 bp to 12,804 bp. All other regions had high coverage made up of bidirectional reads in each of the contigs. As mentioned above, more *de novo* assemblies are required to obtain higher coverage which could increase the confidence in base calls.
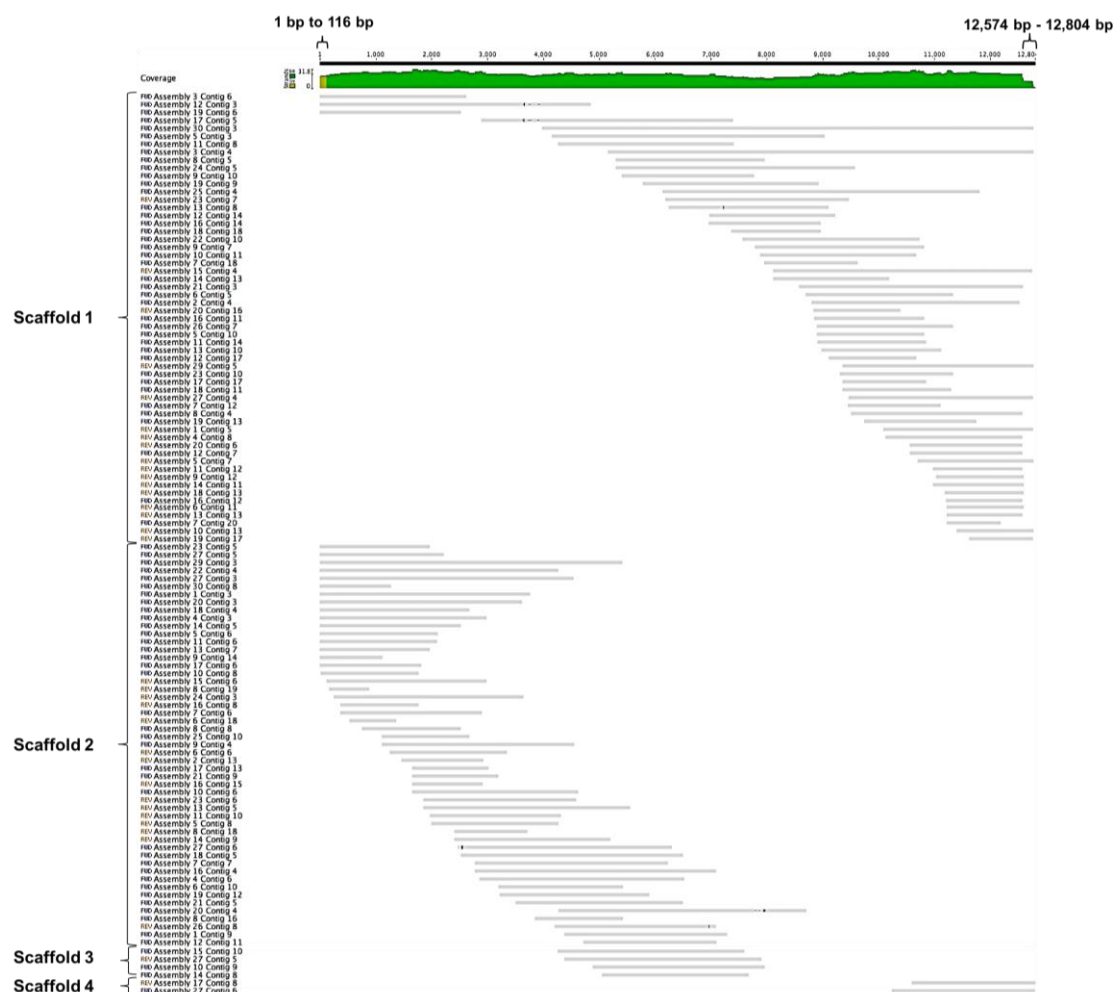
Figure 3.13: Geneious alignment of LNYV H19 subgroup II scaffolds and scaffolds contains the LNYV contigs. Gray: LNYV contigs. Green: bidirectional reads. Brown: unidirectional reads. FWD: forward direction. REV: reverse direction. Numbers at top the figure indicate nucleotide position

3.3.9   Pathway B: Reference assembly

Reference assembly utilises a reference sequence to construct and guide the sequence assembly (Dudley and Karczewski 2013). LNYV H33 subgroup I contigs identified from the *de novo* assembly were reassembled by reference assembly using the published LNYV genome as the reference. The H33 reference assembly shows similar results to the pathway A H33 alignment in section 3.3.8 (Figure 3.14). There was low coverage and unidirectional reads in each of the contigs from nucleotide position 1 to 287, the sequence in this region encodes the 3' leader and N gene (Figure 3.14). The nucleotide sequence between 2,419 bp to 3,807 bp region encoding the P gene and 4b gene contained unidirectional reads in each of the contigs. There was a decrease of coverage from position 12,235 to 12,800, which encodes the L gene and 5' leader. All other regions had high coverage and bidirectional reads.
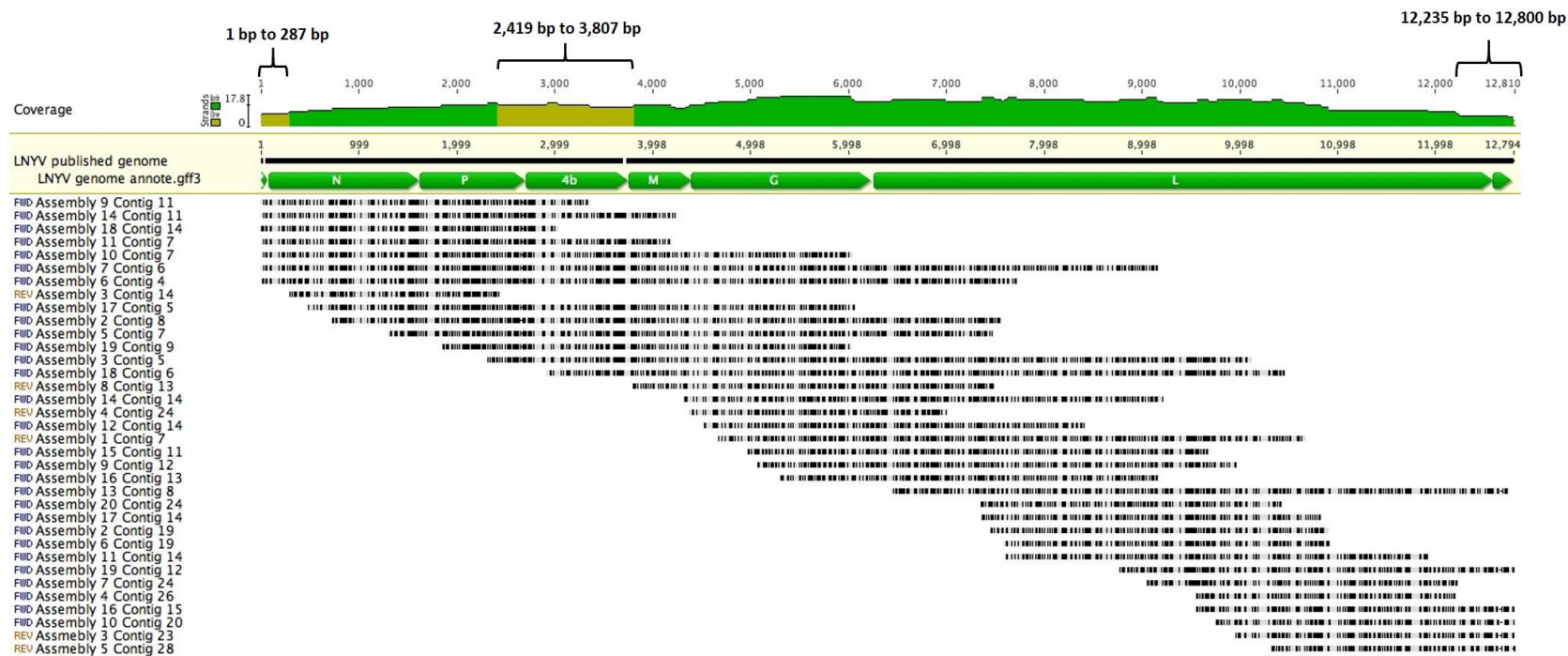
Figure 3.14: Reference assembly of LNYV H33 subgroup I contigs to the published genome. Gray: LNYV contigs. Green: bidirectional reads. Brown: unidirectional reads. FWD: forward direction. REV: reverse direction. Numbers at top the figure indicate nucleotide position. Black colour within the alignment indicate where the nucleotides differ from the published sequence.

LNYV H19 subgroup II contigs identified from the *de novo* assembly were also reassembled by reference assembly and similar results to the pathway A H19 alignment were obtained. Figure 3.15 shows that the region from nucleotide position 1 to 140 had unidirectional reads in each of the contigs while there was a decrease in coverage from 12,598 bp to 12,851 bp. All other regions had high coverage and bidirectional reads in each of the contigs. A higher number of *de novo* assemblies are required to obtain higher coverage and bidirectional reads. The subgroup I and subgroup II consensus genome sequences obtained from reference assembly were compared with the consensus genome sequence obtained from the pathway A to determine which consensus sequence would be used for the subsequent analysis.
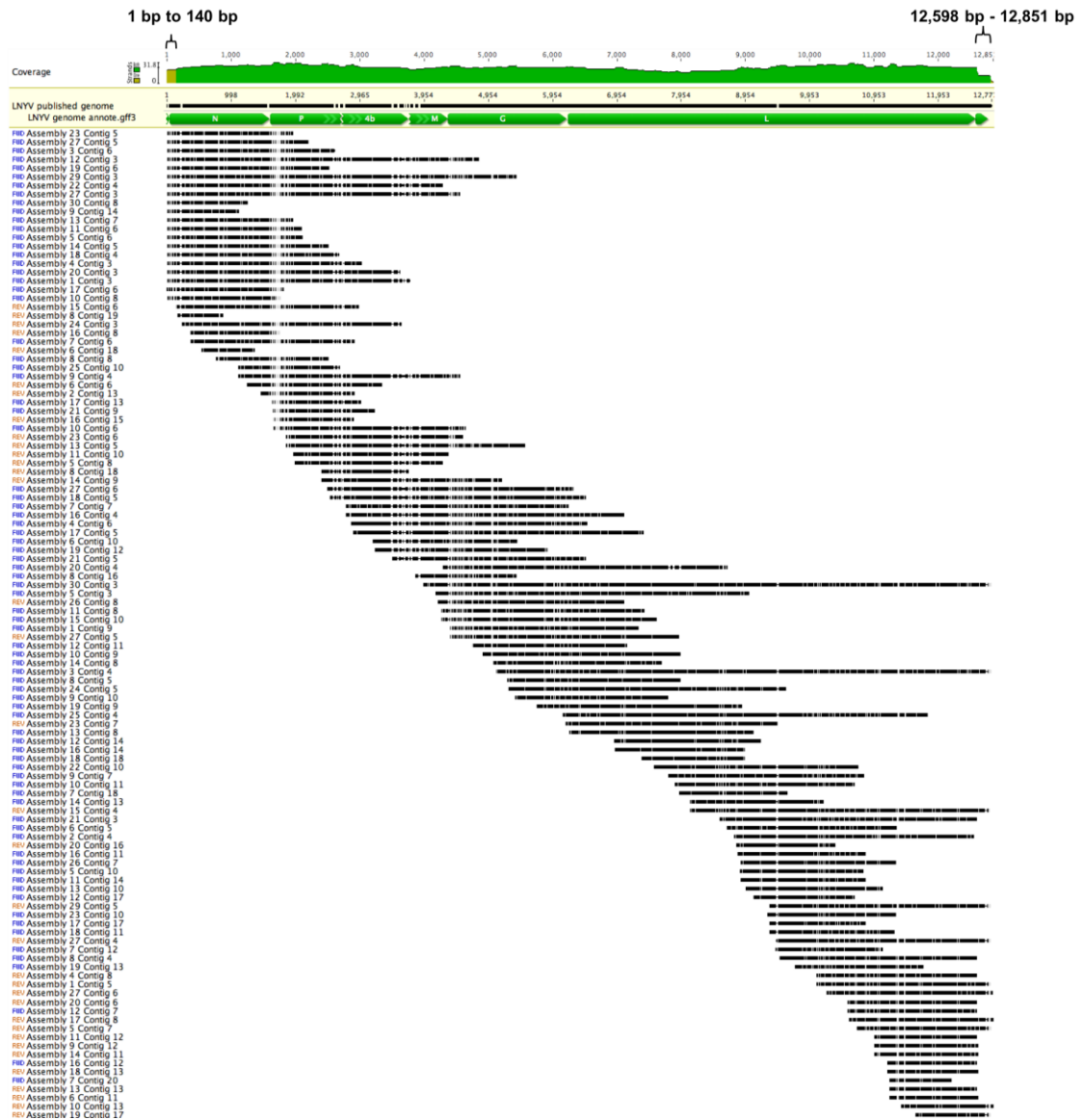
Figure 3.15: Reference assembly of LNYV H19 subgroup II contigs to the published genome. Gray: LNYV contigs. Green: bidirectional reads. Brown: unidirectional reads. FWD: forward direction. REV: reverse direction. Numbers at top the figure indicate nucleotide position. Black colour within the alignment indicate where the nucleotides differ from the published sequence.

3.3.10 Comparison of consensus genome sequence from pathway A and pathway B

The LNYV H33 subgroup I and H19 subgroup II consensus genome sequences from pathway A were compared with the consensus genome sequences from pathway B. The MUSCLE alignment of the H33 consensus sequences obtained from pathways A and B is shown in Figure 3.16a. This figure shows that the pathway B consensus sequences had ten additional nucleotides at beginning of the 3' leader sequence and three additional nucleotides at the end of the 5' leader sequence, indicating that these consensus sequences were not identical. These additional 13 nucleotides were also identified in the published LNYV subgroup I genome sequence. More *de novo* assemblies for this isolate may resolve this issue. Therefore, it was more reliable to use the consensus sequence from the reference assembly (pathway B) for the subsequent analysis. Figure 3.16b shows the MUSCLE alignment of H19 consensus sequences obtained from pathway A and B. This figure shows that both consensus sequences are 100% identical. Therefore, the choice of pathway does not affect the analysis. The H19 consensus sequence obtained from the pathway B was used for the subsequent analysis to keep the analyses consistent with the subgroup I analysis.

Figure 3.16: MUSCLE alignment of H33 and H19 consensus sequence from pathway A and pathway B. a) H33 alignment. b) H19 alignment. Light and dark green: 100% pairwise identity. Numbers at top the figure indicate nucleotide position.

3.3.11 LNYV genome sequence annotation and analysis

The LNYV H33 subgroup I and H19 subgroup II genome sequences from pathway B were annotated with the 3' leader, genes, ORFs and 5' leader using the published LNYV genome sequence as reference (Figure 3.17). The number of nucleotides in the 5'UTR, gene, 3'UTR and the number of amino acids in the ORF for each gene in H33, published and H19 genome sequence is shown in Table 3.8. The table shows that the number of nucleotides and amino acids is identical in the H33, published and H19 G and L genes. For the M and N genes, the number of nucleotides and amino acids is identical between subgroup I and the published sequences. The published P gene and H33 4b gene has an

additional nucleotide. The subgroup II P, 4b and M genes vary in the number of nucleotides. The total number of nucleotides within the H33, published and H19 genomes is 12,779 nt, 12,807 nt and 12,804 nt, respectively. The GC content for published and H33 genomes is 42.9%, while the H19 genome contained a GC content of 43.2%. The significance of the nucleotide and GC content variations in the subgroup II to the protein structures is unknown.
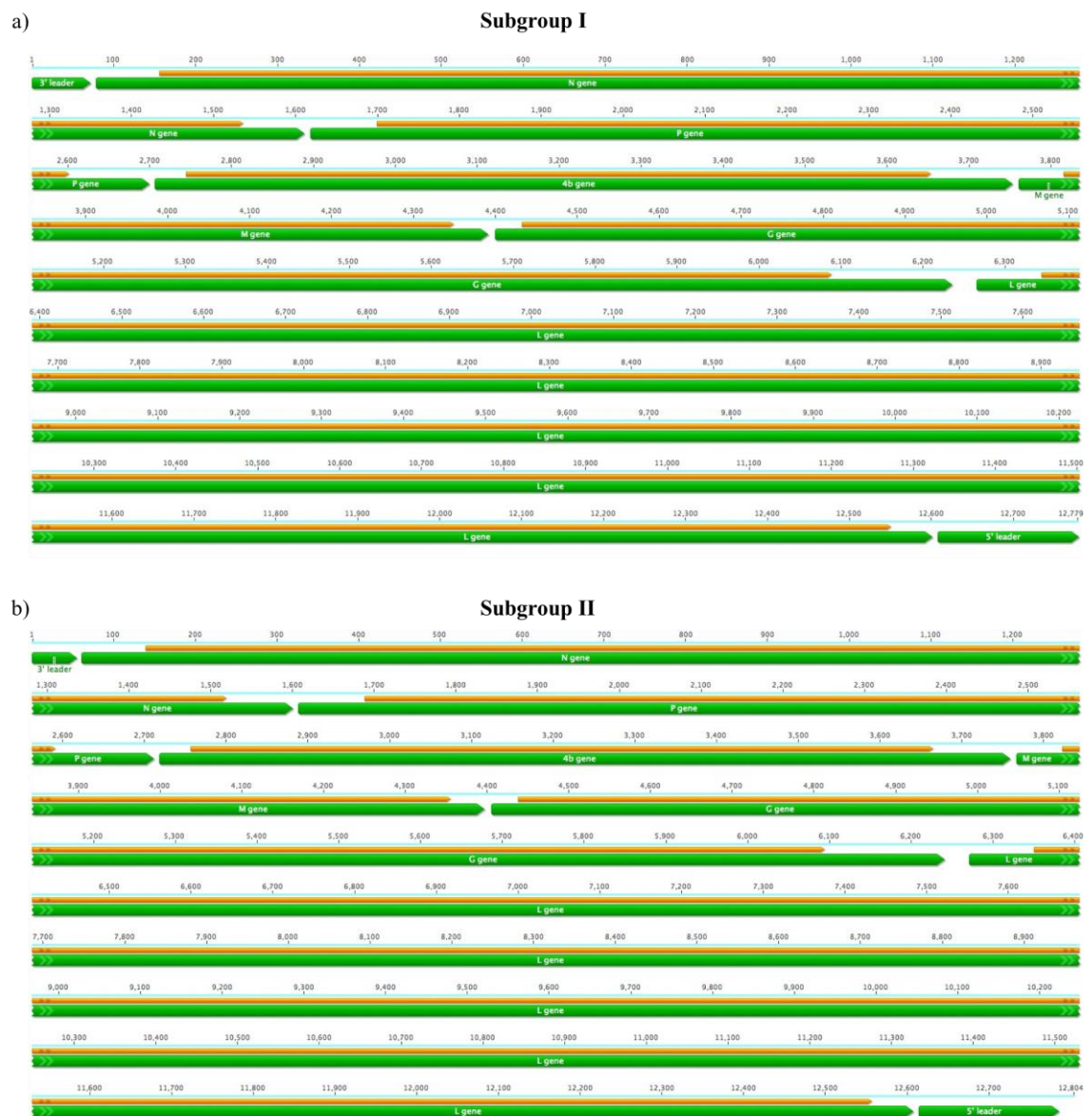
Figure 3.17: Annotated H33 and H19 genome sequences from pathway B. a) H33 annotation. b) H19 annotation. Green region: genes. Orange region: ORF. Black region: sequence. Numbers at top the figure indicate nucleotide position.

