An investigation of the potato *eIF4E* isoforms as targets for non-transgenic CRISPR/Cas9 genome editing for viral resistance

by

Rebecca Jane Hurst

Thesis presented in partial fulfilment of the requirements for the degree of

Master of Science (Plant Biotechnology)

at

Stellenbosch University

Institute for Plant Biotechnology, Department of Genetics, Faculty of Science

Supervisor: **Professor James Lloyd**

Co-Supervisors: **Professor Johan Burger** and **Dr Manuela Campa**

December 2022

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third-party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Rebecca Jane Hurst

December 2022

Acknowledgements

I would like to acknowledge and give thanks to the following individuals and institutions:

To **Stellenbosch University's Postgraduate Scholarship Program** for their financial support.

To **Professor James Lloyd**, thank you for being all that I could've hoped for in a supervisor. Your support and faith in me were always uplifting and given exactly when I needed it. Thank you for your wisdom and advice. I do not think I would've made it here with any other supervisor.

To **Professor Johan Burger** and **Dr Manuela Campa**, thank for welcoming me into the 'CRISPy Crew' and for always offering me guidance as I navigated this new field of research. The knowledge I gained from you both was foundational in this work.

To **Dr Kenneth Oberlander**, you were an unexpected teacher in this journey. Thank you for your assistance with my phylogenetic work, and for your patience with me as I grappled with the new concepts.

To my colleagues at the **Institute for Plant Biotechnology**, I am immensely grateful for the continuous, and unconditional support I have received from you all. The benefits of the culture of cooperation in this laboratory is something I will never forget. I feel blessed to have started my scientific career surrounded by the wealth of knowledge and friendship you provided.

To my **mum**, thank you for always encouraging me to follow my curiosity - without it I would not be here.

To my **dad**, thank you for showing me what it means to work hard at your passions. I have always admired your work ethic and I hope I mirror you in that way.

Abstract

Solanum tuberosum (potato) is an important food source in Southern Africa. Viral infection of potato plants leads to decreased tuber size and number, and symptomatic tubers are often unfit for consumption. Increased prevalence of the viruses Potato Virus Y (PVY) and Potato Leaf Roll Virus (PLRV) in South Africa threatens the country's food security. The *eukaryotic translation initiation factor 4E* (*eIF4E)* gene family encode proteins that are involved in native and viral RNA translation. This mechanism is essential for viral survival and the *eIF4E* family are promising susceptibility factors that can be manipulated to confer viral resistance in plants. Expression of the potato *eIF4E* isoforms – *eIF4E-1, eIF4E-2, eIF(iso)4E* – and the related gene *new RNA cap binding protein* (*nCBP)*, was investigated in tubers of PLRV-infected plants and compared to that of tubers from healthy plants. No significant difference was observed between the samples which may indicate that expression differences are tissue-specific, rather than stress-induced. A Bayesian maximum clade credibility tree was created to elucidate the emergence of *eIF4E* isoforms across plant evolution. Division of the plant RNA cap-binding proteins into two distinct groups - eIF4E and nCBP –occurred in the common ancestor of all land plants. The *eIF4E*-ancestor divided into *eIF4E* and *eIF(iso)4E* in at least the angiosperms and possibly as far back as the vascular plants. The further division of *eIF4E* into *eIF4E-1* and *eIF4E-2* only occurred recently, in an ancestor of the Solanaceous family. During infection, the essential interaction between the viral genome-linked protein (VPg) and the host translation machinery is facilitated by three amino acid residues in the eIF4E cap-binding pocket. Amino acid alignments of the VPg binding region of the potato eIF4E proteins indicate that all have the potential for interaction with the viral protein. This implies that the VPg binding capacity of all these genes would need to be disrupted to engineer complete resistance to PVY and PLRV. To begin the process of creating *eIF4E-1, eIF4E-2* and *eIF(iso)4E* knockout mutants in potato, single guide RNAs were designed for all three genes. Additionally, a single guide RNA that is capable of simultaneously knocking out *eIF4E-1* and *eIF4E-2* was designed. Each of these guides were transcribed *in vitro* and complexed with Cas9. *In vitro* efficacy assays demonstrated that all ribonucleoproteins could induce double stranded breaks to the target genes. Transformation of the CRISPR/Cas9 ribonucleoproteins into protoplasts provides a transgene-free method of *eIF4E* gene editing. The isolation of viable potato protoplasts was established, and tissue culture of these cells yielded micro-calli.

Opsomming

Solanum tuberosum (aartappel) is ‘n belangrike voedselbron in suiderlike Afrika. Virus infeksie in aartappelplante lei tot ‘n afname in knol grootte en hoeveelheid en simptomatiese knolle is dikwels ongeskik vir verbruik. ‘n Verhoogde voorkoms in Suid-Afrika van die virusse “Potato Virus Y (PVY)” en “Potato Leaf Roll Virus (PLRV)” dreig die land se voedselsekuriteit. Die *eIF4E*-geenfamilie kodeer vir proteïene wat betrokke is by inheemse asook virale RNA-translasie. Hierdie meganisme is noodsaaklik vir virale oorlewing, dus is die manipulasie van die *eIF4E*-familie belowende vatbaarheidsfaktore om virale weerstand aan plante te verleen. Die uitdrukking van die aartappel *eIF4E* isovorme – *eIF4E-1, eIF4E-2, eIF(iso)4E* – en die verwante geen *nCBP*, is in knolle van PLRV-geïnfekteerde plante ondersoek en vergelyk met dié uitdrukking in knolle van gesonde plante. Geen beduidende verskil is tussen die monsters waargeneem nie, wat kan aandui dat die uitdrukkingsverskille weefselspesifiek is, eerder as stresgeïnduseerd. ‘n “Bayesian maximum clade credibility tree” filogenetiese boom is geskep om die opkoms van die *eIF4E* isovorme oor plant evolusie toe te lig. Die verdeling van die plantkap-bindende proteïene in drie afsonderlike groepe - *eIF4E, eIF(iso)4E* en *nCBP* – het in die gemeenskaplike voorouer van alle landplante plaasgevind. Die verdere verdeling van *eIF4E* in *eIF4E-1* en *eIF4E-2* het meer onlangs plaasgevind in ‘n voorouer van die Solanaceous-familie. Tydens infeksie word die noodsaaklike interaksie tussen die viral genome-linked protein (VPg) en die gasheer se translasiemasjienerie deur drie aminosuurreste in die eIF4E kap-bindende sak fasiliteer. Alignments van die aminosure van die VPg-bindingsgebied van twee afsonderlike klades dui aan dat al die gene in beide groepe potensiaal het vir interaksie met die virale proteïen. Hierdie impliseer dat die VPg-bindingskapasiteit van al hierdie gene ontwrig sal moet word om volledige weerstand teen PVY en PLRV te verkry. Om *eIF4E-1, eIF4E-2* en *eIF(iso)4E* mutante in aartappels te skep, is enkelgids-RNAs vir elke geen ontwerp. Verder is ‘n enkelgids-RNA ontwerp om gelyktydig *eIF4E-1* en *eIF4E-2* te muteer. Elkeen van hierdie gidse is *in vitro* getranskribeer en gekompleks met Cas9. *In vitro* doeltreffendheidstoetse het getoon dat al hierde ribonukleoproteïene dubbelstringbreuke aan die teikengene kan veroorsaak. Transformasie van die CRISPR/Cas9 ribonukleoproteïene in protoplaste bied ‘n transgeenvrye metode van *eIF4E* geenredigering. Die isolasie van lewensvatbare aartappelprotoplaste is vasgestel en weefselkultuur van hierdie selle het mikro-kallus opgelwer.

Table of contents

List of figures

Chapter 2:

Chapter 3:

collect at the solutions' interface, indicated with the red box **(D).** However, when conducted incorrectly, all protoplasts rupture due to the high osmotic potential of the solution **(E)**.

Figure 4.6 Viable protoplasts were quantified in a haemocytometer and suspended in alginate lenses **(A)**. The alginate lenses were left to solidify at room temperature until a clear convex shape was observed. **(B)** Micro-calli formation was visible in the alginate lenses after 3 weeks.

Figure 4.7 A diagrammatic representation of the *Solanum tuberosum eIF4E-1*, *eIF4E-2* and *eIF(iso)4E* gene structure. The polymerase chain reaction amplicon locations used in this research are indicated. The single guide RNA locations for each gene are labelled as 'Guide A' or Guide B'. The target site of the *eIF4E-1* and *eIF4E-2* double knockout guide is labelled as 'DKO guide'. The location of the viral genome linked protein (VPg) biding amino acids on *eIF4E-1* is also indicated.

List of tables

Chapter 2:

Chapter 3:

Chapter 4:

List of abbreviations

µg	microgram
µL	microlitre
µM	micromolar
BAP	6-benzylaminopurine
bp	base pair
CAF	Central Analytical Facilities
Cas	clustered regularly interspaced short palindromic repeats -associated
cDNA	complementary deoxyribonucleic acid
CDS	coding sequence
CRISPR	clustered regularly interspaced short palindromic region
crRNA	clustered regularly interspaced palindromic repeat regions ribonucleic acid
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
DSB	double stranded deoxyribonucleic acid break
EDTA	ethylenediaminetetraacetic acid
eF	elongation factor
eIF	eukaryotic translation initiation factor
g	gram
gDNA	genomic deoxyribonucleic acid
K	lysine
kb	kilobases
LB	lysogeny broth
LBA	lysogeny broth agar
M	molar
mg	milligram
min	minute
mL	millilitre
mM	millimolar
mRNA	messenger ribonucleic acid
NAA	1-naphthaleneacetic acid
NCBI	National Centre for Biotechnology Information
nCBP	new ribonucleic acid cap-binding protein
NEB	New England Biolabs
nM	nanomolar

NTP	nucleoside triphosphate
PAM	protospacer adjacent motif
PCR	polymerase chain reaction
PDS	phytoene desaturase
PLRV	Potato Leaf Roll Virus
PVY	Potato Virus Y
R	arginine
RNA	ribonucleic acid
RNP	ribonucleoprotein
rpm	revolutions per minute
RT	room temperature
RT-qPCR	Real-time quantitative polymerase chain reaction
sec	seconds
sgRNA	single guide ribonucleic acid
SNP	single nucleotide polymorphism
T_a	annealing temperature
TALENs	transcription-activator like effector nucleases
TB	terrific broth
TILLING	Targeting Induced Local Lesions in Genomes
T_m	melting temperature
tracrRNA	trans-acting clustered regularly interspaced palindromic repeat regions ribonucleic acid
U	enzyme units
V	volts
v/v	volume for volume
v/v/v	volume for volume for volume
vol	volume
VPg	viral genome-linked protein
w/v	weight for volume
ZF	zinc finger
ZFN	zinc finger nucleases

Chapter 1: Introduction

Potato is a staple food source across the world and its importance will only increase as the global human population increases. Viral infection significantly decreases potato tuber yields and threatens global food security. *Eukaryotic translation initiation factor 4E (eIF4E)* genes encode a family of plant proteins that are involved in RNA translation. This gene family has been identified as promising targets for engineering virus resistance in plants as the translation complexes they are associated with also facilitate translation of the viral genome. There are three *eIF4E* isoforms in potato, and a related new RNA cap-binding protein called *nCBP*, which all arose through gene duplications during plant evolution. The points at which these duplication events occurred have yet to be established. Some of the isoforms have been demonstrated to be involved in the replication of viral genomes, however the roles of all potato isoforms during viral infection are still unknown. Genome editing of potato protoplasts with CRISPR/Cas9 ribonucleoproteins provides a promising possible solution to examine this, and to potentially engineer *eIF4E*-mediated viral resistance in potato plants.

1.1 Aims and objectives

This study had 4 aims which are outlined below, along with the objective steps that were taken to achieve them:

1. To gain more knowledge about the differential expression patterns of the potato *eIF4E* genes under viral stress conditions.
    a. Real-time quantitative polymerase chain reactions will be conducted to assess the expression of *eIF4E-1*, *eIF4E-2*, *eIF(iso)4E* and *nCBP* in Potato Leaf Roll Virus-infected and healthy potato plants.
2. To assemble a phylogenetic tree that predicts the emergence of the different *eIF4E* isoforms in plants.
    a. Gene sequences from representative species of major plant phyla will be collected and aligned.
    b. A phylogenetic tree will be created using the Bayesian Interference model.
3. To design and assess CRISPR/Cas9-mediated *eIF4E* gene editing *in vitro* in preparation for future *in vivo* experiments.
    a. Two single guide RNAs will be designed for each gene-of-interest.
    b. CRISPR/Cas9 ribonucleoproteins will be assembled, and assays screening their efficacy will be conducted *in vitro.*
4. Establish a procedure for potato protoplast culture.
    a. The creation and isolation of potato protoplasts will be established and optimised.
    b. Viable protoplasts will be cultured until micro-calli formation.

1.2 Overview of chapters

Chapter 2 of this thesis is a review of the current literature surrounding the topics of this study. The role of the *eIF4E* gene family in both native and viral translation is described. Progress that has been made towards engineering *eIF4E*-mediated virus resistance in plants is reviewed in depth. Traditional and new breeding techniques are discussed and evaluated, with particular focus on CRISPR/Cas9 ribonucleoprotein delivery into protoplasts for genome editing of potato.

Chapter 3 of this thesis is the first research chapter. In this, the expression of the potato *eIF4E* gene family under Potato Leaf Roll Virus stress is compared to that of healthy plants. A phylogenetic study is also conducted which elucidates the evolution of the various *eIF4E* isoforms in plants.

Chapter 4 forms the second research chapter of this thesis. It describes the work done to establish the foundations for transgene-free CRISPR/Cas9 genome editing of the potato *eIF4E* genes. The process of designing single guide RNAs for the gene targets is described, as well as the screening of these guides *in vitro* for efficacy in DNA cleavage. The establishment of a potato protoplast isolation procedure is also outlined, as well as culture of these protoplasts to the micro-callus stage of development.

Chapter 5 concludes this thesis by providing a summary of the findings, as well as outlining the future research prospects that were identified through this work.

Chapter 2: Literature review

2.1 The importance of potatoes

Solanum tuberosum (potato) is the most consumed non-cereal crop globally (Zhang et al., 2017). The potato tuber is calorie-dense and vitamin-rich, and many populations across the world rely on this plant as a staple food source. South African potato production is valued at R8 billion annually, with 8-10% of the total jobs in the South African agricultural industry being related to potato production (Potato South Africa, www.potatoes.co.za). The varied climatic conditions across South Africa's nine provinces mean that potato crops are planted at different times of the year, ensuring constant tuber supply. In addition to national consumption, a large portion of the South African crop is exported internationally. South Africa is the largest exporter of potatoes in Africa with Mozambique, Angola, Zimbabwe, and Zambia receiving 95.3% of the exported tubers (Department of Agriculture, Forestry and Fisheries, 2013).

The global human population is expected to reach 9.8 billion people by 2050, with the majority of this growth being concentrated in Africa (United Nations, Department of Economic and Social Affairs, 2015). Countries will increasingly rely on the stable production of cheap and calorie-dense foods, such a potato, to prevent wide-spread hunger. Disease and abiotic stress threaten the stability of the potato industry and it is essential that in coming years new solutions to these problems are found.

2.2 Threats to potato production

Climate change is the greatest driver of abiotic stress globally. It is predicted to result in more extreme seasonal weather patterns, extended periods of drought, and flash flooding (International Panel on Climate Change, 2014). All these factors already negatively impact crop production and thus threaten global food security (McKersie, 2015; Ray et al., 2019). Potato plants respond poorly to changes in the climatic conditions in which they are grown which negatively effects tuber development, and results in yield losses (Dahal et al., 2019). To maintain stability of potato production, it is essential that future crop cultivars are strengthened against the effects of abiotic stress.

Pests and pathogens are also a growing threat to the stability of the potato industry. New species of pathogens are continuously emerging and previously successful resistance strategies are being overcome (Kapsa, 2008). Potato is threatened by several pathogen species that cause decreases in tuber yield. Fungal species in the *Phytophthora* and *Alternaria* genera cause blight diseases which can destroy a full year's harvest if infection spreads to underground tissues. Bacterial pathogens also cause devastating disease in potatoes with *Pectobacterium* and *Dickeya* species currently being the most prevalent threat (Kapsa, 2008). Furthermore, viral infection is also common within potato farms. The viral family Potyviridae is the largest family of plant-infecting viruses and contains multiple viruses that infect potato plants (Gibbs and Ohshima, 2010). The management of viral pathogens is particularly challenging as viruses mutate quickly and viral pathology can change significantly over a short time span (Visser and Bellstedt, 2015). The use of

biotechnological tools to create fully resistant crop cultivars is a promising management strategy that could result in lasting protection against potyviruses.

2.3 The potato genome

Commercial potato varieties are clones of *Solanum tuberosum ssp. tuberosum*. Duplication of a heterozygous diploid through breeding with 2n gametes resulted in this species being a highly heterozygous autotetraploid (Iwanaga and Peloquin, 1982). This complex genome means that commercial potatoes are difficult to crossbreed, and thus potato plants are usually multiplied asexually through clonal propagation. This is advantageous as it results in vigorous early development of the plant which ensures larger plants at maturity and higher yields, as well as ensuring consistent output from each generation. Relying solely on asexual reproduction, however, limits genetic variation within potato populations (Fasoula, 2002). This lack of allelic variety often means that potato populations respond poorly when either cultivation conditions change, or when the population is placed under a selection pressure. Attempts to introduce favourable genes into potato species through conventional breeding methodologies are inefficient due to the tetraploid genome, and thus novel strategies need to be implemented.

To utilise biotechnological advances to strengthen potatoes against future challenges, a complete and annotated genome sequence is required. A full genome sequence for a crop species allows breeding to be genotype-driven which increases the chances of producing informed, beneficial genetic variation based on sequence data (Visser et al., 2014). Genome sequences can also be mined to identify agronomically important genes which can be manipulated to improve agriculturally valuable traits. The autotetraploid nature of potatoes means that it is important for the consensus sequence of the potato genome to be derived from multiple potato genomes to truly represent the species' genetic variability and for the genome sequence to be a useful tool for researchers (Kyriakidou et al., 2020).

By 2011, two complete potato genome sequences had been published: a homozygous doubled monoploid and a heterozygous diploid (The Potato Genome Consortium, 2011). While this aided potato research at the time, the release of tetraploid genome sequences in subsequent years has been most useful as this is more similar to commercialised cultivars (Uitdewilligen et al., 2015; Kyriakidou et al., 2020). Research into expanding the polyploid potato sequence data is active and recent developments in DNA sequencing technologies have increased the speed at which these are becoming available to researchers (Kyriakidou et al., 2019). Complete tetraploid genomes of the commercial potato cultivars Otava (Sun et al., 2022) and Solyntus (van Lieshout et al., 2020) have recently been published and add to the resources available to potato researchers. A remaining shortfall in this collection of genome data, however, is the lack of a complete genome sequence for the cultivar Désirée which is widely used in potato research. In 2020 a comprehensive single nucleotide polymorphism (SNP) map for the Désirée cultivar was published on the National Centre for Biotechnology Information (NCBI) database (accession number: PRJNA507597; Sevestre et al., 2020). This

SNP map will be a valuable tool for researchers working with Désirée until the time of the publication of the full Désirée genome sequence, but until then the lack of complete genome sequence will hamper efforts in genetically modifying this cultivar.

2.4 Plant RNA viruses

A major threat to the stability of the potato industry is infection with pathogens that decrease yield and produce symptoms that make tubers unfit for market (Dahal et al., 2019). The RNA viruses Potato Virus Y (PVY) and Potato Leaf Roll Virus (PLRV) pose the greatest threat to potato crops due to their global prevalence and the severity of their symptoms (Loebenstein and Gaba, 2012).

2.4.1 Potato Virus Y

Potato Virus Y is a member of the Potyviridae family which is the largest group of plant-infecting pathogens. Symptoms of PVY include chlorotic patches on leaves, decreased tuber size and number, and most significantly necrosis of tubers (Beczner et al., 1984). The virus is spread both vertically (from parent to progeny, or through propagation) and horizontally (through an insect vector; Lacomme and Jacquot, 2017). Sixty-five aphid species have been demonstrated to act as vectors for PVY and have been linked to infection outbreaks globally (Sridhar et al., 2022). The virus exists in a non-persistent manner in these vectors, meaning it survives only on the aphid's mouthpiece and does not complete any part of its lifecycle while on this host (Lacomme and Jacquot, 2017). This makes pesticides ineffective in preventing transmission as the virus is acquired and lost by the vector over a very short period. The emergence of recombinant PVY strains in the last decade has increased the prevalence of this disease globally (Visser et al., 2012). In South Africa, PVY is most often found in seed potato farms (Kruger and van der Waals, 2020) which produce tubers annually to initiate the new year's crop. Infected seed potatoes will transmit PVY to all progeny and outbreaks of the disease can ruin an entire year's harvest. PVY has a positive sense, single stranded RNA genome of approximately 9.7 kilobases (kb). Within this genome, 10 open reading frames encode 11 proteins, including a viral genome-linked protein (VPg) which caps the 5' end of the RNA genome (Kirchner et al., 2014).

2.4.2 Potato Leaf Roll Virus

Potato Leaf Roll Virus is a member of the Luteoviridae family. Infection by this virus causes rolling of leaves, stunted root growth, decreased tuber size and necrosis of the tubers (Harrison, 1984; Radcliffe and Ragsdale, 2002). Similar to PVY, PLRV is spread both horizontally and vertically. However, unlike PVY, PLRV exists in its aphid vector in a persistent manner, meaning that the virus must be ingested by the aphid and then move into the salivary glands before transmission to a plant (Katis et al., 2007). Thirteen aphid species have been identified globally that can transmit PLRV (Sridhar et al., 2022). The virus has a single stranded, positive-sense RNA genome which is approximately 5.8 kb. Eight open reading frames encode 7 proteins which also includes a VPg which, as with PVY, caps the 5'end of the RNA genome (Taliansky et al., 2003).

2.5 Plant anti-viral mechanisms

Plant viral-resistance genes are classified as either dominant or recessive according to their mechanism of inheritance. Dominant resistance genes encode a host receptor that recognise and bind a viral ligand. This interaction initiates a cascade of signalling events through the infected cell which leads to inhibition of virus establishment (Martin et al., 2003). Several attempts have been made to breed dominant resistance genes into commercial potatoes through traditional outcrosses (Davidson, 1980; Bradshaw and MacKay, 1994), however, the high mutation rate of viruses mean this resistance can be quickly overcome (Funke et al., 2017), limiting their usefulness.

Viral genomes encode very few proteins and must rely on the host's cellular machinery to replicate. Recessive viral-resistance genes encode these host factors and resistant alleles contain mutations that inhibit the virus' parasitism (Fraser, 1986; Robaglia and Caranta, 2006). A major cellular function that viruses cannot perform independently is transcription and translation of their genomes, and the subsequent protein assembly. These processes are essential for virus replication and survival and the host genes involved are, therefore, key potential susceptibility factors that may be manipulated to confer viral resistance (Robaglia and Caranta, 2006).

2.6 Eukaryote translation initiation factor-dependent viral infection

2.6.1 The eukaryotic translation initiation factor complex

Translation of eukaryotic messenger RNA (mRNA) is facilitated by the eukaryotic translation initiation factor (eIF) complex. This complex consists of multiple subunits that bind to each end of the mRNA strand and to a ribosome, positioning the mRNA for translation and subsequent protein assembly. The key mRNA-binding subunits within this complex are poly-A binding protein (PABP), which binds to the poly-A tail on the mRNA's 3' end, and the eukaryotic translation initiation factor 4E (eIF4E), which binds to the 5' mRNA cap (Gingras et al., 1999). Binding of an mRNA strand by eIF4E initiates assembly of the rest of the eIF complex.

Multiple *eIF4E* isoforms exist in several plant species. The nomenclature of these isoforms varies between species. The focus of this research is on potato *eIF4E*s and for this reason I will following the naming conventions of this species from here on. In potato three isoforms have been identified: *eIF4E-1*, *eIF4E-2* and *eIF(iso)4E*. The potato proteins eIF4E-1, eIF4E-2 and eIF(iso)4E perform a crucial role for native protein translation as at least one version of the eIF4E protein is needed for translation to occur. While eIF4E-1 and eIF4E-2 bind to eIF4G during eIF complex assembly (Figure 2.1 A), eIF(iso)4E binds eIF(iso)4G (Figure 2.1 B). Three possible versions of the eIF complex, therefore, can form to facilitate translation: one with eIF4E-1 bound to eIF4G, another with eIF4E-2 bound to eIF4G, and lastly, one with eIF(iso)4E bound to eIF(iso)4G. All three iterations of this complex function identically.

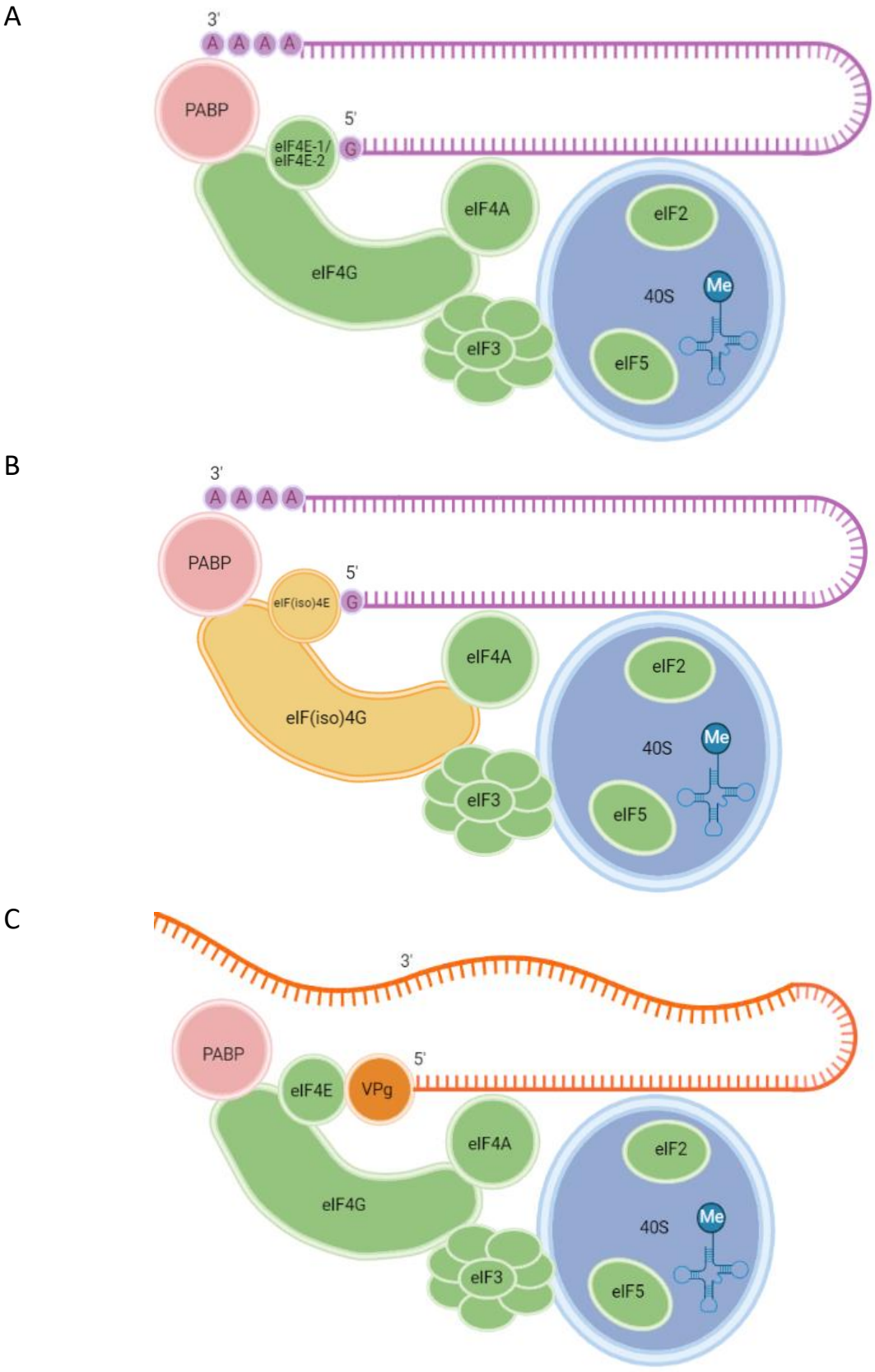


Figure 2.1 The eukaryotic translation initiation factor (eIF) complex binds to messenger RNA and ribosomes for protein synthesis. **(A)** The eIF subunits eIF4E-1 or eIF4E-2 associate with eIF4G and the 5' cap of the messenger RNA to initiate complex assembly. **(B)** Alternatively, another isoform, eIF(iso)4E, can bind eIF(iso)4G within the eIF complex. **(C)** During

viral infection, the viral genome-linked protein (VPg) will outcompete the host messenger RNA for binding to one of the three possible eIF4E isoforms for viral genome translation.