Table 3.8: Summary of nucleotides and amino acid in each gene.

Genes	Gene regions	Subgroup I		Subgroup II
		H33	Published	H19
N gene	5'UTR (nt)	78	78	78
	Gene (nt)	1533	1533	1539
	ORF (amino acid excluding stop codon)	459	459	459
	3'UTR (nt)	78	78	84
P gene	5'UTR (nt)	81	81	81
	Gene (nt)	1081	1082	1104
	ORF (amino acid excluding stop codon)	300	300	300
	3'UTR (nt)	100	101	123
4b gene	5'UTR (nt)	38	38	38
	Gene (nt)	1047	1046	1041
	ORF (amino acid excluding stop codon)	302	302	302
	3'UTR (nt)	103	102	97
M gene	5'UTR (nt)	55	55	56
	Gene (nt)	631	631	631
	ORF (amino acid excluding stop codon)	177	177	177
	3'UTR (nt)	45	45	44
G gene	5'UTR (nt)	33	33	33
	Gene (nt)	1836	1836	1836
	ORF (amino acid excluding stop codon)	551	551	551
	3'UTR (nt)	150	150	150
L gene	5'UTR (nt)	79	79	79
	Gene (nt)	6336	6336	6336
	ORF (amino acid excluding stop codon)	2068	2068	2068
	3'UTR (nt)	53	53	53
GC content		42.9%	42.9%	43.2%
Total genome (nt)		12,779	12,807	12,804

3.3.12   Phylogenetic analysis

Phylogenetic analysis of nucleic acid or amino acid sequences assists in comprehending the evolutionary relationships of organisms (Choudhuri 2014). NJ and ML tree models are among the commonly used tree-building models. Both phylogenetic analyses were used on the complete LNYV genomes from pathway B along with the available cytorhabdovirus and nucleorhabdovirus genomes from the NCBI database. Very similar results were obtained for both trees (Figure 3.18 and 3.19). Both Figure 3.18 and 3.19 show two distinct clades for cytorhabdoviruses and nucleorhabdoviruses. The cytorhabdovirus clade was divided into two sister clades. One clade consisted of viruses that infect monocots, CBDaV, BYSMV and NCMV, while the other had mostly viruses from dicots and viruses from insects whose hosts are unknown (Figure 3.18 and 3.19).

Within the predominately dicot infecting clade, the assembled LNYV genome clustered together with the published genome, as expected. The two LNYV subgroup I genomes are closer to each other than they are to the subgroup II genome (Figure 3.18 and 3.19). It also shows subgroup II genome branched off earlier than subgroup I. When comparing to other cytorhabdoviruses, LNYV is most closely related to LYMoV (bootstrap value of 100) and with a monophyletic origin (Figure 3.18 and 3.19) suggesting a common ancestor. The bootstrap of 100% indicates that there is very high agreement or confidence in the clade (Holmes 2003). This relationship has been observed before (Bejerman et al. 2015; Higgins et al. 2016a; Higgins et al. 2016b; Yang et al. 2016)

Pairwise percent sequence identity was obtained to determine the similarity between two sequences while pairwise distance helps to determine the evolutionary distance. The average percent identity of nucleorhabdoviruses was between 26.1% - 58.8% and cytorhabdoviruses were between 29.6% - 93.3% (Figure 3.18). This high percentage of identity >90% for cytorhabdoviruses was due to the high sequence similarity between the three LNYV genomes. The LNYV genome percentage identity was between 80.3% - 93.3% (Figure 3.18). The pairwise distance for nucleorhabdoviruses was between 0.655 – 2.747 and cytorhabdoviruses were between 0.068 – 2.271 (Figure 3.19). The LNYV genome pairwise distance was between 0.068 – 0.242 (Figure 3.19). The lowest pairwise distance is 0.068 and is the distance between LNYV subgroup I genomes, indicating that these two genomes are closest relative to the others.
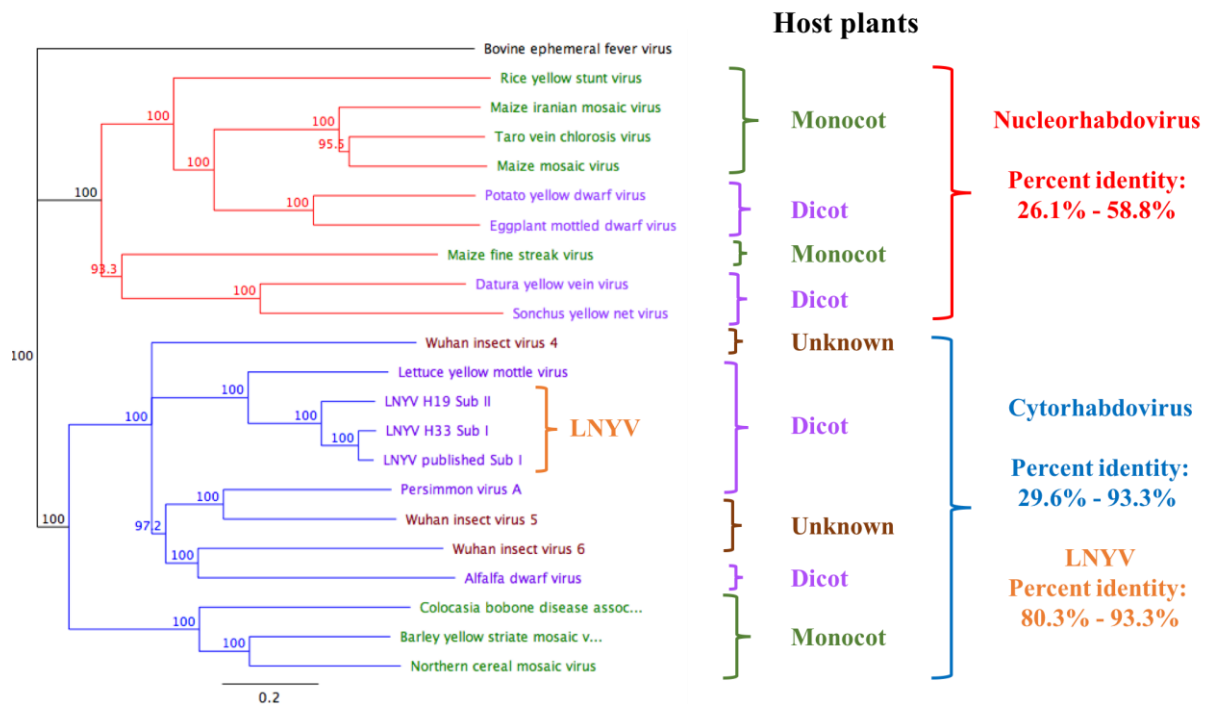
Figure 3.18: Neighbour joining of LNYV genome sequences with the available the cytorhabdovirus and nucleorhabdovirus genome sequences with an outgroup of Bovine ephemeral fever virus. Matched with the host plants Yang et al. (2016) and percent identity with 1,000 bootstrap replicates. The number of substitutions per site was denoted by the scale.
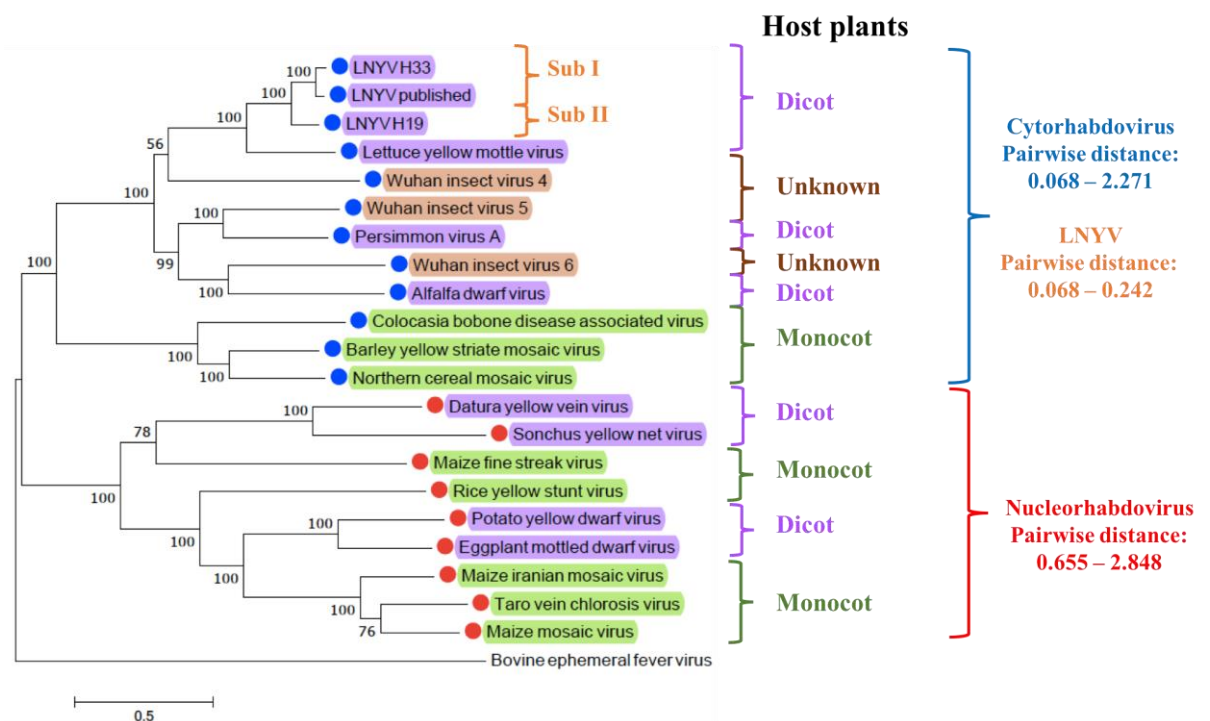
Figure 3.19: Maximum likelihood tree of cytorhabdovirus, nucleorhabdovirus and LNYV genome sequences with 1000 bootstrap replicates with an outgroup of Bovine ephemeral fever virus using GTR + G + I model. The number of substitutions per site was denoted by the scale.

ML analysis was performed on all available N gene sequences of LNYV. Figure 3.20 shows two discrete clades corresponding to LNYV subgroup I and subgroup II with a 99% bootstrap value. It also shows that subgroup I appeared before subgroup II. As expected H33 and H19 N genes derived form the whole genomes assembled in this study were identical to that of published NZ6 and NZ1, respectively, since these were the source isolates for the genome sequence. Isolates from New Zealand and Australia form separate clades. Further, within the subgroup I clade, the isolates form each region forms a separate sub-clade. Pairwise distances for subgroup I was between 0 – 0.228 and for subgroup II was between 0.002 – 0.218 (Figure 3.20). This indicated that the distance within the subgroups was low.

ML analysis was also performed on the N gene amino acid sequences of the published LNYV isolates, H19 and H33 isolates. Figure 3.21 shows that the LNYV isolates were separated into two clades: subgroup I, subgroup II with AU9 appearing to be most closely related to the common ancestor. The pairwise distance for subgroup I in the amino acid phylogeny was between 0 – 0.019 and subgroup II was between 0 – 0.01 (Figure 3.21). These distances were very low when compared to the nucleotide phylogeny, indicating greater conservation of amino acid sequence than nucleotide sequence.
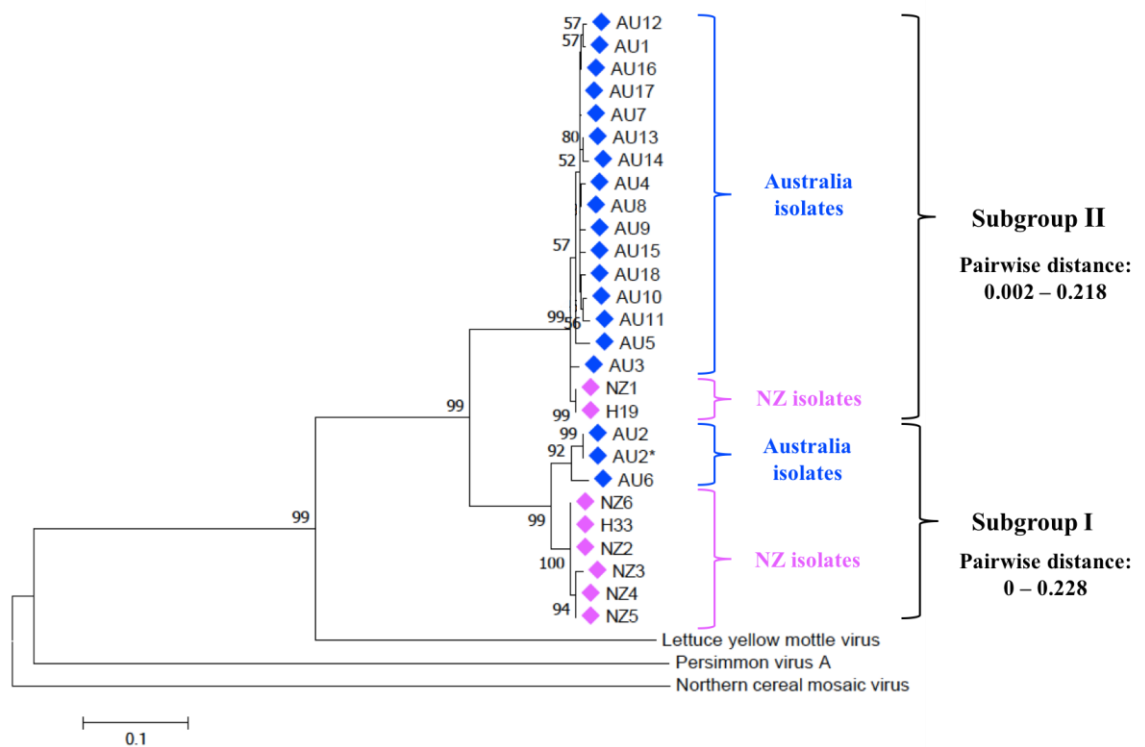
Figure 3.20: The maximum likelihood tree of LNYV subgroup I and subgroup II N gene nucleotide sequences using the Tamura-3-paramter + G model. The outgroup was Northern cereal mosaic virus with 1000 bootstrap replicates. Bootstrap of >50 were shown in the nodes. The number of substitutions per site was denoted by the scale.
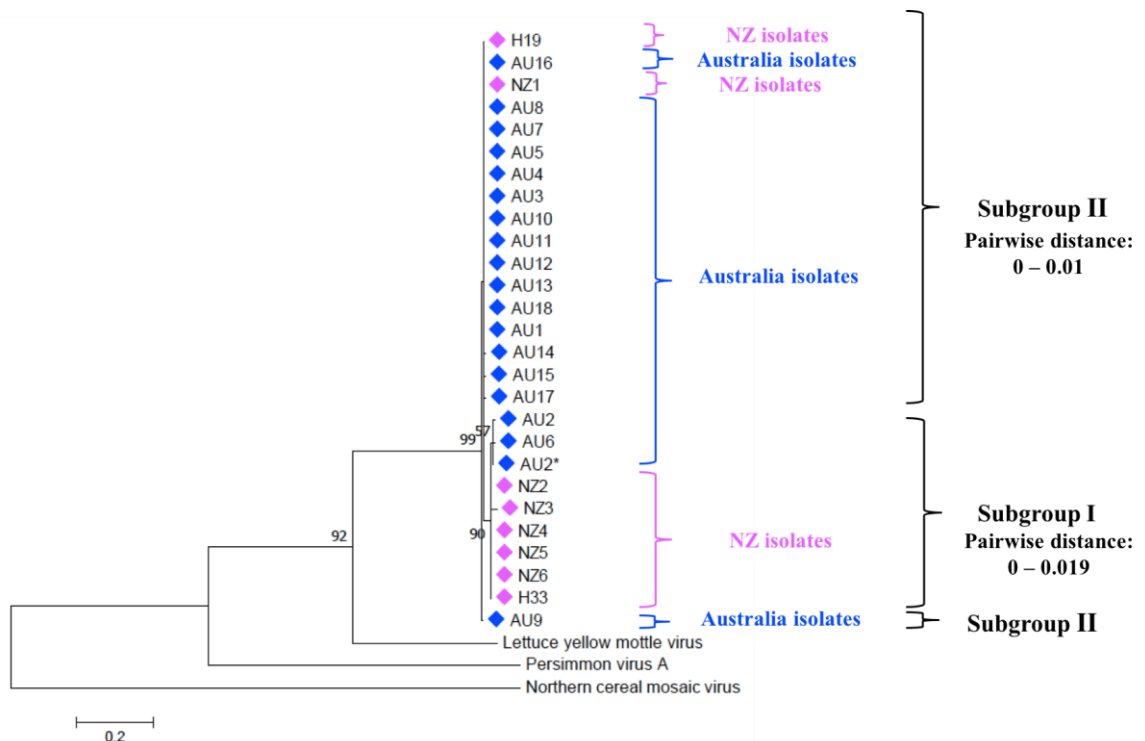
Figure 3.21: The maximum likelihood tree of LNYV subgroup I and subgroup II N gene amino acid sequences. The outgroup was Northern cereal mosaic virus with 1000 bootstrap replicates using the LG + G model. Bootstrap of >50 were shown in the nodes. The number of substitutions per site was denoted by the scale.

3.3.13 Glycoprotein analysis

Protein secondary (2D) and tertiary (3D) structures can be predicted based on the amino acid sequence, which in turn can help with predicting protein function (Gromiha 2010). The glycoprotein (G) gene from subgroup I (H33 and published LNYV) and subgroup II (H19) isolates were analysed. The predicted 5' leader, polyadenylation site, putative signal peptide, glycosylation site, peptide recognition sequence and 3' trailer sites were identified in G gene sequences of H33, published LNYV and H19 based on the information from Callaghan (2005) (Figure 3.22). This figure shows that polyadenylation sequence 'AUUAAA' was identified 95 nt downstream of published G stop codon (position 1,784 nt) and this sequence was identified in two locations (5' leader and near the centre of the gene) in the H19 G gene sequence. However, this sequence was not identified at all in the H33 G gene sequence. A potential polyadenylation signal sequence 'AUUGAA' was identified in H33 G gene sequence at the position 1,784 nt and another potential signal sequence 'AUUUAA' at the position 1,826 nt was also identified (Figure 3.23). In the H19 G gene sequence, potential polyadenylation signal sequences 'AUUAGA' and 'GUUGAA' were identified in positions 1,700 nt and 1,784 nt, respectively (Figure 3.23). These sequences could the potential polyadenylation signal sequences in H33 and H19 G gene sequences.

The putative signal peptide (25 amino acid) was identified in all three sequences in the same location, but there were differences in the amino acid sequences. Figure 3.24 shows that predicted putative signal peptide sequences of subgroups I and II differ by five amino acids. In contrast, the subgroup I sequences of H33 and the published LNYV putative signal peptide differed by only one amino acid. The recognition sequence for cleavage of the signal peptide, VQG↓V (the arrow is the predicted cleavage site), (Dietzgen et al. 2007) was conserved in all of the sequences (Figure 3.22). The potential glycosylation sequence Asn-X-Ser/Thr (Dietzgen et al. 2007) was identified in three locations in all three sequences (Figure 3.22), one was located near the 5' leader region and two near the centre of the gene. An additional glycosylation site was identified in the H33 G gene sequence near the 3' trailer region.
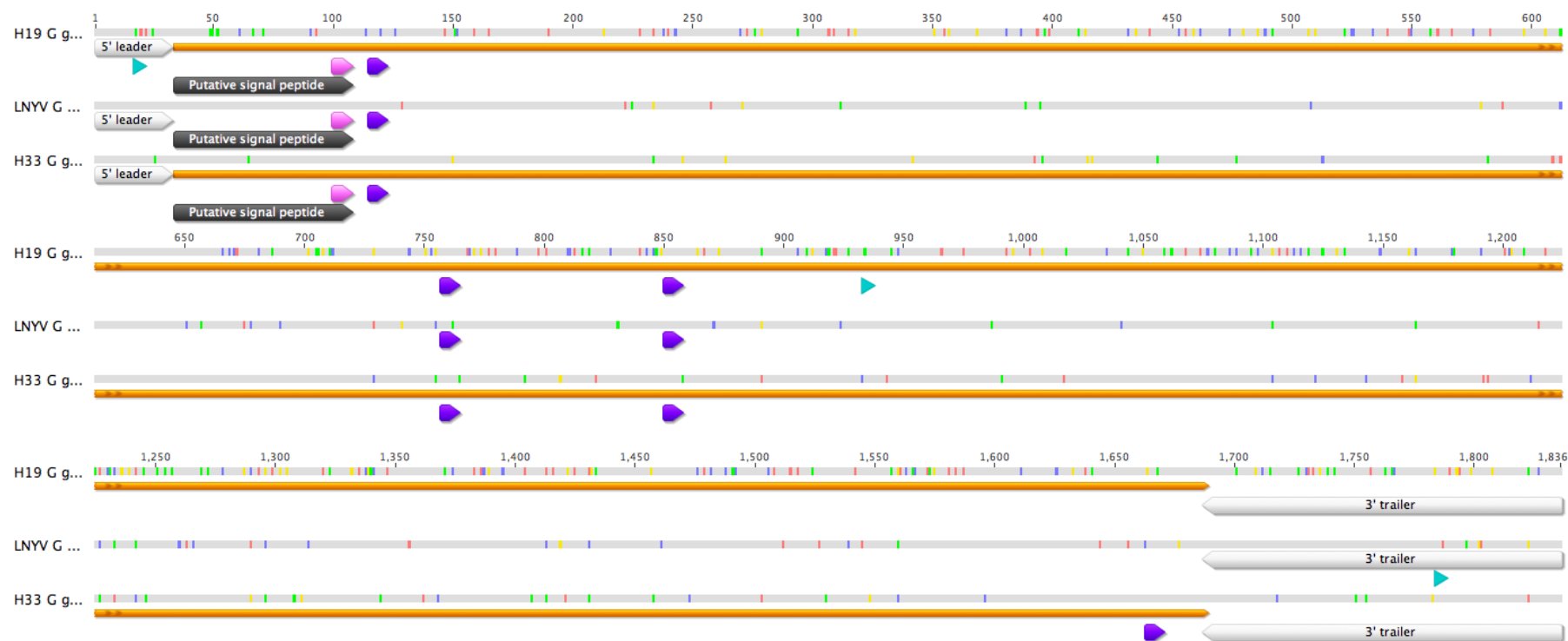
Figure 3.22: The annotated H19 (subgroup II), H33 (subgroup I) and LNYV published (subgroup I) G gene sequences. Coloured nucleotides indicate where the nucleotides differ. Numbers at top the figure indicate nucleotide position.

White region: 5' leader and 3'trailer   Turquoise region: Predicted polyadenylation site   Orange region: ORF

Purple region: Predicted glycosylation site   Dark grey region: Predicted putative signal sequence   Pink region: Predicted peptidase recognition sequence
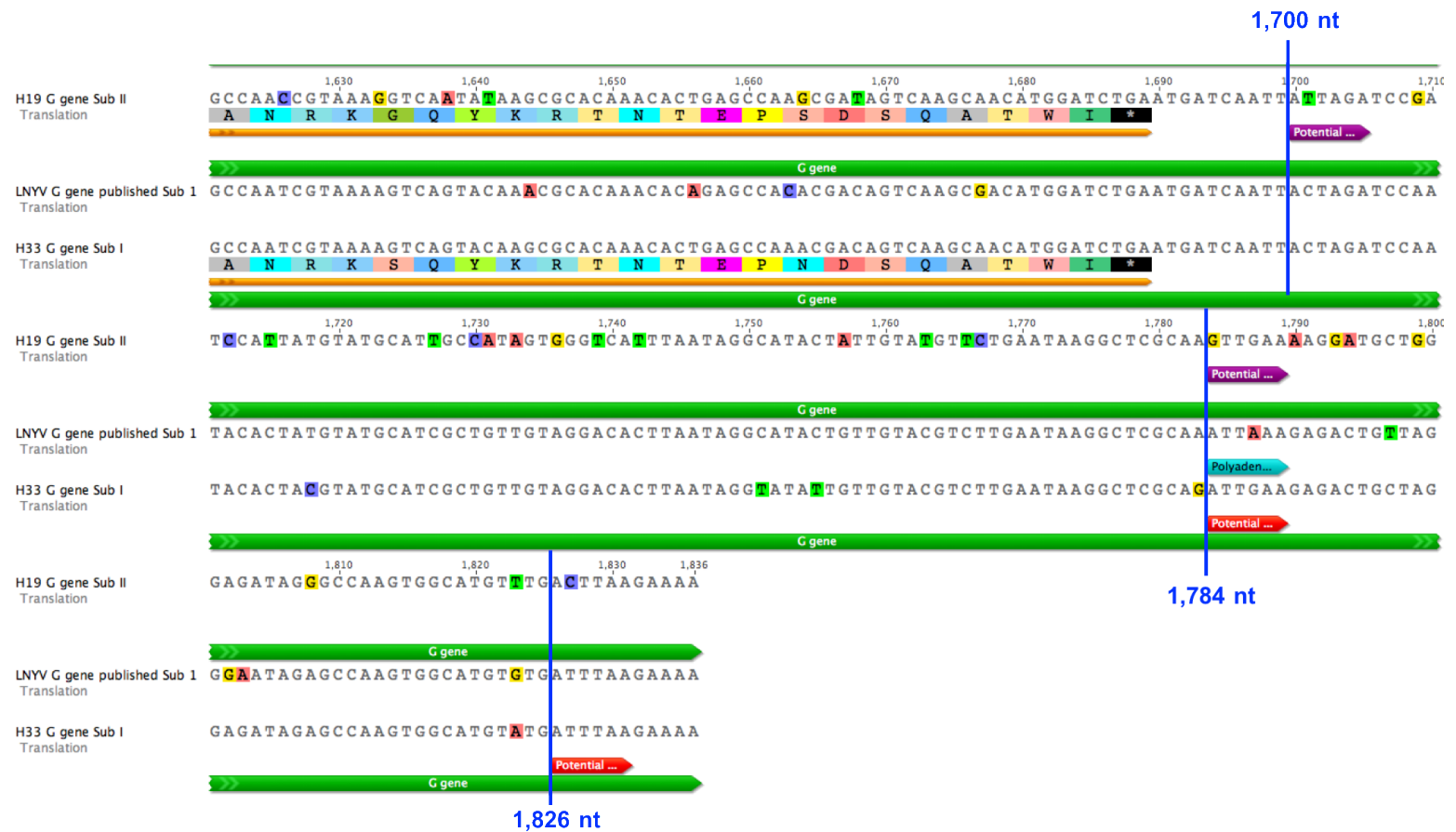
Figure 3.23: Polyadenylation analysis of H19, H33 and published G gene sequences with its nucleotide position.

Turquoise region: Polyadenylation signal of published G gene

Purple region: Potential polyadenylation signal of H33 G gene

Red region: Potential polyadenylation signal of H19 G gene

Green region: G gene

Orange region: Open reading frame
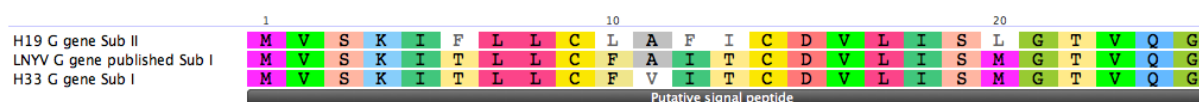
Figure 3.24: The putative signal peptide amino acid sequence of LNYV G gene. The subgroup II (H19) and subgroup I (H33 and LNYV database). Numbers at top the figure indicate amino acid position.

Heptad repeats contain seven amino acids sequence in a pattern of a b c d e f g to generate coiled coils and have been identified in rhabdovirus glycoproteins, including that of LNYV (Callaghan 2005; Chambers et al. 1990; Coll 1995b). The predicted heptad repeat d-a and a-d were identified in the G gene sequences using the data from the study by Callaghan (2005), which consisted of three regions of heptad repeat a-d and two regions of heptad repeat d-a (Figure 3.25). This figure shows the heptad repeat a-d and d-a amino acid sequences were identical between the subgroup I G gene sequences. In contrast, one amino acid was different in each of H19 G gene heptad repeat a-d and d-a sequences when compared to the subgroup I heptad repeat amino acid sequences (Figure 3.25).

Figure 3.25: Heptad repeat a-d and d-a and glycosylation region of the G gene of subgroup I and subgroup II isolates of LNYV. Coloured nucleotides indicate where the nucleotides differ. Numbers at top the figure indicate nucleotide position.

A 2D structure of a protein contains a pattern of alpha helices, beta strands, coils and turns. A 2D structure can be used to predict function, stability and folding of the proteins (Obalinsky 2006). The predicted 2D structures of the LNYV G genes are shown in Figure 3.26. Figure 3.27 and Table 3.9 shows that ten motifs were identified that were different in the 2D structures between subgroup I and subgroup II.

All of these analyses showed that there are differences in nucleotide sequences, amino acid sequences, 2D structures, putative signal peptides, glycosylation sites, polyadenylation sites and heptad a-d and d-a forms in both subgroups.
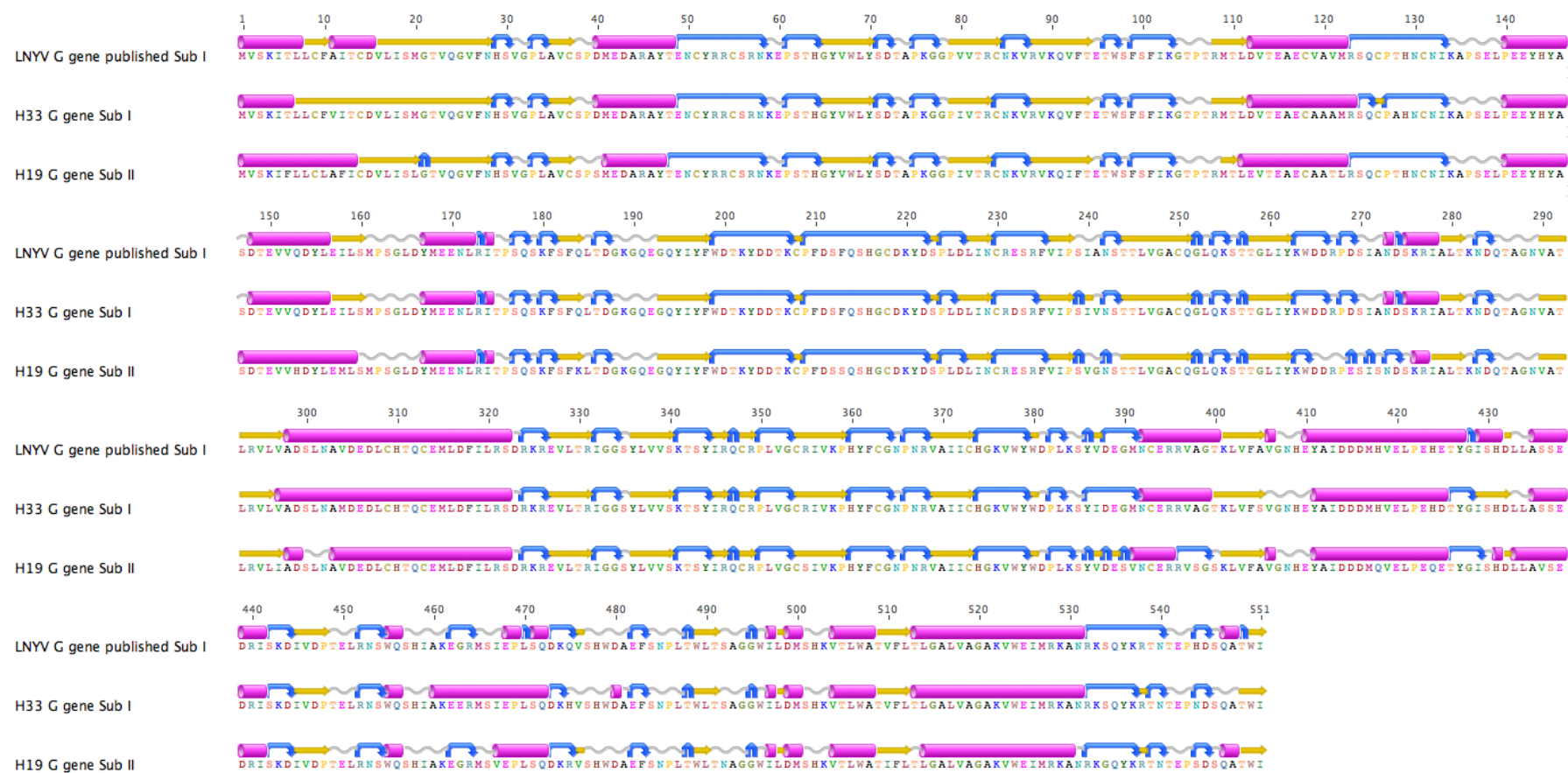
Figure 3.26: Predicted 2D structure present in the glycoprotein gene for LNYV subgroup I and subgroup II. Numbers at top the figure indicate amino acid position.

Alpha helix   Beta strand   Turn   Coil
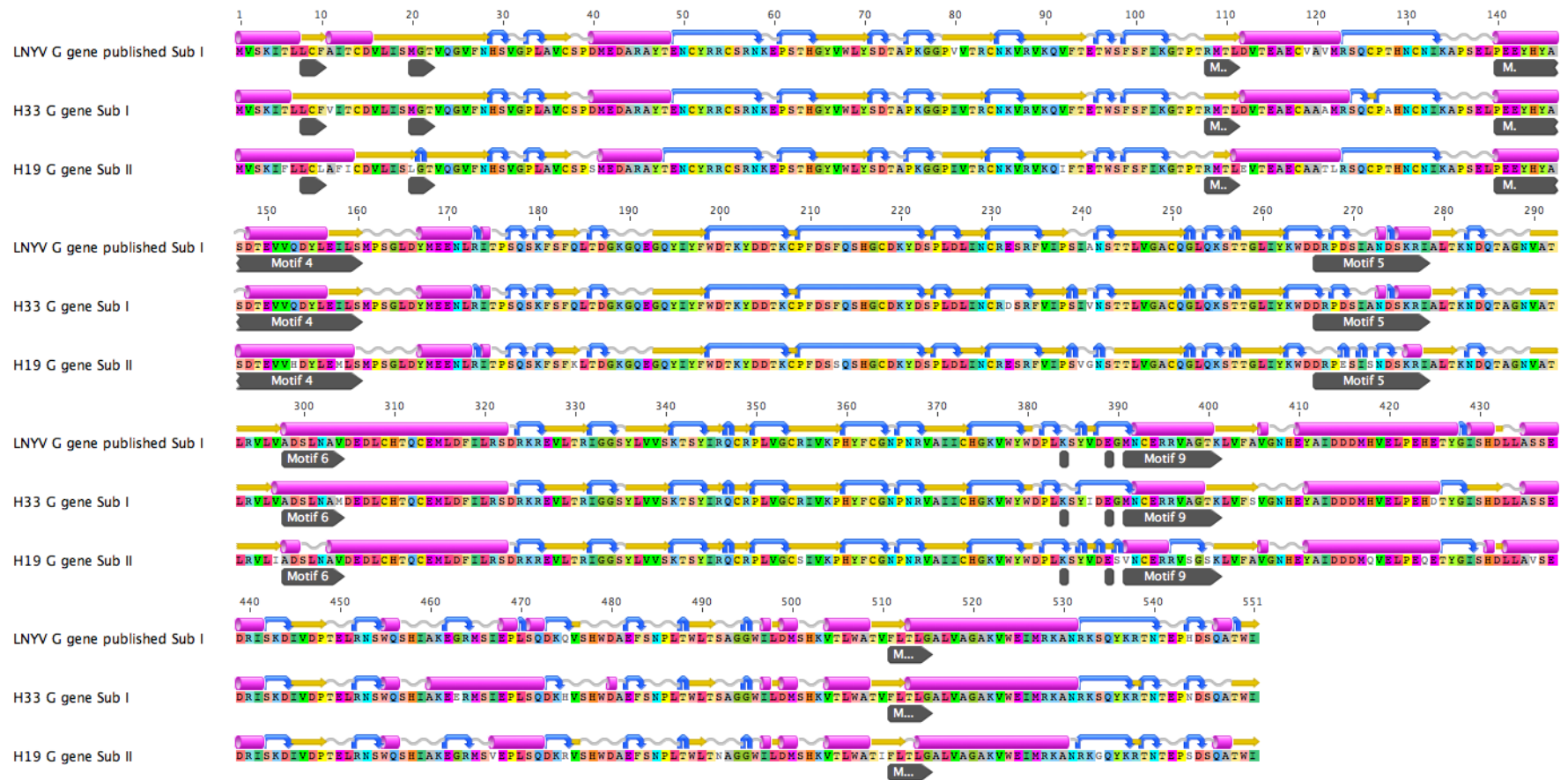
Figure 3.27: Candidate motif differences in the predicted 2D structures of the G protein of subgroup I (H33 and published LNYV accession AJ251533.1) and subgroup II (H19). Numbers at top the figure indicate amino acid position. = Motif

Alpha helix   Beta strand   Turn   Coil

Table 3.9: The summary of 2D structure differences in subgroup I and subgroup II within the seven motifs.

Motif	Sample	2D structure difference
Motif 1	Subgroup I	Beta strand
	Subgroup II	Beta strand + turn + beta strand
Motif 2	Subgroup I	Beta strand
	Subgroup II	Coil + beta strand
Motif 3	Subgroup I	Alpha helix + coil + alpha helix + beta strand
	Subgroup II	Alpha helix + coil
Motif 4	Subgroup I	Turn + coil + turn + coil + alpha helix + turn + alpha helix
	Subgroup II	Coil + turn + coil + turn + coil + turn + coil + alpha helix
Motif 5	Subgroup I	Alpha helix
	Subgroup II	Alpha helix + coil + alpha helix
Motif 6	Subgroup I	Alpha helix
	Subgroup II	Alpha helix + turn
Motif 7	Subgroup I	Beta + alpha helix
	Subgroup II	Beta + coil + alpha helix

3.4 Discussion

The published LNYV subgroup I genome comprises a monopartite, negative sense ssRNA of 12,807 nt (Dietzgen et al. 2006). Presently, the only available sequence data in the database for subgroup II is the N gene, the whole genome and the G gene sequence are unavailable. The G gene encodes the glycoprotein spikes that are present on the surface of the virion envelope (Dietzgen et al. 2007). According to Mann and Dietzgen (2014), plant rhabdoviruses utilise the glycoprotein to permeate into the insect vectors, specifically the insect midgut epithelial cells, by receptor-mediated endocytosis. However, the G protein receptor is yet to be identified. The specificity of the plant viral glycoprotein for its receptor could influence the degree of association and transmission by the insect. Higgins et al. (2016b) hypothesised subgroup II could have a more efficient association with aphids and/or plants than subgroup I allowing it to disperse more rapidly than subgroup I. This has been observed primarily for Australian isolates, as fewer subgroup II isolates from New Zealand have been identified. Based on this hypothesis, in this study the G protein was particularly examined to identify any potential sequence/structure differences that may support the hypothesis of Higgins et al. (2016b). If subgroup II can associate with the insect more efficiently than subgroup I, this may have been the reason for the extinction of subgroup I in Australia.