2.6.2 The eIF4E-VPg interaction

The *eIF4E* gene family plays an essential role in mRNA translation and are, therefore, key susceptibility genes for several plant RNA viruses that rely on this machinery for replication. Susceptibility genes encode host proteins that enable viral infection and thus disruption of these genes could result in viral resistance. For the potyvirus family, a direct association between the VPg and several *Arabidopsis* eIF4E isoforms has been confirmed through *in vitro* yeast-2-hybrid binding assays (Wittmann et al., 1997; Leonard et al., 2000). Similar to the mechanism of native mRNA transcription, binding of PVY VPg by eIF4E initiates assembly of the rest of the eIF complex and translation of the viral genome begins (Figure 2.1 C).

A span of 35 amino acid residues out of a total of 231 amino acids form the eIF4E-1 cap-binding pocket. This cap-binding pocket has been demonstrated to preferentially bind a potyviral VPg over host mRNA (Leonard et al., 2000). For the PVY VPg, this amino acid range has been narrowed down further to four amino acid residues: a tryptophan at position 56, an arginine at position 157, and two lysines at positions 159 and 162 in the potato eIF4E-1 (Coutinho de Olivera et al., 2019). The same VPg-eIF4E interaction has been confirmed to occur during infection with other positive single-stranded RNA viruses in the Caliciviridae (Goodfellow et al., 2005; Hosmillo et al., 2014) and Tombusviridae (Truniger et al., 2008) families. The conservation of this mechanism across these viral families may indicate that this mechanism is conserved between all positive sense RNA viruses that have a 5' VPg, however this needs to be confirmed empirically.

2.6.3 Naturally occurring *eIF4E* mutations confer viral resistance

The eIF4E-PVY protein interaction is essential for the survival of most RNA viruses in plants. Mutations affecting the eIF4E protein that prevent PVY recognition and binding would prevent viral replication and thus confer resistance (Yeam et al., 2007; Wang and Krishnaswamy, 2012). Mutant *eIF4E* alleles in several plant species lead to natural resistance to some RNA viruses. In the close potato relatives *Capsicum annuum* (pepper) and *Solanum lycopersicum* (tomato), the *pvr2* and *pot-1* mutant eIF4E proteins with amino acids changes near the VPg binding location have been associated with potyvirus resistance (Ruffel et al., 2005; Charron et al., 2008). Similarly, in *Cucumis melo* (melon) a mutant *eIF4E* allele confers resistance to Carmovirus, an RNA virus in the Tombusviridae family (Truniger et al., 2008). While most native alleles that provide RNA virus resistance have been linked to *eIF4E,* or its isoform *eIF(iso)4E (*Shopan et al., 2020), mutant *eIF4G* and *eIF(iso)4G* alleles have also been associated with potyvirus resistance in *Oryza sativa* (rice; Albar et al., 2006; Lee et al., 2010). This protein also associates with the viral genome during replication (Figure 2.1) and so a mutation affecting it can also prevent eIF complex formation and confer resistance.

While most naturally occurring *eIF4E* mutants have amino acid changes at the cap-binding site, an *eIF4E* splicing mutant was identified through TILLING of a highly mutagenized tomato population. Inoculation

studies of the mutant line demonstrated that these plants had improved resistance to PVY (Potyvididae) and Pepper Mottle Virus (Tombusviridae) but not Tobacco Etch Virus (Potyvididae; Piron et al., 2010). This indicates that different viruses may utilise the eIF4E protein in different ways and, importantly, an *eIF4E*-mutant cannot be assumed to confer resistance to all viruses within the same family, unless this has been specifically confirmed through inoculation studies.

2.7 Engineering RNA virus resistance through *eIF4E* manipulation

2.7.1 Transgenic approaches to create *eIF4E*-mediated viral resistance

The simultaneous knock-down of *eIF4E-1* and *eIF4E-2* expression in potato through RNAi-induced gene silencing resulted in improved resistance to PVY (Miroshichenko et al., 2020). Interestingly, knockdown of *eIF4E-1* alone did not have the same impact, implying that, in the absence of *eIF4E-1*, the PVY VPg interaction is entirely performed by *eIF4E-2* or *eIF(iso)4E* with no disadvantage to the virus. The impact of the knock-down of *eIF4E-2* alone was not investigated by this group, and this would be an interesting addition to this research in the future. Similar experiments in tomato led to plants displaying a dwarf phenotype that negatively impacted fruit size (Mazier et al., 2011). This was attributed to the decreased efficiency of native translation in the tomato plants. No negative phenotypic changes were reported in the *eIF4E-1* and *eIF4E-2* knock-down potato plants (Miroshichenko et al., 2020) indicating some species differences in this trait. Functional redundancy between the *eIF4E* isoforms and *eIF(iso)4E* mean that all three isoforms can bind native mRNA and initiate translation. In the absence of *eIF4E-1* or *eIF4E-2*, it is likely that *eIF(iso)4E* was able to maintain native translation levels in the potato, but not the tomato plants. No studies have examined interactions between *eIF(iso)4E* and a VPg. All *eIF4E-1 or eIF4E-2* mutants engineered to date have only shown incomplete viral resistance (Miroshichenko et al., 2020; Lucioli et al., 2022; Noureen et al., 2022). It is possible that complete resistance can only be achieved if all three isoforms are knocked out. Unfortunately, a triple knockout is likely to be fatal as the plant would also be unable to translate RNA.

In 2018, Bastet et al. designed and synthesised an *Arabidopsis thaliana eIF4E* gene that contained amino acid mutations which had been shown in other species to maintain native protein functionality, while introducing potyvirus resistance. This synthetic allele was introduced to *Arabidopsis eIF4E-1* and *eIF(iso)4E* single and double mutants. Transformed plants were screened against several potyviruses (Clover Yellow Vein Virus, Turnip Mosaic Virus, Lettuce Mosaic Virus, Plum Pox Virus and Watermelon Mosaic Virus), and decreased viral accumulation was observed in all cases. This wide resistance spectrum is mostly likely due to the combination of the two gene knockouts. Neither the single nor double mutant plants exhibited the dwarfed phenotype oftentimes observed in other plants with *eIF4E* knockouts. The absence of this phenotype is most probably due to the expression of the synthetic resistant allele which maintains native translation levels. This type of allelic resistance is a promising strategy for viral resistance but relies heavily on a transgenic approach which would have regulatory implications for development in crops.

2.7.2 CRISPR/Cas9 mutation of *eIF4E* for viral resistance

In the last year several research groups have published studies that tried to engineer PVY resistance using CRISPR/Cas9-mediated editing of *eIF4E* in potato. Knockout of *eIF4E-1* in potato was achieved by Noureen et al., and Lucioli et al., through DNA-based CRISPR/Cas9 genome editing (Lucioli et al., 2022; Noureen et al., 2022). The *eIF4E-1* mutant plants of both research groups were exposed to PVY and improved viral tolerance was observed in both cases. Unfortunately, complete resistance was not achieved, which can be most likely to be attributed to the virus's ability to utilise *eIF4E-2* or *eIF(iso)4E*.

Parallel to this research in potato, an RNA-based CRISPR/Cas9 knockout of both *eIF4E-1* and *eIF4E-2* in tobacco was achieved (Le et al., 2022). These plants demonstrated significant improvement in PVY tolerance compared to wild type plants, or *eIF4E-1* single knockouts, supporting the hypothesis that a multigenic knockout is needed to confer full resistance to PVY in potato. Whether these single mutants are also resistant to other RNA viruses, such as PLRV, was not examined.

2.8 The role of potato RNA cap-binding proteins in native and viral translation

2.8.1 A novel *eIF4E* isoform?

A new RNA cap-binding protein (nCBP) has recently been identified in several plant species, and it has been suggested that this protein may serve the same function as eIF4E due to their shared ability to bind to a 5' mRNA cap (Keima et al., 2017; Chen et al., 2022). Whether or not nCBP can perform the native role of the eIF4E proteins to bind both mRNA, and eIF4G or eIF(iso)4G in a functional eIF complex has, however, not yet been investigated.

The mRNA-cap binding functionality of nCBP also raises questions regarding its ability to bind viral mRNA, whether the mRNA is coupled with a VPg or not. In tobacco and potato, it has been demonstrated that knock-down of *nCBP* transcript levels decreases the accumulation of Potato Virus X, Potato Virus S and Potato Virus M (Chen et al., 2022). Yeast-2-hybrid assays demonstrated direct protein interactions between nCBP and coat proteins from Potato Virus X and Potato Virus S. Following this research, it was proposed that nCBP is involved in the cell-to-cell movement of viruses in plants (Keima et al., 2017). Beyond this protein research however, little is known about the mechanism(s) underlying nCBP function during infection. Whether nCBP can recognise and bind a viral VPg in a similar manner to eIF4E isoforms is also unknown. This will need to be addressed experimentally if this novel gene is to be considered a target in viral resistance.

2.8.2 A summary of the function of potato RNA cap-binding proteins

To effectively engineer viral resistant potato plants through modification of the *eIF4E* isoforms, it is essential that their roles in native protein translation and viral infection are clearly understood. The addition of nCBP in this discussion further complicates this area. A summation of the function of the eIF4E-1, eIF4E-2, eIF(iso)4E and nCBP proteins in potato are summarised in Table 2.1.

Table 2.1 The functions of the *Solanum tuberosum* (potato) RNA cap-binding proteins in native protein translation and during viral infection.

mRNA cap-binding protein	Function in native translation	Role during viral translation
eIF4E-1	Binds to the 5' mRNA cap, as well as eIF4G in the eIF complex (Gingras et al., 1999)	Binds to the 5' VPg, as well as eIF4G in the eIF complex (Coutinho de Olivera et al., 2019)
eIF4E-2		
eIF(iso)4E	Binds to the 5' mRNA cap, as well as eIF(iso)4G in the eIF complex (Gingras et al., 1999)	Hypothesised that it will bind a VPg in a manner similar to eIF4E-1 or eIF4E-2.
nCBP	Unknown	Facilitates the cell-to-cell movement of virus particles (Keima et al., 2017)

2.9 Biotechnological methods for improving potatoes

2.9.1 Traditional breeding techniques

Traditional plant breeding takes advantage of random, advantageous mutations that occur naturally in a crop, or its close wild relatives. These advantageous alleles are introduced into commercial cultivars through sexual crosses. The complexity of the potato's tetraploid genome makes crossing of this plant a slow and laborious task. Despite this, most changes to commercial potato cultivars have been introduced in this manner (Bradshaw, 2022). While much research has used trans- and cis-genic methods to introduce improved traits into potatoes (Nadakuduti et al., 2018), only one cis-genic potato has been successfully commercialised (Richael, 2020).

2.9.2 New breeding techniques

In recent years, a variety of biotechnologies have been developed that allow the targeted introduction of mutations into a plant's genome (Schaart et al., 2015). These mutations mimic the random mutations that occur naturally in plants and would previously have been identified as potential targets in traditional plant breeding. Because of this similarity, these new technologies have been labelled 'New breeding techniques' (NBTs) and much focus in recent years has been on the use of these techniques for the improvement of crop species. Crop species that do not perform well under traditional breeding strategies - such as potatoes - stand to benefit most from the use of NBTs. While traditional breeding relies on the slow, and targeted, accumulation of mutations, NBTs can engineer advantageous mutations in precise locations over a much shorter time span (Dhugga, 2022).

Four NBTs have recently emerged as the most promising techniques for use in plant breeding. These are meganucleases, zinc finger nucleases (ZFNs), transcription-activator like effector nucleases (TALENs) and Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated (CRISPR/Cas). Each of these

systems use nuclease enzymes to introduce double-stranded DNA breaks (DSB) that are then imperfectly repaired, creating mutations in the target gene region.

Meganucleases are naturally occurring endo-deoxyribonucleases found in a range of eukaryotic organisms (Daboussi, 2006). They are characterised by their large target recognition sites which are typically between 20 and 30 base pairs (bp) long but can span up to 40 bp. The catalytic domain of meganucleases overlaps with the DNA binding site which spans in both directions from the cleavage site. The main advantage of using meganucleases are that they are small proteins which makes their delivery into target cells easier than some alternative methods. They are also highly specific which reduces chance of off-target breaks. Meganucleases are, unfortunately, challenging to engineer *in vitro*. Their catalytic region and binding domain cannot be separated, meaning that they have less flexibility regarding their target range than alternative NTBs.

Zinc finger nucleases are engineered chimeric enzymes consisting of a nuclease, based on the *FokI* restriction enzyme, and a DNA-binding region derived from the Zinc-Finger (ZF) protein. Several ZF recognition sites (each recognising a 3 bp gene region) are combined and associated with a single nuclease. The more ZF sites present, the more specific the binding, however, most commonly 3 ZF sites are combined to create a 9 bp recognition site. The specificity of this technique comes from the fact that the *FokI* restriction enzyme is only active when dimerized, and thus, two ZFNs must combine for a DSB to occur. When using three ZF sites per nuclease the final DNA recognition sequence is, therefore, 18 base pairs (Urnov et al., 2010). Zinc finger nucleases are a flexible, modular systems with several ZF recognition sites available to complex to the *FokI* nuclease. Unfortunately, while the recognition sites of ZFNs are programmable, the specificity of the endonuclease component is low which can result in a high degree of off-target DSBs. Another major disadvantage of ZFNs is the need for DNA triplets already existing on zinc fingers which can limit the span of DNA targets.

TALENS are another type of engineered chimeric endonucleases. The TALEN complex comprises a restriction enzyme coupled with a DNA binding domain from TAL effectors: proteins secreted by *Xanthomonas* bacteria (Cade et al., 2012). The TAL DNA binding region is a 33 to 35 bp repeating region containing a repeat-variable di-residue at residues 12 and 13. This repeat-variable di-residue is variable for each repeat and determines the specificity of the binding. This residue can be further engineered to target other regions, broadening the possible DNA targets. Like ZFNs, TALENs are modular systems giving them more flexibility than meganucleases. However, a major disadvantage of their use is that their effectivity is greatly decreased by DNA methylation which interferes with the endonuclease activity. This means that they cannot be used effectively in methylated segments of the genomes.

Each of the above NBTs have been used successfully in Solanaceous species (Daboussi et al., 2006; Hilioti et al., 2016). However, the high labour and time costs needed for protein engineering component before use

makes these NBTs less appealing when compared to newer genome-editing technologies such as CRISPR/Cas which does not require any protein manipulation.

2.10 CRISPR/Cas genome editing

2.10.1 CRISPR/Cas: a prokaryotic adaptive immune response

Prokaryotic organisms have an adaptive immune system called CRISPR/Cas (Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated) that provides a defence against invading plasmids or viruses (Terns and Terns, 2011). CRISPR/Cas immunity is an RNA based system that occurs in three distinct phases: adaptation, expression, and interference (Figure 2.2). During the adaptation phase, the prokaryote integrates a short fragment of the nucleic acid from the invading plasmid of virus, known as the protospacer, into the CRISPR region of its genome. Integration of this DNA fragment ensures long-term immunity for the prokaryote against the same pathogen. During the expression phase, this CRISPR region is transcribed to form pre-CRISPR RNA (pre-crRNA) which is spliced into mature crRNAs corresponding to individual protospacers. The *Cas* endonuclease gene, which is located on the same operon, is transcribed alongside the crRNAs, and assembled into its tertiary structure. The mature Cas protein associates with a single crRNA to form a CRISPR/Cas complex. During the interference phase the crRNA in the CRISPR/Cas complex guides the endonuclease towards the corresponding region of the invading nucleic acid. Recognition between the CRISPR/Cas complex and the invading nucleic acid occurs in two forms. Firstly, through base-pairing with the crRNA and secondly, through recognition by Cas of a short region near the protospacer sequence known as the Protospacer-Adjacent-Motif (PAM). The Cas protein then creates a double stranded break in the foreign DNA/RNA three base pairs upstream from the PAM site. This cleaved nucleic acid is then degraded by the host cell.

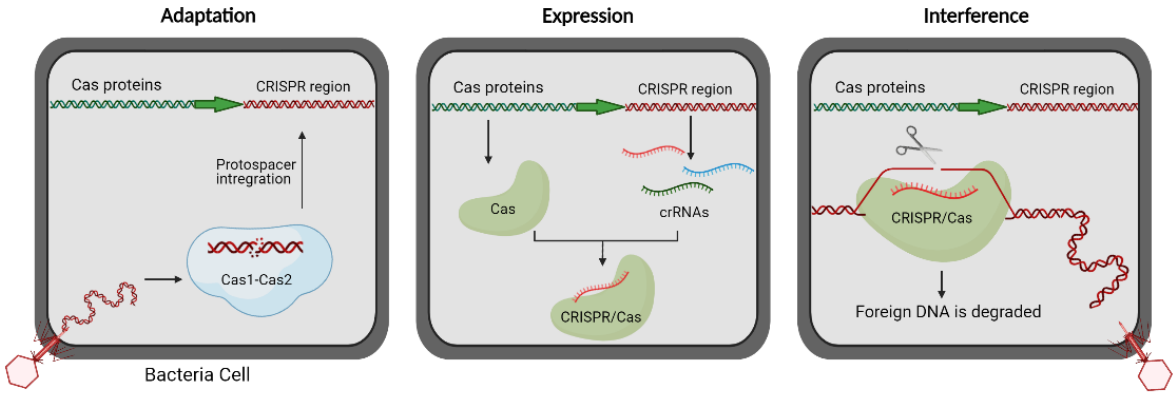


Figure 2.2 Functioning of CRISPR/Cas in prokaryotes in three stages. **Adaptation**: The prokaryote integrates a short fragment of invading genetic material, known as the protospacer, into the CRISPR region of its genome. **Expression**: The protospacers are transcribed and processed into crRNAs (CRISPR-RNAs) which associate with the Cas endonuclease to form the CRISPR/Cas complex**. Interference**: The CRISPR/Cas complex is guided to the invading genome by the

corresponding crRNA. The Cas endonuclease cleaves the foreign nucleic acid, which is then degraded, providing the prokaryote with immunity (Adapted from Knott and Doudna, 2018). Figure created with BioRender.com.

2.10.2 CRISPR/Cas type classification

There are 2 distinct classes of CRISPR/Cas systems found in prokaryotes - Type I and Type II. Type II systems are utilised in biotechnology and are further sub-dived into classes according to the different structures and functions of the Cas protein (Makarova and Koonin, 2015). Type II class I CRISPR/Cas systems utilise Cas3 which has two active domains. In this class, the entire CRISPR/Cas complex is encoded by a single operon. Once assembled, the CRISPR/Cas3 complex must be coupled to an effector complex to function. This system processes pre-crRNA into mature crRNAs (Richter et al., 2012). The type II class II CRISPR/Cas system utilises the Cas9 endonuclease. Cas9 is a bidomain protein with each domain having the ability to act independently (Chylinski et al., 2013). Unlike class I, class II CRISPR/Cas9 systems do not rely on an effector complex for functionality. Another unique property of class II CRISPR/Cas9 systems is the need for two RNA structures to be complexed with the endonuclease for activity (Fonfara et al., 2014). This additional requirement gives these systems a high degree of specificity. Like class I systems, class II CRISPR/Cas systems process pre-crRNA into mature crRNAs. Type II Class III CRISPR/Cas systems utilise Cas10 which contains 4 active domains (Staals et al., 2013). The class III CRISPR/Cas10 system does not process pre-crRNA but instead relies on the products of the class I and class II systems. This limits the effectiveness of this system as it cannot function independently in a cell. The class IV CRISPR/Cas system is the rarest found in nature. It utilises the csf1 protein (also known as Cas12a) which couples with an effector domain for endonuclease functioning. This system does not process its own pre-crRNA, nor does it integrate a CRISPR cassette into host genome.

2.10.3 The development of the CRISPR/Cas system into a biotechnological tool

In the last decade, the type II class II CRISPR/Cas system was manipulated to act as a biotechnological tool that can engineer precise, site-specific genome edits (Jinek et al., 2012). Using this system double stranded DNA breaks are introduced into the target genome which are then repaired by the host (Figure 2.3). These breaks are most frequently repaired through non-homologous end joining, an imperfect molecular mechanism which introduces mutations. The alternative repair mechanism, homology-directed repair, incorporates exogenous DNA between the break ends at this point. This repair mechanism allowing more complex insertions to be created and can be used biotechnologically for targeted indels to be introduced into a eukaryote genome in a diverse and efficient manner.

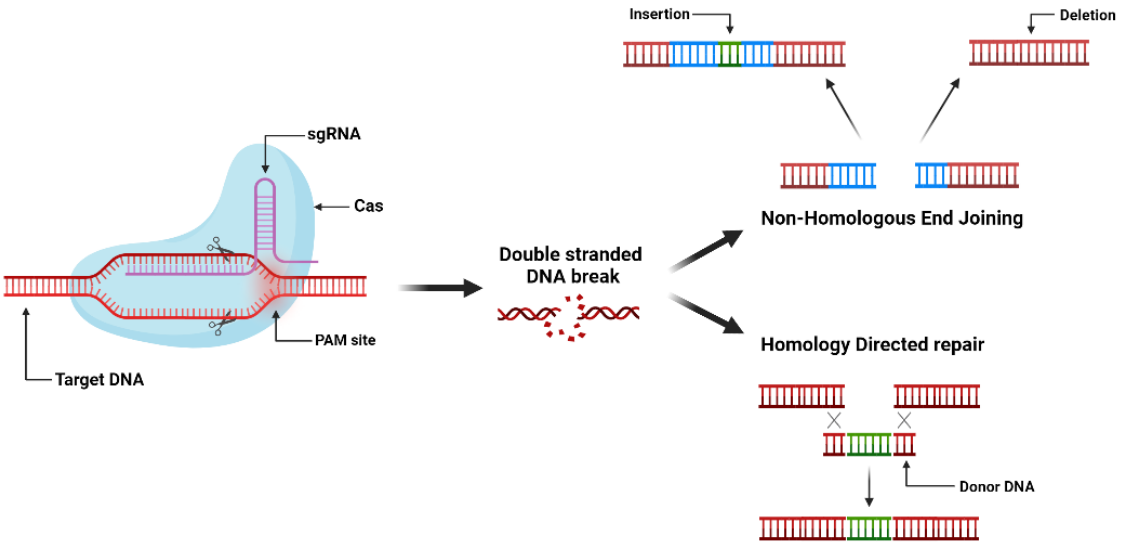


Figure 2.3 Double stranded DNA breaks created by the CRISPR/Cas9 complex can be repaired via two mechanisms: non-homologous end joining or homology-directed repair. Non-homologous end joining can result in imperfect repair and the introduction of random indels. Homology directed repair creates larger insertions defined by a segment of donor DNA that is used as a template for repair. Figure created using BioRender.com.

Originally, the type II class II CRISPR/Cas system was chosen over other types due to the specificity of Cas9 (Knott and Doudna, 2018). Unlike other endonucleases, Cas9 only associates with RNA in the form of a crRNA-tracrRNA (trans-acting crRNA) complex. This dual-RNA requirement decreases the risk of non-specific RNA association with Cas9. The HNH and Ruv-C active domains of Cas9 only create dsDNA breaks when bound to DNA adjacent to an 'NGG' PAM site. For use in gene editing, the system was simplified by hybridising the crRNA-tracrRNA complex into a single guide RNA (sgRNA) through introduction of a linker loop between the two molecules. This was beneficial as only a single RNA molecule needed to be engineered and introduced to the target organism (Jinek et al., 2012).

Recently however, other Cas endonucleases have been utilised for gene editing in both mammals and plants (Zhong et al., 2018). This change is motivated by the limitations of Cas9, which include the restricted target range due to the need for an adjacent PAM site, as well as the large size of the protein which can make cloning challenging. The other Cas proteins also offer functionalities not available with Cas9 (Wada et al., 2022). For example, Cas3 can recognise and excise large regions of a genome with a single guide (Csörgő et al., 2020).

2.10.4 Advancements beyond CRISPR/Cas gene-editing

The ability for the CRISPR/Cas system to introduce precise DSB to a target genome provided expanded opportunities for genome editing in plants. Development of a "dead" Cas9 (dCas9), with no endonuclease activity has further expanded the possibilities of this complex as a biotechnological tool. When coupled with

a sgRNA, dCas9 still migrates to a target genome location. By complexing other enzymes to the CRISPR/dCas9 complex, it is possible to direct other proteins of interest to a highly specific genome location (Figure 2.4). This technology has been used to control gene expression through delivery of activator or repressor domains (Cheng et al., 2013; Gilbert et al., 2013) to specific genes of interest. Fluorescent proteins have also been coupled with CRISPR/dCas9 and have facilitated site specific DNA tagging (Tanenbaum et al., 2014).

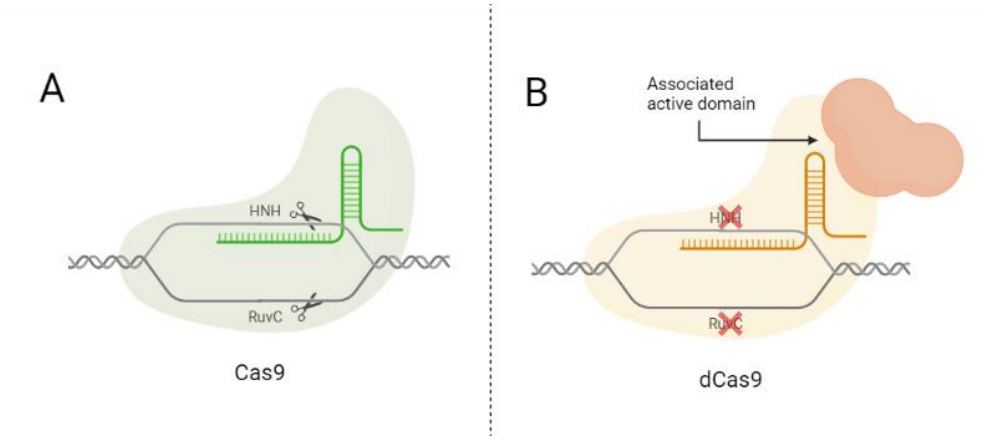


Figure 2.4 Modified versions of the Cas9 endonuclease created through selective inactivation of the catalytic domains. **(A)** The unmodified Cas9 endonuclease with both catalytic domains still active. **(B)** A catalytically dead Cas9 (dCas9) wherein both catalytic domains are inactivated, associated to an alternative active domain. Figure created using Biorender.com.

Recently, a further modification to CRISPR/dCas9 system has allowed site specific SNP mutations to be engineered by coupling a base-editing protein to the complex (Kim et al., 2017). Cytosine deaminase and adenosine deaminase are enzymes that remove an amino group from cytosine and adenine, converting these nucleotides into thiamine and guanosine respectively. This is advantageous as many agriculturally significant traits are influenced by SNPs, while large gene knockouts or knock-ins can have undesirable impacts on the plant phenotypes. For example, knock-down of *eIF4E* in tomato creates semi-dwarfed plants (Mazier et al., 2011). To prevent this phenotypic sacrifice, base editing of *eIF4E* could be utilised to engineer an eIF4E version that retains native function, without VPg binding capacity. This technique holds a lot of potential for the development of better crop varieties in the future (Azameti and Dauda, 2021).

2.10.5 CRISPR/Cas delivery methods

The CRISPR/Cas components can be introduced into a target organism as engineered DNA which is expressed using the host's machinery, as RNA which would need to be translated for Cas9 formation, or in the form of an already active ribonucleoprotein (RNP; Chen et al., 2019). When introduced as DNA, plasmid(s) containing the sgRNA sequence and coding domain sequence for Cas9 are engineered and then transformed into the genome of target cell (Liang et al., 2018; Chen et al., 2019). The expression cassettes must be transcribed and the Cas9 protein needs to be translated and processed into a mature protein using the host's cellular

machinery before genome editing occurs. When this technique is used for crop transformations, transformed progeny are backcrossed with the parental line when possible to remove the foreign DNA (Zhu et al., 2021).

For an RNA-based delivery system, the sgRNA and *Cas9* are transcribed *in vitro* and transformed into the receptor cell biolistically, or through PEG-mediated protoplast transformation (Yue et al., 2021). The Cas9 protein is then synthesised and complexed with the sgRNA, before DSB can be introduced to the cell's genome. Single cells are then regenerated into plants where some of the plants have become mutated in the targeted gene. One advantage of this system is that it is DNA free which is favoured by regulatory authorities and consumers. However, RNA is transient and unstable, and a lower mutation efficiency is observed when this system is used (Chen et al., 2019). It is also not possible to select for edited cells after using this technique so a great deal of screening through DNA sequences is needed which can be an expensive and time-consuming process.

The third option to introduce the CRISPR/Cas elements to a cell is through transformation with a RNP complex (Woo et al., 2015). In this system, a fully functional CRISPR/Cas9 RNP is assembled by *in vitro* transcribing the sgRNA and complexing it with a commercially acquired Cas9 protein before introducing the RNP into the cell. This system is the fastest acting and least taxing on the host cell as no nucleic acid or protein processing is needed. It also shows the lowest frequency of off-target editing (Chen et al., 2019). In plant cells the cell wall inhibits uptake of the CRISPR/Cas RNP so protoplasts - which are cells that have had their cell walls enzymatically or mechanically removed - need to be used. This system is also entirely DNA free system and thus offers the regulatory benefits of the RNA delivery technique.

2.11 RNP transformation of protoplasts for transgene-free genome editing

2.11.1 The benefits of genome editing using RNPs

The transformation of CRISPR/Cas9 RNPs into protoplasts allows for precise modifications to be made to crop species without the use of foreign DNA elements. This method is technically advantageous as little *in vitro* work is needed to construct the CRISPR/Cas9 machinery prior to transformation. Furthermore, no backcrosses are needed to eliminate foreign DNA. In potato, this is highly advantageous as sexual crosses between potatoes are difficult (Fasoula, 2002). Overall, this technique is significantly more efficient, less expensive, and better suited for potato genome editing than DNA- or RNA-based CRISPR/Cas transformations.

2.11.2 Potato protoplast isolation, transformation, and regeneration procedures

A protocol for potato protoplast isolation, transformation, and culture was established by Nicolia et al. (2015) with the express purpose of being used for genome editing techniques. In that study potato protoplasts were used for TALEN-mediated genome editing, however, the procedure for CRISPR/Cas9 is identical in most aspects and several groups have successfully used it for CRISPR/Cas genome editing in potatoes (Andersson et al., 2018; González et al., 2020).

Potato protoplasts are created in a two-step process that involves cell plasmolysis, followed by the enzymatic digestion of the cell wall. Following isolation and PEG-mediated transformation with RNPs, the protoplasts are regenerated in tissue culture. Cell wall regeneration occurs naturally when cells are cultured in the dark in a nutrient enriched medium. While liquid culture of potato protoplasts is possible (Tavazza and Ancora, 1986), encasement in an alginate gel allows refreshment of the regeneration medium without dilution of protoplast concentration (Eeckhaut et al., 2013). This technique is not only favoured in potato protoplast isolation but is common for other crop species such as cabbage (*Brassica oleracea*), canola (*Brassica napus*) and carrot (*Daucus carota*) (Kielkowska and Adamus 2012; Grzebelus and Skop, 2014; Sahab et al., 2018). After the regrowth of the cell walls, protoplasts can be stimulated hormonally to dedifferentiate and multiply to form callus. Shoots and roots can be induced from the callus tissue resulting in whole plant regeneration from a single cell (Cocking, 1972).

2.11.3 Drawbacks of using protoplasts

Protoplast culture is a highly specialised tissue culture technique and, while species-specific protoplast culture protocols have improved the likelihood of success of this procedure, it is often a barrier to progress (Eeckhaut et al., 2013). Contamination of protoplast cultures with microbes or fungi is a common issue that inhibits cell culture and unfortunately, antimicrobial, and antifungal agents are often detrimental to plant development (Herman, 2017).

Somaclonal variation and chimera formation create unexpected phenotypes in plants that are regenerated from protoplasts and thus the plants arising from such a protocol are not uniform. Somaclonal variance is defined as unexpected aneuploidy, chromosomal rearrangement, epigenetic changes, and SNPs that arise during cell culture. A high degree of somaclonal variation has been observed in potato plants following protoplast culture (Fossi et al., 2019) which had negative impacts on agriculturally significant traits such as plant size and overall health. Complex genome sequencing and analysis is needed to identify the changes that arise from this, which is expensive and time-consuming. In plants that can be easily backcrossed to eliminate this variation the largest drawback associated with somaclonal variation is the additional time and effort. In potatoes where clonal propagation is the primary form of reproduction, however, somaclonal variation will persist in subsequent generations and is a more significant issue. At present, the mechanism causing somaclonal variation is unknown and thus conscious avoidance of this issue is unlikely (Fossi et al., 2019). The creation of chimeras is also a negative and unexpected outcome of protoplast culture. Fusion of protoplasts at the early stages of regeneration will result in a chimeric plant and this has been observed in several instances of potato protoplast culture (Fossi et al., 2019; Banfalvi et al., 2020). For genome editing in potato to be successful, the engineered genomic state is required uniformly in the plant and therefore the frequent development of chimeras decreases the efficiency of this procedure. When plants are regenerated from protoplasts a variety of individuals with differing amounts of somaclonal variation will grow. Some of these individuals will contain insignificant genome rearrangements or chimerism and can be selected for

further propagation. The ability to produce a high number of plants through protoplast regeneration is pivotal in the success of this technique as the greater the number of regenerants, the greater the chance of finding individuals with little somaclonal variation.

2.12 References

Albar L, Bangratz-Reyser M, Hébrard E, Ndjiondjop MN, Jones M, Ghesquière A (2006) Mutations in the eIF(iso)4G translation initiation factor confer high resistance of rice to *Rice* Yellow Mottle Virus. Plant J **47:** 47-426

Andersson M, Turesson H, Olsson N, Falt A, Ohlsson P, Gonzalez MN, Samuelsson M, Hofvander P (2018) Genome editing in potato via CRISPR-Cas9 ribonucleoprotein delivery. Physiol Plant **164:** 378–384

Azameti MK, Dauda WP (2021) Base editing in plants: Applications, challenges, and future prospects. Front Plant Sci **12:** 664997

Bánfalvi Z, Csákvári E, Villánya V, Kondrák M (2020) Generation of transgene-free PDS mutants in potato by Agrobacterium-mediated transformation. BMC biotechnol **20:** 25

Bastet A, Lederer B, Giovinazzo N, Arnoux X, German-Retana S, Reinbold C, Brault V, Garcia D, Djennane S, Gersch S, Lemaire O, Robaglia C, Gallois JL (2018) Trans-species synthetic gene design allows resistance pyramiding and broad-spectrum engineering of virus resistance in plants. Plant Biotechnol J **16:** 1569-1581

Beczner L, Horvath J, Romhanyi I, Forster H (1984) Studies on the etiology of tuber necrotic ringspot disease in potato. Potato Res **27:** 339-352

Bradshaw JE, Mackay GR (1994) Breeding strategies for clonally propagated potatoes. In: Potato Genetics

Bradshaw JE (2022) A brief history of the impact of potato genetics on the breeding of tetraploid potato cultivars for tuber propagation. Potato Res **65:** 461-501

Cade L, Reyon D, Hwang WY, Tsai SQ, Patel S, Khayter C, Joung JK, Sander JD, Peterson RT, Yeh JRJ (2012) Highly efficient generation of heritable zebrafish gene mutations using home- and heterodimeric TALENs. Nuc Acids Res **40:** 8001-8010

Charron C, Nicolaï M, Gallois JL, Robaglia CR, Moury B, Palloix A, Caranta C (2008) Natural variation and functional analyses provide evidence for co-evolution between plant eIF4E and potyviral VPg. Plant J **54:** 56-68

Chylinski K, Le Rhun A, Charpentier E (2013) The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. RNA Biol **10:** 726-737

Chen K, Wang Y, Zhang R, Zhang H, Gao C (2019) CRISPR/Cas genome editing and precision plant breeding in agriculture. Annu Rev Plant Biol **70:** 28.1-28.31

Chen R, Yang M, Tu Z, Xie F, Chen J, Luo T, Hu X, Nie B, He C (2022) Eukaryotic translation initiation factor 4E family member nCBP facilitates the accumulation of TGB-encoding viruses by recognizing the viral coat protein in potato and tobacco. Front Plant Sci **13:** 946873

Cheng AW, Wang H, Yang H, Shi L, Katz Y, Theunissen TW, Rangarajan S, Shivalia CS, Dadon DB, Jaenisch R (2013) Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. Cell Res **23:** 1163-1171

Cocking EC (1972) Plant cell protoplasts – isolation and development. Ann Rev Plant Physiol **23:** 29-50

Coutinho de Olivera L, Volpon L, Rahardjo A, Osborne M, Culjkovic-Kraljacic B, Trahan C, Oeffinger M, Kwok B, Borden K (2019) Structural studies of the eIF4E–VPg complex reveal a direct competition for capped RNA: Implications for translation. PNAS **116:** 24056-24065

Csörgő B, León LM, Chau-Ly II, Vasquez-Rifo A, Berry JD, Mahendra C, Crawford ED, Lewis JD, Bondy-Denomy J (2020) A compact cascade-Cas3 system for targeted genome engineering. Nat Methods **17:** 1183-1190

Daboussi F, Stoddard TJ, Zhang F (2006) Engineering meganucleases for precise plant genome modification. In: Advances in new technology for targeted modification of plant genomes

Dahal K, Xiu-Qing L, Tai H, Creelman A, Bizimungu B (2019) Improving potato stress tolerance and tuber yield under a climate change scenario – a current overview. Front Plant Sci **10:** 563

Davidson TMW (1980). Breeding for resistance to virus disease of the potato (*Solanum tuberosum*) at the Scottish Plant Breeding Station. In: Scottish Plant Breeding Station 59th Annual Report

Department of Agriculture, Forestry and Fisheries (2013) A profile of the South African potato market value chain. Directorate: Statistics and Economic Analysis

Dhugga KS (2022) Gene editing to accelerate crop breeding. Front Plant Sci **13:** 889995

Eeckhaut T, Lakshmanan PS, Deryckere D, Van Bockstaele E, Van Huylenbroeck J (2013) Progress in plant protoplast research. Planta **238:** 991-1003

Fasoula DA (2002) The effects of clonal propagation on the genetic improvement of potato. Acta Hort **579:** 71-75

Fonfara I, Le Rhun A, Chylinski K, Makarova KS, Lecrivain AL, Bzdrenga J, Koonin EV, Charpentier E (2014) Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. Nucleic Acids Res **42:** 2577-2590

Fossi M, Amundson K, Kuppu S, Britt A, Comai L (2019) Regeneration of *Solanum tuberosum* plants from protoplasts induces widespread genome instability. Plant Physiol **180:** 78-86

Fraser RSS, Van Loon LC (1986) Genes for resistance to plant viruses. Crit Rev Plant Sci **3:** 257-294

Funke CN, Nikolaeva OG, Green KJ, Tran LT, Chikh-Ali M, Quintero-Ferrer A, Olsen N, Pavek MJ, Crosslin JM, Karasev AV (2017) Strain-specific resistance to Potato Virus Y (PVY) in potato and its effect on the relative abundance of PVY strains in commercial potato fields. Plant Dis **101:** 20-28

Gibbs A, Ohshima K (2010) Potyviruses and the digital revolution. Annu Rev Phytopathol **48:** 205-223

Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, Stern-Ginossar N, Brandman O, Whitehead EH, Doudna JA, Lim WA, Weissman JS, Qi LS (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell **154:** 442-451

Gingras AC, Raught B, Sonenberg N (1999) eIF4 initiation factors: Effectors of mRNA recruitment to ribosomes and regulators of translation. Annu Rev Biochem **68:** 913-963

González MN, Massa GA, Andersson M, Oneto CAD, Turesson H, Storani L, Olsson N, Fält AS, Hofvander P, Feingold SE (2021) Comparative potato genome editing: *Agrobacterium tumefaciens*-mediated transformation and protoplasts transfection delivery of CRISPR/Cas9 components directed to StPPO2 gene. PCTOC **145:** 291-305

Goodfellow I, Chaundhry Y, Gioldasi I, Gerondopoulos A, Natoni A, Labrie L, Lalibetré JF, Roberts L (2005) Calicivirus translation initiation requires an interaction between VPg and eIF4E. EMBO Reports **6:** 968-972

Grzebelus E, Skop L (2014) Effect of β-lactam antibiotics on plant regeneration in carrot protoplast cultures. In Vitro Cell Dev Biol **50:** 568-575

Harrison BD (1984) Potato leafroll virus. In: Descriptions of Plant Viruses

Herman EB (2017) Plant tissue culture contamination: Challenges and opportunities. Acta Hortic **1155:** 33

Hilioti Z, Ganopoulos I, Ajith S, Bossis Y, Tsaftaris A (2016) A novel arrangement of zinc finger nuclease system for *in vivo* targeted genome engineering: The tomato LEC1-LIKE4 gene case. Plant Cell Rep **35:** 2241-2255

Hosmillo M, Chaudhry Y, Kim DS, Goodfellow I, Cho KO (2014) Sapovirus translation requires an interaction between VPg and the cap binding protein eIF4E. J Virol **88:** 12213-12221

International Panel on Climate Change (2014) Climate change 2014 synthesis report

Iwanaga M, Peloquin SJ (1982) Origin and evolution of cultivated tetraploid potatoes via 2n gametes. Theor Appl Genet **61:** 161-169

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual RNA-guided DNA endonuclease in adaptive bacterial immunity. Science **17**: 816-821

Kapsa J (2008) Important threats in potato production and integrated pathogen/pest management. Potato Res **51:** 385-401

Katis N, Tsitsipis JA, Stevens M, Powell G (2007) Transmission of plant viruses. In: Aphids as crop pests

Keima T, Hagiwara-Komoda Y, Hashimoto M, Neriya Y, Koinuma H, Iwabuchi N, Nishida S, Yasuyuki Y, Namba S (2017) Deficiency of the eIF4E isoform nCBP limits the cell-to-cell movement of a plant virus encoding triple-gene-block proteins in *Arabidopsis thaliana*. Sci Rep. **7:** 39678

Kielkowska A, Adamus A (2012) Peptide growth factor phytosulfokine-α stimulates cell divisions and enhances regeneration from *B. oleracea* var. *capitata* L. protoplast culture. J Plant Growth Regul **38:** 931-944

Kim YB, Komor AC, Levy JM, Packer MS, Zhao KT, Liu DR (2017) Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. Nature Biotechnol **35:** 371-377

Kirchner SM, Hiltunen LH, Santala J, Döring TF, Ketola J, Kankaala A, Virtanen E, Valkonen JPT (2014) Comparison of straw mulch, insecticides, mineral oil, and birch extract for control of transmission of Potato Virus Y in seed potato crops. Potato Res **57:** 59-75

Knott GJ, Doudna JA (2018) CRISPR-Cas guides the future of genetic engineering. Science **361**: 866-869

Kruger K, van der Waals, JE (2020) Potato Virus Y and Potato Leafroll Virus management under climate change in sub-Saharan Africa. S Afr J Sci **116:** 11-12

Kyriakidou M, Tai HH, Anglin NL, Ellis D, Stomvik MV (2019) Current strategies of polyploid plant genome sequence assembly. Front Plant Sci **9**: 1660

Kyriakidou M, Anglin NL, Ellis D, Tai HH, Stomvik MV (2020) Genome assembly of six polyploid potato genomes. Nature Sci Data **7**: 88

Lacomme C, Jacquot E (2017) General characteristics of Potato Virus Y (PVY) and its impact on potato production: An overview. In: Potato Virus Y: Biodiversity, pathogenicity, epidemiology and management

Le NT, Tran HT, Bui TP, Nguyen GT, Nguyen DV, Ta DT, Trinh DD, Molnar A, Pham NB, Chu HH, Do PT (2022) Simultaneously induced mutations in *eIF4E* genes by CRISPR/Cas9 enhance PVY resistance in tobacco. Sci **12:** 14627

Lee JH, Muhsin M, Atienza GA, Kwak DY, Kim SM, De Leon TB, Angeles ER, Coloquio E, Kondoh H, Satoh K, Cabunagan RC, Cabauatan PQ, Kikuchi S, Leung H, Choi IR (2010) Single nucleotide polymorphisms in a gene for translation initiation factor (eIF4G) of rice (*Oryza sativa*) associated with resistance to *Rice Tungro Spherical Virus*. MPMI **23:** 29-38

Leonard S, Plante D, Wittmann S, Daigneault N, Fortin MG, Laliberté JF (2000) Complex formation between potyvirus VPg and translation eukaryotic initiation factor 4E correlates with virus infectivity. J Virol **74:** 7730-7737

Liang Z, Chen K, Zhang Y, Liu J, Yin K, Qiu JL, Gao C (2018) Genome editing of bread wheat using biolistic delivery of CRISPR/Cas9 *in vitro* transcripts or ribonucleoproteins. Nat Prot **13:** 413-430

Loebenstein G, Gaba V (2012) Viruses of potato. Adv Virus Res **84:** 209-246

Lucioli A, Tavazza R, Baima S, Fatyol K, Burgyan J, Tavazza M (2022) CRISPR-Cas9 targeting of the eIF4E1 gene extends the Potato Virus Y resistance spectrum of the Solanum tuberosum L. cv. Desirée. Front Micobiol **13:** 873930

Makarova KS, Koonin EV (2016) Annotation and classification of CRISPR-Cas systems. Methods Mol Biol **1311:** 47-75

Martin GB, Bogdanove AJ, Sessa G (2003) Understanding the functions of plant disease resistance proteins. Annu Rev Plant Biol **54:** 23-61

Mazier M, Flamain F, Nicolaï, Sarnette V, Caranta C (2011) Knock-down of both eIF4E1 and eIF4E2 genes confers broad-spectrum resistance against potyviruses in tomato. PLoS One **6:** e29595

McKersie B (2015) Planning for food security in a changing climate. J Exp Bot **66:** 3435-3450

Miroshichenko D, Timerbaev V, Okuneva A, Klementyeva A, Sidorova T, Pushin A, Dolgov S (2020) Enhancement of resistance to PVY in intragenic marker-free potato plants by RNAi-mediated silencing of eIF4E translation initiation factors. PCTOC **140:** 691-705

Nadakuduti SS, Buell CR, Voytas DF, Starker CG, Douches DS (2018) Genome editing for crop improvement – Applications in clonally propagated polyploids with a focus on potato (*Solanum tuberosum L.*). Front Plant Sci **9**: 1607

Nicolia A, Proux-Wéra E, Åhman I, Onkokesung N, Andersson M, Andreasson E, Zhu L (2015) Targeted gene mutation in tetraploid potato through transient TALEN expression in protoplasts. J Biotechnol **204:** 17-24

Noureen A, Khan MZ, Amin I, Zainab T, Mansoor S (2022) CRISPR/Cas9-mediated targeting of susceptibility factor *eIF4E*-enhanced resistance against Potato Virus Y. Front Genet **13:** 922019

Piron F, Nicolaï M, Minoïa S, Piednoir E, Moretti A, Salgues A, Zamir D, Caranta C, Bendahmane A (2010) An induced mutation in tomato eIF4E leads to immunity to two potyviruses. PLos One **5:** e11313

Radcliffe EB, Ragsdale DW (2002) Aphid-transmitted potato viruses: The importance of understanding vector biology. Amer J of Potato Res **79:** 353-386

Ray DK, West PC, Clark M, Gerber JS, Prishchepov AV, Chatterjee S (2019) Climate change has likely already affected global food production. PLoS One **14:** e0217148

Richter C, Gristwood T, Clulow JS, Fineran PC (2012) *In vivo* protein interactions and complex formation in the *Pectobacterium* atro-septicum subtype I-F CRISPR/Cas System. PLoS One **7:** e49549

Richael CM (2020) Development of the genetically modified Innate® potato. Plant Breed Rev **44:** 3

Robaglia C, Caranta C (2006) Translation initiation factors: A weak link in plant RNA virus infection. Trends Plant Sci **11:** 1360-1385

Ruffel S, Gallois JL, Lesage ML, Caranta C (2005) The recessive potyvirus resistance gene *pot-1* is the tomato orthologue of the pepper *pvr2-eIF4E* gene. Mol Genet Genom **274**: 346-353

Sahab S, Hayden MJ, Mason J, Spangenberg G (2018) Mesophyll protoplasts and PEG-mediated transfections: Transient assays and generation of stale transgenic canola plants. In: Methods in molecular biology

Schaart JG, van de Wiel CCM, Lotz LAP, Smulders MJM (2015) Opportunities for products of new plant breeding techniques. Trends Plant Sci **21:** 438-449

Sevestre F, Facon M, Wattebled F, Szydlowski N (2020) Facilitating gene editing in potato: A Single-Nucleotide Polymorphism (SNP) map of the *Solanum tuberosum* L. cv. Desiree genome. Sci Rep **10:** 2045

Shopan J, Lv X, Hu Z, Zhang M, Yang J (2020) Eukaryotic translation initiation factors shape RNA viruses resistance in plants. Hortic Plant **6**: 81-88

Sridhar J, Venkateswarlu V, Shah MA, Kumari N, Raigond B, Bhatnagar A, Choudhary JS, Sharma S, Nagesg M, Chakrabarti SK (2022) Species composition and distribution of the vector aphids of PVY and PLRV in India. Potato Res **65:** 601-617

Staals RH, Agari Y, Maki-Yonekura S, Zhu Y, Taylor DW, van Duijn E, Barendregt A, Vlot M, Koehorst JJ, Sakamoto K, Masuda A, Dohmae N, Schaap PJ, Doudna JA, Heck AJ, Yonekura K, van der Oost J, Shinkai A (2013) Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of Thermus thermophilus. Mol Cell **52:** 135-145

Sun H, Jiao WB, Krause K, Campoy JA, Goel M, Folz-Donahue K, Kukat C, Huettel B, Schneeberger K (2022) Chromosome-scale and haplotype-resolved genome assembly of a tetraploid potato cultivar. Nat Genet **54:** 342-348

Taliansky M, Mayo MA, Barker H (2003) Potato leafroll virus: a classic pathogen shows some new tricks. Mol Plant Pathol **4:** 81-89

Tanenbaum ME, Gilbert LA, Qi LS, Weissman JS, Vale RD (2014) A protein-tagging system for signal amplification in gene expression and fluorescence imaging. Cell **159:** 635-646

Tavazza R, Ancora G (1986) Plant regeneration from mesophyll protoplasts in commercial potato cultivars (Primura, Kennebec, Spunta, Desirée). Plant Cell Rep **5:** 243-246

Terns MP, Terns RN (2011) CRISPR-based adaptive immune systems. Curr Opin Microbiol **14**: 321-327

The Potato Genome Consortium (2011) Genome sequence and analysis of the tuber crop potato. Nature **475**: 189-197

Truniger V, Nieto C, González-Ibeas D, Aranda M (2008) Mechanism of plant eIF4E-mediated resistance against a Carmovirus (Tombusviridae): Cap-independent translation of a viral RNA controlled in cis by an (a)virulence determinant. Plant J **56:** 716-727

Uitdewilligen JGAML, Wolters AA, D'hoop BB, Borm TJA, Visser RGF, van Eck HJ (2015) A next-generation sequencing method for genotyping by-sequencing of highly heterozygous autotetraploid potato. PLoS **8**: e62355

United Nations, Department of Economic and Social Affairs, Population Division (2015) World population prospects: The 2015 revision

Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD (2010) Genome editing with engineered zinc finger nucleases. Nat Rev **11:** 636-646

Van Lieshout N, van der Burgt A, de Vries ME, te Maat M, Eickhold D, Esselink D, van Kaauwen MPW, Kdde LP, Visser RGF, Lindhout P, Finkers R (2020) Solyntus, the new highly contiguous reference genome for potato (*Solanum tuberosum*). G3 Genes Genomes Genet **10:** 3489-3495

Visser JC, Bellstedt DU, Pirie MD (2012) The recent recombinant evolution of a major crop pathogen Potato Virus Y. PLoS ONE **7:** e50631

Visser RGF, Bachem CWB, Borm T, de Boer J, van Eck HJ, Finkers R, van der Linden G, Maliepaard CA, Uitdewilligen JGAML, Voorrips R, Vos P, Wolters AMA (2014) Possibilities and challenges of the potato genome sequence. Potato Res **57:** 327-330

Visser JC, Bellstedt DU (2015) Potato Virus Y in South Africa: Isolate characterization and assessment of potato cultivar resistance. In: Potato and sweet potato in Africa: Transforming the value chains for food and agriculture

Wada N, Osakabe K, Oskabe Y (2022) Expanding the plant genome editing toolbox with recently developed CRISPR–Cas systems. Plant Phys **188:** 1825-1837

Wang A, Krishnaswamy S (2012) Eukaryotic translation initiation factor 4E-mediated recessive resistance to plant viruses and its utility in crop improvement. Mol Plant Pathol **13:** 795-803

Wittmann S, Chatel H, Fortin MG, Laliberté JF (1997) Interaction of the viral protein genome linked of Turnip Mosaic potyvirus with the translational eukaryotic initiation factor (iso) 4E of *Arabidopsis thaliana* using the yeast two-hybrid system. Virol **243:** 84-92

Woo JW, Kim J, Kwon SI, Corvalán C, Cho SW, Kim H, Kim SG, Kim ST, Choe S, Kim JS (2015) DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. Nat Biotechnol **33:** 1162-1165

Yeam I, Cavatorta JR, Ripoll DR, Kan B, Jahn MM (2007) Functional dissection of naturally occurring amino acid substitutions in eIF4E that confers recessive potyvirus resistance in plants. Plant Cell **19**: 2913-2928

Yue JJ, Yuan JL, Wu FH, Yuan YH, Cheng QW, Hsu CT, Lin CS (2021) Protoplasts: From isolation to CRISPR/Cas genome editing application. Front Genome Ed **3:** 717017

Zhang H, Xu F, Wu Y, Hu H, Dai X (2017) Progress of potato staple food research and industry development in China. J Integr Agric **16**: 2924-2932

Zhong Z, Zhang Y, You Q, Tang X, Ren Q, Liu S, Yang X, Ren Q, Liu S, Yang L, Wang Y, Zheng X, Le Y, Zhang Y, Qi Y (2018) Plant genome editing using FnCpf1 and LbCpf1 nucleases at redefined and altered PAM sites. Mol Plant **11:** 999-1002

Zhu H, Li C, Gao C (2021) Applications of CRISPR-Cas in agriculture and plant biotechnology. Nat Rev **21:** 661-667

Chapter 3: Investigating the evolution and expression patterns of plant *eIF4E* isoforms.

3.1 Introduction

Eukaryotic translation is facilitated by the eukaryotic translation initiation factor (eIF) complex which positions a messenger RNA (mRNA) strand near a ribosome for protein assembly. The eIF4E subunit within this complex is the polypeptide that directly binds the 5' end of the mRNA strand and initiates formation of the rest of the complex (Gingras et al., 1999).

Multiple *eIF4E* isoforms have been identified in plant species (Patrick and Browning, 2012; Kropiwnicka et al., 2015), and these are often further divided into *eIF4E,* and *eIF(iso)4E* isoforms. A related RNA-cap binding protein called new cap binding protein (nCBP) has also been identified. These proteins have a degree of functional redundancy which raises questions about why this gene multiplication was maintained during evolution. It is possible, for example, that the isoforms are differentially expressed across tissue types, temporally during plant development, or when exposed to differing biotic or abiotic stresses. Tissue-specific differences in expression between *eIF4E* and *eIF(iso)4E* were observed in *Arabidopsis thaliana* (Rodriguez et al., 1998). That study found that *eIF4E* was expressed ubiquitously, whereas *eIF(iso)4E* expression was concentrated in root and flowering tissues. Replication of some RNA virus genomes utilizes the plant host's eIF complex. In these cases, viral RNA encoding a viral genome linked protein (VPg) also binds to eIF4E (Leonard et al., 2000). Expression studies of these isoforms in any other plants species across different tissues, or under biotic and abiotic stresses, has not been conducted.