The present study is to the first to compare the LNYV subgroup I and subgroup II genomes. These are also the first genomes to be determined for New Zealand isolates of LNYV. The phylogenetic analyses of the genome and N gene sequences were conducted to examine the evolutionary relationships with other plant rhabdoviruses as well as between LNYV isolates. The glycoprotein was also analysed to identify changes in amino acid sequence, glycosylation, polyadenylation, signal sequence and 2D structure between the subgroups to determine the association. This research is the first to study this association.

RNA integrity analysis using agarose gel electrophoresis and bioanalyzer
RNA integrity assessment is important, prior to sequencing of the samples because degraded samples could decrease the read coverage (Davila et al. 2016; Sheng et al. 2017). The separation of ribosomal RNA (28S rRNA and 18S rRNA) is examined for RNA degradation by using non-denaturing agarose gel electrophoresis and a bioanalyzer

(Aranda et al. 2012; Schroeder et al. 2006; Thermo Fisher Scientific 2017). These methods were utilised to determine the LNYV subgroup RNA integrity for genome sequencing in the present study. The results from the non-denaturing electrophoresis revealed that the ribosomal RNA of H33 (subgroup I) samples were partially degraded. Some of the ribosomal RNA of H19 (subgroup II) samples were extensively degraded. However, the bioanalyzer results of the same H33 samples showed that ribosomal RNA were non-degraded and the same H19 samples were partially degraded. These results illustrated that the agarose gel electrophoresis and bioanalyzer had contradicting results. It could signify that the non-denaturing gel electrophoresis had pseudo RNA degradation or bioanalyzer had pseudo RNA integrity- this will be explained in the following section in more detail.

The non-denaturing electrophoresis method causes smeared bands of non-degraded RNA samples due to the formation of RNA secondary structure (Mitra 2004; Thermo Fisher Scientific 2017). Hybridisation of RNA complementary sequences to form secondary structures affects the migration of ribosomal RNA, resulting in smearing (Carson et al. 2012). Hence, it can be concluded that in the current study there was pseudo RNA degradation in the non-denaturing gel electrophoresis and the results obtained from the gel electrophoresis were inaccurate; a denaturing gel system would likely have given more accurate results. The present research further proved and verified that the bioanalyzer is the most efficient method to determine the degree of RNA degradation prior to sequencing by Illumina. The bioanalyzer also provides an RNA integrity number (RIN) to quantify the RNA integrity of the samples.

RIN value utilises Bayesian statistics to predict the RNA integrity and the degree of degradation ranges from RIN 1 to RIN 10, degraded and non-degraded, respectively (Schroeder et al. 2006). The recommended RIN value to obtain plant virus genome sequence was > 7 (Otago Genomics & Bioinformatics Facility 2016). In the present study, the LNYV H33 samples that were selected for sequencing had RIN values of 6.8 and 7 for sample 1 and sample 4, respectively. Among the H19 samples, sample 1 had the highest RIN value of 5 and was selected for the subgroup II genome sequencing. In spite of the relatively low RIN value for the H19 sample and H33 sample, the genome sequence obtained from this sample had good quality reads and high coverage. A recently published study conducted by Maina et al. (2017) had successfully obtained a complete genome of

cucurbit aphid-borne yellows virus with high number of reads and 31,129× coverage by Illumina sequencing with a RIN value of 2. The RIN value measures the integrity of the total RNA using the plant ribosomal RNA as the measure (Babu and Gassmann 2011) but does not measure the target mRNA or viral RNA. The LNYV RIN value and RNA integrity were unknown by using the bioanalyzer or gel electrophoresis. The current LNYV genome study, supported by Maina et al. (2017), verified that a low RIN value may not determine the quality or coverage of the plant virus genome and is unreliable to determine integrity of the plant RNA. It requires another technology to accurately determine the integrity of the target virus rather than the host plant.

RNA concentration comparison of NanoVue and Qubit

RNA concentration can be measured using various instruments that included NanoVue spectrophotometer and Qubit fluorometer for Illumina sequencing (Gonella-Diaza et al. 2017; Otago Genomics & Bioinformatics Facility 2016). Both devices were used to determine the RNA concentration of the H19 and H33 samples in the current research. The concentration results from the NanoVue and Qubit were compared. The H19 samples had similar concentrations using both instruments, while the H33 samples had a higher concentration using the NanoVue than the Qubit. This result signified that either the NanoVue overestimated the H33 sample concentration or the Qubit had underestimated the concentration. According to the study by Tuffaha (2008), the spectrophotometer is more likely to overestimate the concentration is to underestimate, because the spectrophotometer is incapable of distinguishing between the nucleic acids. Presence of both nucleic acids results in a higher concentration in the spectrophotometer. The Qubit florometer was considered to be more accurate and sensitive than the spectrophotometer because the dye is specific to either DNA or RNA and the presence of the other nucleic acid does not impact the determination of concentration (Anonymous 2010; DeLong 2013). Therefore, the samples that had the highest concentration from the Qubit was tested for LNYV by using RT-PCR analysis to finalise the samples for the genome sequencing.

LNYV detection by RT-PCR analysis

RT-PCR is currently one of the preeminent RNA detection methods and used for LNYV detection. The H18 (subgroup II) pooled samples were LNYV negative, even though it was tested to be infected with LNYV in sections 2.3.4.2.1, 2.3.4.2.2a and 2.3.4.2.3a

multiple times. The false negative result could have been due to the uneven distribution of LNYV within the infected plant leaves; the portion of leaf used for RNA extraction may have had a low LNYV titre. This was further discussed in detail in section 2.4. The finalised H33 (subgroup I) and H19 (subgroup II) samples were then tested using RT-PCR analysis and were LNYV positive. The RT-PCR analysis was the final verification of LNYV, prior to the genome sequencing. The LNYV subgroup I and II samples that had high concentration, RIN value, purity and were LNYV positive by RT-PCR were selected for genome sequencing by Illumina.

Fast QC analysis

Several bioinformatics software programs are available to verify the Illumina sequences for contamination and low-quality sequences (Zhou et al. 2014). These factors can significantly influence the subsequent sequence analysis. The software used for the quality analysis includes FastQC, Fastx-Toolkit, PRINSEQ and NSQC Toolkit (Zhou et al. 2014). NZGL provided the FastQC report for the LNYV sequence reads in the current research. The five critical criteria from the FastQC reports were mean quality scores, per sequence quality scores, per sequence GC content, per base N content and sequence duplication levels. The mean Phred score also known as the quality score for both subgroups were > 30. The Phred scores ranged from 10 (90% base call accuracy) to 50 (99.999% base call accuracy). A Phred score of 30 signified that there was 99.9% of base call accuracy and the probability of inaccurate base call was 1 in 1000 sequence reads (Illumina 2011a). A high Phred score of >30 indicated that there was a low probability of errors in the nucleotide sequences in both LNYV subgroups reads.

The per sequence quality score criterion determines the distribution of the sequences with a mean Phred score. It assists in identifying if a subset of the sequences contained a low Phred score . Approximately 98.75% of the LNYV sequence reads comprised > 30 Phred score and around 1.25% of the subset contained < 30 Phred score in the present study. Based on the preceding research, the low-quality sequences can be rectified by filtration or trimming of sequence terminals to obtain good quality sequences (Gandhi and Scaria 2016). Therefore, when the *de novo* assembly was conducted to obtain LNYV genome, the 5' and 3' terminals were trimmed to obtain better quality sequences.

The per sequence GC content indicates the mean GC content and the distribution of GC content across the sequences (Gandhi and Scaria 2016). The mean GC content for LNYV subgroup reads was between 42% and 43%. The reads obtained from the extract contained predominately of *N. glutinosa* or lettuce reads than LNYV. So, the mean GC value may not represent the expected LNYV genome GC content. The LNYV sequence reads had an approximately bell-shaped distribution. A non-bell-shaped distribution indicates contamination in the cDNA library (Korpelainen et al. 2014). The results from the current research illustrated that the cDNA library was uncontaminated and these reads were reliable to obtain the LNYV genome sequences for both subgroups.

The per base N content determines the percentage of N base call (unknown nucleotide) at each of the read position (Gandhi and Scaria 2016). The maximum percentage of N in the current LNYV subgroup reads was 0.2%. The percentage should ideally < 5% at any position because a high percentage of N implies insignificant confidence in determining the nucleotide at a location which increases the bias of the sequence reads (Babraham Institute Enterprise n.d.-a). The low percentage of N in the LNYV sequence reads signified very low bias and the reads were valid for genome analysis.

Sequence duplication levels indicate the percentage of duplicated sequences in the data set (Gandhi and Scaria 2016). The highest duplication level obtained from the H33 forward reads. High duplication level occurs frequently in a RNA-seq libraries due to the highly expressed genes (Delhomme et al. 2014). It was unable to identify the overrepresented sequences because those sequences were not provided separately.

Overall, the H33 and H19 sequence reads had high-quality sequences and were uncontaminated. These reads were reliable for assembly with trimming to identify the genome sequence for both subgroups.

Comparison of pathway A and pathway B consensus genome sequences
In pathway A, *de novo* assembly was used to generate contigs, which were then aligned to to generate scaffolds. The complete LNYV genome sequence was obtained as the consensus of this alignment. Pathway B was the reference assembly of the LNYV contigs; the original contigs were re-assembled with the published LNYV genome as the reference sequence and the consensus sequence take as the final LNYV genome sequence. The

sequence coverage was relatively similar for both pathways. In both subgroups, there was low coverage near the 3' leader and 5' leader even though 20 – 30 million reads were used for analysis, further supporting LNYV as a low titre virus. According to Kesanakurti et al. (2016), a higher number of reads is needed to detect low titre viruses. Therefore, more reads are required to obtain higher coverage at the termini of the LNYV genome from both subgroups.

The comparison of pathway A and pathway B subgroup I consensus sequences showed that the consensus sequences were not identical. There were an additional 13 nucleotides in the consensus sequence obtained from pathway B, and these 13 nucleotides were also identified in the published genome. Hence, this consensus sequence is more accurate and more reliable. The subgroup II consensus sequences were 100% identical in both pathways. The advantage of *de novo* assembly is that new DNA sequences can be obtained that are not present in the reference sequence. The reference assembly advantages are gaps within sequences can be predicted and contaminated sequence can be removed (Kumar and Eng 2015). According to Seshasayee (2015), a combination of *de novo* and reference assembly can be benefited by the advantages of both assemblies. Therefore, the pathway B consensus sequences were used for the subsequent analysis for both subgroups.

Phylogenetic analysis of the LNYV genome

The whole-genome phylogenetic analysis of organisms is considered to be highly efficient to determine the evolutionary relationship between organisms than specific genes (Savva et al. 2003). The genomes of cytorhabdoviruses, nucleorhabdoviruses and LNYV were evaluated using NJ and ML (Figure 3.18 and 3.19). The outgroup for both phylogenies was the Bovine ephemeral fever virus. The cytorhabdoviruses and nucleorhabdoviruses had two distinct clades. This result supports other phylogenetic analyses of plant rhabdovirus L and N gene (Bejerman et al. 2015; Dietzgen et al. 2006; Ghosh et al. 2008; Higgins et al. 2016a; Huang et al. 2003; Ito et al. 2013; Li et al. 2015; Massah et al. 2008; Revill et al. 2005; Zhai et al. 2014). The L gene, N gene and genome analysis showed clearly that nucleorhabdoviruses and cytorhabdoviruses can be distinguished by their sequence variation. It also implied that there was similarity within each genus. The nucleotide differences between cytorhabodovirus and

nucleorhabdoviruses are likely to be responsible for distinctive replication sites of cytoplasm and nucleus, respectively.

Amalgamation of plant virus genome phylogeny with host plant type could assist understanding the virus relationships (Figure 3.18). It was recognised in both phylogenetic trees that CBDaV, BYSMV and NCMV (cytorhabdoviruses) infectmonocots and formed a monophyletic clade. Whereas cytorhabdoviruses infecting unknown hosts (WIV4, WIV5 and WIV6) and cytorhabdoviruses infecting dicots (LNYV, LYMoV, PeVA and ADV) formed a monophyletic clade. This type of clustering by host type was not observed for the nucleorhabdoviruses. Higgins et al. (2016a) had shown that planthoppers transmitted the CBDaV, BYSMV and NCMV. The ADV and LNYV was transmitted by aphids while PeVA and LYMoV is transmitted by an unknown vector (Higgins et al. 2016a). Coalescing the genome, plant hosts and vectors results, it was hypothesised that there was a substantial association with the monocot infecting cytorhabdoviruses and planthoppers, and also dicot infecting cytorhabdoviruses and aphids. The host plants for WIV4, WIV5 and WIV6 are unknown (Li et al. 2015). A detailed study was not conducted on each of these viruses; however, based on the current genomic data, there is a high likelihood that these viruses could be infecting dicots rather than monocots because they were located in the dicot infecting virus clade. More cytorhabdovirus genomes are required to confirm this hypothesis and the association with hosts and vectors.

The genome analysis of LNYV along with other cytorhabdoviruses could establish the evolutionary history of the virus. The NJ and ML phylogenies showed that LYMoV is the closest related organism to LNYV. This was observed in other N and L gene analyses (Bejerman et al. 2015; Higgins et al. 2016a; Higgins et al. 2016b; Ito et al. 2013). The NJ and ML phylogenies from the current study also showed that subgroup II emerged earlier than subgroup I. This result contradicts with the nucleotide N gene phylogenetic analysis from the current study and previous study. For instance, according to Higgins et al. (2016b) the nucleotide N gene phylogenetic analysis and BEAST analysis shows that subgroup I had emerged earlier than subgroup II. Therefore, more LNYV genome sequences are required to determine the evolutionary history of LNYV subgroups more accurately.

The present study also showed that there was a high similarity within the subgroup I genomes with definitive differences in nucleotides between the subgroups. The similarity of sequences between LNYV subgroup I genomes was 93.3% and similarity between the subgroups was 80.3% (Figure 3.18). Another study conducted by Klerks et al. (2004) showed that SCV also has two subgroups (group I and group II). The L gene similarity within the same subgroups of that virus was 98% and between the subgroups was 89% (Klerks et al. 2004). When comparing with LNYV, the SCV L gene had few sequence differences within and between the subgroups than the LNYV genome. Since the SCV subgroup genomes were not available, it was impossible to compare the SCV subgroup genomes to the LNYV genomes. In the current study, the percentage of sequence similarity for nucleorhabdovirus genomes was 26.1% to 58.8% and cytorhabdovirus genomes was 29.6% to 93.3% (Figure 3.18). Bias in the results could cause the very high sequence similarity for the cytorhabdoviruses due to the presence of three LNYV genomes forming a single cluster and other viruses containing only one genome. Overall, it can be concluded that LNYV subgroup I and subgroup II genomes were distinctive to other viruses.

Phylogenetic analysis of N gene

Phylogenetic analysis was conducted on H33, H19 and all the available N gene sequences. The nucleotide phylogeny showed the separation of subgroups with the subgroup I lineage appearing slightly earlier than subgroup II lineage (Figure 3.20). This result is supported by the BEAST analysis, which suggested that subgroup I and subgroup II appeared 150 years ago and 75 years ago, respectively (Higgins et al. 2016b). Therefore, a detail understanding of subgroup I may help to determine the origin of LNYV.

Phylogenetic analysis of amino acid sequences is an efficient method to establish the evolutionary history of an organism and genes due to the conservation of amino acid sequences (Nei and Kumar 2000). The LNYV N gene amino acid sequence phylogenetic analysis also showed that there were two distant clades of subgroup I and subgroup II and that that AU9 isolate appears to be most closely related to the common ancestor. The AU9 isolate belongs to subgroup II and was collected from an infected sowthistle in South Australia (Callaghan and Dietzgen 2005). This seems to contradict the findings of Higgins et al (2016b) since they reported subgroup I appeared before subgroup II while

the current results support the LNYV genome phylogenetic analysis. South Australia could be the origin of LNYV emergence, the virus may have spread from there to the rest of Australia and into New Zealand. More samples are required from both Australia and New Zealand to more fully understand the origin of LNYV.

Glycoprotein analysis

Plant rhabdoviruses have glycoproteins that protrude on the surface of the virions. This protein is thought to bind to midgut receptors of the insect vector to penetrate into the epithelial cells. Subsequently, viruses translocate from the midgut to the salivary glands for transmission (Dietzgen et al. 2016). Glycoprotein analysis of the LNYV subgroups could determine if one of the subgroups has a higher association and transmissibility by aphids than the other subgroup. Identification of variations in the 2D structure, amino acid sequences and other characteristics that included polyadenylation sites, putative signal peptide, glycosylation sites, heptad a-d form, heptad d-a form and peptide recognition sequence between the subgroups could determine the difference in this association. These features were identified in New Zealand subgroup I and subgroup II G gene sequences. The putative signal peptide sequence consisted of 25 amino acids with a peptide recognition sequence also known as the signal peptide sequence VQG↓V (the arrow is the cleavage site) (Callaghan 2005). This was identified in H33, H19 and the published LNYV G sequences, located in the same region near the 5' leader (Table 3.10). The 25 amino acid sequence was hydrophobic in the amino-terminal (N-terminal) region and the peptide recognition sequence could be a signal peptide to the endoplasmic reticulum, which will undergo post-translation modification by glycosylation (Dietzgen et al. 2007; Dietzgen et al. 2006). The subgroup II differs by five amino acids and one amino acid was different within the subgroup I 25 amino acid putative signal peptide sequence. The significance of this difference is unknown. According to Bejerman et al. (2015) the LNYV hydrophilicity plot of the G protein hydrophobic region is similar to LYMoV and ADV hydrophilicity plot. The signal peptide sequence for LYMoV and ADV were LSD↓F and IGD↓R, respectively (Bejerman et al. 2015; Heim et al. 2008). This confirmed that the hydrophilicity in the G gene N-terminal region was conserved among these plant rhabdoviruses, but the signal peptide sequence itself was not. LYMoV and ADV is transmitted by an unknown vector and aphids, respectively (Higgins et al. 2016a). It does not seem to be insect specific. There is no detailed research conducted on the G gene in other plant rhabdoviruses. According to Shao et al. (2016) glycoprotein

stable signal peptide is important in viral entry of Arenaviruses into the host cells. Since the signal peptide sequences were identical for both subgroups of LNYV, the viral entry into the plant hosts and vectors could be same.

The G protein is glycosylated and the glycosylation site was analysed to identify variation between the subgroups. The predicted glycosylation sequence for LNYV is Asn-X-Ser/Thr (X is any amino acid) (Callaghan 2005). This sequence was identified in three locations in H33, published LNYV and the H19 G sequences (Table 3.10). One glycosylation site was near the N-terminus and two were near the centre of the protein. These were conserved for subgroup I and subgroup II, but there was an additional glycosylation site in the H33 sequence near the C-terminal region (Table 3.10). The conserved regions were identified in the previous study by Dietzgen et al. (2006). The LYMoV G gene also had the same glycosylation sequence and four predicted sites were identified (Heim et al. 2008). Six potential glycosylation sites were identified in ADV, NCMV, SYNV and MIMMV with the same amino acid sequence (Bejerman et al. 2015; Goldberg et al. 1991; Massah et al. 2008; Tanno et al. 2000). This indicated that the amino acid glycosylation sequence is conserved, with between three to six sites in plant rhabdovirus G proteins. N-linked-glycosylation influences the secretion of proteins and cell surface proteins (Shakin-Eshleman et al. 1996). It occurs in the Asn-X-Ser/Thr sequence, but glycosylation does not occur at all sites that contain this sequence (Shakin-Eshleman et al. 1996). In the present study, there was variation in the number of glycosylation sites within subgroup I since H33 had an extra predicted site than the published sequence. It was not ascertained at which sites that the glycosylation occurs nor if the addition of a glycosylation site has an effect in aphid association in subgroup I than subgroup II. Glycosylation could impact the transmission of viruses. For instance, when Beet western yellows virus (BWYV) was deglycosylated by PNGase F or α-D-galactosidase had resulted in inhibition of *Myzus persicae* transmission (Seddas and Boissinot 2006). This shows the importance of glycosylation in virus transmission by aphids. Hence, having an extra glycosylation could increase the transmission efficiency and enabling the virus to survive in New Zealand. Since the extra glycosylation site was not present in the published subgroup I sample from Australia, it could reduce the transmission efficiency and may have led to subgroup I extinction in Australia.

Polyadenylation is the accumulation of a poly(A) tail to the 3' terminus of RNA and contains a series of adenosines (Li et al. 2014). In viruses, it is synthesised by transcription of poly(U) from a template strand. It is required for the stability of the RNA and translation (Li et al. 2014). In the majority of instances, polyadenylation site consists of a polyadenylation signal (Carter and Saunders 2013). The LNYV polyadenylation signal in the G gene is 'AUUAAA' and is located 95nt downstream of stop codon (Callaghan 2005). In the current study, this sequence was identified at two sites in H19 and was not located near the stop codon (Table 3.10). However, this sequence was not identified in the H33 sequence (Table 3.10). Two potential polyadenylation sequences were identified in the H33 G gene sequence. The 'AUUGAA' sequence was located in the same location as previously identified LNYV polyadenylation signal but there is one nucleotide difference while 'AUUUAA' potential polyadenylation sequence was five nucleotides away from the end of mRNA. So, it is unlikely that 'AUUUAA' is the polyadenylation signal. Two potential polyadenylation sequences were also identified in the H19 G gene sequence. When comparing to the previously identified LNYV polyadenylation signal, 'GUUGAA' potential polyadenylation signal differs by two nucleotides but was located in the same position while 'AUUAGA' potential signal differ by one nucleotide but was located in different position. Therefore, GUUGAA could be the polyadenylation signal in H19 G gene sequence. PCR and sequencing of the mRNA 3'UTR might help identify the polyadenylation signal sequence in H33 and H19.

Heptad-repeats contain seven amino acid sequences in the order a-b-c-d-e-f-g to form coiled-coil structures (Coll 1995b). Position a and d are hydrophobic or neutral (Chambers et al. 1990). These repeats were identified and studied in the glycoprotein in many viruses including rhabdoviruses. The two forms of heptad-repeats were a-d and d-a in the LNYV G protein (Callaghan 2005). In the present study, the heptad-repeats a-d were located in four locations within the published LNYV G protein sequence (Table 3.10). There were more amino acid differences between the subgroups in all heptad-repeats than between the subgroup I G protein sequences. These heptad repeats were recognised in other nucleorhabdovirus G protein sequences but not identified in other cytorhabdoviruses yet. A total of three heptad repeats were identified in the IMMV G protein and four repeats were identified in RYSV (Luo and Fang 1998; Massah et al. 2008). Luo and Fang (1998) suggested that these repeats could be related to the viral fusion into the host membrane. The amino acid differences between the subgroups may

cause differences in the efficiency with which LNYV subgroups fuse with the host membrane, affecting replication or transmission efficiency. Mutation of these specific regions in each subgroup may provide some insight. Therefore, mutational analysis on the glycoprotein sequence of both subgroups, combined with aphid transmission studies should be conducted in the future to determine if mutation impacts the transmission or replication of LNYV in aphids.

Protein secondary structure comprises alpha helices, beta sheets, strands and loops based on the primary structure (Obalinsky 2006). The 2D structure of the glycoprotein was predicted in order to identify differences between LNYV subgroup I and subgroup II G protein sequences. A total of ten motifs were identified that were different between the subgroups. Among those motifs, six had an extra turn or coil with an alpha helix or beta sheet in subgroup II that was absent from subgroup I sequences (Table 3.9). However, according to Kocincová et al. (2017), analysis of the secondary structure of the protein is unable to predict the protein characteristics because spatial information is essential for such evaluation. It requires a tertiary structure analysis by X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy (Berg et al. 2002). Currently, the 2D structure and 3D structure of plant rhabdovirus G proteins, including LNYV, has not been documented. There are glycoprotein 3D structures from non-plant rhabdoviruses, namely vesicular stomatitis virus (VSV) and chandipura virus (CV), obtained by X-ray crystallography (Baquero et al. 2017; Baquero et al. 2015; Roche et al. 2006; Roche et al. 2007). In another study, the RYSV G protein was compared with non-plant rhabdovirus G protein and was very dissimilar (Luo and Fang 1998). Hence, it is likely that the LNYV G protein would also be dissimilar to the currently available 3D structures of non-plant rhabdovirus G proteins. The tertiary structure of the G protein from each subgroup should be determined by x-ray crystallography or NMR. Comparison of these would help determine if there are any biologically relevant differences that will affect the replication or transmission efficiency of LNYV subgroups.

Summary

Table 3.10 is the summary of genome analysis, phylogenetic analysis and glycoprotein analysis conducted on the LNYV published, H33 and H19 genomes. It shows that total number of nucleotides in the genome of both subgroups vary between 28 nt and the

genome structure is identical. The evolutionary history of LNYV subgroup was same in both NJ and ML phylogenies of LNYV genomes which showed that subgroup II had emerged earlier than subgroup I. The percentage identity within the subgroup I genome was higher than between the subgroup I and subgroup II while the pairwise distance within the subgroup I genome was lower than between the subgroup I and subgroup II (Table 3.10).

The LNYV G protein sequence analysis focused on six characteristics (Table 3.10). These were the putative signal peptide, the peptide recognition sequence, glycosylation sites, polyadenylation sites, heptad a-d form and heptad d-a form. When the putative signal peptide sequence was compared to the published LNYV, the subgroup II had a high or number of amino acid differences than subgroup I. For both subgroups, the putative signal peptide sequences were located near the 5' leader. The peptide recognition sequence VQG↓V and glycosylation sequence Asn-X-Ser/Thr were also identified in both subgroups. The H33 G gene sequence had an additional glycosylation sites than the published LNYV and H19 sequences. The polyadenylation sequence was identified in one site of the published LNYV sequence, three sites in H19 sequence and appeared absent from the H33 sequence. Heptad a-d and d-a forms were identified in subgroup I and subgroup II; the subgroup II sequence had a higher number of amino acid differences than subgroup I. All of these differences could cause variation in the association between the LNYV subgroup and in the midgut of the aphid vector. Mutation analysis needs to be carried out to determine if any of these characters have a role in LNYV replication and transmission.

Table 3.10: Summary results of genome analysis, phylogenetic analysis and glycoprotein analysis on the LNYV subgroup I and subgroup II genomes.

Analysis		Subgroup I		Subgroup II
		LNYV published	H33	H19
Genome analysis	**Total genome (nt)**	12,807 nt	12,779 nt	12,804 nt
	Genome structure	Same genome structure – all three LNYV genomes consist of six genes, ORF in each gene, 3' leader and 5' leader		
	GC content	42.9%	42.9%	43.2%
Phylogenetic analysis	**Neighbour joining genome phylogenetic analysis**	Evolutionary history: subgroup II appeared before subgroup I		
		Percent identity within subgroup I genomes: 93.3%		Percent identity between LNYV published and H19: 80.5% Percent identity between H33 and H19: 80.3%
	Maximum likelihood genome phylogenetic analysis	Evolutionary history: subgroup II appeared before subgroup I		
		Pairwise distance within subgroup I genomes: 0.068		Pairwise distance between LNYV published and H19: 0.234 Pairwise distance between H33 and H19: 0.242
Glycoprotein analysis	**Putative signal peptide**	Located near 5' leader		
		Baseline	1 out 25 amino acids were different when compared to published LNYV G gene	5 out 25 amino acids were different when compared to published LNYV G gene
	Peptide recognition sequence	VQG↓V		
	Glycosylation sequence	Asn-X-Ser/Thr		
		3 glycosylation sites	4 glycosylation sites	3 glycosylation site
	Polyadenylation signal sequence	AUUAAA	AUUAAA not identified	AUUAAA
		Near stop codon (position 1,784 nt)		Not near the stop codon
	Potential polyadenylation signal sequence	Does not apply	AUUGAA position 1,784 nt	GUUGAA position 1,784 nt
	Heptad a-d form	4 sites		
		Baseline	1 amino acid differences in 1 site	1 amino acid difference in 3 sites
	Heptad d-a form	2 sites		
		Baseline	1 amino acid difference in 1 sites	1 amino acid difference in 2 sites

Chapter 4

Final Discussion

Chapter 4 Final discussion

LNYV is the type species of the genus *Cytorhabdovirus*. This virus was initially identified in lettuce in 1954 by (Stubbs and Grogan 1963a). LNYV infection of lettuce has a devastating effect previously resulting in catastrophic crop loss of up to 100% in Victoria, Australia (Stubbs and Grogan 1963a). Economic losses of 15% – 20%, specifically from lettuce production in Australia have been reported (Dietzgen et al. 2007).

Rapid identification of LNYV during the early stages of infection is imperative to control the dispersion and reduce further impact on lettuce production. The LNYV population consists of two subgroups (subgroup I and subgroup II) that were identified by phylogenetic analysis of the N gene sequence from Australian and New Zealand isolates (Callaghan and Dietzgen 2005; Higgins et al. 2016b). ELISA and RT-PCR analysis can be used for the identification of LNYV (Dietzgen et al. 2007). While, a rapid diagnosis of the subgroups has not been developed, hitherto. It is important to determine the subgroups more rapidly and efficiently.

The preliminary objective of this study was to a develop diagnostic test to detect LNYV subgroups by using RT-PCR analysis and RT-PCR-RFLP. The previously designed LNYV primers BCNG1/BCNG2 and LNYV_440F/LNYV_1185R primers were tested to detect LNYV. These primers were tested on symptomatic samples to detect LNYV and subgroups. The TPS procedure with RT-PCR was tested on known LNYV infected samples.

Diagnosis of LNYV using the BCNG1/BCNG2 and LNYV_440F/LNYV_1185R primers

LNYV was detected by primers that were previously designed using N gene sequences; these included the BCNG1/BCNG2 and LNYV_440F/LNYV_1185R primers (Callaghan and Dietzgen 2005; Higgins et al. 2016b). These primers were tested and re-optimised on known infected samples. It was found that the LNYV_440F/LNYV_1185R primers were more efficient than the BCNG1/BCNG2 primers because LNYV_440F/LNYV_1185R primers amplified a smaller PCR product and partially degraded RNA was less likely to result in a false negative result. In section 2.4, this aspect was discussed in detail.

LNYV subgroup diagnosis using RT-PCR and RT-PCR-RFLP analysis

This research is the initial study to develop diagnostic methods for detection of LNYV subgroups by using endpoint RT-PCR and RT-PCR-RFLP analysis. LNYV subgroup diagnosis was established by designing subgroup-specific and non-subgroup specific primers. The primers were tested and the conditions were optimised using the previously confirmed LNYV infected samples from the study conducted by Higgins et al. (2016b). However, if the RT-PCR method is used exclusively, this approach requires each sample to be tested twice, once for each subgroup. Therefore, RT-PCR-RFLP diagnosis was developed using the *Mae*III restriction enzyme to detect each subgroup simultaneously, albeit in a two-step process as this digest is post RT-PCR. The optimised primers were subsequently used to test potentially infected samples. A total of three subgroup I, six subgroup II samples were identified in symptomatic lettuce samples from the Auckland and Canterbury regions from 2011 and 2017, respectively (Fletcher et al. 2017; Higgins et al. 2016b). Two samples were LNYV subgroup unknown, indicating the subgroup specific primers require further optimisation. The Canterbury subgroup identification in the current study was further used by Fletcher et al. (2017) in a lettuce virus disease report to determine the LNYV subgroup in Canterbury. This shows the importance of the currently developed subgroup detection methods for disease surveillance, which was further discussed in section 2.4.

The LNYV subgroup diagnostic method can be further improved by the development of an RT-qPCR assay with high resolution melting analysis to detect and quantify LNYV subgroups using the subgroup specific primers developed in this research. End-point RT-PCR consists of several limitations; it is less precise, has poor resolution, it is laborious, unquantifiable and require post PCR analysis when compared to RT-qPCR (Gochhait et al. 2007). RT-qPCR with high resolution melting analysis can overcome these limitations. A rapid, quantifiable, high precision and sensitive diagnostic method are crucial for the detection of LNYV in economically important crops and native plants to control the spread of the disease.

LNYV was reported to infect the *Sonchus hydrophilus* and puha (*Sonchus kirkii*), which are endemic to Australia and New Zealand, respectively (Francki et al. 1989; Higgins et al. 2016b). According to FloraBase (n.d.), *S. hydrophilus* is not a threatened species in Australia. There was only been a limited amount of research conducted on the LNYV

infection of *S. hydrophilus*. *S. kirkii* is an at risk or declining plant species in New Zealand. It has a limited geographical distribution and was last reported on the Great Island, Auckland and northern part of North Island (Figure 4.1) (Cameron 2000). *S. oleraceus* and *S. asper* seemed to out-compete *S. kirkii* due to the latter's prolonged maturation period and low dispersion by heavy seeds (Cameron 2000). Moreover, LNYV infection of *S. kirkii* could further promote this decline and possibly complete extinction of this endemic plant. Plant viruses caused few native plants to be endangered in New Zealand. For instance, *Sicyos australis* is an endangered native species in New Zealand; it has been speculated that plant viruses may have caused the decrease in population during the past 50 years (Delmiglio and Pearson 2006). Therefore, a rapid and efficient LNYV detection method is essential to identify and reduce further decline of the population. It is also imperative to determine if one of the subgroups infects *S. kirkii* and *S. hydrophilus* more efficiently than the other subgroup. Such information could be used to develop an effective conservation method to reduce the infection incidence and spread of the disease. According to Fletcher et al. (2017), there is concern that New Zealand may contain a virulent strain of LNYV which could have come from Australia. This strain is likely to be the subgroup II because it was hypothesised by Higgins et al. (2016b), subgroup II could have more optimum relationship with the plant hosts and insects. Subgroup II could have out-competed subgroup I which may have led to the extinction of subgroup I in Australia. More surveys of these plant species are essential to improve understanding of the LNYV subgroup infection incidence in these plants.
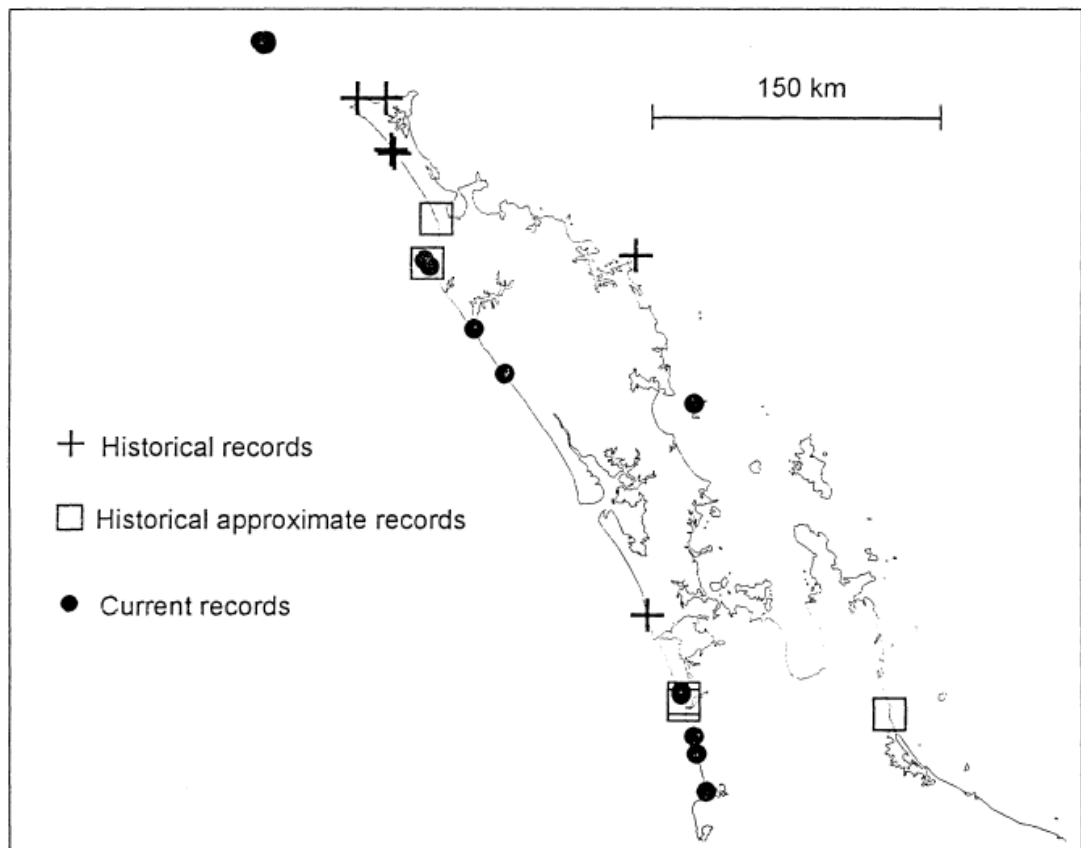
Figure 4.1: *S. kirkii* distribution in North Island, New Zealand (Cameron 2000).

Diagnosis of LNYV using the TPS procedure

The TPS procedure is a crude rapid virus release method and is combined with RT-PCR to detect various plant viruses including LNYV (Thomson and Dietzgen 1995). In the current study, known LNYV infected samples were tested using the TPS procedure with RT-PCR. It was unable to detect LNYV or subgroups using this procedure and could be due to the presence of PCR inhibitors.

LNYV genome sequence and glycoprotein analysis

The second aim of this study was to sequence the entire genome of LNYV subgroup I (H33) and subgroup II (H19) from New Zealand isolates by Illumina sequencing. Previously, the only LNYV genome available was from a subgroup I isolate from Australia (Dietzgen et al. 2006). Phylogenetic analysis was conducted on the genomes together with available plant rhabdovirus genomes to elucidate evolutionary relationships. Nucleotide and amino acid phylogenetic analyses of N gene sequences were also performed with publicly available LNYV N gene sequences and the N gene

sequences extracted from the New Zealand LNYV genomes. Finally, the G gene sequences of the published LNYV, H33 and H19 were analysed to determine the differences in nucleotides, amino acids and 2D structures with various characteristics of the G protein. These characteristics are the putative signal peptide, the peptide recognition sequence, glycosylation, polyadenylation sequence and heptad a-d or d-a form. Differences between the subgroups for each characteristic may impact the association of the LNYV subgroup with the insect vector.