Genes encoding eIF4E proteins are promising targets in engineering virus-resistant plants due to the importance of their role in viral genome translation (Robaglia and Caranta, 2006; Shopan et al., 2020). New *eIF4E* isoforms are being discovered in plant species as genome sequences become increasingly available (Patrick et al., 2014; Patrick et al., 2018). For *eIF4E*-mediated viral resistance engineering to be successful, all *eIF4E* isoforms that can interact with a viral genome need to be identified and modified, as functional redundancy between isoforms means that repression of a single isoform leads to only partial immunity (Mazier et al., 2011; Lucioli et al., 2022). This was demonstrated recently in potato and tobacco when *eIF4E-1* mutant plants showed improved, but incomplete, resistance to potato virus Y (Miroshinchenko et al., 2020; Le et al., 2022).

Complete genome sequencing provides a reliable source of information about the presence of different isoforms in species. For example, *Solanum lycopersicum* is known to contain genes encoding four RNA cap-binding proteins, eIF4E-1, eIF4E-2, eIF(iso)4E, and nCBP (Labaron et al., 2016). Until recently only *eIF4E-1*, *eIF(iso)4E* and *nCBP* genes had been identified in potato (*Solanum tuberosum*), but a potato genomic DNA

sequence similar to tomato *eIF4E-2* is present in the National Center for Biotechnology Information (NCBI) database (NCBI number: LOC102605001).

Recent years have seen vast expansion of sequence data for many evolutionarily diverse plant species. Publication of genomes from green, red, and brown algae (Nozaki et al., 2007; Cock et al., 2010; Zhang et al., 2021) and early land plants, such as moss, liverworts and ferns (Lang et al., 2018; Huang et al., 2022; Zhang et al., 2022), mean we can now form an improved overview of the evolution of specific genes. There are still some gaps in this evolutionary picture, however, as few genomes representing key evolutionary groups, such the brown algae and gymnosperms, have been published. Despite these gaps, the evolutionary knowledge that can be generated is an invaluable tool in biotechnological research. In this study I assembled a phylogenetic tree showing the evolution of the *eIF4E* gene family across a collection of representative species of many major evolutionary groups. In addition, I examined the expression of the RNA cap-binding genes in potato to identify any differential expression patterns that exist in in tubers from healthy plants compared with those infected with Potato Leaf Roll Virus.

3.2 Materials and methods

3.2.1 Suppliers

The chemical reagents used in this research were supplied by Sigma Aldrich (Missouri, USA), Bio-Rad (California, USA) and Merck Millipore (Massachusetts, USA). Molecular kits and enzymes were obtained from Thermo Fischer Scientific (Massachusetts, USA). Primers were synthesized by Inqaba Biotech (Pretoria, South Africa).

3.2.2 Acquisition of Potato Leaf Roll Virus-infected tubers

Tubers from *Solanum tuberosum* cv Markies plants that had tested positive for Potato Leaf Roll Virus (PLRV) were obtained from Plantovia (Pretoria, South Africa), and were a kind gift of Ms. Anel Espach. The tubers were stored at 4 °C after harvesting and shipped overnight on ice. Commercially bought tubers (cv Valor) were used as healthy control samples and were stored at room temperature (RT) prior to RNA extractions.

3.2.3 RNA isolation

For RNA isolation, tuber samples were frozen with liquid nitrogen and ground to a powder using a pestle and mortar. RNA was extracted from these samples following the method of Logemann et al. (1986). Two hundred mg of tissue was homogenized in 800 µL of sterile guanidine buffer (8 M guanidine hydrochloride, 20 mM 4-morpholineethansulfonic acid, 20 mM EDTA, 50 mM β-mercaptoethanol, pH 7, KOH). After centrifugation for 10 min at 15 300 x g and 4 °C, the supernatant was added to 1 vol of a 25:24:1 (v/v/v) phenol/chloroform/isoamyl alcohol solution. Samples was centrifuged for 45 min as described above, and the resulting aqueous phase was mixed with 0.7 vol of 4 °C cooled ethanol and 0.2 vol of room temperature (RT) 1 M acetic acid. After overnight incubation at -20 °C, the precipitate was harvested by centrifugation for 10 min as described above. The pellet was washed with 3 M RT sodium acetate, followed by a final 10-min

centrifugation, and a final wash with RT 70% (v/v) ethanol. After evaporation of the residual ethanol the RNA pellet was resuspended in 20 µL of water.

3.2.5 Complementary DNA synthesis

RNA was converted to complementary DNA (cDNA) with the RevertAid First Stand cDNA Synthesis Kit (Thermo Fischer Scientific). Prior to cDNA synthesis RNA samples were DNase-treated to remove any genomic DNA contamination. DNase-treatment reactions were set up as follows: 1 µg RNA, 2 U DNase I, and 1X Reaction Buffer with volume set to 10 µL with water. Following a 30 min incubation at 37 °C, 1 µL of 50 mM EDTA was added. Samples were then incubated for 10 min at 65 °C. Samples were cooled on ice before 1 µL of Oligo $(dT)_{18}$ primer was added. The samples were again incubated at 65 °C for an additional 5 min. The cDNA synthesis reaction was then set up containing the following: 1X Reaction Buffer, 20 U RiboLock RNase Inhibitor, 1 mM dNTP mix and 200 U RevertAid M-MuLV RT. Samples were incubated for 60 min at 42 °C, followed by termination of the reaction by incubation at 70 °C for 5 min.

3.2.6 Detection of Potato Leaf Roll Virus

To confirm that the commercial tuber samples obtained were PLRV-positive polymerase chain reactions (PCRs) were conducted using PLRV diagnostic primers (Forward: 5' ATGGGTACGGTCGTGGTTAAAGG 3'; Reverse: 5 'CTATTTGGGGTTCTGCAAAGC 3') and Phusion DNA Polymerase (Thermo Fisher Scientific). The diagnostic primers amplify a region within the virus's coat protein gene and were kindly gifted by Ms. Mandi Engelbrecht (Department of Genetics, Stellenbosch University). Fifty µL reactions contained the equivalent of 25 ng RNA, 2 mM of each primer, 1X Phusion™ Plus PCR Master Mix and 1X Phusion™ GC Enhancer. These reactions were incubated in a thermocycler with initial denaturation occurring at 98 °C for 30 sec, followed by 30 cycles of 98 °C for 10 sec, 60 °C for 30 sec and 72 °C for 15 sec, and final extension for 5 min at 72 °C.

3.2.7 Separation and visualisation of amplicons by agarose gel electrophoresis

PCR amplification was confirmed through separation of the PCR products on a 1% (w/v) agarose-Tris borate EDTA gel containing 0.005% (v/v) ethidium bromide with visualisation under ultraviolet light using the GelDoc imaging system (Thermo Fischer Scientific). The GeneRuler Express DNA ladder (Thermo Fischer Scientific) was included in each gel electrophoresis as a molecular weight marker.

3.2.8 Real-time quantitative polymerase chain reaction primer design

Primers to amplify regions of *eIF4E-1*, *eIF4E-2*, and *eIF(iso)4E* cDNA for real-time quantitative polymerase chain reactions (RT-qPCRs) were designed using PrimerBLAST (https://www.ncbi.nlm.nih.gov/ tools/primer-blast/). Primers to amplify regions within *nCBP* were taken from Lucioli et al. (2022). Primers for the reference gene elongation factor 1-α (*eF1α*), which is stably expressed in *S. tuberosum* plants under potyviral infection, were taken from Yin et al. (2021; Table 3.1).

Table 3.1 Primer sequences used to amplify complementary DNA sequences from *Solanum tuberosum* for real-time quantitative polymerase chain reactions.

Gene	Primer Name	Sequence (5' – 3')	Amplicon length (bp)
eIF4E-1	eIF4E1_Q_F	ACGGCGTCGTATTTGGGGAAAGA	119
	eIF4E1_Q_R	CGAAGTGAGCTTCCCCAAGCAGT	
eIF4E-2	eIF4E2_Q_F	GAGGAAGGAGAGATCGTGGG	110
	eIF4E2_Q_R	GGGTTATCGAACCAAAATGTCCA	
eIF(iso)4E	eIF(iso)4E_Q_F	ACGGAGATTCCGCCGGTCGC	116
	eIF(iso)4E_Q_R	CAAGCGGCGCCTTGTTTCGG	
nCBP	nCBP_Q_F	TGGAAGTGACACCGGAGAAG	93
	nCBP_Q_R	GATCCTCGGCGGCTATTGAG	
eF1α	eF1a_Q_F	AGATTGACAGGCGTTCAGGTAAGG	130
	eF1a_Q_R	ATGGTGGGTATTCAGCAAAGGTC	

3.2.9 Real-time quantitative polymerase chain reactions

An RT-qPCR master-mix was made for each gene target. Master-mixes consisted of 1X PowerUp™ SYBR™ Green Master Mix (Thermo Fisher Scientific), 50 nM of each of the respective forward and reverse primers, the cDNA equivalent of 2.5 ng of RNA. Reaction volumes were set to 15 µL with water. The reactions were placed into a 96-well plate and were run in the Quant Studio 3 Applied Biosystems RT-qPCR instrument from Thermo Fisher Scientific with the standard thermocycler settings (Table 3.2).

Table 3.2 Thermocycler settings for the real-time quantitative polymerase chain reactions.

Stage	Step	Temp (°C)	Duration (sec)	Ramp rate (°C/sec)	Cycles	Data collection
Hold	UDG activation	50	120	1.6	Hold	Off
	Dual-Lock™ DNA polymerase	95	120	1.6	Hold	Off
PCR	Denaturation	95	15	1.6	X 40	Off
	Annealing/Extension	60	60	1.6		On
Melt curve	1	95	15	1.6	Hold	Off
	2	60	60	1.6	Hold	Off
	3 (Dissociation)	95	15	0.15	Hold	On

3.2.10 Expression analysis

The relative fold change for each target gene was calculated using ΔCT and ΔΔCT values. Five biological replicates, each with 4 technical replicates, were used for both the healthy and PLRV-positive samples. The mean transcript levels of the genes were compared using a one-way analysis of variance followed by the *post hoc* Bonferroni multiple comparisons test (XLSTAT, 2022). All data analysis was conducted in-line with the 'Minimum Information for Publication of Quantitative Real-Time PCR Experiments' (Bustin et al., 2009).

3.2.11 Collection of sequences for phylogenetic tree construction

The *S. tuberosum* eIF4E-1 amino acid sequence was used as the query sequence in tBLASTn and BLASTp searches against the NCBI (https://www.ncbi.nlm.nih.gov/) and Phytozome (https://phytozome-next.jgi.doe.gov/) databases. In the case of some species where *nCBP* sequences were not found using this methodology, "new cap binding protein" was used as a query in searches of the NCBI database. The coding sequences (CDS) for the BLAST results were downloaded when possible. If the CDS was unavailable, the mRNA sequences were taken, and the open reading frame of these sequences were found using the NCBI's ORFfinder (https://www.ncbi.nlm.nih.gov/orffinder/).

3.2.12 Nucleotide sequence alignment and phylogenetic tree construction

Automatic alignment of the *eIF4E* sequences was created using CLUSTALW (Thompson et al., 1994) on the CIPRES Gateway (version 3.3; https://www.phylo.org/index.php/) using the default alignment settings. This was manually adjusted using BioEdit (Hall, 1999) to correct any alignment errors. Phylogenetic analyses of DNA sequence alignments were then conducted on the CIPRES Gateway version 3.3 using MrBayes version 3.2.6 (Ronquist et al., 2012). The resulting phylogenetic tree was rooted at the midpoint. Bayesian posterior support probabilities were calculated and included at each branching point. The tree topology was assessed to formulate a hypothesis for the evolution of the *eIF4E* gene family in plants.

3.3 Results

3.3.1 Confirmation of Potato Leaf Roll Virus infection

RNA was successfully isolated from both putative PLRV infected tubers and healthy control tubers. After DNase treatment, the RNA samples were converted into cDNA. This was used to confirm infection through PCR amplification using PLRV diagnostic primers and visualization of the PCR products through DNA gel electrophoresis (Figure 3.1). The virus was present in all five tuber samples, and no amplification was observed when using cDNA from a healthy tuber sample.

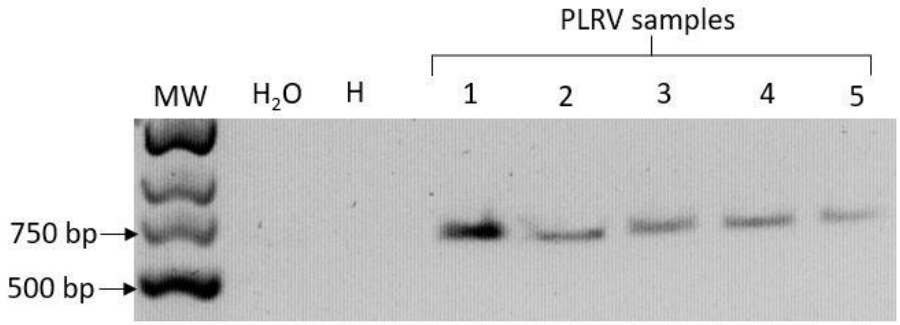


Figure 3.1 The polymerase chain reaction products generated through screening of putatively Potato Leaf Roll Virus-positive tuber samples with diagnostic primers separated on a 1% (w/v) agarose-Tris Borate EDTA gel using gel

electrophoresis. Complementary DNA from a putatively healthy tuber was also tested (H), as well as a water negative control (H_2O). The full gel image is shown in Supplementary Figure 3.1.

3.3.2 eIF4E gene family transcript levels analysis

Transcript levels of *eIF4E-1*, *eIF4E-2*, *eIF(iso)4E* and *nCBP* in healthy and PLRV-infected tubers were obtained through RT-qPCR using the ΔCt and ΔΔCt calculation method (Figure 3.2). This data was normalized using the *eF1α* reference gene transcript levels. The PLRV-infected samples showed a mean relative fold change in expression of 3.803, 2.378, 1.582 and 9.942 for *eIF4E-1*, *eIF4E-2, eIF(iso)4E* and *nCBP* respectively when compared to the healthy controls. Following one way analysis of variance, the Bonferroni-holm post hoc test showed that this increase was not statistically significant for any of the genes.

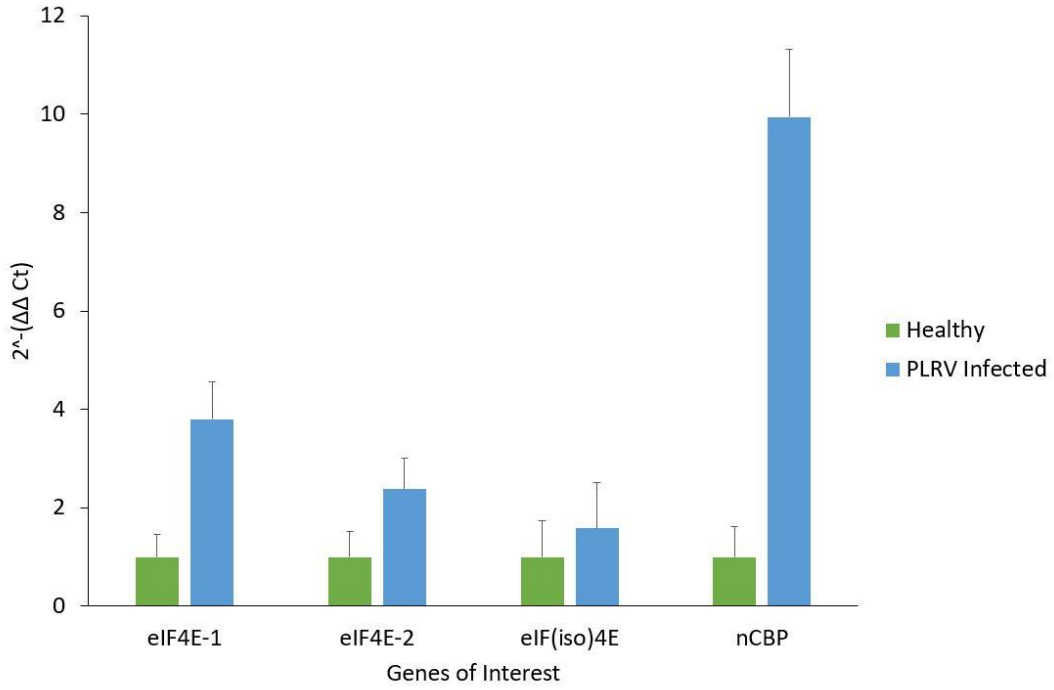


Figure 3.2 The relative fold change in expression ($2^{\Delta\Delta Ct}$) for *eIF4E-1*, *eIF4E-2*, *eIF(iso)4E* and *nCBP* in Potato Leaf Roll Virus-infected tubers compared to healthy controls samples (n=5). Error bars indicate standard errors of the mean.

3.3.3 Phylogenetic tree construction of the *eIF4E* gene family

One hundred and twenty accessions, representing 34 species were collected through the tBLASTn, and BLASTp database searches (Supplementary table 3.1). These species included the *S. tuberosum* close relatives *Solanum lycopersicum*, *Capsicum annuum*, *Nicotiana benthamiana* and *Nicotiana tabacum*, as well as representative species of many major land plant, green algal and red algal phyla. Sequences from zebrafish (*Danio rerio*) were included as a eukaryotic outgroup. The gene names given by the different databases were highly inconsistent. For the purposes of this investigation the accessions were relabeled with alphabetical differentiators. The original gene name annotation can be found in Supplementary table 3.1.

Initial nucleotide and amino acid alignments were created using the CLUSTAL W algorithm. The nucleotide alignment was manually adjusted using amino acid translations as a guide to correct any frame-shifting gap insertions. The resulting alignment - which was 1332 nucleotides long - was used to create phylogenetic trees using Bayesian Inference. The MrBayes analysis used separate model-averaged partitions for each codon position and gamma corrections for among-site rate variability and was run for 2×10^8 generations. MrBayes output diagnostics indicated convergence on, and adequate sample sizes from, the posterior probability distribution. The midpoint-rooted maximum clade credibility tree from multiple converged runs is presented in Figure 3.3, with posterior probabilities > 50% indicated on each branch.

Midpoint rooting of the tree divided the sequences *Chondus crispus* B and *Porphyra umbilicus* B from the rest of the accessions. The remainder of the tree divided the RNA cap-binding proteins into two distinct clades containing the *nCBP* (Figure 3.3 A) and *eIF4E* (Figure 3.3 B) sequences, respectively. Within each of these two groups the *Danio rerio* outgroup sequences were separate from the plant sequences. The *nCBP* clade mirrors the currently accepted land plant phylogeny. Clade B containing *eIF4E* sequences closely approximates the known Archaeplastida phylogeny, with successive clades of red and green algae sister to the land plant lineage. The non-vascular land plants within this clade form individual branches (Figure 3.3 B) and multiple moss-specific gene duplications are present in *Physcomitrium patens* and *Sphagnum fallax*. A duplication of the *eIF4E*-ancestor into separate *eIF(iso)4E* and *eIF4E* clades is observed in at least the angiosperms and possibly as far back as the vascular plants (Figure 3.3 C and D). As was the case with the *nCBP* clade, these two clades mirror the known evolutionary history of these groups. A further gene duplication occurred in the ancestor of the sampled Solanaceae species, as these sequences further divide into individual *eIF4E-1* and *eIF4E-2* clades (Figure 3.3 E and F).

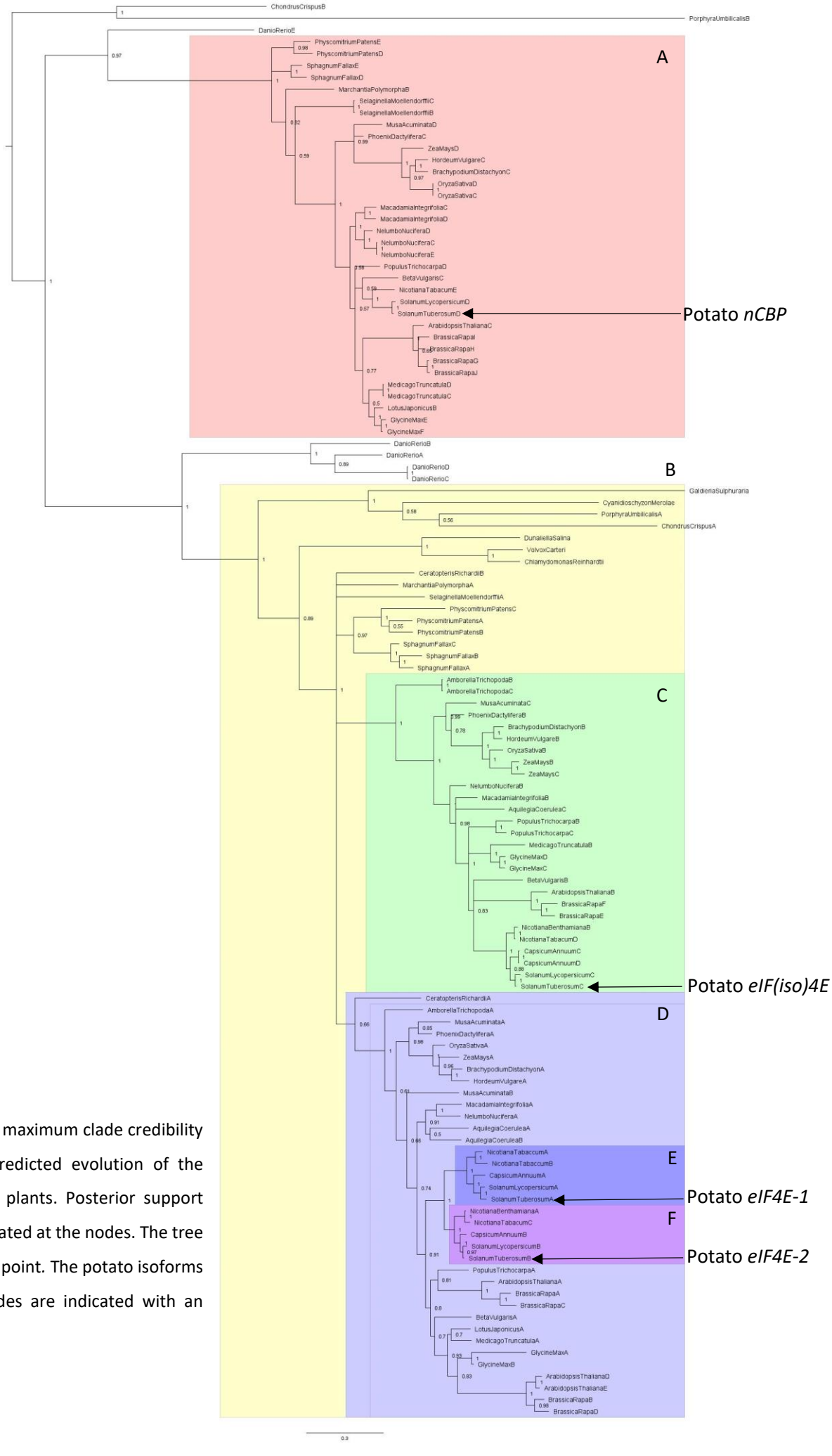


Figure 3.3 A Bayesian maximum clade credibility tree depicting the predicted evolution of the *eIF4E* gene family in plants. Posterior support probabilities are indicated at the nodes. The tree was rooted at the midpoint. The potato isoforms in each of these clades are indicated with an arrow.

3.3.4 Analysis of the VPg-interacting amino acid residues of the different phylogenetic clades

The amino acid sequence alignment for each of the phylogenetic tree's major clades were assessed for similarity in the region containing the known eIF4E-1 VPg binding site (Figure 3.4). The critical residues in this protein interaction are an arginine at amino acid position 157 (R157), and two lysines at amino acid positions 159 and 162 (K159 and K162; Coutinho de Oliveira et al., 2019) where this numbering is based on the potato eIF4E-1 amino acid sequence. In the clade containing the potato eIF4E sequences (Figure 3.3 C; Figure 3.4) the R157 residue is 100% conserved. The K159 and K162 residues are conserved in most sequences, and if not, were substituted with an arginine. The exception to this are *Hordeum vulgare* A and *Lotus japonicus* A which have an asparagine and an alanine in position 159 respectively. The clade containing the potato eIF(iso)4E sequence (Figure 3.3 B; Figure 3.4) contains proteins where the R157 and K162 residues are conserved, except for *Brachypodium distachyon* A which has a lysine-to-arginine substitution at position 162. In this clade the residue at position 159 is more variable and an arginine substitution in present in many families. In the Solanaceae family a serine substitution is present at position 159 for all sequences except *Nicotiana tabacum* C, and in the Fabaceae family a tryptophan substitution is conserved. The clade containing the potato nCBP sequence (Figure 3.3 A; Figure 3.4) demonstrates complete conservation of the R157 residue. The K159 residue is, however, substituted for an asparagine in all sequences except for *Brassica rapa* G, *Brassica rapa* J and *Beta vulgaris* C in which it is substituted by a serine. In all sequences within this clade the K162 residue is substituted for an isoleucine.

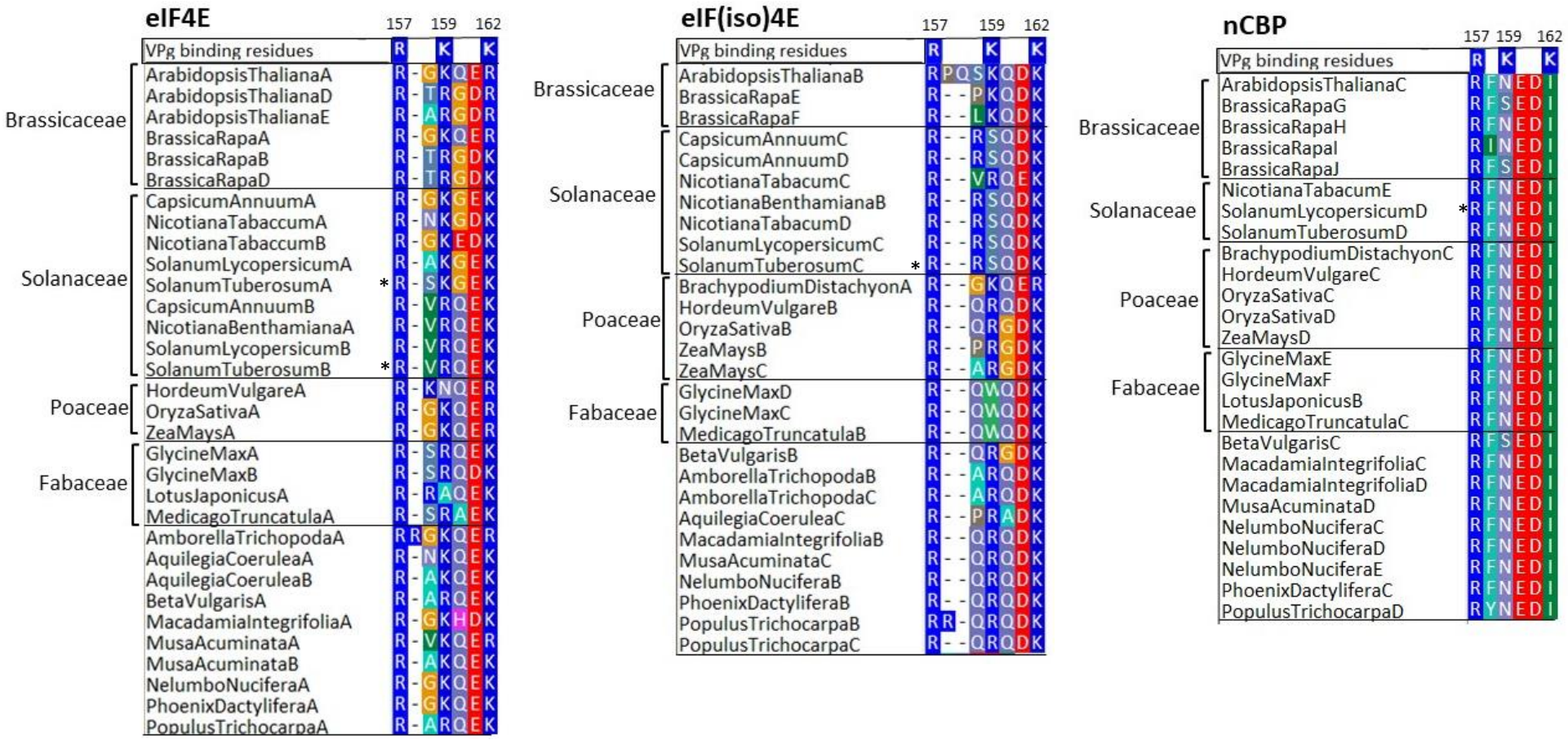


Figure 3.4 The amino acid sequences around the known eIF4E-1 VPg-binding site were aligned for the angiosperm members of the phylogenetic clades containing the potato eIF4E, eIF9iso)4E and nCBP sequences to assess for similarity. The potato sequences are indicated with an asterisk.

3.4 Discussion

In potato *eIF4E-1*, *eIF4E-2* and *eIF(iso)4E* are functionally redundant (Nellist et al., 2013). Expression studies in *Arabidopsis thaliana* and *Zea mays* indicate that the different isoforms are differentially expressed across tissue types and during different stages of plant development (Rodriguez et al., 1998; Dinkova and de Jiménez, 1999). Plant transcriptomes also vary greatly under viral stress, and it is possible that changes in the expression of these genes would differ under these conditions as well (Whitham et al., 2006). In this research the expression of these three *eIF4E* isoforms, as well as the *nCBP* RNA cap-binding protein, was studied in tubers from healthy and PLRV-infected potato plants. No significant difference was found between the healthy and PLRV-infected samples which indicates that the expression of these genes does not change under this biotic stressor. The validity of this data is questionable, however, due to differences between the sample groups. The cultivar and age of the PLRV-infected and healthy samples were different, and the tubers from both these groups were stored in different conditions post-harvest. This was because it proved difficult to acquire infected tubers and, when they were received in July 2022, the supplier did not have uninfected tubers of the same cultivar to act as a control. It was not possible to source tubers of that cultivar from elsewhere before the expression analysis was performed. When they arrived the PLRV-infected tubers were also dormant, which would affect the degree to which the virus would accumulate in this tissue (Gugerli and Gehriger, 1980). All these variables would impact the state of the transcriptome and thus the healthy tuber samples used were not suitable as a comparative control, nor were the PLRV-infected tubers a reliable indicator of the transcriptome of a PLRV-infected plant. This analysis, however, did allow for the optimization of primers in RT-qPCR reactions which can be use in future experiments. The PLRV-infected tubers have since sprouted and will be grown in soil so that RNA can be isolated from a range of tissue types of actively growing plants to repeat this experiment in other tissues. Healthy plants of the same cultivar which are of the same age should also be grown in these experiments as a more reliable comparative control. It would also be interesting to conduct aphid-inoculation studies with healthy potato plants to generate newly infected samples and assess if any expression differences can be observed as infection progresses.

Bayesian Inference of the *eIF4E* gene family showed a major division between plant *nCBP* (Figure 3.3 A) and *eIF4E/eIF(iso)4E* sequences (Figure 3.4 B). The positions of the red and green algal sequences that were included in this study indicate that the divide of the ancestral RNA cap-binding protein into nCBP and the eIF4E-ancestor is ancient and occurred substantially before the emergence of land plants. Within clade B, single gene copies are present in the non-vascular plants *Marchantia polymorpha* and *Selaginella moellendorffi* while three versions of *eIF4E* are present in the bryophytes *Physcomitrium patens* and *Sphagnum fallax*. This indicates that these isoforms evolved within these species individually, possibly through genome duplication events which are known to have occurred in *P. patens* (Rensing et al., 2007). The angiosperm sequences within the *eIF4E*-ancestor clade further divide into individual *eIF4E* (Figure 3.4 C) and *eIF(iso)4E* (Figure 3.4 D) groups, indicating that this split occurred somewhere between the common

ancestor of all vascular plants and the emergence of angiosperms. The absence of any gymnosperm representatives and the uncertain placement of *Ceraptopteris* and *Selaginella* isoforms in this tree means that a more precise placement between these two evolutionary events cannot be inferred. The later duplication of *eIF4E* into *eIF4E-1* (Figure 3.4 E) and *eIF4E-2* (Figure 3.4 F) occurred deep within the eudicotyledonous lineage, possibly in the common ancestor of the Solanaceae.

The increased cellular complexity of land plants compared to their aquatic ancestors could explain the need for multiple, functionally redundant proteins. Transcriptome analysis in *A. thaliana* has demonstrated that *eIF4E* genes are differentially expressed in different plant tissues (Rodriguez et al., 1998) and this may be true for all land plants. The further divergence into *eIF4E-1* and *eIF4E-2* within the Solanaceae sequences, however, is more puzzling. Expression analysis of differences between these two isoforms in different tissue types, and under stressed conditions would be helpful in determining the different role these proteins have in Solanaceous plants, but analysis of knockout mutants will be needed for functional analysis. The separation of the potato *eIF(iso)4E* and *eIF4E* genes into separate clades in the phylogenetic tree mirrors the differences between these proteins in the eIF complex. Within this complex, the eIF4E proteins and eIF(iso)4E protein associate with eIF4G and eIF(iso)4G respectively. We can presume that this difference emerged after the divergence of these group, and that these associations would be conserved within all members of each of the two clades.

The posterior probability values for the split of *eIF4E*, *eIF(iso)4E* and *nCBP* were 1.00, 1.00, and 0.97 respectively, which indicates strong support for these branches. The duplication of *eIF4E* in Solanaceous plants are also supported by a posterior probability value of 1.00. Each of the major clades contain branches with posterior support values below 0.95 which indicates that these branches are weakly supported and may not be a true representation of the evolution of these sequences. It is possible, however, that errors in the alignment of these sequences biased the Bayesian analysis and resulted in misleadingly low posterior support values (Jones, 2008). Both the CLUSTAL W alignment and Bayesian Interference tree were created without modifications to the standard parameters. Adjustment to these to compensate for sequence variables such GC-content improve the confidence with which we look at these branches. The use of several alignment programs to obtain a consensus alignment prior to Bayesian analysis has also been demonstrated to improve reliability of a tree (Golubchik et al., 2007), and this could also be considered should this work be replicated in the future. With that being said, it is unlikely that these methods would have a significant impact on the overall topology of the tree. The phylogenetic patterns that exist within each of the tree's three major clades closely approximate each other, and the known evolution of land plants and their closest relatives. This means that there is a high degree of confidence that this branching pattern is a true depiction of the evolutionary events.

A major limitation of the phylogenetic tree produced in this study is the absence of any gymnosperm representatives. The genome sequences of several gymnosperms have recently been published (Huang et al., 2022; Liu et al., 2022) and these were included in the BLAST searches for *eIF4E* relatives. These searches identified no complete gene sequences that resembled the *eIF4E* genes in any significant way. Some partial sequences were identified, but it was not possible to use these as the sequence lengths were too short to be aligned for use in a meaningful phylogenetic analysis. One complete gene sequence was identified from *Pinus taeda* (NCBI accession number: BQ699082.1) but it was too dissimilar to the other sequences to be included. It is unlikely that this absence of sequences is due to deletion of this gene family in gymnosperms because the *eIF4E* genes perform an essential function in translation. It is more likely, therefore, that incomplete genome records resulted in these sequences being absent. To elucidate a more precise time point for the divergence of *eIF4E* and *eIF(iso)4E* this analysis will need to be repeated once more gymnosperm representatives can be included.

The role of the different cap-binding proteins in viral translation can be implied through analysis of the amino acid sequences in the VPg-binding region of these genes (Figure 3.4). The presence of amino acid residues in eIF4E or nCBP proteins that are known to interact with a VPg indicate that this protein would need to be disrupted during viral resistance engineering. Protein interaction studies elucidated that three amino acid residues in the potato eIF4E-1 cap-binding region interact with the PVY VPg (Coutinho de Oliveira et al., 2019). In all potato eIF4E isoforms the R157 and K162 residues are conserved, however, in eIF4E-2 there is a conservative K159R substitution in eIF4E-2 and a K159S substitution in eIF(iso)4E (Figure 3.5). Lysine and arginine are both polar amino acids with positively charged sidechains and therefore would be expected to similarly interact with the VPg (Betts and Russel, 2003). Based on this it can be hypothesised that the eIF4E-2-VPg interaction will be similar to that of the eIF4E-1-VPg interaction. Potato eIF(iso)4E contains a serine at position 159, but an arginine in position 158. Both serine and arginine are polar amino acids and, despite their different sidechain charges, may be able to conservatively substitute for each other (Betts and Russel, 2003). It is also possible that the R158 present in eIF(iso)4E could perform the same interaction with the VPg as the L159. Both these possibilities need to be examined empirically. The R157 residue is conserved in potato nCBP (Figure 3.5), however, a lysine-to-asparagine and lysine-to-isoleucine substitutions occurred in position 159 and 162. Neither asparagine, nor isoleucine have a positively charged side chain. Despite this, asparagine is polar and might be able to substitute for lysine in the protein-VPg interaction. Isoleucine on the other hand is aliphatic and so is very unlikely to be able to interact with the VPg. Given these differences it is unlikely that nCBP would be able to interact with a VPg like its eIF4E relatives. Studies into the role of this protein during infection demonstrate that is involved in viral cell-to-cell movement (Keima et al., 2017), rather than viral genome translation and based on this data it is unlikely that it would need to be mutated for viral resistance to be achieved. The amino acid differences between eIF4E-1 and the three other proteins

described above are conserved in all the solanaceous species examined in this study (Figure 3.4) and thus we can infer that the VPg binding capability of one isoform will be conserved across of this family.

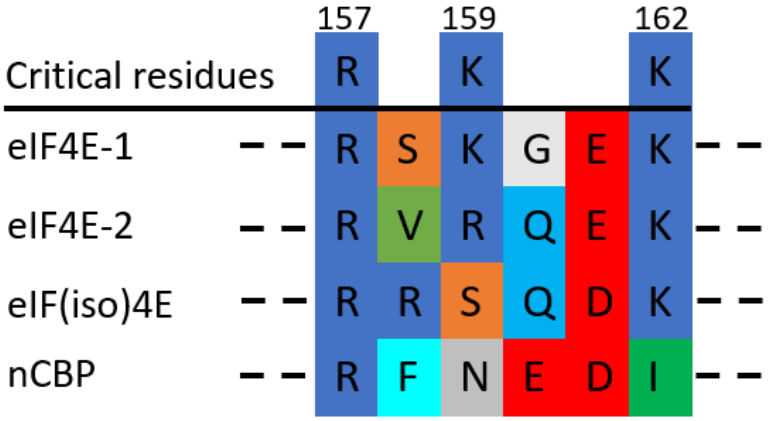


Figure 3.5 Alignment of the amino acid sequences of the known Potato Virus Y VPg interaction site in potato eIF4E-1 and equivalent sites in potato eIF4E-2, eIF(iso)4E, and nCBP. The arginine in position 157 is conserved across all four proteins. The lysine in position 159 of eIF4E-1 is substituted for an arginine, serine, and phenylalanine in eIF4E-2, eIF(iso)4E and nCBP respectively. In eIF4E-1, eIF4E-2 and eIF(iso)4E, the lysine in position 162 is present, however in nCBP it is substituted with an isoleucine.

To conclude, this phylogenetic resource provides a valuable tool when designing experiments to create *eIF4E*-mediated resistance in several species as it can be used to predict the VPg binding capabilities of an eIF4E protein. The separation of the *eIF4E*, *eIF(iso)4E* and *nCBP* clades align with what is known regarding their role in both native translation and during viral infection. Expression studies of these genes in different tissue types in healthy and virus-stressed plants in the future would help elucidate the specific roles these genes play in Solanaceous plants.

3.5 References

Betts MJ, Russel RB (2003) Amino acid properties and consequences of substitutions. In: Bioinformatics for Geneticists. **14:** 289-316

Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffi MW, Shipley GL, Vandesompele J, Wittwer CT (2009) The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. Cli Chem **55:** 611-622

Cock JM, Sterck L, Rouzé P, Scornet D, Allen AE, Amoutzias G, Anthourd V, Artiguenave F, Aury JM, Badger JH, Beszteri B, Billiau K, Bonnet E, Bothwell JH, Bowler C, Boyen C, Brownlee C, Carrano CJ, Charrier B, Cho GY, Coelho SM, Collén J, Corre E, Da Silva C, … Winker P (2010) The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. Nature **3:** 617-621

Coutinho de Olivera L, Volpon L, Rahardjo A, Osborne M, Culjkovic-Kraljacic B, Trahan C, Oeffinger M, Kwok B, Borden K (2019) Structural studies of the eIF4E–VPg complex reveal a direct competition for capped RNA: Implications for translation. PNAS **116:** 24056-24065

Dinkova TD, de Jiménez ES (1999) Differential expression and regulation of translation initiation factors -4E and -iso4E during maize germination. Physiol Plant **107:** 419-425

Gingras AC, Raught B, Sonenberg N (1999) eIF4 initiation factors: Effectors of mRNA recruitment to ribosomes and regulators of translation. Annu Rev Biochem **68:** 913-963

Golubchik T, Wise MJ, Easteal S, Jermiin LS (2007) Mind the gaps: Evidence in estimates of multiple sequence alignments. Mol Biol Evol **24:** 2433-2442

Gugerli P, Gehriger W (1980) Enzyme-linked immunosorbent assay (ELISA) for the detection of Potato Leaf Roll Virus and Potato Virus Y in potato tubers after artificial break of dormancy. Pot Res **23:** 353-359

Hall TA (1999) A user-friendly biological sequence alignment editor and analysis program for Windows. Nucleic Acids Symp Ser **41:** 95-98

Huang X, Wang W, Gong T, Wickell D, Kuo LY, Zhang X, Wen J, Kim H, Lu F, Zhao H, Chen S, Li H, Wu W, Yu C, Chen S, Fan W, Chen S, Bao X, Li L, Zhang D, Jiang L, Yan X, Liao Z, Zhou G, … Li G (2022) The flying spider-monkey tree fern genome provides insights into fern evolution and arborescence. Nat Plants **8:** 500-512

Jones G (2008) On the reliability of Bayesian posterior clade probabilities in phylogenetic analysis. Nat Prec doi.org/10.1038/npre.2008.1494.1

Keima T, Hagiwara-Komoda Y, Hashimoto M, Neriya Y, Koinuma H, Iwabuchi N, Nishida S, Yasuyuki Y, Namba S (2017) Deficiency of the eIF4E isoform nCBP limits the cell-to-cell movement of a plant virus encoding triple-gene-block proteins in *Arabidopsis thaliana*. Sci Rep. **7:** 39678

Kropiwnicka A, Kuchta K, Lukaszewicz M, Kowalska J, Jemielity J, Ginalski K, Darzynkiewicz E, Zuberek J (2015) Five eIF4E isoforms from Arabidopsis thaliana are characterized by distinct features of cap analogs binding. BBRC **456:** 47-52

Lang D, Ullich KK, Murat F, Fuchs J, Jenkins J, Haas FB, Piednoel M, Gundlach H, Van Bel M, Meyberg R, Vives C, Morata J, Symeonidi A, Hiss M, Muchero W, Kamisugi Y, Saleh O, Blanc G, Decker EL, van Gessel N, Grimwood J, Hayes RD, Graham SW, Gunter EL, … Rensig SA (2018) The Physcomitrella patens chromosome-scale assembly reveals moss genome structure and evolution. Plant J **93:** 515-533

Lebaron C, Rosado A, Sauvage C, Gauffier C, German-Retana SS (2016) A new eIF4E1 allele characterized by RNAseq data mining is associated with resistance to PVY in tomato albeit with a low durability. J Gen Viol **97:** 3063-3072

Le NT, Tran HT, Bui TP, Nguyen GT, Nguyen DV, Ta DT, Trinh DD, Molnar A, Pham NB, Chu HH, Do PT (2022) Simultaneously induced mutations in *eIF4E* genes by CRISPR/Cas9 enhance PVY resistance in tobacco. Sci **12:** 14627

Leonard S, Plante D, Wittmann S, Daigneault N, Fortin MG, Laliberté JF (2000) Complex formation between potyvirus VPg and translation eukaryotic initiation factor 4E correlates with virus infectivity. J Virol **74:** 7730-7737

Liu Y, Wang S, Li L, Yang T, Dong S, Wei T, Wu S, Liu Y, Gong Y, Feng X, Ma J, Chang G, Huang J, Yang Y, Wang H, Liu M, Xu Y, Liang H, Yu J, Cai Y, Zhang Z, Fan Y, Mu W, Sahu SK, … Zhang S (2022) The Cycas genome and the early evolution of seed plants. Nat Plants **8:** 389-401

Logemann L, Schell J, Willmitzer L (1987) Improved method for the isolation of RNA from plant tissues. Anal Biochem **163**: 16-20

Lucioli A, Tavazza R, Baima S, Fatyol K, Burgyan J, Tavazza M (2022) CRISPR-Cas9 targeting of the eIF4E1 gene extends the Potato Virus Y resistance spectrum of the Solanum tuberosum L. cv. Desirée. Front Micobiol **13:** 873930

Mazier M, Flamain F, Nicolaï, Sarnette V, Caranta C (2011) Knock-down of both eIF4E1 and eIF4E2 genes confers broad-spectrum resistance against potyviruses in tomato. PLoS One **6:** e29595

Miroshichenko D, Timerbaev V, Okuneva A, Klementyeva A, Sidorova T, Pushin A, Dolgov S (2020) Enhancement of resistance to PVY in intragenic marker-free potato plants by RNAi-mediated silencing of eIF4E translation initiation factors. PCTOC **140:** 691-705

Nellist CF, Qian W, Jenner CE, Moore JD, Zhang S, Wang X, Briggs WH, Barker GC, Sun R, Walsh JA (2013) Multiple copies of eukaryotic translation initiation factors in *Brassica rapa* facilitate redundancy, enabling diversification through variation in splicing and broad-spectrum virus resistance. Plant J **32:** 12389

Nozaki H, Takano H, Misumi O, Terasawa K, Matsuzaki M, Maruyama S, Nishida K, Yagisawa S, Kuroiwa H, Tanaka K, Sato N, Kuroiwa T (2007) A 100%-complete sequence reveals unusually simple genomic features in the hot-spring alga *Cyanidioschyzon merolae.* BMC Biol **5:** 28

Patrick RM, Browning KS (2012) The eIF4F and eIF(iso)4F complexes of plants: An evolutionary perspective. Comp Funct Genomics **2012:** 287814

Patrick RM, Mayberry LK, Choy G, Woodard LE, Liu JS, White A, Mullen RA, Tanavin TM, Latz CA, Browning KS (2014) Two *Arabidopsis* loci encode a novel eukaryotic initiation factor 4E isoforms that are functionally distinct from the conserved plant eukaryotic initiation factor 4E. Plant Physiol **164:** 1820-1830

Patrick RM, Lee JCH, Teetsel JRJ, Yang SH, Choy GS, Browning KS (2018) Discovery and characterization of conserved binding of eIF4E 1 (CBE1), a eukaryotic translation initiation factor 4E-binding plant protein. J Biol Chem **44:** 17240-17247

Rensing SA, Ick J, Fawcett JA Lang D, Zimmer A, Van de Peer Y, Reski R (2007) An ancient genome duplication contributed to the abundance of metabolic genes in the moss *Physcomitrella patens*. BMC Evol Biol **7**: 130

Robaglia C, Caranta C (2006) Translation initiation factors: A weak link in plant RNA virus infection. Trends Plant Sci **11:** 1360-1385

Rodriguez CM, Freire MA, Camilleri C, Robaglia C (1998) The *Arabidopsis thaliana* cDNAs for eIF4E and eIF(iso)4E are not functionally equivalent for yeast complementation and are differentially expressed during plant development. Plant J **14:** 465-473

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19:** 1572-1574

Shopan J, Lv X, Hu Z, Zhang M, Yang J (2020) Eukaryotic translation initiation factors shape RNA viruses resistance in plants. Hortic Plant **6**: 81-88

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res **22:** 4673-4680

Whitham SA, Yang C, Goodin MM (2006) Global impact: elucidating plant responses to viral infection. MPMI **16:** 1207-1215

Yin Z, Xie F, Michalak K, Zhang B, Zimnoch-Guzowska E (2021) Reference gene selection for miRNA and mRNA normalization in potato in response to Potato Virus Y. Mol Cell Probes **55:** 101691

Zhang X, Cvetkovska M, Morgan-Kiss R, Hüner NPA, Smith DR (2021) Draft genome sequence of Antarctic green alga *Chlamydomonas* sp. UWO241. Cell Press **24:** 102048

Zhang J, Fu XX, Li RQ, Zhao X, Liu Y, Li MH, Zwaenepoel A, Ma H, Goffinet B, Guan YL, Xue JY, Liao YY, Wang QF, Wang QH, Wang JY, Zhang GQ, Wang ZW, Jia Y, Wang MZ, Dong SS, Yang JF, Jiao YN, Guo YL, Kong HZ, … Chen ZD (2022) The hornwort genome and early land plant evolution. Nat Plants **6:** 107-118

Supplementary material

Supplementary Table 3.1 Complete list of accessions included in the nucleotide and amino acid alignments for phylogenetic tree construction.

Supplementary Figure 3.1 The original DNA gel showing PCR amplification from PLRV-infected tuber cDNA samples with PLRV-diagnostic primers.

Genus species	Common name/Description	Accession label	Database and accession number	Database annotation
Amborella trichopoda	Amborella	AmborellaTrichopodaA	XM_011628365.2	eIF4E-1
		AmborellaTrichopodaB	XM_006858875.3	eIF variant
		AmborellaTrichopodaC	XM_011629857.2	nCBP
Aquilegia coerulea	Colorado blue columbine	AquilegiaCoeruleaA	Aqcoe1G208100.1	eIF4E-related
		AquilegiaCoeruleaB	Aqcoe6G163900.1	eIF4E-related
		AquilegiaCoeruleaC	Aqcoe2G381800.1	eIF(iso)4E
Arabidopsis thaliana	Arabidopsis	ArabidopsisThalianaA	NM_001344112.1	eIF4E-1b
		ArabidopsisThalianaB	NM_117914.4	eIF4E-1c
		ArabidopsisThalianaC	NM_102695.3	eIF4E
		ArabidopsisThalianaD	AC068667.6	eIF(iso)4E
		ArabidopsisThalianaE	AF028809.1	nCBP
Beta vulgaris	Beetroot	BetaVulgarisA	EL10Ac8g19813.1	eIF4E
		BetaVulgarisB	EL10Ac3g04952.1	eIF(iso)4E
		BetaVulgarisC	XM_010695511.3	nCBP
Brachypodium distachyon	Purple false brome	BrachypodiumDistachyonA	Bradi2g61950.2	eIF4E
		BrachypodiumDistachyonB	Bradi3g28340.1	eIF4E related
		BrachypodiumDistachyonC	Bradi1g67380.1	nCBP
Brassica rapa	Field mustard	BrassicaRapaA	Brara.H00906.1	eIF4E related
		BrassicaRapaB	Brara.G00756.1	eIF4E-like
		BrassicaRapaC	Brara.A00926.1	eIF4E-like

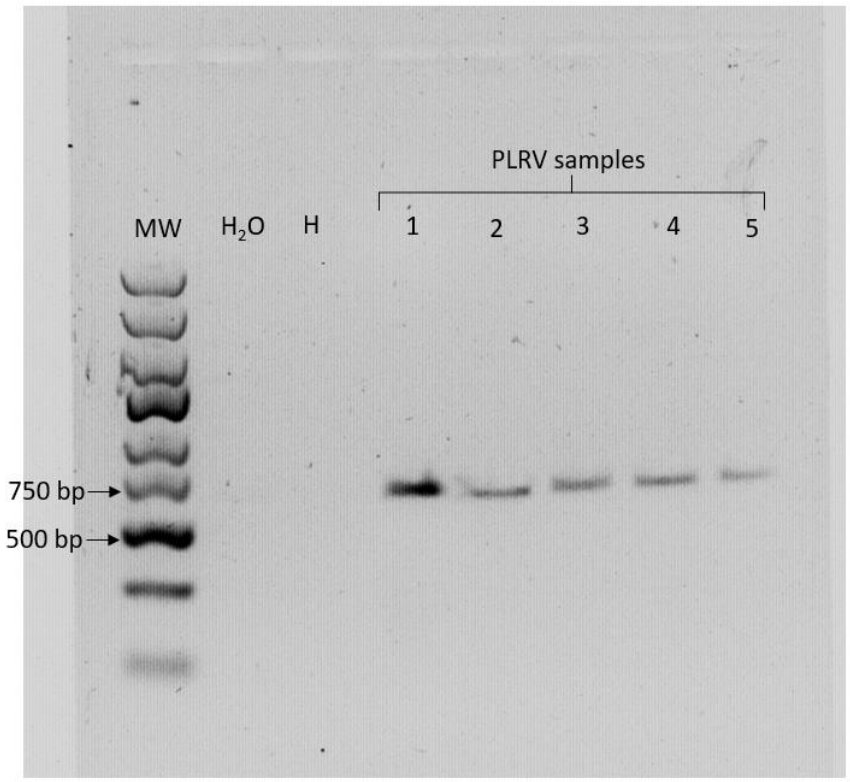
Brassica rapa	Field mustard	BrassicaRapaD	Brara.I02873.1	eIF4E
		BrassicaRapaE	Brara.D00815.1	eIF4E
		BrassicaRapaF	Brara.E01392.1	eIF(iso)4E
		BrassicaRapaG	Brara.J01750.1	eIF(iso)4E
		BrassicaRapaH	Brara.C00800.1	eIF4E Isoform B
		BrassicaRapaI	Brara.B00758.1	eIF4E Isoform B
		BrassicaRapaJ	XM_009122770.3	nCBP
Capsicum annuum	Bell pepper	CapsicumAnnuumA	MZ041789.1	eIF4E-1
		CapsicumAnnuumB	XM_016702482.2	eIF4E-2
		CapsicumAnnuumC	FN824361.1	eIF(iso)4E
		CapsicumAnnuumD	JW091118.1	nCBP
Ceratopteris richardii	Triangle waterfern	CeratopterisRichardiiA	Ceric.28G004300.1	eIF4E
		CeratopterisRichardiiB	Ceric.19G028500.1	eIF4E-like
Chlamydomonas reinhardtii	Unicellular green algae	ChlamydomonasReinhardtii	Cre03.g199900.t1.2	eIF4E
Chondrus crispus	Irish moss (red algae)	ChondrusCrispusA	XM_005710101.1	eIF4E-type 1
		ChondrusCrispusB	XM_005709999.1	eIF4E-type 2
Cyanidioschyzon merolae	Unicellular red algae	CyanidioschyzonMerolae	XM_005537163.1	eIF4E
Danio rerio	Zebra fish	DanioRerioA	NM_131733.1	eIF4E-1C
		DanioRerioB	NM_131454.1	eIF4E-1A
		DanioRerioC	NM_001017851.2	eIF4E-1B
		DanioRerioD	XM_021480575.1	eIF4E-type 2
		DanioRerioE	NM_001014815.2	eIF4E-type 3
Dunaliella salina	Unicellular green micro algae	DunaliellaSalina	NM_001004589.1	eIF4E-1C-isoform