Plant virus subgroups more likely originate from one common ancestor. For instance, it was observed in the phylogenetic analysis of CMV that each subgroup had one common ancestor and strains in each subgroup had diverged from one origin (Roossinck et al. 1999). The N gene amino acid phylogenetic tree showed that amino acid sequence of AU9 appears to be closely related to the common ancestor, indicating at least the origin of LNYV subgroup II. The AU9 isolate is a subgroup II isolate that was collected from South Australia (Callaghan and Dietzgen 2005). According to Higgins et al. (2016b), BEAST analysis showed that subgroup II emerged 75 years after subgroup I and subgroup II seems to have dispersed rapidly, which could have resulted in subgroup I extinction in Australia. Could the AU9 isolate be the origin of the rapid dispersal?

The AU9 nucleotide sequence does not occupy the same ancestral position in the nucleotide phylogenetic tree. This could have been caused by the different evolution trajectory of the amino acid and nucleotide sequences. According to Nei and Kumar (2000) amino acid and nucleotide sequences can evolve independently with amino acid sequences being considered more reliable because they evolve at a much slower rate. However, the complete genome and other gene sequences from the AU9 isolate are not available. Therefore, it is not known if this observation is specific to the AU9 N protein sequence, or if this might be observed for the other viral genes. Future research should be conducted to obtain and examine more complete genomes or genes from different locations in Australia and New Zealand to further understand the population structure and origin of subgroup I and subgroup II from both countries. This analyses should also include the native host species. Higgins et al. (2016b) hypothesised that LNYV could have been present in the ancestral species of *S. hydrophilus* and *S. kirkii*. Evolution of LNYV could be analysed more accurately with the addition of LNYV infected native species in the phylogenetic analysis.

The glycoprotein analysis of LNYV subgroups could help determine if there a subgroup specific difference in the association between LNYV and its aphid vector. This may explain the possible extinction of subgroup I in Australia, and the rapid dispersal of subgroup II since subgroup II may have more efficient at replicating in the insect hosts, or more efficient at attaching to the insect (Higgins et al. 2016b). In the present study, ten motifs were identified in the 2D structure of the G protein there were different between the subgroups. Additional differences in the putative signal peptide amino acid sequence, glycosylation sites, polyadenylation sequence and heptad repeat amino acid sequences were also identified and are discussed in detail in section 3.4. This was preliminary research and requires further extensive studies to determine if these changes have an impact on the association with the aphids. For instance, an earlier study was conducted on TSWV glycoprotein and its transmission to host plants through a vector - thrips (Sin et al. 2005). Similar to plant rhabdoviruses, TSWV replicates in the midgut, adheres to the epithelial cells and translocates to the salivary glands. The study showed that nonsynonymous mutation in the glycoprotein sequence of TSWV resulted in the non-transmissible virus by thrips to plants without affecting the viral assembly. Frameshift and or nonsense mutation in the glycoprotein sequence resulted in non-transmissible and non-functional viral assembly (Sin et al. 2005). This study shows the importance of glycoproteins in the virus life cycle. Thus, changes within the G sequence could drastically impact the transmissibility and acquisition of LNYV by the aphid. Future research should focus on obtaining 3D structures of LNYV subgroup I and subgroup II G protein. The 3D structural analysis can be combined with mutation analysis to understand which mutations could affect the function of the protein.

Conclusion

The current research has enabled diagnosis of LNYV subgroups in infected plants more rapidly by RT-PCR and RT-PCR-RFLP than the previous diagnosis using RT-PCR followed by sequencing and phylogenetic analysis. Rapid diagnosis is crucial when LNYV infects the agricultural crops and native plants. The BCNG1/2 and LNYV_440F/LNYV_1185R primers were tested on known LNYV infected samples and conditions were re-optimised to enable diagnosis of LNYV. These primers were tested on symptomatic samples from Harrisville, Tuakau and Canterbury to detect LNYV and subgroups. The subgroup specific primers were able to detect most LNYV subgroups but

it still requires further optimisation to enable the detection of a wider LNYV subgroup population. The TPS procedure with RT-PCR had failed to detect LNYV from known infected samples and can be concluded as an inefficient method.

In the future, control methods of each subgroup could be developed, which require rapid diagnosis to reduce further the spread of the disease and catastrophic impact to the crops. The LNYV subgroup I and subgroup II genomes from New Zealand were also obtained during this research. A detailed and in depth genome analysis in the future could provide some insight into controlling each LNYV subgroup specifically. The LNYV G gene and protein sequences of subgroup I and subgroup II were analysed. There were differences in 2D structure and various characteristics between the two subgroups. However, it was unable to determine if the differences enable subgroup II to be more efficient than subgroup I.

Reference

Albrechtsen SE. Testing Methods for Seed-transmitted Viruses: Principles and Protocols [Internet]. Oxfordshire, UK: CABI Publising; 2006. Available from: https://books.google.co.nz/books?id=cpq4waJu6McC

Almeida AV, Neuzil P, Manz A. DNA sequencing in droplets. 18th International Conference on Miniaturized Systems for Chemistry and Life Sciences, MicroTAS 2014; 2014. p. 1018-1020.

Ammar ED, Tsai CW, Whitfield AE, Redinbaugh MG, Hogenhout SA. 2009. Cellular and Molecular Aspects of Rhabdovirus Interactions with Insect and Plant Hosts. Annual Review of Entomology 54:447-468.

Andrews S. FastQC [Internet]. Cambridgeshire, UK: Babraham Institute. Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Anonymous. Accurate and sensitive quantitation of nucleic acids, even at low concentrations. BioProbes [Internet]. (62):20-22. Available from: https://www.thermofisher.com/content/dam/LifeTech/migration/en/filelibrary/support/bioprobes/bioprobes-62.par.14365.file.dat/bioprobes-62-qubit.pdf

Aranda PS, LaJoie DM, Jorcyk CL. Bleach gel: A simple agarose gel for analyzing RNA quality. Electrophoresis [Internet].33(2):366-369. Available from: http://dx.doi.org/10.1002/elps.201100335

Ari Ş, Arikan M. 2016. Next-Generation Sequencing: Advantages, Disadvantages, and Future. In: Hakeem KR, Tombuloğlu H, Tombuloğlu G, editors. Plant Omics: Trends and Applications. Cham: Springer International Publishing. p. 109-135.

Arrigo SJ. 1994. Assessment of Therapy Effectiveness: Infectious Disease. In: Mullis KB, Ferré F, Gibbs RA, editors. The Polymerase Chain Reaction. Boston, MA: Birkhäuser Boston. p. 344-356.

Babraham Institute Enterprise. Per Base N Content [Internet]. Babraham Institute Enterprise Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/Help/3 Analysis Modules/6 Per Base N Content.html

Babraham Institute Enterprise. Per Sequence GC Content [Internet]. Babraham Institute Enterprise Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/Help/3 Analysis Modules/5 Per Sequence GC Content.html

Babraham Institute Enterprise. Per Sequence Quality Scores [Internet]. Babraham Institute Enterprise Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/Help/3 Analysis Modules/3 Per Sequence Quality Scores.html

Babu B, Washburn BK, Ertek TS, Miller SH, Riddle CB, Knox GW, Ochoa-Corona FM, Olson J, Katırcıoğlu YZ, Paret ML. 2017. A field based detection method for Rose rosette virus using isothermal probe-based Reverse transcription-recombinase polymerase amplification assay. Journal of Virological Methods 247:81-90.

Babu S, Gassmann M. Assessing integrity of plant RNA with the Agilent 2100 Bioanalyzer [Internet]. California, United States: Agilent Technologies. Available from: http://www.agilent.com/cs/library/applications/5990-8850EN.pdf

Baker KK, Ramsdell DC, Gillett J. 1985. Electron microscopy: Current applications to plant virology. Plant Disease 69(1):85-91.

Balasubramanian S. 2015. Solexa sequencing: Decoding genomes on a population scale. Clinical Chemistry 61(1):21-24.

Bald-Blume N, Bergervoet J, Maiss E. 2017. Development of a molecular assay for the detection of Cucumber mosaic virus and the discrimination of its subgroups I and II. Journal of Virological Methods 243:35-43.

Bale JS. 1991. Implications of Cold Hardiness for Pest Management. In: Lee RE, Denlinger DL, editors. Insects at Low Temperature. Boston, MA: Springer US. p. 461-498.

Baquero E, Albertini AA, Raux H, Abou-Hamdan A, Boeri-Erba E, Ouldali M, Buonocore L, Rose JK, Lepault J, Bressanelli S. 2017. Structural intermediates in the fusion-associated transition of vesiculovirus glycoprotein. The EMBO Journal 36(5):679-692.

Baquero E, Albertini AA, Raux H, Buonocore L, Rose JK, Bressanelli S, Gaudin Y. 2015. Structure of the low pH conformation of Chandipura virus G reveals important features in the evolution of the vesiculovirus glycoprotein. PLoS Pathogens 11(3):e1004756.

Bawden F, Nixon H. 1951. The application of electron microscopy to the study of plant viruses in unpurified plant extracts. Microbiology 5(1):104-109.

Bayés M, Heath S, Gut IG. 2012. Applications of Second Generation Sequencing Technologies in Complex Disorders. In: Cryan JF, Reif A, editors. Behavioral Neurogenetics. Berlin, Heidelberg: Springer Berlin Heidelberg. p. 321-343.

Behncken G. 1983. A disease of chickpea caused by lettuce necrotic yellows virus. Australasian Plant Pathology 12(4):64-65.

Bejerman N, Giolitti F, de Breuil S, Trucco V, Nome C, Lenardon S, Dietzgen RG. 2015. Complete genome sequence and integrated protein localization and interaction map for alfalfa dwarf virus, which combines properties of both cytoplasmic and nuclear plant rhabdoviruses. Virology 483:275-283.

Berg JM, Tymoczko JL, Stryer L. 2002. Three-Dimensional Protein Structure Can Be Determined by NMR Spectroscopy and X-Ray Crystallography. Biochemistry. 5 ed. New York, USA: W H Freeman.

Bernardo V, Ribeiro Pinto LF, Albano RM. Gene expression analysis by real-time PCR: Experimental demonstration of PCR detection limits. Analytical Biochemistry [Internet].432(2):131-133. Available from: http://www.sciencedirect.com/science/article/pii/S000326971200485X

Berthiaume F, Morgan JR. Methods in Bioengineering: 3D Tissue Engineering [Internet]. Massachusetts, USA: Artech House; 2010. Available from: https://books.google.co.nz/books?id=4Y3S67RnRgIC

Bester R, Jooste AE, Maree HJ, Burger JT. 2012. Real-time RT-PCR high-resolution melting curve analysis and multiplex RT-PCR to detect and differentiate grapevine leafroll-associated virus 3 variant groups I, II, III and VI. Virology Journal 9(1):219-230.

Blawid R, Silva J, Nagata T. 2017. Discovering and sequencing new plant viral genomes by next-generation sequencing: description of a practical pipeline. Annals of Applied Biology 170(3):301-314.

Boakye DB, Randles JW. 1974. Epidemiology of lettuce necrotic yellows virus in South Australia. III. Virus transmission parameters, and vector feeding behaviour on host and non-host plants. Australian Journal of Agricultural Research 25(5):791-802.

Boonham N, Kreuze J, Winter S, van der Vlugt R, Bergervoet J, Tomlinson J, Mumford R. 2014. Methods in virus diagnostics: from ELISA to next generation sequencing. Virus Research 186:20-31.

Boyanton BL, Rushton JR. 2010. Molecular Techniques in Hematopathology. In: Crisan D, editor. Hematopathology: Genomic Mechanisms of Neoplastic Diseases. Totowa, NJ: Humana Press. p. 1-38.

Bozzola JJ, Russell LD. Electron Microscopy: Principles and Techniques for Biologists [Internet]. Toronoto, Canada: Jones and Bartlett; 1999. Available from: https://books.google.co.nz/books?id=zMkBAPACbEkC

Breitbart M, Rohwer F. Here a virus, there a virus, everywhere the same virus? Trends in Microbiology [Internet].13(6):278-284. Available from: http://www.sciencedirect.com/science/article/pii/S0966842X05001083

Brown TA. Genomes (Second edition) [Internet]. Oxford, UK: Wiley-Liss; 2002. Available from: https://www.ncbi.nlm.nih.gov/books/NBK21122/

Brunstein J. 2013. PCR: the basics of the polymerase chain reaction. Medical Laboratory Observer 45(4):32-35.

Callaghan B. 2005. Sequence analysis and variability study of Lettuce necrotic yellows virus. [Queensland, Australia]: The University of Queensland. p. 285.

Callaghan B, Dietzgen R. 2005. Nucleocapsid gene variability reveals two subgroups of Lettuce necrotic yellows virus. Archives of Virology 150(8):1661-1667.

Cameron E. 2000. Native sow thistle, Sonchus kirkii, rediscovered in the Auckland Region. Auckland Botanical Society Journal 55(1):21-24.

Carson S, Miller H, Witherow DS. Molecular Biology Techniques: A Classroom Laboratory Manual [Internet]. London, Uk: Academic Press; 2012. Available from: https://ebookcentral.proquest.com/lib/aut/detail.action?docID=802466.

Carter J, Saunders V. Virology : principles and applications [Internet]. Spain: John Wiley & Sons Ltd; 2013. Available from: https://ebookcentral.proquest.com/

Carver M, Woolcock LT. 1986. The introduction into Australia of biological control agents of Hyperomyzus lactucae (L.) (Gomoptera: Aphididae). Australian Journal of Entomology 25(1):65-69.

Cattoli G, Monne I. 2009. Molecular Diagnosis of Avian Influenza. In: Capua I, Alexander DJ, editors. Avian Influenza and Newcastle Disease: A Field and Laboratory Manual. Milano: Springer Milan. p. 87-111.

Cerritelli SM, Crouch RJ. 2009. Ribonuclease H: the enzymes in eukaryotes. The FEBS Journal 276(6):1494-1505.

Chambers P, Pringle CR, Easton AJ. 1990. Heptad repeat sequences are located adjacent to hydrophobic regions in several types of virus fusion glycoproteins. Journal of General Virology 71(12):3075-3080.

Chen YP, Siede R. 2007. Honey Bee Viruses. Advances in Virus Research. Academic Press. p. 33-80.

Choudhuri S. Bioinformatics for Beginners [Internet]. Oxford: Academic Press; 2014. Available from: https://www.sciencedirect.com/science/book/9780124104716

Chu P, Francki R. 1982. Detection of lettuce necrotic yellows virus by an enzyme-linked immunosorbent assay in plant hosts and the insect vector. Annals of Applied Biology 100(1):149-156.

Coll J. 1995a. The glycoprotein G of rhabdoviruses. Archives of Virology 140(5):827-851.

Coll JM. 1995b. Heptad-repeat sequences in the glycoprotein of rhabdoviruses. Virus Genes 10(2):107-114.

Conzelmann K-K. 1998. Nonsegmented negative-strand RNA viruses: genetics and manipulation of viral genomes. Annual Review of Genetics 32(1):123-162.

Da Poian A, Carneiro F, Stauffer F. 2005. Viral membrane fusion: is glycoprotein G of rhabdoviruses a representative of a new class of viral fusion proteins? Brazilian Journal of Medical and Biological Research 38(6):813-823.

Davila JI, Fadra NM, Wang X, McDonald AM, Nair AA, Barbara RC, Wu X, Blommel JH, Jen J, Rumilla KM. 2016. Impact of RNA degradation on fusion detection by RNA-seq. BMC Genomics 17(1):814-823.

Delhomme N, Mähler N, Schiffthaler B, Sundell D, Mannapperuma C, Hvidsten T, Street N. Guidelines for RNA-Seq data analysis. Epigenesys protocol [Internet].67:1-24. Available from: https://www.epigenesys.eu/images/stories/protocols/pdf/20150303161357_p67.pdf

Delmiglio C, Pearson M. 2006. Effects and incidence of Cucumber mosaic virus, Watermelon mosaic virus and Zucchini yellow mosaic virus in New Zealand's only native cucurbit, Sicyos australis. Australasian Plant Pathology 35(1):29-35.

DeLong E. Microbial Metagenomics, Metatranscriptomics, and Metaproteomics [Internet]. Oxford, UK: Elsevier Science; 2013. Available from: https://ebookcentral.proquest.com/lib/aut/detail.action?docID=1414035.

Deng Y-M, Spirason N, Iannello P, Jelley L, Lau H, Barr IG. 2015. A simplified Sanger sequencing method for full genome sequencing of multiple subtypes of human influenza A viruses. Journal of Clinical Virology 68:43-48.

Dervan A, Shendure J. 2017. Chapter 3 - The State of Whole-Genome Sequencing A2 - Ginsburg, Geoffrey S. In: Willard HF, editor. Genomic and Precision Medicine (Third Edition). Boston: Academic Press. p. 45-62.

Deuter P, White N, Putland D. n.d. Critical temperature thresholds Case study Lettuce. Queensland, Australia: Department of Agriculture, Fisheries and Forestry.

Dieffenbach C, Lowe T, Dveksler G. 1993. General concepts for PCR primer design. PCR Methods Applications 3(3):30-37.

Dietzgen RG. 2003. Monitoring of tospoviruses by real time polymerase chain reaction. Sydney, Australia: Department of Primary Industries & Horticulture Australia

Dietzgen RG. 2011. Cytorhabdovirus. In: Tidona C, Darai G, editors. The Springer Index of Viruses. New York, USA: Springer New York. p. 1709-1713.

Dietzgen RG, Callaghan B, Campbell PR. 2007. Biology and genomics of lettuce necrotic yellows virus. Plant Viruses 1(1):85-92.

Dietzgen RG, Callaghan B, Wetzel T, Dale JL. 2006. Completion of the genome sequence of Lettuce necrotic yellows virus, type species of the genus Cytorhabdovirus. Virus Research 118(1):16-22.

Dietzgen RG, Kondo H, Goodin MM, Kurath G, Vasilakis N. The family Rhabdoviridae: mono- and bipartite negative-sense RNA viruses with diverse genome organization and common evolutionary origins. Virus Research [Internet].227:158-170. Available from: http://www.sciencedirect.com/science/article/pii/S0168170216304579

Dietzgen RG, Mann KS, Johnson KN. 2016. Plant virus–insect vector interactions: Current and potential future research directions. Viruses 8(11):303-324.

Dietzgen RG, Twin J, Talty J, Selladurai S, Carroll M, Coutts B, Berryman D, Jones R. 2005. Genetic variability of Tomato spotted wilt virus in Australia and validation of real time RT-PCR for its detection in single and bulked leaf samples. Annals of Applied Biology 146(4):517-530.

Digiaro M, Elbeaino T, Martelli GP. 2007. Development of degenerate and species-specific primers for the differential and simultaneous RT-PCR detection of grapevine-infecting nepoviruses of subgroups A, B and C. Journal of Virological Methods 141(1):34-40.

Dijkstra J, Jager CP. 1998. Gel Double-Diffusion Test in Plates. Practical Plant Virology: Protocols and Exercises. Berlin, Heidelberg: Springer Berlin Heidelberg. p. 341-347.

Dorak MT. Genetic Association Studies: Background, Conduct, Analysis, Interpretation [Internet]. New York, USA: Taylor & Francis Group; 2017. Available from: https://books.google.co.nz/books?id=1TEoDQAAQBAJ

Dudley JT, Karczewski KJ. Exploring Personal Genomics [Internet]. Oxford, UK: Oxford University Press; 2013. Available from: https://books.google.co.nz/books?id=eH9pAgAAQBAJ

Dutta D, Mandal C, Mandal C. Unusual glycosylation of proteins: Beyond the universal sequon and other amino acids. Biochimica et Biophysica Acta (BBA) - General Subjects [Internet].1861(12):3096-3108. Available from: http://www.sciencedirect.com/science/article/pii/S0304416517302854

Edwardson JR, Christie RG. CRC Handbook of Viruses Infecting Legumes [Internet]. Florida, USA: Taylor & Francis; 1991. Available from: https://books.google.co.nz/books?id=_juxCOW7LkgC

Egerton RF. Physical Principles of Electron Microscopy: An Introduction to TEM, SEM, and AEM [Internet]. Boston, MA: Springer US; 2005. Available from: https://doi.org/10.1007/b136495

Fabre A-L, Colotte M, Luis A, Tuffet S, Bonnet J. 2014. An efficient method for long-term room temperature storage of RNA. European Journal of Human Genetics 22(3):379-385.

Fagan L, McLachlan A, Till C, Walker M. 2010. Synergy between chemical and biological control in the IPM of currant-lettuce aphid (Nasonovia ribisnigri) in Canterbury, New Zealand. Bulletin of Entomological Research 100(2):217-223.

Feng Y, Zhang Y, Ying C, Wang D, Du C. 2015. Nanopore-based fourth-generation DNA sequencing technology. Genomics, Proteomics & Bioinformatics 13(1):4-16.

Fenner F. 1983. The Florey Lecture, 1983: biological control, as exemplified by smallpox eradication and myxomatosis. Proceedings of the Royal Society of London B: Biological Sciences 218(1212):259-285.

Fereres A, Moreno A. Behavioural aspects influencing plant virus transmission by homopteran insects. Virus Research [Internet].141(2):158-168. Available from: http://www.sciencedirect.com/science/article/pii/S0168170208004383

Fields BN, Knipe DM, Howley PM. Fields' Virology [Internet]. Wolters Kluwer Health/Lippincott Williams & Wilkins; 2007. Available from: https://books.google.co.nz/books?id=5O0somr0w18C

Fletcher J, France CM, Butler RC. 2005. Virus surveys of lettuce crops and management of lettuce big-vein disease in New Zealand. New Zealand Plant Protection 58:239-244.

Fletcher J, Walker M, Davidson M, Paull S, Palmer A. 2017. Outdoor lettuce virus disease project 2016-2018 Year 1 report. Auckland, New Zealand: Plant & Food Research.

FloraBase. Sonchus hydrophilus Boulos Native Sowthistle [Internet]. Western Australia, Australia: Department of Biodiversity, Conservation and Attractions. Available from: https://florabase.dpaw.wa.gov.au/browse/profile/9367

Fox JG, Barthold S, Davisson M, Newcomer CE, Quimby FW, Smith A. The Mouse in Biomedical Research: Normative Biology, Husbandry, and Models [Internet]. California, UK: Academic Press; 2006. Available from: https://books.google.co.nz/books?id=Gi6rIvx7Ni4C

Franca LT, Carrilho E, Kist TB. 2002. A review of DNA sequencing techniques. Quarterly Reviews of Biophysics 35(2):169-200.

Francki R, Randles J, Dietzgen R. Lettuce necrotic yellows virus. AAB Descriptions of Plant Viruses [Internet]. (343). Available from: http://www.dpvweb.net/dpv/showadpv.php?dpvno=343

Fry P, Close R, Procter C, Sunde R. 1973. Lettuce necrotic yellows virus in New Zealand. New Zealand Journal of Agricultural Research 16(1):143-146.

Gandhi S, Scaria V. The Hitchhiker's Guide to Whole Exome Analysis [Internet]. New Delhi, India: Research in Genomics; 2016. Available from: https://books.google.co.nz/books?id=FAKADQAAQBAJ

Gardner RC, Howarth AJ, Hahn P, Brown-Luedi M, Shepherd RJ, Messing J. 1981. The complete nucleotide sequence of an infectious clone of cauliflower mosaic virus by M13mp7 shotgun sequencing. Nucleic Acids Research 9(12):2871-2888.

Gelderblom HR. 1996. Structure and classification of viruses. Chapter 41: Structure and classification of viruses,. Galveston, TX: University of Texas Medical Branch at Galveston.

Geneious. Geneious 6.0 [Internet]. Biomatters Ltd. Available from: https://assets.geneious.com/documentation/geneious/GeneiousManual6.0.pdf

Gerlier D, Lyles DS. 2011. Interplay between innate immunity and negative-strand RNA viruses: towards a rational model. Microbiology and Molecular Biology Reviews 75(3):468-490.

Ghosh D, Brooks RE, Wang R, Lesnaw J, Goodin MM. 2008. Cloning and subcellular localization of the phosphoprotein and nucleocapsid proteins of Potato yellow dwarf virus, type species of the genus Nucleorhabdovirus. Virus Research 135(1):26-35.

Gochhait S, Bukhari SI, Bamezai RNK. 2007. mRNA Quantitation Using Real Time PCR. In: Varma A, Oelmüller R, editors. Advanced Techniques in Soil Microbiology. Berlin, Heidelberg: Springer Berlin Heidelberg. p. 53-72.

Goldberg K-B, Modrell B, Hillman BI, Heaton LA, Choi T-J, Jackson A. 1991. Structure of the glycoprotein gene of sonchus yellow net virus, a plant rhabdovirus. Virology 185(1):32-38.

Gonella-Diaza AM, da Silva Andrade SC, Sponchiado M, Pugliesi G, Mesquita FS, Van Hoeck V, de Francisco Strefezzi R, Gasparin GR, Coutinho LL, Binelli M. Oviductal transcriptional profiling of a bovine fertility model by next-generation sequencing. Genomics Data [Internet].13:27-29. Available from: http://www.sciencedirect.com/science/article/pii/S2213596017300600

Graham SC, Assenberg R, Delmas O, Verma A, Gholami A, Talbi C, Owens RJ, Stuart DI, Grimes JM, Bourhy H. 2008. Rhabdovirus matrix protein structures reveal a novel mode of self-association. PLoS Pathogens 4(12):e1000251.

Grisoni M, Marais A, Filloux D, Saison A, Faure C, Julian C, Theil S, Contreras S, Teycheney P-Y, Roumagnac P. 2017. Two novel Alphaflexiviridae members revealed by deep sequencing of the Vanilla (Orchidaceae) virome. Archives of Virology 162(12):3855-3861.

Gromiha MM. Protein Bioinformatics: From Sequence to Function [Internet]. New Delhi, India: Elsevier Science; 2010. Available from: https://ebookcentral.proquest.com/lib/aut/detail.action?docID=691414.

Haff LA. 1994. Improved quantitative PCR using nested primers. Genome Research 3(6):332-337.

Harrison B, Crowley N. 1965. Properties and structure of lettuce necrotic yellows virus. Virology 26(2):297-310.

Hartl D, Ruvolo M. Genetics [Internet]. Massachusetts, USA: Jones & Bartlett Learning; 2012. Available from: https://books.google.co.nz/books?id=5G9MxI9BXm0C

Heather JM, Chain B. 2016. The sequence of sequencers: The history of sequencing DNA. Genomics 107(1):1-8.

Heim F, Lot H, Delecolle B, Bassler A, Krczal G, Wetzel T. 2008. Complete nucleotide sequence of a putative new cytorhabdovirus infecting lettuce. Archives of Virology 153(1):81-92.

Hernandez-Rodriguez P, Ramirez AG. Polymerase Chain Reaction: Types, Utilities and Limitations [Internet]. INTECH Open Access Publisher; 2012. Available from: https://www.intechopen.com/books/polymerase-chain-reaction/polymerase-chain-reaction-types-utilities-and-limitations

Higgins CM, Bejerman N, Li M, James AP, Dietzgen RG, Pearson MN, Revill PA, Harding RM. 2016a. Complete genome sequence of Colocasia bobone disease-associated virus, a putative cytorhabdovirus infecting taro. Archives of Virology 161(3):745-748.

Higgins CM, Chang W-L, Khan S, Tang J, Elliott C, Dietzgen RG. 2016b. Diversity and evolutionary history of lettuce necrotic yellows virus in Australia and New Zealand. Archives of Virology 161(2):269-277.

Holmes S. 2003. Bootstrapping phylogenetic trees: theory and methods. Statistical Science 18(2):241-255.

Houldcroft CJ, Beale MA, Breuer J. 2017. Clinical and biological insights from viral genome sequencing. Nature Reviews Microbiology 15(3):183-192.

Hu A, Colella M, Tam JS, Rappaport R, Cheng S-M. 2003. Simultaneous detection, subgrouping, and quantitation of respiratory syncytial virus A and B by real-time PCR. Journal of Clinical Microbiology 41(1):149-154.

Huang Y, Zhao H, Luo Z, Chen X, Fang R-X. 2003. Novel structure of the genome of Rice yellow stunt virus: identification of the gene 6-encoded virion protein. Journal of General Virology 84(8):2259-2264.

Hull R. Plant Virology (Fifth Edition) [Internet]. Boston: Academic Press; 2014. Available from: http://www.sciencedirect.com/science/book/9780123848710

Husemann P, Stoye J. 2009. Phylogenetic Comparative Assembly. In: Salzberg SL, Warnow T, editors. Algorithms in Bioinformatics: 9th International Workshop, WABI 2009, Philadelphia, PA, USA, September 12-13, 2009. Proceedings. Berlin, Heidelberg: Springer Berlin Heidelberg. p. 145-156.

ICTV. International Committee on Taxonomy Virus (ICTV): 2016 Release [Internet]. London, UK. Available from: http://www.ictvonline.org/virusTaxonomy.asp?msl_id=30

Illumina. Illumina Sequencing Technology [Internet]. California, USA: Illumina. Available from: https://www.illumina.com/documents/products/techspotlights/techspotlight_sequencing.pdf

Illumina. Quality Scores for Next-Generation Sequencing [Internet]. California, USA: Illumina. Available from: https://www.illumina.com/documents/products/technotes/technote_Q-Scores.pdf

Illumina. 2011b. TruSeq™ Sample Preparation Best Practices and Troubleshooting Guide. California, USA: Illumina.

Illumina. Coverage depth recommendations [Internet]. California, USA: Illumina. Available from: https://www.illumina.com/science/education/sequencing-coverage.html

Innis M, Gelfand D. 1999. Optimization of PCR: Conversations between Michael and David. PCR Applications. San Diego: Academic Press. p. 3-22.

Innis M, Gelfand DH. 1990. Optimisation of PCRs. PCR Protocols. San Diego: Academic Press. p. 3-12.

Integrated DNA Technologies. What characteristics make up the most efficient PCR primers? [Internet]. California, USA: Integrated DNA Technologies. Available from: http://www.idtdna.com/pages/decoded/decoded-articles/ask-alex/decoded/2012/01/10/what-characteristics-make-up-the-most-efficient-pcr-primers-

Ito T, Suzaki K, Nakano M. 2013. Genetic characterization of novel putative rhabdovirus and dsRNA virus from Japanese persimmon. Journal of General Virology 94(8):1917-1921.

Jackson AO, Dietzgen RG, Goodin MM, Bragg JN, Deng M. 2005. Biology of Plant Rhabdoviruses. Annual Review of Phytopathology 43:623-660.

Jeong J-j, Ju H-j, Noh J. 2014. A Review of Detection Methods for the Plant Viruses. Research in Plant Disease 20(3):173-181.

Kalaisekar A, Padmaja PG, Bhagwat VR, Patil JV. Insect Pests of Millets: Systematics, Bionomics, and Management [Internet]. London, UK: Elsevier Science; 2017. Available from: https://ebookcentral.proquest.com/lib/AUT/detail.action?docID=4773782.

Kaur N, Hasegawa DK, Ling K-S, Wintermantel WM. 2016. Application of genomics for understanding plant virus-insect vector interactions and insect vector control. Phytopathology 106(10):1213-1222.

Kavak P, Ergüner B, Üstek D, Yüksel B, Sağıroğlu MŞ, Güngör T, Alkan C. 2016. Improving Genome Assemblies Using Multi-platform Sequence Data. In: Angelini C, Rancoita PMV, Rovetta S, editors. Computational Intelligence Methods for Bioinformatics and Biostatistics: 12th International Meeting, CIBB 2015, Naples, Italy, September 10-12, 2015, Revised Selected Papers. Cham: Springer International Publishing. p. 220-232.

Kennedy S, Oswald N. PCR Troubleshooting and Optimization: The Essential Guide [Internet]. Norfolk, UK: Caister Academic Press; 2011. Available from: https://books.google.co.nz/books?id=oXoUkTSbnFgC

Kesanakurti P, Belton M, Saeed H, Rast H, Boyes I, Rott M. 2016. Screening for plant viruses by next generation sequencing using a modified double strand RNA extraction protocol with an internal amplification control. Journal of Virological Methods 236:35-40.

Khalifa ME, Varsani A, Ganley AR, Pearson MN. 2016. Comparison of Illumina de novo assembled and Sanger sequenced viral genomes: A case study for RNA viruses recovered from the plant pathogenic fungus Sclerotinia sclerotiorum. Virus Research 219:51-57.

Klerks M, Lindner J, Vaškova D, Špak J, Thompson J, Jelkmann W, Schoen C. 2004. Detection and tentative grouping of Strawberry crinkle virus isolates. European Journal of Plant Pathology 110(1):45-52.

Kochanowski B, Udo R. Quantitative PCR protocols [Internet]. New Jersey, USA: Humana Press; 1999. Available from: http://www.springer.com/gp/book/9781592592623

Kocincová L, Jarešová M, Byška J, Parulek J, Hauser H, Kozlíková B. 2017. Comparative visualization of protein secondary structures. BMC Bioinformatics 18(2):23-35.

Komínek P, Glasa M, Komínková M. 2009. Analysis of multiple virus-infected grapevine plant reveals persistence but uneven virus distribution. Acta Virologica 53(4):281-285.

Korpelainen E, Tuimala J, Somervuo P, Huss M, Wong G. RNA-seq Data Analysis: A Practical Approach [Internet]. New York, USA: CRC Press; 2014. Available from: https://ebookcentral.proquest.com/lib/aut/detail.action?docID=1589041.

Kumar D, Eng C. Genomic Medicine: Principles and Practice [Internet]. New York, USA: Oxford University Press; 2015. Available from: https://books.google.co.nz/books?id=9vZwBAAAQBAJ

Lee HH, Morse SA, Olsvik Ø. Nucleic Acid Amplification Technologies: Application to Disease Diagnosis [Internet]. Massachusetts, USA: Eaton publishin; 1997. Available from: http://www.springer.com/gp/book/9780817639211

Lee SVB, Abdul R. 2012. Discriminatory Power of Agarose Gel Electrophoresis in DNA Fragments Analysis. In: Magdeldin S, editor. Gel Electrophoresis - Principles and Basic. InTech.

Lemey P, Salemi M, Vandamme AM. The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing [Internet]. Cambridge, UK: Cambridge University Press; 2009. Available from: https://books.google.co.nz/books?id=C47QjT2XEY0C

Levesque-Sergerie J-P, Duquette M, Thibault C, Delbecchi L, Bissonnette N. 2007. Detection limits of several commercial reverse transcriptase enzymes: impact on the low-and high-abundance transcript levels assessed by quantitative RT-PCR. BMC Molecular Biology 8(1):93-111.

Li CX, Shi M, Tian J-H, Lin X-D, Kang Y-J, Chen L-J, Qin X-C, Xu J, Holmes EC, Zhang Y-Z. 2015. Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. eLife 4:1-26.

Li W, Zhang Y, Zhang C, Pei X, Wang Z, Jia S. 2014. Presence of poly (A) and poly (A)-rich tails in a positive-strand RNA virus known to lack 3′ poly (A) tails. Virology 454-455:1-10.

Life Science. PCR Troubleshooting [Internet]. California, USA: Life Science Available from: http://www.bio-rad.com/en-au/applications-technologies/pcr-troubleshooting - gel2

Lima JAA, Nascimento AKQ, Radaelli P, Purcifull DE. 2012. Serology applied to plant virology. In: Al-Moslih M, editor. Serological diagnosis of certain human, animal and plant diseases. InTech. p. 170.

Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M. 2012. Comparison of next-generation sequencing systems. BioMed Research International 2012:1-11.

Lodish H, Berk A, Zipursky S, Matsudaira P, Baltimore D, Darnell J. 2000. Viruses: Structure, Function, and Uses. Molecular Cell Biology, 4th Edition New York, USA: WH Freeman.

Lorenz TC. 2011. Polymerase chain reaction: basic protocol plus troubleshooting and optimization strategies. Journal of Visualized Experiments(63):e3998.

Lorsch J. Laboratory Methods in Enzymology: DNA [Internet]. Massachusetts, USA: Elsevier Science; 2013. Available from: https://ebookcentral.proquest.com/lib/aut/detail.action?docID=1375440

Luo Z, Fang R. 1998. Structure analysis of the rice yellow stunt rhabdovirus glycoprotein gene and its mRNA. Archives of Virology 143(12):2453-2459.

Mackay IM, Arden KE, Nitsche A. 2002. Real-time PCR in virology. Nucleic Acids Research 30(6):1292-1305.

Maina S, Edwards OR, de Almeida L, Ximenes A, Jones RA. 2017. Analysis of an RNA-seq Strand-Specific Library from an East Timorese Cucumber Sample Reveals a Complete Cucurbit aphid-borne yellows virus Genome. Genome Announcements 5(19):e00320-17.

Malmanger B, Harrington C, Savage S. 2012. Comparison of Agilent 2100 Bioanalyzer and Caliper Life Sciences GX II in Functionality, Total RNA Scoring Algorithms and Reproducibility to Evaluate Total RNA Integrity. Journal of Biomolecular Techniques 23:S49-S50.

Mandahar CL. Multiplication of RNA Plant Viruses [Internet]. Dordrecht, Netherlands: Springer Netherlands; 2006. Available from: https://doi.org/10.1007/1-4020-4725-8

Mann KS, Bejerman N, Johnson KN, Dietzgen RG. 2016a. Cytorhabdovirus P3 genes encode 30K-like cell-to-cell movement proteins. Virology 489:20-33.

Mann KS, Dietzgen RG. 2014. Plant rhabdoviruses: new insights and research needs in the interplay of negative-strand RNA viruses with plant and insect hosts. Archives of Virology 159(8):1889-1900.

Mann KS, Johnson KN, Carroll BJ, Dietzgen RG. 2016b. Cytorhabdovirus P protein suppresses RISC-mediated cleavage and RNA silencing amplification in planta. Virology 490:27-40.

Mann KS, Johnson KN, Dietzgen RG. 2015. Cytorhabdovirus phosphoprotein shows RNA silencing suppressor activity in plants, but not in insect cells. Virology 476:413-418.

Markoulatos P, Siafakas N, Moncany M. 2002. Multiplex polymerase chain reaction: a practical approach. Journal of Clinical Laboratory Analysis 16(1):47-51.

Martin KM, Dietzgen RG, Wang R, Goodin MM. 2012. Lettuce necrotic yellows cytorhabdovirus protein localization and interaction map, and comparison with nucleorhabdoviruses. Journal of General Virology 93(4):906-914.

Martinez N, Ribeiro EA, Leyrat C, Tarbouriech N, Ruigrok RW, Jamin M. 2013. Structure of the C-terminal domain of Lettuce Necrotic Yellows Virus phosphoprotein. Journal of Virology 87(17):9569-9578.