Galdieria sulphuraria	Unicellular red algae	GaldieriaSulphuraria	XP_005703964.1	eIF4E
Glycine max	Soybean	GlycineMaxA	Glyma.15G089800.1	eIF4E-related
		GlycineMaxB	Glyma.13G222400.1	eIF4E-related
		GlycineMaxC	Glyma.20G085800.1	eIF(iso)4E
		GlycineMaxD	Glyma.10G136100.1	eIF(iso)4E
		GlycineMaxE	Glyma.16G153800.1	eIF4E Isoform B
		GlycineMaxF	Glyma.02G072700.1	eIF4E Isoform B
Hordeum vulgare	Barely	HordeumVulgareA	HORVU3Hr1G113940.1	eIF4E
		HordeumVulgareB	HORVU1Hr1G039260.2	eIF(iso)4E
		HordeumVulgareC	HORVU4Hr1G061500.4	eIF4E related
Lotus japonicus	Birdsfoot trefoil	LotusJaponicusA	Lj6g0019784.1	eIF4E
		LotusJaponicusB	Lj2g0024720.1	eIF4E Isoform B
Macadamia integrifolia	Macadamia	MacadamiaIntegrifoliaA	XM_042623963.1	eIF4E-1-like
		MacadamiaIntegrifoliaB	XM_042624013.1	eIF4E-like
		MacadamiaIntegrifoliaC	XM_042657589.1	nCBP-like
		MacadamiaIntegrifoliaD	XM_042660014.1	nCBP
Marchantia polymorpha	Common liverwort	MarchantiaPolymorphaA	Mapoly0005s0006.1	eIF(iso)4E
		MarchantiaPolymorphaB	Mapoly0034s0012.1	eIF4E Isoform B
Medicago truncatula	Barrelclover	MedicagoTruncatulaA	Medtr2g018260.1	eIF4E-like
		MedicagoTruncatulaB	Medtr1g045130.1	eIF(iso)4E
		MedicagoTruncatulaC	Medtr6g029230.1	eIF4E Isoform B
		MedicagoTruncatulaD	XM_003618903.4	nCBP
Musa acuminata	Banana	MusaAcuminataA	GSMUA_Achr3T02580_001	eIF4E-related

Musa acuminata	Banana	MusaAcuminataB	GSMUA_Achr4T12820_001	eIF4E-like
		MusaAcuminataC	GSMUA_Achr5T12940_001	eIF(iso)4E
		MusaAcuminataD	XM_018823025.1	nCBP
Nelumbo nucifera	Sacred lotus	NelumboNuciferaA	XM_010274360.2	eIF4E-1-like
		NelumboNuciferaB	XM_010264130.2	eIF4E-like
		NelumboNuciferaC	XM_010256118.2	nCBP-like
		NelumboNuciferaD	XM_010259333.2	nCBP-like
		NelumboNuciferaE	XM_010256118.2	nCBP-like
Nicotiana benthamiana	Tobacco	NicotianaBenthamianaA	DQ393833.1	eIF4E
		NicotianaBenthamianaB	KC625578.1	eIF(iso)4E
Nicotiana tabaccum	Tobacco	NicotianaTabacumA	NM_001326163.1	eIF4E-1
		NicotianaTabacumB	KC625591.1	eIF4E-like
		NicotianaTabacumC	MN896999.1	eIF4E-2
		NicotianaTabacumD	XM_016592736.1	eIF(iso)4E
		NicotianaTabacumE	XM_016592736.1	nCBP
Oryza sativa	Rice	OryzaSativaA	Os01g73880.1	eIF4E
		OryzaSativaB	Os10g32970.1	eIF4E-related
		OryzaSativaC	Os03g15590.1	eIF4E Isoform B
		OryzaSativaD	XM_015776024.2	nCBP
Phoenix dactylifera	Date palm	PhoenixDactyliferaA	XM_008777059.3	eIF4E-1-like
		PhoenixDactyliferaB	XM_008803026.4	eIF4E-like
		PhoenixDactyliferaC	XM_039116335.1	nCBP-like
Physcomitrium patens	Earth moss	PhyscomitriumPatensA	Pp3c11_1600V3.1	eIF4E-like

Physcomitrium patens	Earth moss	PhyscomitriumPatensB	Pp3c2_32310V3.1	eIF(iso)4E
		PhyscomitriumPatensC	Pp3c1_5840V3.1	eIF(iso)4E
		PhyscomitriumPatensD	Pp3c26_10900V3.1	nCBP-like 1
		PhyscomitriumPatensE	Pp3c4_24310V3.1	nCBP-like 2
Populus trichocarpa	Cottonwood	PopulusTrichocarpaA	Potri.011G077200.1	eIF4E-like
		PopulusTrichocarpaB	Potri.010G066700.1	eIF(iso)4E
		PopulusTrichocarpaC	Potri.008G171100.1	eIF(iso)4E-like
		PopulusTrichocarpaD	Potri.013G057000.1	nCBP
Porphyra umbilicalis	Purple laver (seaweed)	PorphyraUmbilicalisA	Pum2109s0002.1	eIF4E
		PorphyraUmbilicalisB	Pum2061s0001.1	eIF(iso)4E
Selaginella moellendorffii	Spikemoss	SelaginellaMoellendorffiiA	XM_002985418.2	nCBP
		SelaginellaMoellendorffiiB	XM_002985418.1	Unannotated mRNA
		SelaginellaMoellendorffiiC	XM_024688741.1	nCBP
Solanum lycopersicum	Tomato	SolanumLycopersicumA	NM_001247530.2	eIF4E-1
		SolanumLycopersicumB	NM_001320649.1	eIF4E-2
		SolanumLycopersicumC	NM_001247843.2	eIF(iso)4E
		SolanumLycopersicumD	XM_004249251.4	nCBP
Solanum tuberosum	Potato	SolanumTuberosumA	NM_001288431.1	eIF4E-1
		SolanumTuberosumB	MT828875.1	eIF4E-2
		SolanumTuberosumC	MT828876.1	eIF(iso)4E
		SolanumTuberosumD	XM_006351298.2	nCBP
Sphagnum fallax	Bog moss	SphagnumFallaxA	Sphfalx10G029900.1	eIF4E
		SphagnumFallaxB	Sphfalx04G055600.1	eIF4E

Sphagnum fallax	Bog moss	SphagnumFallaxC	Sphfalx11G023700.1	eIF4E
		SphagnumFallaxD	Sphfalx18G101700.1	eIF4E isoform B
		SphagnumFallaxE	Sphfalx06G003400.1	eIF4E isoform B
Volvox carteri	Colonial green algae	VolvoxCarteri	Vocar.0035s0007.1	eIF4E
Zea mays	Maize	ZeaMaysA	Zm00001d041682_T001	eIF4E
		ZeaMaysB	Zm00001d032775_T002	eIF(iso)4E-related
		ZeaMaysC	Zm00001d014065_T001	eIF(iso)4E
		ZeaMaysD	NM_001138869.1	nCBP



Chapter 4: Establishing the foundations for non-transgenic CRISPR/Cas9 genome editing of *Solanum tuberosum*

4.1 Introduction

Potato is the most consumed non-cereal crop globally (Zhang et al., 2017) and is a staple food source across Southern Africa. Potato pathogens greatly threaten food security as infection can significantly reduce tuber yield, and symptomatic tubers often are not fit for consumption (Dahal et al., 2019). Potato Virus Y (PVY) and Potato Leaf Roll Virus (PLRV) are two prevalent RNA viruses that pose a threat to the outputs of potato farming globally (Loebenstein and Gaba, 2012). Both diseases can be spread vertically which is a significant problem in potatoes which oftentimes use seed potatoes to establish a new crop (Lacomme and Jacquot, 2017). New infection can also arise in the plants via contact with aphid vectors and previous pest-management strategies are becoming inefficient as new pathogen and vector species emerge (Sridhar et al., 2022). Some Solanaceous species demonstrate natural resistance to these viruses (Ruffel et al., 2005; Charron et al., 2008), but a naturally occurring allele leading to resistance has yet to be identified in potato. New plant-breeding techniques, such as CRISPR/Cas9 - which edit a plants genome in a precise manner - provide a solution to these problems and have the potential to be used to create virus resistant plants without the use of transgenic technology (Schaart et al., 2016; Robertson et al., 2022).

Due to the limitations of the size of a viral genome, viruses rely on host machinery for essential functions such as genome transcription and translation, and protein assembly. Host genes that encode proteins that perform these functions are known as susceptibility genes and are often targets for engineering viral resistance in plants (Robaglia and Caranta, 2006). The eukaryotic translation initiation factor 4E (eIF4E) is a host protein that directly associates with the viral genome-linked protein (VPg) of some RNA viruses to facilitate translation of the viral genome. Because of the importance of this association for viral replication, the *eIF4E* gene is a common target for resistance-engineering experiments (Wang and Krishnaswamy, 2012).

In potato there are three eIF4E isoforms: *eIF4E-1*, *eIF4E-2* and *eIF(iso)4E*. Gene knock-down of these isoforms have resulted in improved resistance to Potato Virus Y (Miroshichenko et al., 2020), however complete resistance has not yet been achieved. Functional redundancy between these three isoforms means that all can be utilised by the virus during infection and mutations in all alleles of all three genes is needed to confer complete resistance to the virus. As these proteins still perform an essential function for potato translation, the knockout of all three would most likely be lethal.

Knockout of *eIF4E-1* in potato was recently achieved through CRISPR/Cas9 editing (Lucioli et al., 2022; Noureen et al. 2022). As expected, inoculation studies showed improved, but not complete, resistance to PVY. Both studies used a transgenic CRISPR/Cas9 system for plant transformation when creating the

knockouts. The CRISPR/Cas9 complex can be introduced to a plant cell in a non-transgenic way through PEG-mediated transformation of a preassembled ribonucleoprotein into protoplasts (Woo et al., 2015). The regeneration of mature plants from potato protoplasts has been achieved by several groups (Tavazza and Ancora, 1986; Nicolia et al., 2015) however, the procedure is highly complex and requires specialised tissue culture technique. Because of this complexity, this tends to be the limiting step for establishment of this procedure. This chapter describes efforts to establish protoplast isolation and regeneration and to design guide RNAs to target the potato *eIF4E* gene family

4.2 Materials and methods

4.2.1 Suppliers and DNA sequencing

The chemical reagents used in this project were supplied by Sigma Aldrich (Missouri, USA), Bio-Rad (California, USA) and Merck Millipore (Massachusetts, USA). Enzymes used in tissue culture work were from Duchefa Biochemie (Haarlem, The Netherlands). Kits and enzymes for molecular biological procedures were obtained from New England Biolabs (NEB; Massachusetts, USA), Promega Corp (Wisconsin, USA), Invitrogen (Massachusetts, USA), Zymo Research (California, USA), and Thermo Fischer Scientific (Massachusetts, USA). Primers and oligonucleotides were synthesized by Inqaba Biotech (Pretoria, South Africa). DNA sequencing was conducted by The Central Analytical Facilities (CAF) at Stellenbosch University.

4.2.3 Primer design for the amplification of gene-of-interest fragments for sequencing

Primers to amplify regions of *eIF4E-1*, *eIF4E-2* and *eIF(iso)4E* from *Solanum tuberosum* cv Désirée genomic DNA (gDNA) were designed using their respective accessions obtained from the National Centre for Biotechnology Information (NM_001288431.1, NM_001288408.1, and NM_001288204.1) using the program PrimerBLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). Two sets of primers were designed for *eIF4E-2*: the first spanning a region within exon 1 and a second spanning across exons 2 and 3. The resulting amplicons obtained through polymerase chain reaction (PCR) are referred to as eIF4E-2_A and eIF4E-2_B respectively. Primers to amplify a region within the *Phytoene Desaturase* (*PDS*) gene were taken from Bánfalvi et al. (2020; Table 4.1)

Table 4.1 Primer sequences used to amplify genomic DNA fragments from *Solanum tuberosum.*

Gene	Primer Name	Sequence (5' – 3')	Amplicon length (bp)	Annealing temperature (T_a; °C)
eIF4E-1	eIF4E-1_F	TGAACCATGTCGCTTATTCTGC	477	67
	eIF4E-1_R	AAAATGACAAACGTGGCCGT		
eIF4E-2	eIF4E-2_A_F	ACTGAAATCCATAACCAACCCCA	513	67
	eIF4E-2_A_R	TGAGGAGTGACGTTTACAGCA		
	eIF4E-2_ B_F	TGGAATTAGTTGAGGTGCGG	702	67
	eIF4E-2_ B_R	CGTATGTATCCCATTTAGTGACCCA		
eIF(iso)4E	eIF(iso)4E_F	AGGTGGGTAGAAGCTTTTCTTAGAT	542	66
	eIF(iso)4E_R	CACATTTCGGAGATAGTGGAACA		
PDS	PDS_F	TTTCCCCGAAGCTTTACCCG	532	69
	PDS_R	ATCTGTCACCCTATCCGGCA		

4.2.4 Genomic DNA extraction

Genomic DNA was extracted from 150 mg of potato leaf tissue using the *Quick*-DNA™ Plant/Seed Miniprep Kit (Zymo Research) according to the manufacturer's instructions.

4.2.5 Polymerase chain reaction amplification of gene fragments and ligation into pJET1.2/blunt

Portions of the *eIF4E-1*, *eIF4E-2, eIF(iso)4E* and *PDS* genes were amplified from gDNA using polymerase chain reactions with Q5® High-Fidelity DNA polymerase (NEB). An individual master-mix was created for each gene. The master-mixes contained: 1X Q5® reaction buffer, 0.2 mM per dNTP, 2 µM of each respective primer, 1% Q5® DNA polymerase, 500 ng of template DNA and ddH_2O up to 50 µL per reaction. The reactions proceeded in a thermocycler set as follows: initial denaturation at 98 °C for 30 sec, followed by 30 cycles of 98 °C for 10 sec; T_a °C for 30 sec and 72 °C for 25 sec, with final extension for 2 min at 72 °C. T_a values were calculated using the NEB T_m Calculator (https://tmcalculator.neb.com /#!/main; Table 3.1).

4.2.6 Agarose gel electrophoresis separation of DNA

PCR amplification was confirmed through separation of the amplicons on a 1% (w/v) agarose-Tris borate EDTA gel containing 0.005% (v/v) ethidium bromide with visualisation under ultraviolet light using the GelDoc imaging system (ThermoFischer Scientific). The GeneRuler Express DNA ladder from ThermoFischer Scientific was included in each gel electrophoresis as a molecular weight marker.

4.2.7 Purification of polymerase chain reaction amplicons

Amplicons were purified from the PCR reaction mixtures using the Wizard® SV Gel and PCR Clean-Up System (Promega Corp.) according to the manufacturer's instructions. These fragments were ligated into pJET1.2/blunt (ThermoFischer Scientific) according to the manufacturer's protocol.

4.2.8 Media preparation for bacterial transformation and culture

The following media was prepared and autoclaved for bacterial transformation. Lysogeny broth agar (LBA; 10 g L^{-1} tryptone, 5 g L^{-1} yeast extract, 10 g L^{-1} NaCl and 15 g L^{-1} bacteriological agar), liquid lysogeny broth (LB; 10 g L^{-1} tryptone, 5 g L^{-1} yeast extract, 10 g L^{-1} NaCl) and terrific broth (TB; 11.8 g L^{-1} tryptone, 23.6 g L^{-1} yeast extract, 9.4 g L^{-1} K_2HPO_4, 2.2 g L^{-1} KH_2PO_4, 4 mL L^{-1} glycerol).

4.2.9 Preparation of chemically competent *Escherichia coli* cells

Competent *Escherichia coli* (*E. coli*) DH5α cells were prepared using an adapted version of the method outlined in Inoue et al. (1990). DH5α cells were grown overnight on LBA plates at 37 °C. A single colony was used to inoculate 3 mL of LB which was grown overnight at 37 °C with shaking at 200 rpm. This culture was used to inoculate 200 mL of LB which was grown in the same conditions until an OD_{600} of 0.5 was reached. The culture was incubated on ice for 10 min, before harvesting the cells by centrifugation at 1430 x g for 10 min at 4 °C. The resulting pellet as re-suspended in 20 mL of ice-cold TB and incubated on ice for 10 min. The samples were then centrifuged at 1430 x g for 10 min at 4 °C and the pellet re-suspended in 10 mL ice-cold TB containing 70 µL L^{-1} dimethyl sulfoxide. Aliquots were frozen in liquid nitrogen.

4.2.10 Transformation of plasmids into *Escherichia coli*

The pJET-gene constructs were separately transformed into competent *E. coli* DH5α cells using the standard heat-shock transformation method. Five µL of each plasmid was added to 50 µL of thawed competent cells and incubated on ice for 30 min. Samples were heat shocked at 42 °C for 45 sec before being placed back on ice for 2 min. To each reaction 250 µL of LB media was added and cells were incubated for 30 min at 37 °C, shaking at 200 rpm. Fifty µL of each transformation was plated onto selective LBA containing 100 µg mL^{-1} ampicillin and grown overnight at 37 °C.

To confirm transformation, colony PCR using the pJET1.2/blunt forward sequencing primer (5'-CGACTCACTATAGGGAGAGCGGC-3') and the gene reverse primers (Table 3.1) was conducted using GoTaq® DNA Polymerase (Promega Corp). An individual master-mix was created for each gene fragment. The master-mixes contained 1X Green GoTaq® reaction buffer (containing 7.5 mM $MgCl_2$), 0.3 mM dNTP mix, 0.3 µM of each respective primer, 0.4% GoTaq® DNA polymerase and ddH_2O up to 50 µL reaction. A single bacterial colony was used for each reaction. The reactions proceeded in a thermocycler set as follows: initial denaturation at 95 °C for 3 min, followed by 30 cycles of 95 °C for 30 sec; T_a °C for 30 sec and 72 °C for 30 sec, with final extension at 72 °C for 5 min. T_a values for these reactions were calculated using the Promega Biomath Calculator (https://worldwide.promega.com/ resources/tools/ biomath/) and were 65.5 °C, 63.5 °C , 63 °C ,65 °C, and 66 °C for *eIF4-1*, *eIF4E-2*_A, *eIF4E-2*_B, *eIF(iso)4E* and *PDS* respectively. Amplicons were separated via gel electrophoresis as before.

4.2.11 Plasmid isolation and sequencing

Bacterial colonies containing the pJET-gene vector were used to inoculate 5 mL LB media and grown overnight at 37 °C, shaking at 200 rpm. Plasmids were subsequently isolated using the Wizard® Plus SV Minipreps DNA Purification System (Promega Corp., USA) according to the manufacturer's instructions and sent for sequencing.

4.2.12 Single guide RNA design

DNA sequences of the gene-of-interest PCR amplicons were imported into CLC Sequence Viewer (Qiagen, Hilden, Germany) and aligned to the respective coding sequence (CDS) accessions obtained from the NCBI (Accession numbers: JN831440.1 (*eIF4E-1*), JN564590.1 (*eIF4E-2*) and FN666437.1 (*eIF(iso)*4E). The *PDS* fragment sequence was aligned to the *PDS* sequence published in Bánfalvi et al. (2020; NCBI accession number: LOC102577582). Alignments were assessed for similarity.

Single guide RNAs (sgRNAs) for *eIF4E-1*, *eIF4E-2* and *eIF(iso)4E* were designed using the platforms CRISPR RGEN (Park et al., 2015), CRISPR-P (Lei et al., 2014), and CRISPOR (Concordet and Haeussler, 2018). Potential off-target activity of each guide was assessed with up to three nucleotide mismatches of the target using Cas-OFFinder within the CRISPR RGEN program (Bae et al., 2014).

4.2.13 pGEM-scaffold-sgRNA construction

A pair of complementary oligonucleotides were designed for each sgRNA with the addition of a 5'-CGTC-3' and a 5'-AAAC-3' sequence at the start of the forward and reverse oligonucleotides, respectively, to allow for unidirectional annealing into the pGEM-scaffold plasmid after digestion with *BbsI* (Figure 4.1 A). The oligonucleotide pairs were annealed in a 50 µL reaction consisting of 3 µM ATP, 3 µM CTP, 3 µM GTP, 3 µM TTP and 5 µL of the NEB 2.1 buffer. Reactions were incubated at 95 °C for 4 min and then placed in 500 mL of water at 70 °C which was left at room temperature until cool. A pGEM®-T Easy plasmid, already containing a sgRNA scaffold sequence downstream of the T7 initiation site (pGEM-scaffold; Dijkerman, 2021) was a kind gift of Mr. Alex Dijkerman (Department of Genetics, Stellenbosch University). This plasmid was linearized though digestion with *BbsI* and each oligonucleotide pair was individually ligated into the vector using T4 DNA ligase following the manufacturer's instructions (Promega Corp.; Figure 4.1 B). The ligated plasmid was transformed into DH5α *E. coli* using the heat-shock transformation method described above. *E. coli* were plated onto LBA plates containing 100 µg mL^{-1} ampicillin and grown overnight at 37 °C.

The pGEM-scaffold-sgRNA plasmids were isolated using the Wizard® Plus SV Miniprep DNA Purification System (Promega Corp.) and sequenced using an SP6 primer.

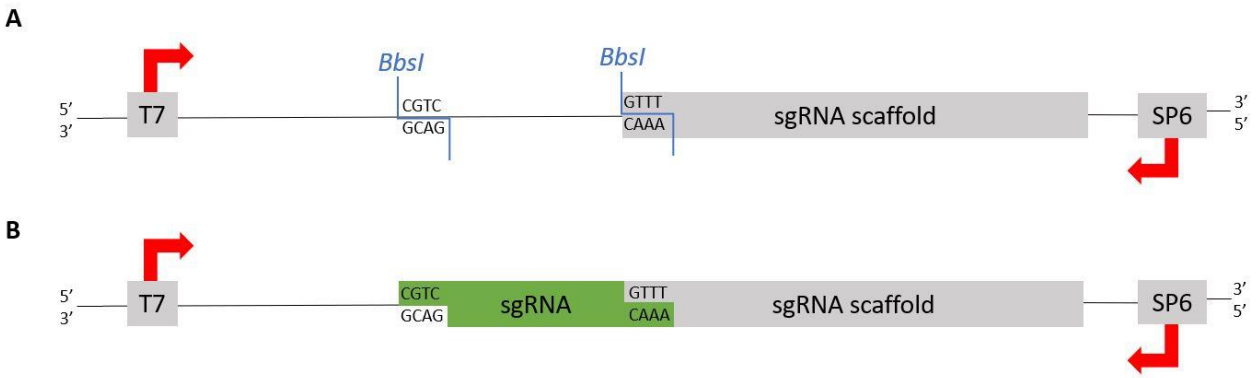


Figure 4.1 Diagrammatic representation of the cloning of the synthesised single guide RNA oligonucleotides into pGEM-scaffold. **(A)** The pGEM-scaffold plasmid was cut with *BbsI,* leaving unique overhangs which were exploited for **(B)** the unidirectional insertion of the annealed single guide RNA oligonucleotides.

4.2.14 Single guide RNA transcription and *in vitro* analysis

To transcribe the sgRNAs *in vitro* the pGEM-scaffold-sgRNA plasmids were linearized by digestion with *EcoRI* and purified using the Zymoclean Gel DNA Recovery Kit (Zymo Research) according to the manufacturer's instructions. A T7 RNA transcription reaction was conducted using the HiScribe™ T7 High Yield RNA Synthesis Kit (NEB). The following reaction was set up: 1 µg linearized plasmid template, 1.5 µL 10X reaction buffer, 1.5 µL ATP (100 mM), 1.5 µL UTP (100 mM), 1.5 µL GTP (100 mM), 1.5 CTP (100 mM) and 1.5 µL T7 RNA polymerase mix. The reaction volume was made to 20 µL with water. After incubation overnight at 37 °C the reactions were treated with 5µL DNase and the sgRNA transcripts were purified using the RNA Clean and Concentrator kit (Zymo Research) according to the manufacturer's instructions. Purified sgRNAs were separated via agarose gel electrophoresis as previously described to confirm transcription.

In vitro assays were conducted for each sgRNA to confirm the *in silico* prediction that the guides would complex with Cas9 and cleave the gene targets without non-specific endonuclease activity. Reactions were set up as follows: 1 µg Cas9 protein (Invitrogen TrueCut™ v2), 1 µg sgRNA, 2 µL Cas9 reaction buffer (0.2 M HEPES pH 7.5, 0.1 M MgCl, 5 mM dichlorodiphenyltrichloroethane, 1.5 M KCl), 250 ng purified PCR amplicon, with volume set to 20 µL with water. Three negative controls were included, the first without sgRNA, the second without sgRNA and Cas9, and the third with a non-target DNA fragment. Tubes were incubated for 1 hour and 10 min at 37 °C followed by denaturation of the enzyme at 65 °C for 10 min. Reactions were separated via agarose gel electrophoresis as previously described.

4.2.15 Plant tissue culture

Wild type *Solanum tuberosum* cv. Désirée plants were propagated from pre-existing tissue culture stocks in plastic Magenta tissue culture vessels containing approximately 65 mL of medium A (2.2% (w/v) MS basal medium without vitamins, 4.4 mM sucrose, 0.8% (w/v) agar, pH 5.6). Plantlets were placed in a growth room

at 23 °C under a 16/8-hour light/dark photoperiod (50 µmol photons min^2 sec^{-1} of photosynthetically active radiation) under cool white, fluorescent lights (Osram, L 58V/740).

4.2.16 Protoplast isolation

Protoplasts were isolated, transformed and regenerated according to the protocol described in Nicolia et al. (2015), with some minor modifications. Thirty of the newest leaves were excised from 4–6-week-old tissue culture grown plants and placed abaxial side down in Petri dishes containing 20 mL medium B (Addendum 1). The dishes were incubated in the dark at 4 °C for 24 hours. Leaves were cut into 1-2 mm wide slices with a scalpel and medium B was replaced with 25 mL of plasmolysis solution (0.5 M sorbitol) before being incubated in the dark at room temperature (RT) for 30 min. The plasmolysis solution was then replaced with 30 mL of cell wall digest solution (medium C; Addendum 1). Plates were incubated in the dark at RT for 14 hours on a Belly Dancer™ Orbital Platform Shaker (IBI Scientific™) at the slowest setting. Following this, the solution was sieved through a 100 µM filter, which had been pre-wetted with 5 mL of wash solution (Addendum 1). The remaining protoplasts were rinsed from the Petri dish with 20 mL wash solution and filtered. This solution was centrifuged in a swinging bucket rotor (Allegra™ X-22R Centrifuge SX 4250) at 50 x g for 5 min with minimum acceleration and deceleration. The supernatant was removed, and the pellet re-suspended in 4 mL wash solution. This solution was layered on top of 6 mL of ice-cold 21% (w/v) sucrose solution. Tubes were centrifuged at 50 x g for 15 min as above. Viable protoplasts at the solutions' interface were transferred into 3 mL of transformation buffer 1 (0.5 M mannitol, 15 mM $MgCl_2 \cdot 6H_2O$, 0.5% (w/v) MES) for counting.

4.2.17 Assessment of protoplast density

Protoplast density (number of viable protoplasts mL^{-1}) was quantified by counting numbers in 10 µL aliquots of the protoplast solution using a haemocytometer under a light microscope.

4.2.18 Suspension of protoplasts in alginate lenses for regeneration

Protoplasts were harvested by centrifugation at 50 x g for 10 min. The resulting pellet was re-suspended in medium E (Addendum 1) to a concentration of approximately 1.6 x 10^6 protoplasts mL^{-1}. The cell suspension was mixed with an equal volume of sodium alginate solution (0.4 M sorbitol, 2.8% (w/v) sodium alginate) and as added in 500 µL drops onto Petri dishes containing 10 mL of solid setting agar (0.4 M sorbitol, 50 mM $CaCl_2 \cdot 2H_2O$, 8 g L^{-1} agar) and left at RT for 2 hours. Three mL of floating solution (0.4 M sorbitol, 50 mM $CaCl_2 \cdot 2H_2O$) was added to each plate to release the solid alginate lenses from the surface of the solid setting agar. The lenses were transferred to Petri dishes containing 20 mL of medium E and placed in a dark growth room at 25 °C for seven days.

4.2.19 Micro-callus generation

After seven days in the dark, the plates were wrapped in white paper and moved into a light growth room (16/8-hour light/dark photoperiod (50 µmol photons min^{-2} sec^{-1} of photosynthetically active radiation) under cool white, fluorescent lights (Osram, L 58V/740)). Plates were exposed to full light after three weeks. Fresh medium E was provided every two weeks.