Massah A, Izadpanah K, Afsharifar A, Winter S. 2008. Analysis of nucleotide sequence of Iranian maize mosaic virus confirms its identity as a distinct nucleorhabdovirus. Archives of Virology 153(6):1041-1047.

McLean G, Wolanski B, Francki R. 1971. Serological analysis of lettuce necrotic yellows virus preparations by immunodiffusion. Virology 43(2):480-487.

Méndez-García C, Bargiela R, Martínez-Martínez M, Ferrer M. 2018. Chapter 2 - Metagenomic Protocols and Strategies A2 Metagenomics. London, UK: Academic Press. p. 15-54.

Metzker ML. 2010. Sequencing technologies—the next generation. Nature Reviews Genetics 11(1):31-46.

Mitra S. Sample Preparation Techniques in Analytical Chemistry [Internet]. New Jersey, USA: John Wiley & Sons, Inc.; 2004. Available from: http://dx.doi.org/10.1002/0471457817

Mittal B, Chaturvedi P, Tulsyan S. 2013. Restriction Fragment Length Polymorphism A2 - Maloy, Stanley. In: Hughes K, editor. Brenner's Encyclopedia of Genetics (Second Edition). San Diego: Academic Press. p. 190-193.

Mullan B, Fanning LJ, Shanahan F, Sullivan DG. 2002. RT-PCR for the Assessment of Genetically Heterogenous Populations of the Hepatitis C Virus. In: O'Connell J, editor. RT-PCR Protocols. Totowa, NJ: Humana Press. p. 171-188.

Muthaiyan MC. Principles and Practices of Plant Quarantine [Internet]. New Delhi, India: Allied Publishers; 2009. Available from: https://books.google.co.nz/books?id=23zJygAACAAJ

Naidu R, Hughes J. 2003. Methods for the detection of plant virus diseases. Plant Virology in Sub Saharan Africa:233-253.

Nakano K, Shiroma A, Shimoji M, Tamotsu H, Ashimine N, Ohki S, Shinzato M, Minami M, Nakanishi T, Teruya K. 2017. Advantages of genome sequencing by long-read sequencer using SMRT technology in medical area. Human Cell 3:149-161.

Narayanasamy P. Molecular Biology in Plant Pathogenesis and Disease Management: Disease Development [Internet]. New York, USA: Springer, Dordrecht; 2008. Available from: https://doi-org.ezproxy.aut.ac.nz/10.1007/978-1-4020-8245-0

Narayanasamy P. Microbial Plant Pathogens-Detection and Disease Diagnosis: Viral and Viroid Pathogens [Internet]. Dordrecht: Springer Netherlands; 2011a. Available from: https://doi-org.ezproxy.aut.ac.nz/10.1007/978-90-481-9754-5

Narayanasamy P. Microbial plants pathogens-detection and disease diagnosis: [Internet]. New York, USA: Springer; 2011b. Available from: https://doi-org.ezproxy.aut.ac.nz/10.1007/978-90-481-9754-5

Nei M, Kumar S. Molecular Evolution and Phylogenetics [Internet]. New York, USA: Oxford University Press; 2000. Available from: https://books.google.co.nz/books?id=0nt-qaAfIbAC

Nixon C. 2015. How valuable is that plant species? Application of a method for enumerating the contribution of selected plant species to New Zealand's GDP. Wellington, New Zealand: Ministry for Primary Industries.

O'Loughlin GT, Chambers TC. The systemic infection of an aphid by a plant virus. Virology [Internet].33(2):262-271. Available from: http://www.sciencedirect.com/science/article/pii/0042682267901456

O'Connell J. 2002. The Basics of RT-PCR. In: O'Connell J, editor. RT-PCR Protocols. Totowa, NJ: Humana Press. p. 19-25.

Obalinsky TR. Protein Folding: New Research [Internet]. New York, USA: Nova Science Publishers; 2006. Available from: https://books.google.co.nz/books?id=TxJBx_S_ahQC

Otago Genomics & Bioinformatics Facility. Sample requirements [Internet]. Otago, New Zealand: Otago Genomics & Bioinformatics Facility. Available from: http://www.otago.ac.nz/genomics/project-sample-info/requirements/index.html

Parrella G, Greco B. 2016. Sequence variation of block III segment identifies three distinct lineages within Eggplant mottled dwarf virus isolates from Italy, Spain and Greece. Acta Virologica 60(1):100-105.

Pelt-Verkuil Ev, Belkum Av, Hays JP. Principles and Technical Aspects of PCR Amplification [Internet]. Leiden, Netherlands: Springer Netherlands; 2008. Available from: https://doi-org.ezproxy.aut.ac.nz/10.1007/978-1-4020-6241-4

Pestana EA, Belak S, Diallo A, Crowther JR, Viljoen GJ. 2010. Traditional PCR. Early, rapid and sensitive veterinary molecular diagnostics - real time PCR applications. Dordrecht: Springer Netherlands. p. 9-25.

Pfaffl MW. 2006. Relative quantification. In: Dorak MT, editor. Real-time PCR. New York, USA: Taylor and Francis Group.

Posthuma K, Adams A, Hong Y, Kirby M. 2002. Detection of Strawberry crinkle virus in plants and aphids by RT-PCR using conserved L gene sequences. Plant Pathology 51(3):266-274.

Puli'uvea C, Khan S, Chang W-L, Valmonte G, Pearson MN, Higgins CM. 2017. First complete genome sequence of vanilla mosaic strain of Dasheen mosaic virus isolated from the Cook Islands. Archives of Virology 162(2):591-595.

Raiol T, Agustinho DP, Simi KCR, Maria de Souza Silva C, Walter ME, Silva-Pereira I, Brígido M. 2014. Transcriptome Analysis Throughout RNA-seq. In: Passos GA, editor. Transcriptomics in Health and Disease. Cham: Springer International Publishing. p. 49-68.

Randles JW, Carver M. 1971. Epidemiology of lettuce necrotic yellows virus in South Australia. II. Distribution of virus, host plants, and vectors. Crop and Pasture Science 22(2):231-237.

Reddy SR, Reddy SM. Essentials of virology [Internet]. Jodhpur, India: Scientific publichers 2012. Available from: https://books.google.co.nz/books?id=pbs4DwAAQBAJ

Redinbaugh MG, Hogenhout SA. 2005. Plant Rhabdoviruses. In: Fu ZF, editor. The World of Rhabdoviruses. Berlin, Heidelberg: Springer Berlin Heidelberg. p. 143-163.

Regenmortel MHVV, Mahy BWJ. Desk Encyclopedia of Plant and Fungal Virology [Internet]. Oxford, UK: Academic Press; 2010. Available from: https://books.google.co.nz/books?id=8Mnnr-_mtIwC

Reiter M, Ptatfl MW. 2011. RT-PCR optimization strategies. PCR Troubleshooting and Optimization: The Essential Guide. Norfolk, UK: Caister Academic Press.

Revill P, Trinh X, Dale J, Harding R. 2005. Taro vein chlorosis virus: characterization and variability of a new nucleorhabdovirus. Journal of General Virology 86(2):491-499.

Rhoads A, Au KF. 2015. PacBio sequencing and its applications. Genomics, Proteomics & Bioinformatics 13(5):278-289.

Roche S, Bressanelli S, Rey FA, Gaudin Y. 2006. Crystal structure of the low-pH form of the vesicular stomatitis virus glycoprotein G. Science 313(5784):187-191.

Roche S, Rey FA, Gaudin Y, Bressanelli S. 2007. Structure of the prefusion form of the vesicular stomatitis virus glycoprotein G. Science 315(5813):843-848.

Roe BA. Frederick Sanger (1918–2013). Genome Research [Internet].24(4):xi-xii. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3975070/

Roossinck MJ, Zhang L, Hellwald K-H. 1999. Rearrangements in the 5′ nontranslated region and phylogenetic analyses of cucumber mosaic virus RNA 3 indicate radial evolution of three subgroups. Journal of Virology 73(8):6752-6758.

Rosner A, Maslenin L, Spiegel S. 1997. The use of short and long PCR products for improved detection of prunus necrotic ringspot virus in woody plants. Journal of Virological Methods 67(2):135-141.

Rychlik W, Spencer W, Rhoads R. 1990. Optimization of the annealing temperature for DNA amplification in vitro. Nucleic Acids Research 18(21):6409-6412.

Sastry KS. Plant Virus and Viroid Diseases in the Tropics: Volume 1: Introduction of Plant Viruses and Sub-Viral Agents, Classification, Assessment of Loss, Transmission and Diagnosis [Internet]. Dordrecht: Springer Netherlands; 2013a. Available from: https://doi.org/10.1007/978-94-007-6524-5

Sastry KS. Seed-borne plant virus diseases [Internet]. New Delhi, India: Springer India; 2013b. Available from: https://doi.org/10.1007/978-81-322-0813-6

Savva G, Dicks J, Roberts IN. 2003. Current approaches to whole genome phylogenetic analysis. Briefings in Bioinformatics 4(1):63-74.

Schrader C, Schielke A, Ellerbroek L, Johne R. 2012. PCR inhibitors–occurrence, properties and removal. Journal of Applied Microbiology 113(5):1014-1026.

Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, Lightfoot S, Menzel W, Granzow M, Ragg T. 2006. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Molecular Biology 7(1):3-17.

Seddas P, Boissinot S. Glycosylation of beet western yellows virus proteins is implicated in the aphid transmission of the virus. Archives of Virology [Internet].151(5):967-984. Available from: https://doi.org/10.1007/s00705-005-0669-8

Sela N, Lachman O, Reingold V, Dombrovsky A. 2013. A new cryptic virus belonging to the family Partitiviridae was found in watermelon co-infected with Melon necrotic spot virus. Virus Genes 47(2):382-384.

Seshasayee ASN. Bacterial Genomics [Internet]. Delhi, India: Cambridge University Press; 2015. Available from: https://ebookcentral.proquest.com/lib/aut/detail.action?docID=1983202.

Shakin-Eshleman SH, Spitalnik SL, Kasturi L. 1996. The amino acid at the X position of an Asn-X-Ser sequon is an important determinant of N-linked core-glycosylation efficiency. Journal of Biological Chemistry 271(11):6363-6366.

Shao J, Liu X, Ly H, Liang Y. 2016. Characterization of the Glycoprotein Stable Signal Peptide in Mediating Pichinde Virus Replication and Virulence. Journal of Virology 90(22):10390-10397.

Sharma PC, Chaudhary S. 2016. NGS-Based Expression Profiling of HSP Genes During Cold and Freeze Stress in Seabuckthorn (Hippophae rhamnoides L.). In: Asea AAA, Kaur P, Calderwood SK, editors. Heat Shock Proteins and Plants. Cham: Springer International Publishing. p. 309-327.

Sheng Q, Vickers K, Zhao S, Wang J, Samuels DC, Koues O, Shyr Y, Guo Y. Multi-perspective quality control of Illumina RNA sequencing data analysis. Briefings in Functional Genomics [Internet].16(4):194-204. Available from: http://dx.doi.org/10.1093/bfgp/elw035

Simko I. 2016. High-resolution DNA melting analysis in plant research. Trends in Plant Science 21(6):528-537.

Sin S-H, McNulty BC, Kennedy GG, Moyer JW. 2005. Viral genetic determinants for thrips transmission of Tomato spotted wilt virus. Proceedings of the National Academy of Sciences of the United States of America 102(14):5168-5173.

Singh VK, Govindarajan R, Naik S, Kumar A. 2000. The effect of hairpin structure on PCR amplification efficiency. Molecular Biology Today 1(3):67-69.

Smith KM. 1974. Testing for Viruses: Indicator Plants. Plant Viruses. Boston, MA: Springer US. p. 163-168.

Stadhouders R, Pas SD, Anber J, Voermans J, Mes TH, Schutten M. 2010. The effect of primer-template mismatches on the detection and quantification of nucleic acids using the 5′ nuclease assay. The Journal of Molecular Diagnostics 12(1):109-117.

Stagnati L, Soffritti G, Lanubile A, Busconi M. 2017. Comparison of six methods for the recovery of PCR-compatible microbial DNA from an agricultural biogas plant. Applied Microbiology and Biotechnology 9(101):3907-3917.

Stubbs LL, Grogan RG. Lettuce Necrotic Yellows Virus. Nature [Internet].197(4873):1229-1229. Available from: http://dx.doi.org/10.1038/1971229a0

Stubbs LL, Grogan RG. 1963b. Necrotic yellows: a newly recognized virus disease of lettuce. Crop and Pasture Science 14(4):439-459.

Stussi-Garaud C, Haeberle A-M, Ritzenthaler C, Rohfritsch O, Lebeurier G. 1994. Electron microscopy of plant viruses. Biology of the Cell 80(2-3):147-153.

Sward R. 1990. Lettuce necrotic yellows rhabdovirus and other viruses infecting garlic. Australasian Plant Pathology 19(2):46-51.

Tanno F, Nakatsu A, Toriyama S, Kojima M. 2000. Complete nucleotide sequence of Northern cereal mosaic virus and its genome organization. Archives of Virology 145(7):1373-1384.

Teakle D. Virus and viroid diseases and their control [Internet]. New South Wales, Australia: Rockvale Publications; 1997. Available from: https://www.appsnet.org/Publications/Brown_Ogle/30_Control-virus_&_viroids (DT).pdf

Tennant P, Fermin G. Virus Diseases of Tropical and Subtropical Crops [Internet]. Oxfordshire, UK: CABI; 2015. Available from: https://books.google.co.nz/books?id=f84_CwAAQBAJ

Thermo Fisher Scientific. Real-time PCR handbook [Internet]. Massachusetts, USA: Thermo Fisher Scientific,. Available from: https://www.thermofisher.com/content/dam/LifeTech/Documents/PDFs/PG1503 -PJ9169-CO019861-Update-qPCR-Handbook-branding-Americas-FLR.pdf

Thermo Fisher Scientific. Is Your RNA Intact? Methods to Check RNA Integrity [Internet]. Massachusetts, USA: ThermoFIsher Scientific. Available from: https://www.thermofisher.com/nz/en/home/references/ambion-tech-support/rna-isolation/tech-notes/is-your-rna-intact.html

Thomson D, Dietzgen RG. 1995. Detection of DNA and RNA plant viruses by PCR and RT-PCR using a rapid virus release protocol without tissue homogenization. Journal of Virological Methods 54(2-3):85-95.

Tuffaha MSA. Phenotypic and Genotypic Diagnosis of Malignancies: An Immunohistochemical and Molecular Approach [Internet]. Baden-Württemberg, Germany: John Wiley & Sons; 2008. Available from: https://books.google.co.nz/books?id=bo3hzJixbjIC

Tuffereau C, Fischer S, Flamand A. 1985. Phosphorylation of the N and M1 proteins of rabies virus. Journal of General Virology 66(10):2285-2289.

Valencia CA, Pervaiz MA, Husami A, Qian Y, Zhang K. Next generation sequencing technologies in medical genetics [Internet]. New York, USA: Springer New York; 2013. Available from: https://doi.org/10.1007/978-1-4614-9032-6

Voelkerding KV, Dames SA, Durtschi JD. 2009. Next-generation sequencing: from basic research to diagnostics. Clinical Chemistry 55(4):641-658.

Walker GW, P: Stufkens, M; Wringt, P; Fletcher, J; Curtis, C; MacDonald, D; Winkler, S; Qureshi, S; Walker, M; James, D & Davis, S. 2005. Integrated Pest and Disease Management (IPM) for outdoor lettuce. Gisborne, New Zealand: New Zealand Institute for Crop & Food Research Limited No. 1467.

Walker PJ, Dietzgen RG, Joubert DA, Blasdell KR. 2011. Rhabdovirus accessory genes. Virus Research 162(1):110-125.

Wang Q, Ma X, Qian S, Zhou X, Sun K, Chen X, Zhou X, Jackson AO, Li Z. Rescue of a Plant Negative-Strand RNA Virus from Cloned cDNA: Insights into Enveloped Plant Virus Movement and Morphogenesis. PLOS Pathogens [Internet].11(10):e1005223. Available from: https://doi.org/10.1371/journal.ppat.1005223

Wetzel T, Dietzgen R, Dale J. 1994a. Genomic organization of lettuce necrotic yellows rhabdovirus. Virology 200(2):401-412.

Wetzel T, Dietzgen R, Geering A, Dale J. 1994b. Analysis of the nucleocapsid gene of lettuce necrotic yellows rhabdovirus. Virology 202(2):1054-1057.

Whyte LG, Greer CW. 2005. Molecular Techniques for Monitoring and Assessing Soil Bioremediation. Monitoring and Assessing Soil Bioremediation. Berlin, Heidelberg: Springer Berlin Heidelberg. p. 201-231.

Woiwod IP, Rothery P, Zhou X. 2000. Analysis of Population Fluctuations in the Aphid Hyperomyzus Lactucae and the Moth Perizoma Alchemillata. In: Perry JN, Smith RH, Woiwod IP, Morse DR, editors. Chaos in Real Data: The Analysis of Non-Linear Dynamics from Short Ecological Time Series. Dordrecht: Springer Netherlands. p. 97-120.

Wylie SJ, Jones MGK. 2011. The complete genome sequence of a Passion fruit woodiness virus isolate from Australia determined using deep sequencing, and its relationship to other potyviruses. Archives of Virology 156(3):479-482.

Xiao H, Kim W-S, Meng B. 2015. A highly effective and versatile technology for the isolation of RNAs from grapevines and other woody perennials for use in virus diagnostics. Virology Journal 12(1):171-186.

Yadav N, Khurana SP. 2016. Plant Virus Detection and Diagnosis: Progress and Challenges. In: Shukla P, editor. Frontier Discoveries and Innovations in Interdisciplinary Microbiology. New Delhi: Springer India. p. 97-132.

Yaffe H, Buxdorf K, Shapira I, Ein-Gedi S, Zvi MM-B, Fridman E, Moshelion M, Levy M. 2012. LogSpin: a simple, economical and fast method for RNA isolation from infected or healthy plants and other eukaryotic tissues. BMC Research Notes 5(1):45-53.

Yan T, Zhu J-R, Di D, Gao Q, Zhang Y, Zhang A, Yan C, Miao H, Wang X-B. 2015. Characterization of the complete genome of Barley yellow striate mosaic virus reveals a nested gene encoding a small hydrophobic protein. Virology 478:112-122.

Yang F, Wang G, Xu W, Hong N. 2017. A rapid silica spin column-based method of RNA extraction from fruit trees for RT-PCR detection of viruses. Journal of Virological Methods 247:61-67.

Yang X, Huang J, Liu C, Chen B, Zhang T, Zhou G. 2016. Rice Stripe Mosaic Virus, a Novel Cytorhabdovirus Infecting Rice via Leafhopper Transmission. Frontiers in Microbiology 7:1-13.

Yourno J. 1992. A method for nested PCR with single closed reaction tubes. Genome Research 2(1):60-65.

Zhai Y, Miglino R, Sorrentino R, Masenga V, Alioto D, Pappu H. 2014. Complete genomic characterization of eggplant mottled dwarf virus from agapanthus sp. by deep sequencing and de novo assembly. Journal of Plant Pathology 96(3):585-588.

Zhao L, Feng C, Wu K, Chen W, Chen Y, Hao X, Wu Y. 2017. Advances and prospects in biogenic substances against plant virus: A review. Pesticide Biochemistry and Physiology 135:15-26.

Zhou Q, Su X, Ning K. 2014. Assessment of quality control approaches for metagenomic data analysis. Scientific Reports 4(6957).