4.3 Results

4.3.1 Amplification of fragments from *eIF4E* genes from potato genomic DNA

To examine if the target gDNA sequences from the plants used in these experiments are identical to published sequence data, amplicons containing the putative targets were generated by PCR. Each of these fragments were ligated individually into the pJet1.2/blunt plasmid. Ligation was confirmed by colony PCR before the plasmids were sent for sequencing (Figure 4.2). The sequences of these amplicons were compared to the sequence found at corresponding NCBI accessions and were found to be identical.

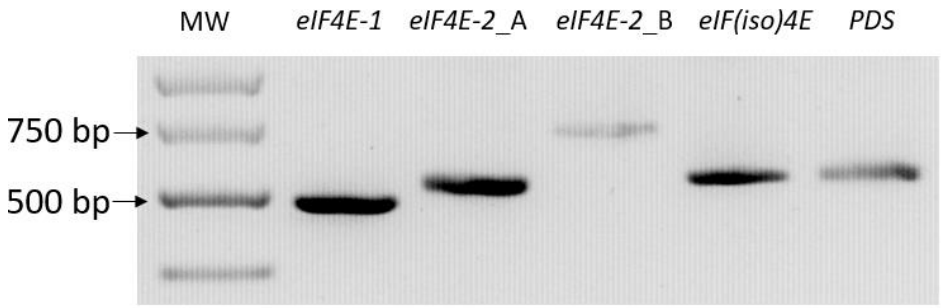


Figure 4.2 Purified polymerase chain reaction amplicons from the *eIF4E* and *PDS* genes separated on a 1% (w/v) agarose-Tris Borate EDTA gel.

4.3.2 Single guide RNA design

Potential sgRNAs were identified for each of the *eIF* gene targets using three different sgRNA prediction programs (Table 4.2). The gene sequences previously obtained from CAF for the genes of interest were used as the query sequence. CRISPR-P failed to identify the *eIF4E-1* and *eIF4E-2* sequences and thus no guides were identified by this program for those genes. Each of the three programs reported quality scores for all potential guides. CRISPR-RGEN recommended a minimum quality score of 66% for the guides to be considered usable while no guidance regarding interpretation of the quality scores were given by the other two programs. All guides that had a quality score below this threshold on CRISPR-RGEN were excluded.

Table 4.2 Summary of all results obtained from three single guide RNA prediction programs for the *eIF4E* gene family.

Program	Gene target	Total number of guides identified	Quality score range (%)
CRISPR-RGEN	*eIF4E-1*	27	58.0 – 76.5
	eIF4E-2	68	48.2 – 84.0
	eIF(iso)4E	24	54.1 – 74.1
CRISPR P	*eIF4E-1*	0	-
	eIF4E-2	0	-
	eIF(iso)4E	24	1.86 – 61.5
CRISPOR	*eIF4E-1*	27	18.0 – 71.0
	eIF4E-2	68	17.0 – 80.0
	eIF(iso)4E	24	14.0 – 82.0

Guides were screened for off-target matches in the *S. tuberosum* genome. Any guide that had potential to direct Cas9 to a non-target DNA site was excluded. Of the remaining guides, two per gene with the highest quality scores across all three programs were chosen for *in vitro* transcription and analysis (Table 4.3).

Table 4.3 Single guide RNAs designed to target *eIF4E-1, eIF4E-2, eIF4E-1* + *eIF4E-2,* and *eIF(iso)*4E and their quality control scores (%) from the programs CRISPR-RGEN, CRISPR P and CRISPOR.

Gene target	sgRNA name	sgRNA sequence (5’ – 3’)	Quality score (%)		
			CRISPR-RGEN	CRISPR P	CRISPOR
eIF4E-1	eIF4E-1-A	AGTCGTTAGTGTCCGGGCTA	69.8	-	46.0
	eIF4E-1-B	GTGGAGCAGTCGTTAGTGTC	62.9	-	41.0
eIF4E-2	eIF4E-2-A	AAAAAATAGCTCTGTGGACC	69.2	-	22.0
	eIF4E-2-B	ATGATACGGCCTCGTCTTTA	82.3	-	57.0
eIF(iso)4E	eIF(iso)4E-A	GACTGCTACGAGCAGCCGAA	71.9	40.9	60.0
	eIF(iso)4E-B	GTCAGAAGATATATGTGGAG	72.5	31.6	57.0
eIF4E-1 + eIF4E-2	eIF4E-1+2	GATCCTGTATGTGCCAATGG	70.3	-	65.0

The selected guides were aligned to the respective gene sequences obtained from CAF and the CDS obtained from the NCBI. All 7 guides showed 100% identity to both sequences. Bánfalvi et al. (2020) have previously successfully knocked out the *PDS* gene in potatoes using CRISPR-Cas9 and it was decided to use this gene as a potential positive control.

Oligonucleotide pairs for each guide RNA sequence were synthesised and cloned into the pGEM-scaffold plasmid after linearisation with *BbsI*. The pGEM-scaffold-sgRNA plasmids were sequenced to confirm insertion of the sgRNA sequence next to the scaffold sequence and T7 transcription start site. For all gene targets, both sgRNA sequences were correctly ligated into the vectors (Figure 4.3).

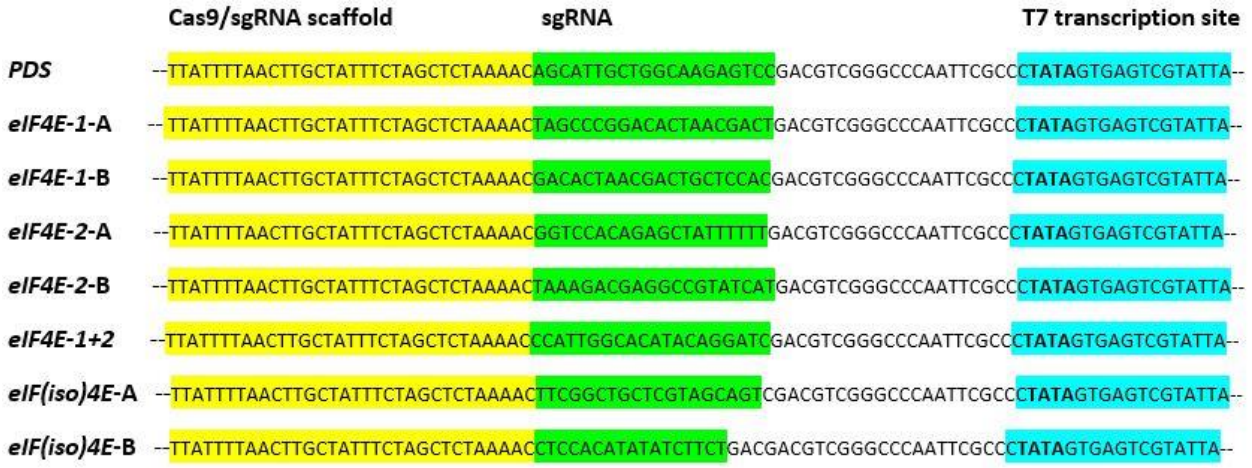


Figure 4.3 Nucleotide sequences obtained from the Central Analytics Facility of the pGEM-scaffold-sgRNA plasmid. Cas9/sgRNA scaffold sequences, single guide RNA sequences and T7 transcription sites are highlighted in yellow, green, and blue respectively.

4.3.3 Single guide RNA transcription and *in vitro* assays

To assess the accuracy of the *in silico* analysis in identifying sgRNAs that will complex with Cas9 and cleave the target site, an *in vitro* cleavage experiment was performed. For this sgRNAs were transcribed *in vitro* and complexed individually with Cas9. The resulting RNPs were incubated with PCR amplicons containing the target sequences, which were obtained from the pJET-gene vectors. The expected sizes for each amplicon before and after cleavage are shown in Table 4.4.

Table 4.4 Expected DNA fragment sizes after cleavage by Cas9 for each DNA target and single guide RNA pair.

Gene target	**Guide**	**Expected DNA fragment sizes after Cas9 cleavage (bp)**	
eIF4E-1	eIF4E-1+2	Uncut	477
		Cut	348 and 129
	eIF4E-1-A	Uncut	477
		Cut	373 and 104
	eIF4E-1-B	Uncut	477
		Cut	366 and 111
eIF4E-2 (exon 2/3)	eIF4E-1+2	Uncut	702
		Cut	504 and 198
	eIF4E-2-A	Uncut	702
		Cut	468 and 234
eIF4E-2 (exon 1)	eIF4E-2-B	Uncut	513
		Cut	344 and 169
eIF(iso)4E	eIF(iso)4E-A	Uncut	542
		Cut	367 and 175
	eIF(iso)4E-A	Uncut	542
		Cut	388 and 154
PDS	Bánfalvi _2020-PDS	Uncut	535
		Cut	286 and 249

Each sgRNA was tested against DNA containing its target sequence as well as a non-target DNA fragment to assess whether non-specific cleavage occurred. The non-target DNA fragments were the *eIF4E* isoform that showed the most similar sequence to the target site and, therefore, would be most likely to be erroneously cleaved. *eIF4E-1* and *eIF4E-2* were screened against each other. While *eIF(iso)4E* was screened against both *eIF4E-1* and *eIF4E-2.* The guide eIF4E-1+2, which is designed to target both *eIF4E-1* and *eIF4E-2* was screened against *eIF(iso)4E.* For *PDS*, *eIF4E-1* was used as a random non-target control. Controls without the sgRNA complexed to the Cas9 were also conducted.

After incubation DNA fragments were separated via gel electrophoresis (Figure 4.4). All the sgRNAs designed successfully cleaved the target DNA and no cleavage was observed in the non-target DNA control samples. Two other negative control reactions – the first without the sgRNA, and second without the sgRNA and Cas9 showed no cleavage. The smaller of the cleaved fragments is not visible as it is masked by the sgRNA fragments which lie at the bottom of the gels.

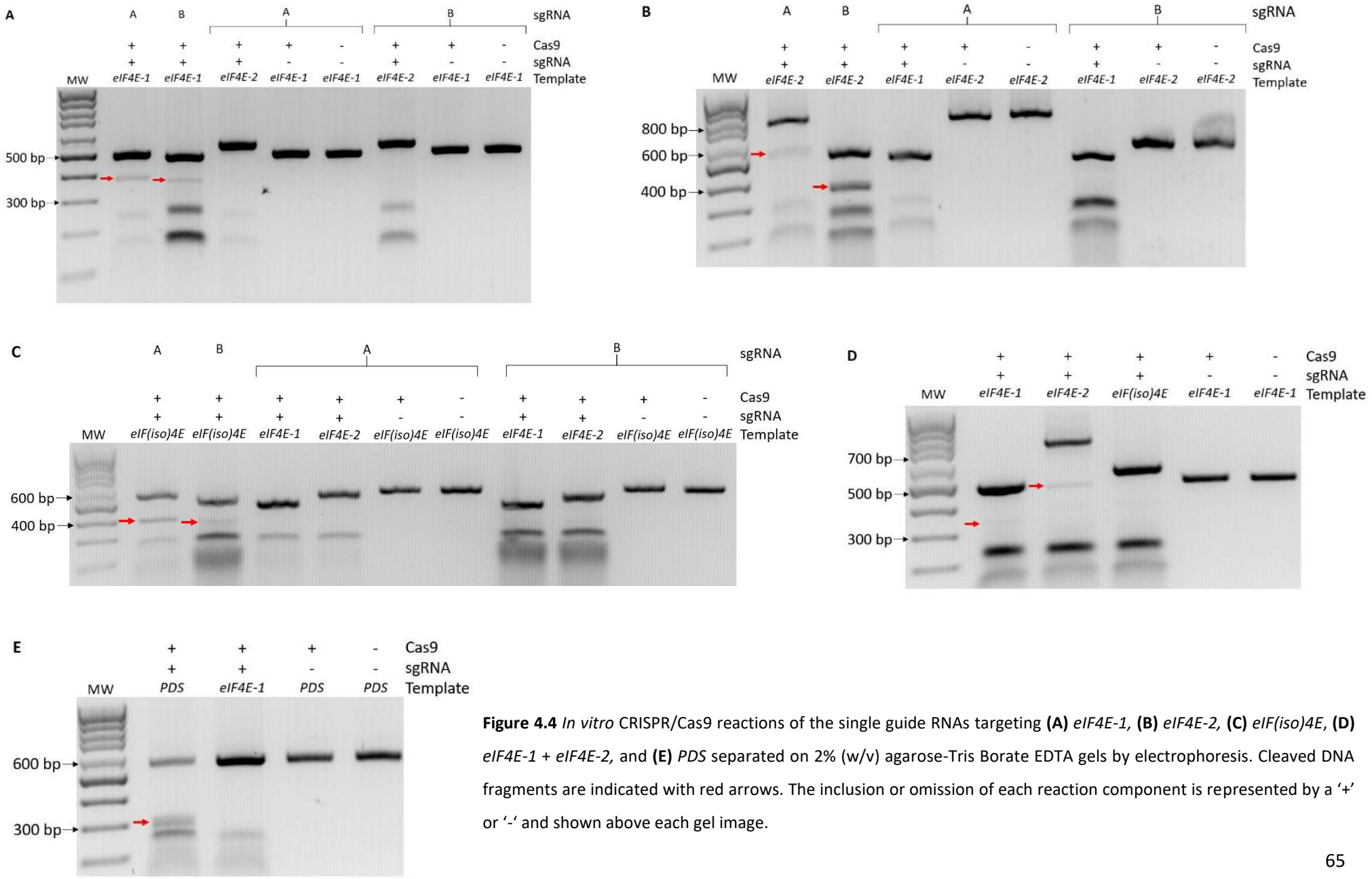


Figure 4.4 *In vitro* CRISPR/Cas9 reactions of the single guide RNAs targeting **(A)** *eIF4E-1,* **(B)** *eIF4E-2,* **(C)** *eIF(iso)4E*, **(D)** *eIF4E-1* + *eIF4E-2,* and **(E)** *PDS* separated on 2% (w/v) agarose-Tris Borate EDTA gels by electrophoresis. Cleaved DNA fragments are indicated with red arrows. The inclusion or omission of each reaction component is represented by a '+' or '-' and shown above each gel image.

4.3.4 Protoplast isolation and micro-callus regeneration

Protoplasts were successfully isolated from the leaf material of *S. tuberosum* cv. Désirée plants propagated in tissue culture through enzymatic digestion. Viable protoplasts were separated from burst protoplasts and other cell debris through separation on a sucrose-cushion. The average yield of viable protoplasts ranged from 1.64 x 10^6 protoplasts g^{-1} leaf tissue to 2.31 x 10^6 protoplasts g^{-1} leaf tissue (Figure 4.5).

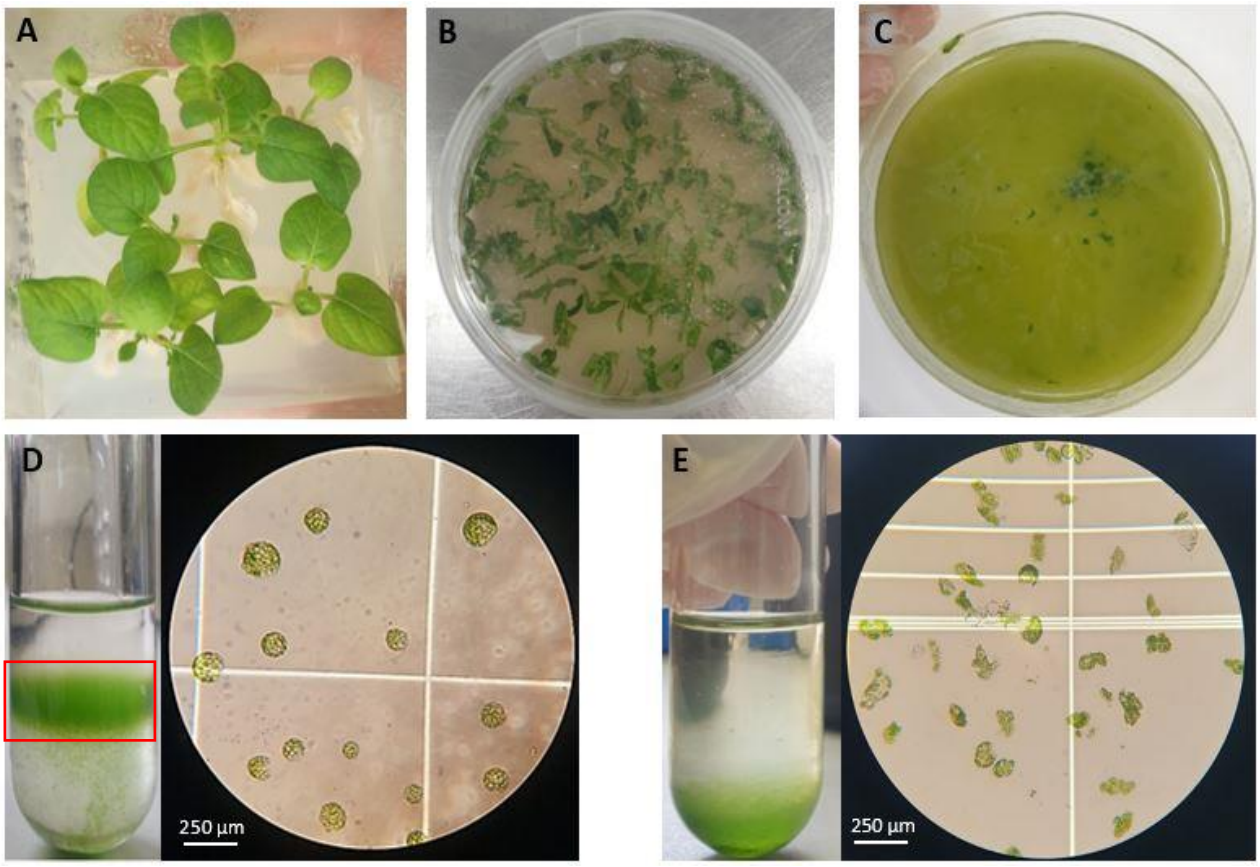


Figure 4.5 (A) Leaves of *Solanum tuberosum* plants from tissue culture grown plants were harvested for protoplast isolation. **(B)** Leaves were sliced into 1-2 mm strips prior to overnight enzymatic digestion. Protoplasts were harvested from the digested solution **(C)** by centrifugation on a sucrose-cushion. When successful, the sucrose-cushion allows separation of viable protoplasts from the cell debris and the cells collect at the solutions' interface, indicated with the red box **(D).** However, when conducted incorrectly, all protoplasts rupture due to the high osmotic potential of the solution **(E).**

Viable protoplasts were suspended in alginate lenses at a concentration of 1.6 x 10^6 protoplasts mL^{-1}. Once solidified, the lenses were transferred into growth medium where the protoplasts formed micro-calli which were visible within 3 weeks of cultivation (Figure 4.6).

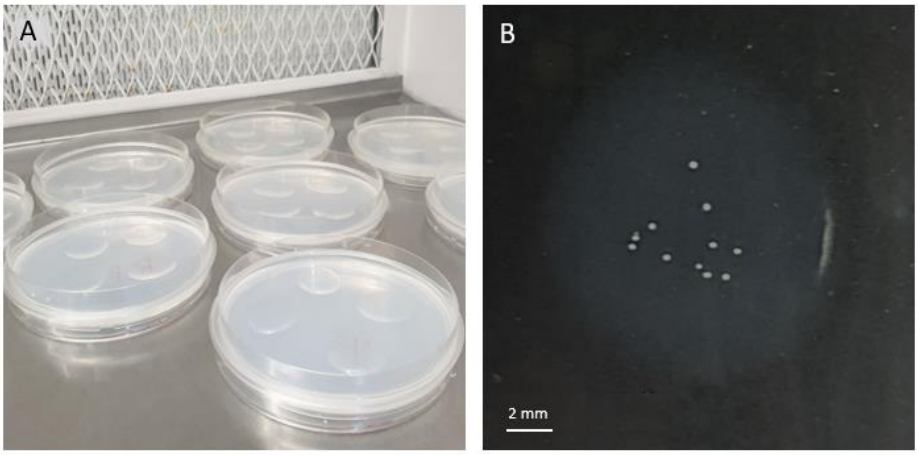


Figure 4.6 Viable protoplasts were quantified in a haemocytometer and suspended in alginate lenses **(A)**. The alginate lenses were left to solidify at room temperature until a clear convex shape was observed. **(B)** Micro-calli formation was visible in the alginate lenses after 3 weeks.

4.4 Discussion

To try and engineer resistance to RNA viruses through CRISPR/Cas9-induced disruption to the *eIF4E* genes, the location of the sgRNA target needs to be carefully selected. Repair of the resulting dsDNA break should introduce an indel mutation to the targeted *eIF4E* genes. Introduction of a premature stop-codon, a frame-shift mutation that alters the amino acid composition of the protein, or disruption to the amino acid sequence near the VPg binding location would all create an eIF4E protein that was unrecognisable to the VPg. To achieve this the sgRNA must target a site before or close to the amino acid residues that are involved in VPg biding. In eIF4E-1 three amino acids (R157, K159, and K162) have been demonstrated to be directly involved in the binding of the PVY VPg during infection (Coutinho de Oliveira et al., 2019). The R157 and K162 residues are conserved across all three eIF4E isoforms in potato, however, the lysine residue in position 159 in eIF4E-1 is substituted by an arginine in eIF4E-2, and a serine in eIF(iso)4E. The structure of the genes encoding the three potato *eIF4E* isoforms are similar. Each contains 5 exons and 4 introns, with the introns between exon 1 and 2, and exon 3 and 4 being the largest in all three genes (Figure 4.7). The CDS of potato *eIF4E-1*, *eIF4E 2* and *eIF(iso)4E* are 696, 666 and 609 nucleotides long, respectively. The sequences encoding the amino acid residues that make up the VPg binding location of eIF4E-1, and the putative binding locations of eIF4E-2 and eIF(iso)4E, are all located in exon 3 of the genes. The guides selected for this research targeted the genes-of-interest near this location to maximise the chance that the resulting indel would disrupt this binding site (Figure 4.7).

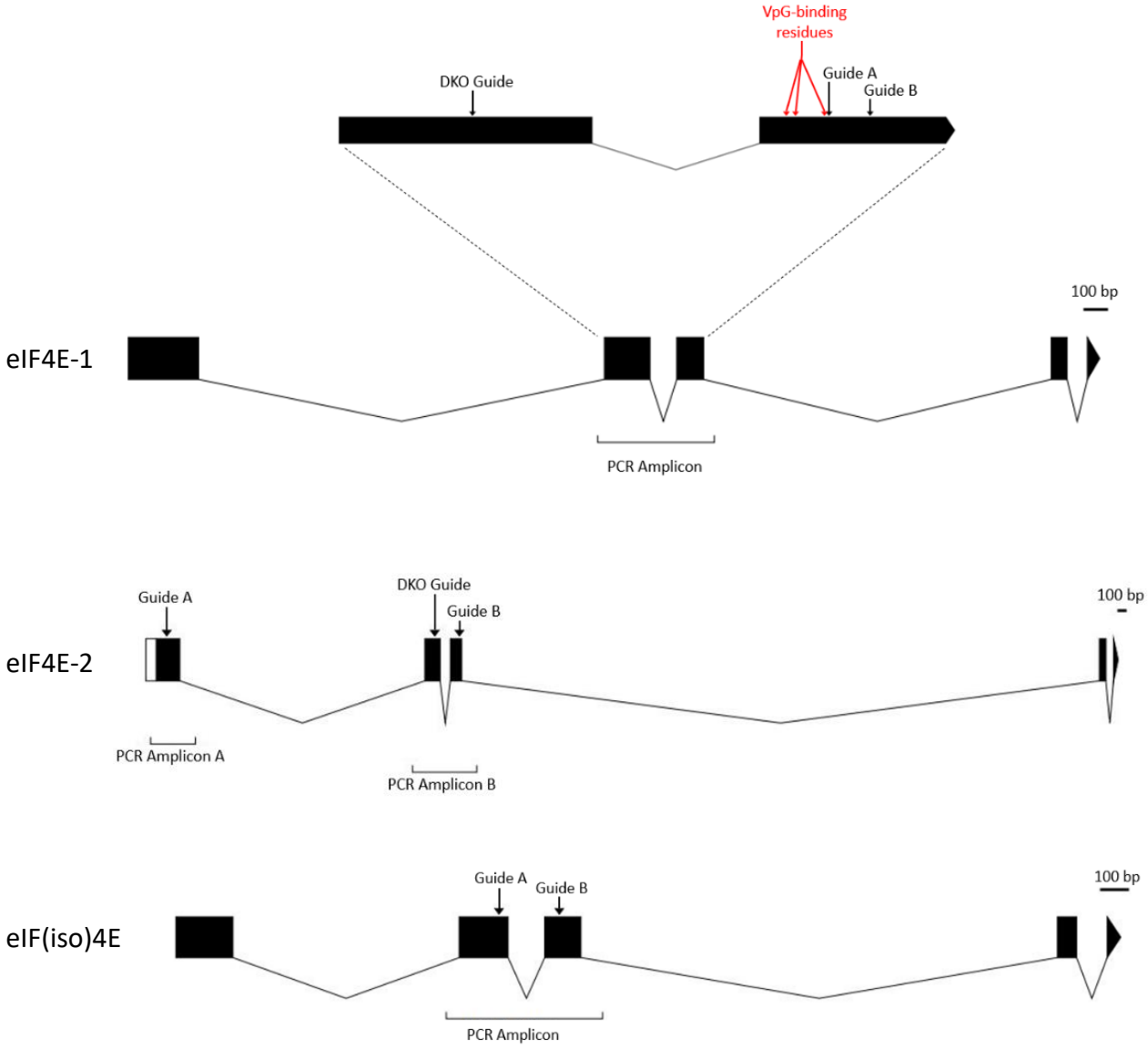


Figure 4.7 A diagrammatic representation of the *Solanum tuberosum eIF4E-1*, *eIF4E-2* and *eIF(iso)4E* gene structure. The polymerase chain reaction amplicon locations used in this research are indicated. The single guide RNA locations for each gene are labelled as 'Guide A' or Guide B'. The target site of the *eIF4E-1* and *eIF4E-2* double knockout guide is labelled as 'DKO guide'. The location of the viral genome linked protein (VPg) biding amino acids on *eIF4E-1* is also indicated.

Successful editing of potato eIF4E-1 was recently achieved by two research groups using CRISPR/Cas9 (Lucioli et al., 2022; Noureen et al., 2022). In both cases the resulting mutations conferred some degree of resistance to PVY. Interestingly, both research groups chose sgRNAs that targeted exon 1 of *eIF4E-1*. This decision was based on the high possibility that an indel introduced close to the 5' end of the gene would most likely result

in complete gene knockout. With this sgRNA target location, it is unlikely that a smaller, synonymous indel would inhibit VPg binding.

While guide location is an important factor to consider during sgRNA design, another important consideration is the predicted suitability of the sequence as a guide. Online sgRNA design platforms identify all sequences in a query that preceed a PAM site as potential guides. As this process is non-variable, and non-descriminatry, the sgRNAs outputs for the same query sequence are identical across all platforms. To help researchers identify the best guide from this output the program calculates a quality score for each guide. This indicates how suitable the guide is when considering three critera: the GC content, the guides base-pairing score (Liang et al., 2016), and the likelihood that the guide will match with non-target sites in the genome (Bae et al., 2014). Unfortunaetly, the weighting of these factors is program dependant and quality scores allocated to a single guide can vary widely across several programs. The programs CRISPOR and CRISPR RGEN were also not designed specifically for plant genome editing and as a result their algorthims may introduce inaccuracies in their assessements of these guides. This was the case with the the sgRNA options examined in this study (Table 4.2). This variabilty greatly limits the usefulness of this feature and, as a result, guides must be screened using different methods before being considered suitable for use *in vivo* (Uniyal et al., 2019). To further refine the number of guides considered for *in planta* genome editing all guide sequence identified were compared with the potato genome to identity any off-target matches, and only guides with no significant similarity to other sequences in the genome were considered further. The two guides for each gene that showed the highest average quality scores across all programs were chosen for *in vitro* analysis of on-target efficiency in *in vitro* assays (Table 4.3). As the coding sequences for *eIF4E-1* and *eIF4E-2* are very similar a guide was also designed that could mutate both genes simultaneously. The location that this guide targetted was severly restricted in as it had to be in a position that was identical between the two genes. Only one region preceeding a PAM site met this criterion and this sequence also demonstrated a low chance of off-target activity and so was also selected for *in vitro* assessment.

The *in vitro* assays showed all guides were able to successfully form secondary sgRNA structure, bind to Cas9, and guide the endonuclease to the target DNA region. This assay provides a binary answer regarding guide efficacy but does not indicate if one guide is more efficient than another. This can only be assessed *in vivo* and would be the next step in the progress of this research. Another limitation of this assay is the narrowness of the off-target activity screening. The *in vitro* assay only screens against the DNA sequence provided. To examine the specificity further, sequences from the isoform most similar to the target were also screened for each guide to examine if they became cleaved. In all cases no off-target activity was seen in these controls. While off-target screening was conducted *in silico*, *in planta* screening is needed to provide certainty that off-

target activity does not occurring. Off-target changes to the genome can be detected using sanger or high-throughput sequencing, restriction fragment length polymorphism analysis (Koo et al., 2015).

Before mutated plants can be generated using these guides, regeneration of protoplasts to form callus and mature plants must be established. Although regeneration protocols have been published for potato, establishment of such a system can still be difficult and requires a high degree of tissue culture expertise (Reed and Bargmann, 2021). This study optimised the foundational steps of isolating a high enough number of viable protoplasts to be used for transformation. The yield achieved was in the similar range to that reported in other studies, including one describing genome editing of potato plants regenerated from protoplasts (Nicolia et al., 2015). More recent studies describing potato genome editing have improved that protocol and increased protoplast yield (Moon et al., 2021). This method can be used in future experiments to maximise the likelihood of obtaining regenerated plants with limited somaclonal variation.

It was possible to regenerate micro-calli from protoplasts in this study, which is the first step towards regenerating mature plantlets. Growth of these micro-calli into a size suitable for root and shoot induction, unfortunately, did not occur. Protoplast culture procedures are highly species and cultivar specific (Reed and Bargmann, 2021) . Although the protocol was developed using the Désirée cultivar that was also used in this study, other research groups have achieved improved results by altering the Nicolia et al., 2015 method (Moon et al., 2021; Andreasson et al., 2022) indicating that laboratory-specific optimisation is needed. To further this research, each variable that may be preventing calli maturation needs to be individually assessed. A review of different plant protoplast culture techniques indicates three variables that have a consistent impact on growth of callus following regeneration from protoplasts (Andreasson et al., 2022). These are the hormone composition in the culture medium, the protoplast density, and the volume of the tissue culture vessel. In my studies micro-calli were cultured in a medium containing 0.4 mg L^{-1} of the cytokinin 6-Benzylaminopurine (BAP) and 1 mg L^{-1} of the auxin 1-Naphthaleneacetic acid (NAA; medium E, Addendum 1). Traditionally, a 1:1 molar ratio of cytokinins to auxins is used for callus culture (Skoog and Miller, 1957). While we achieved callus formation using the 0.4:1 ratio, investigating whether increasing the cytokinin concentration may prove beneficial and should be considered. The remaining two variables are related to the rate at which the protoplasts are able to condition the regeneration medium. Plant cells produce conditioning factors (CFs) that aid in the proliferation of cultured cells (Stuart and Street, 1971). Usually, a higher cell density leads to more successful cell culture as CFs accumulate more easily. In the current study protoplasts were suspended in alginate lenses which were then placed into liquid medium. This maintains a constant cell density while the external liquid medium can be refreshed periodically. The protoplasts will condition both the alginate lenses and the liquid medium surrounding them. While Nicolia et al., 2015 recommend a concentration of 5 x 10^4 protoplasts mL^{-1}, they later revised this to 8 x 10^4 protoplasts mL^{-1}

(Nicolia et al., 2020). This increase to cell density would improve the rate at which the protoplasts condition the liquid medium. In contrast, Moon et al., 2021, which focused on optimising the Nicolia et al., 2015 protocol to increase protoplast yield and callus regeneration, decreased this concentration to 4 x 10^3 protoplasts mL^{-1}. They argued that a lower cell density in the alginate lenses allows more space for the micro-calli to develop and results in healthier callus for shoot and root induction. Both arguments are valid and future experiments should examine all three concentrations. A final variable that may affect the success of callus development from protoplasts is the volume of the tissue culture vessel the alginate lenses are placed into. A smaller volume would allow the accumulation of CFs to achieve optimal levels earlier, however, too small a vessel may starve the cells of sufficient hormones and nutrients needed for development. While the original methodology shows success when the lenses are placed into 20 mL of regeneration medium in a standard Petri dish (Nicolia et al., 2015), other research reports improved proliferation when the lenses are placed in 4 mL of regeneration medium in 6-well tissue culture plates (Andreasson et al., 2022). Culture of the protoplasts in petri dishes, 6-well plates, and 12-well plates will be assessed in the future to determine which results in the highest efficiency of protoplast to micro-calli induction.

To conclude, two guide RNAs per gene of interest have been designed *in silico* and screened *in vitro* to assess on-target efficiency. Each of these guides has been deemed suitable for RNP transformation to assess *in vivo* activity. If successful, mutant plants with disrupted eIF4E activity may show a phenotype that is resistant to infection from multiple RNA viruses. Continuation of this work is, however, currently limited by the establishment of callus growth. Several variables have been identified that may improve the current protoplast culture procedures, and each of these should be investigated further. RNP transformation of protoplasts is a highly advantageous method of introducing the CRISPR/Cas9 genome editing method to plants and if successfully established in the future, will allow the production of precisely genome-edited RNA virus-resistant potato plants.

4.5 References

Andreasson E, Kieu NP, Zahid MA, Carlsen FM, Marit L, Sandgrind S, Peterson BL, Zhu LH (2022) Invited mini-review research topic: Utilization of protoplasts to facilitate gene editing in plants: Schemes for *in vitro* shoot regeneration from tissues and protoplasts of potato and rapeseed: Implications of bioengineering such as gene editing of broad-leaved plants. Front Genome Ed **4:** 780004

Bae S, Park J, Kim JS (2014) Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics **30:** 1473-1475

Bánfalvi Z, Csákvári E, Villánya V, Kondrák M (2020) Generation of transgene-free PDS mutants in potato by Agrobacterium-mediated transformation. BMC biotechnol **20:** 25

Charron C, Nicolaï M, Gallois JL, Robaglia CR, Moury B, Palloix A, Caranta C (2008) Natural variation and functional analyses provide evidence for co-evolution between plant eIF4E and potyviral VPg. Plant J **54:** 56-68

Concordet JP, Haeussler M (2018) CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. Nucleic Acids Research **46:** W242–W245

Coutinho de Olivera L, Volpon L, Rahardjo A, Osborne M, Culjkovic-Kraljacic B, Trahan C, Oeffinger M, Kwok B, Borden K (2019) Structural studies of the eIF4E–VPg complex reveal a direct competition for capped RNA: Implications for translation. PNAS **116:** 24056-24065

Dahal K, Xiu-Qing L, Tai H, Creelman A, Bizimungu B (2019) Improving potato stress tolerance and tuber yield under a climate change scenario – a current overview. Front Plant Sci **10:** 563

Dijkerman A (2022) Genome editing in bread wheat using CRISPR/Cas9. MSc thesis, Faculty of Science, Stellenbosch University, South Africa

Inoue H, Nojima H, Okayama H (1990) High efficiency transformation of *Escherichia coli* with plasmids. Gene **96:** 23-28

Koo T, Lee J, Kim JS (2015) Measuring and reducing off-target activities of programmable nucleases including CRISPR/Cas9. Mol Cells **38:** 475-481

Lacomme C, Jacquot E (2017) General characteristics of Potato Virus Y (PVY) and its impact on potato production: An overview. In: Potato Virus Y: Biodiversity, pathogenicity, epidemiology and management

Le NT, Tran HT, Bui TP, Nguyen GT, Nguyen DV, Ta DT, Trinh DD, Molnar A, Pham NB, Chu HH, Do PT (2022) Simultaneously induced mutations in *eIF4E* genes by CRISPR/Cas9 enhance PVY resistance in tobacco. Sci **12:** 14627

Lei Y, Lu L, Liu HY, Sen L, Xing F, Chen LL (2014) CRISPR-P: A web tool for synthetic single-guide RNA design of CRISPR-system in plants. Mol Plant **7:** 1494-1496

Liang G, Zhang H, Lou D, Yu D (2016) Selection of highly efficient sgRNAs for CRISPR-Cas9-based plant genome editing. Sci Rep **6:** 21451

Loebenstein G, Gaba V (2012) Viruses of potato. Adv Virus Res **84:** 209-246

Lucioli A, Tavazza R, Baima S, Fatyol K, Burgyan J, Tavazza M (2022) CRISPR-Cas9 targeting of the eIF4E1 gene extends the Potato Virus Y resistance spectrum of the Solanum tuberosum L. cv. Desirée. Front Micobiol **13:** 873930

Miroshichenko D, Timerbaev V, Okuneva A, Klementyeva A, Sidorova T, Pushin A, Dolgov S (2020) Enhancement of resistance to PVY in intragenic marker-free potato plants by RNAi-mediated silencing of eIF4E translation initiation factors. PCTOC **140:** 691-705

Moon K, Park J, Park S, Lee H, Cho H, Min S, Park Y, Jeon J, Kim H (2021) A more accessible, time-saving, and efficient method for *in vitro* plant regeneration from potato protoplasts. Plants **10:** 781

Nicolia A, Proux-Wéra E, Åhman I, Onkokesung N, Andersson M, Andreasson E, Zhu L (2015) Targeted gene mutation in tetraploid potato through transient TALEN expression in protoplasts. J Biotechnol **204:** 17-24

Nicolia A, Fält A, Hofvander P, Andersson A (2021) Protoplast-based method for genome editing in triploid potato. Methods Mol Biol **2264:** 177-186

Noureen A, Khan MZ, Amin I, Zainab T, Mansoor S (2022) CRISPR/Cas9-mediated targeting of susceptibility factor *eIF4E*-enhanced resistance against Potato Virus Y. Front Genet **13:** 922019

Park J, Bae S, Kim JS (2015) Cas-Designer: A web-based tool for choice of CRISPR-Cas9 target sites. Bioinformatics **31:** 4014-4016

Reed KM, Bargmann BOR (2021) Protoplast regeneration and its use in new plant breeding technologies. Front Genome Ed **3:** 734951

Robaglia C, Caranta C (2006) Translation initiation factors: a weak link in plant RNA virus infection. Trends Plant Sci **11:** 1360-1385

Robertson G, Burger J, Campa M (2022) CRISPR/Cas-based tools for the targeted control of plant viruses. Mol Plant Pathol **00:** 1-18

Ruffel S, Gallois JL, Lesage ML, Caranta C (2005) The recessive potyvirus resistance gene *pot-1* is the tomato orthologue of the pepper *pvr2-eIF4E* gene. Mol Genet Genom **274**: 346-353

Schaart JG, van de Wiel CCM, Lotz LAP, Smulders MJM (2016) Opportunities for products of new plant breeding techniques. Tren 438-449ds Plant Sci **5:** 438-449

Sridhar J, Venkateswarlu V, Shah MA, Kumari N, Raigond B, Bhatnagar A, Choudhary JS, Sharma S, Nagesg M, Chakrabarti SK (2022) Species composition and distribution of the vector aphids of PVY and PLRV in India. Potato Res **65:** 601-617

Skoog F, Miller CO (1957) Chemical regulation of growth and organ formation in plant tissues cultured in vitro. Symp Soc Exp Biol **11**: 118–130

Stuart R, Street HE (1971) Studies on the growth in culture of plant cells: Further studies on the conditioning of culture media by suspension of *Acer pseudoplantanus* L. cells. J Exp Biol **22:** 96-106

Tavazza R, Ancora G (1986) Plant regeneration from mesophyll protoplasts in commercial potato cultivars (Primura, Kennebec, Spunta, Desirée). Plant Cell Rep **5:** 243-246

Uniyal AP, Mansotra K, Yadav SK, Kumar V (2019) An overview of designing and selection of sgRNAs for precise genome editing by the CRISPR-Cas9 system in plants. 3 Biotech **9:** 223

Wang A, Krishnaswamy S (2012) Eukaryotic translation initiation factor 4E-mediated recessive resistance to plant viruses and its utility in crop improvement. Mol Plant Pathol **13:** 795-803

Woo JW, Kim J, Kwon SI, Corvalá C, Cho SW, Kim H, Kim SG, Kim ST, Choe S, Kim JS (2015) DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. Nat Biotechnol **33:** 1162-1164

Zhang H, Xu F, Wu Y, Hu H, Dai X (2017) Progress of potato staple food research and industry development in China. J Integr Agric **16**: 2924-2932

Zhang C, Zarka KA, Zarka DG, Whitworth JL, Douches DS (2021) Expression of the tomato *pot-1* gene confers Potato Virus Y (PVY) resistance in susceptible potato varieties. Amer J of Potato Res **98:** 42-50

Addendum 1

Macronutrient stock

Component	Concentration
KNO_3	74.0 g L^{-1}
$MgSO_4 \cdot 7H_2O$	49.2 g L^{-1}
KH_2PO_4	3.4 g L^{-1}

Micronutrient stock

Component	Concentration
H_3BO_3	1.5 g L^{-1}
$MnSO_4 \cdot H_2O$	5.0 g L^{-1}
$ZnSO_4 \cdot 7H_2O$	1.0 g L^{-1}
$Na_2MoO_4 \cdot 2H_2O$	1.2 g L^{-1}
$CuSO4 \cdot 5H_2O$	12.0 mg L^{-1}
$CoCl_2 \cdot 6H_2O$	12.0 mg L^{-1}
KI	380.0 mg L^{-1}

Iron stock

Component	Concentration
Na_2EDTA	1.4 g L^{-1}
$FeSO_4 \cdot 7H_2O$	1.9 g L^{-1}

Sugars stock

Component	Concentration
Sorbitol	6.25 g L^{-1}
Sucrose	6.25 g L^{-1}
D(-)Fructose	6.25 g L^{-1}
D(-)Ribose	6.25 g L^{-1}
D(+)Xylose	6.25 g L^{-1}
D(+)Mannose	6.25 g L^{-1}
L(+)Rhamnose monohydrate	6.25 g L^{-1}
D(+)Cellobiose	6.25 g L^{-1}
Myo-Inositol	2.50 g L^{-1}

Wash solution pH 5.6 (KOH)

Component	Concentration
Macronutrient stock	10.00 mL L^{-1}
Micronutrient stock	1.00 mL L^{-1}
Iron stock	10.00 mL L^{-1}
$CaCl_2$ stock (2 M)	3.00 mL L^{-1}
NaCl	14.03 g L^{-1}
NAA stock (2g L^{-1})	1.00 mL L^{-1}
BAP stock (1g L^{-1})	500.00 µL L^{-1}

Medium B pH 5.6 (KOH)

Component	**Concentration**
Murashige and Skoog basal medium	2.7 g L^{-1}
Casein hydrolysate	100.0 mg L^{-1}
NAA stock (2g L^{-1})	1.0 mL L^{-1}
BAB stock (1g L^{-1})	500.0 µL L^{-1}

Medium E pH 5.6 (KOH)

Component	**Concentration**
Macronutrient stock	10.00 mL L^{-1}
Micronutrient stock	1.00 mL L^{-1}
Iron stock	10.00 mL L^{-1}
Sugars stock	20.00 mL L^{-1}
Organic acids stock	10.00 mL L^{-1}
$CaCl_2$ stock (2 M)	1.25 mL L^{-1}
Casein hydrolysate	500.00 mg L^{-1}
Glucose	33.70 g L^{-1}
Mannitol	30.92 g L^{-1}
Bovine serum albumin	1.00 g L^{-1}
NAA stock (2g L^{-1})	500.00 µL L^{-1}
BAP stock (1g L^{-1})	400.00 µL L^{-1}

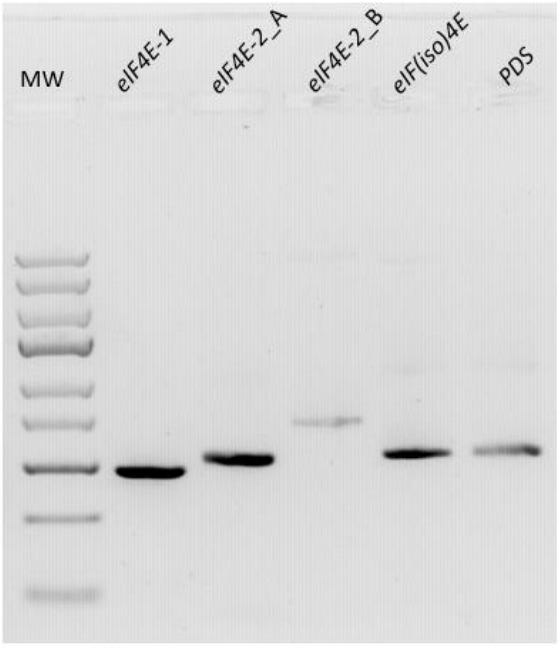
Medium C pH 5.6 (KOH)

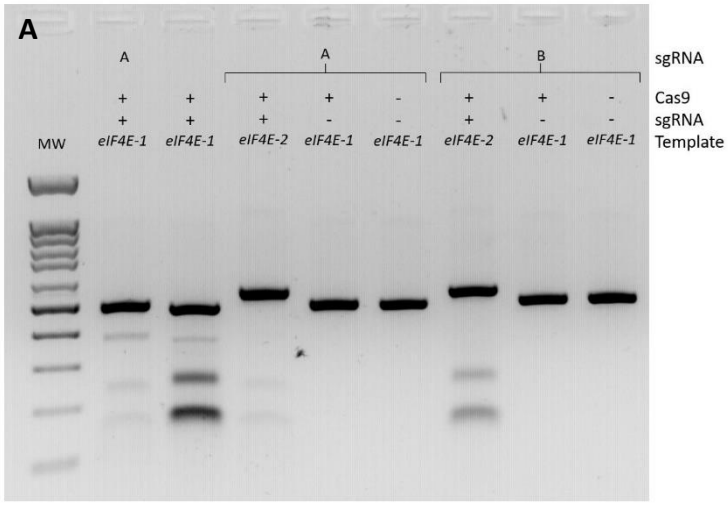
Component	**Concentration**
Macronutrient stock	10.00 mL L^{-1}
Micronutrient stock	1.00 mL L^{-1}
Iron stock	10.00 mL L^{-1}
Sugars stock	20.00 mL L^{-1}
Organic acids stock	10.00 mL L^{-1}
Casein hydrolysate	500.00 mg L^{-1}
Glucose	36.95 g L^{-1}
Mannitol	37.35 g L^{-1}
PVP 10	20.00 g L^{-1}
NAA stock (2g L^{-1})	500.00 µL L^{-1}
BAP stock (1g L^{-1})	400.00 µL L^{-1}
Macerozyme	2.00 g L^{-1}
Cellulase RS	10.00 g L^{-1}
$CaCl_2$ stock (2 M)	3.00 mL L^{-1}

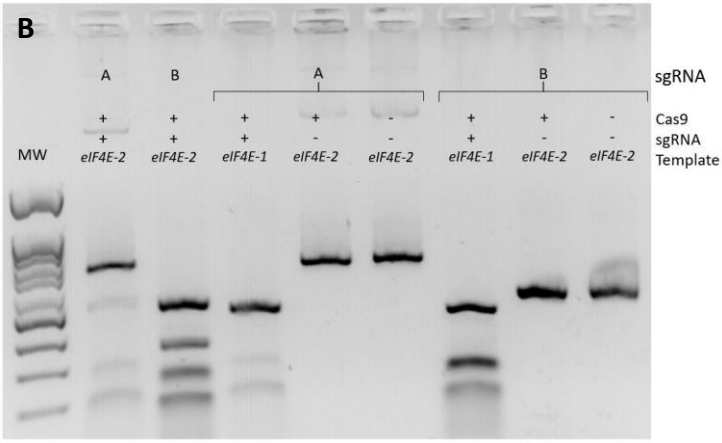
Supplementary material

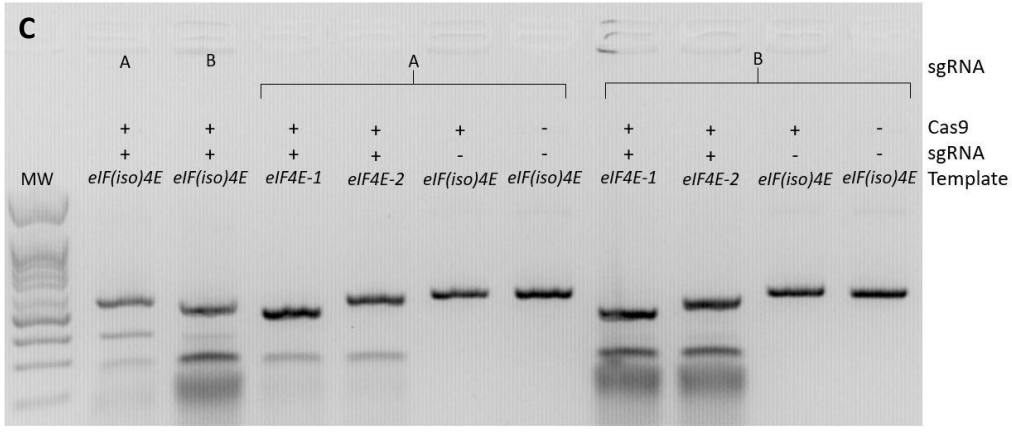
Supplementary Figure 4.1 Full size DNA gels showing gel electrophoresis of purified PCR amplicons from the *eIF4E* and *PDS* genes.

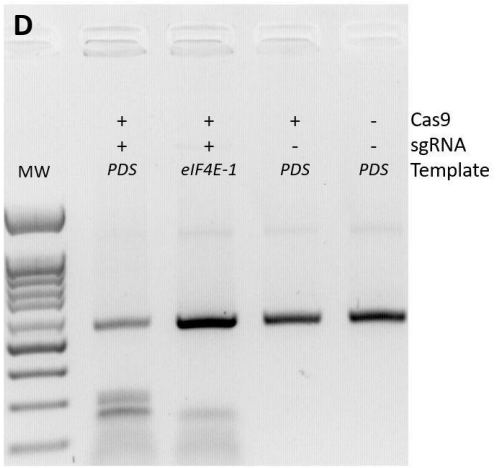
Supplementary Figure 4.2 Full size gel images for the *in vitro* CRISPR/Cas9 reactions of the sgRNAs targeting **(A)** *eIF4E-1,* **(B)** *eIF4E-2,* **(C)** *eIF(iso)4E*, **(D)** *eIF4E-1* + *eIF4E-2,* and **(E)** *PDS*. The inclusion or omission of each reaction component is represented by a '+' or '-' and shown above each gel image.

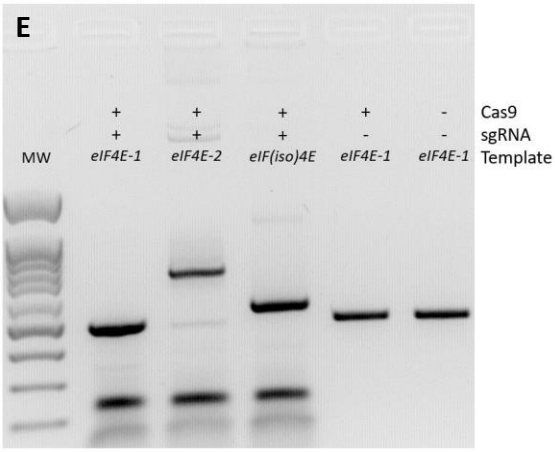












Chapter 5: Conclusion

5.1 Summary of research findings

The focus of this research was to examine members of the plant *eukaryotic translation initiation factor 4E* (*eIF4E)* gene family as they are promising susceptibility factors to confer viral resistance in plants. Phylogenetic analysis of these RNA cap-binding proteins elucidated an ancient divide into three major groups – eIF4E, eIF(iso)4E and nCBP – in the common ancestor of all land plants (Figure 3.3). The duplication of *eIF4E* into *eIF4E-1* and *eIF4E-2* was identified to have occurred deep within the eudicotyledonous lineage in the common ancestor of Solanaceous plants. The increased cellular complexity of land plants might be responsible for the conservation of multiple isoforms of proteins with similar functions in these plants. It is likely that these isoforms are differentially expressed in different tissue types, or under stressed conditions, although this needs to be confirmed in plants outside of *A. thaliana*. Transcript analysis of the potato RNA cap-binding proteins in tubers showed no significant differences between PLRV-infected and healthy plants (Figure 3.2). This data may indicate that expression differences between these genes are tissue-specific rather than induced by biotic stress. More research into expression of the genes in different tissue types of these plants will need to be conducted before this hypothesis can be confirmed. Three major clades representing *eIF4E*, *eIF(iso)4E* and *nCBP* were identified by phylogenetic analysis (Figure 3.3). Amino acid sequence alignments of the angiosperm members of each of these clades identified conservation of a region in most members of the *eIF4E* and *eIF(iso)4E* groups which is known to be important for potyviral VPg binding (Figure 3.4). The phylogenetic tree and amino acid sequence alignments provide a valuable tool for *eIF4E*-mediated resistance engineering as it helps identify isoforms within a species that can would be predicted to interact with the VPg. CRISPR/Cas9 RNPs can be transformed into potato protoplasts to introduce double stranded breaks to the *eIF4E* genes and potentially create mutations that would inhibit a VPg interacting with the proteins. Two single guide RNAs were designed for each of the potato *eIF4E* isoforms, as well as a guide that will simultaneously target *eIF4E-1* and *eIF4E-2* (Table 4.3). *In silico* analysis of each of the guides showed low potential for off-target genome editing and all guides were transcribed *in vitro* and demonstrated to complex with Cas9. On-target DNA excision was catalyzed *in vitro* for each RNP complex (Figure 4.4) and thus all guides were deemed suitable for *in vivo* transformations. The isolation of viable potato protoplasts for transformation was established and optimized (Figure 4.5). Protoplasts transformation and culture can lead to the development of mature plants regenerated from a single mutant cell. The regeneration of callus from protoplasts proved challenging and was only successful up until the micro-callus stage of development (Figure 4.6). For this methodology to be successful the complete procedure for protoplast culture from single cells to mature plantlets needs to be achieved.

5.2 Future prospects

The PLRV-infected tubers what were sampled in this study have been sprouted and will be planted in the upcoming months. Tissue samples will be harvested from these plants, and from healthy controls, to create samples for expression analysis. The RT-qPCRs that were conducted in this research will then be repeated across all tissue types to determine how expression of the *eIF4E* isoforms differ. It would also be beneficial to this research to acquire other potato plants that are infected with PVY, and other prevalent plant viruses, and replicate these expression studies in these plants to determine is virus-specific expression patterns can be identified.

It is paramount that regeneration of potato protoplasts into plantlets be achieved for this non-transgenic engineering of *eIF4E*-mediated resistance to proceed. It is common in protoplast research for this stage of the procedure to produce a bottleneck as much optimization is needed, alongside development of the tissue culture skills of the researcher. Possible adjustments to the regeneration protocol described in chapter 4 are being investigated with advice from the authors of the original protocol (Nicolia et al., 2015) to determine the limiting factor.

Once protoplast regeneration has been achieved with wild-type plants, transformations of each sgRNA-Cas9 complex designed in this research can begin. The *eIF4E-1*, *eIF4E-2*, and *eIF(iso)4E* single mutants, and the *eIF4E-1/eIF4E-2* double mutant would need to be subjected to inoculation studies with PVY and PLRV virus to determine if the induced mutations confer resistance. Additional *eIF4E-1/eIF(iso)4E* and *eIF4E-2/eIF(iso)4E* double mutants can also be created through back-to-back protoplast isolations, transformations, and regenerations with a single guide at a time. These mutants will also be screened against both viruses. All mutants generated will be assessed for any undesirable phenotypic changes that may arise from the gene knockouts or protoplast regeneration procedures.

It is most likely that complete resistance against any RNA virus, will only be achieved if all three susceptible *eIF4E* isoforms are knocked out. To prevent fatality in the triple mutants an *eIF4E* allele engineered not to bind to viral VPgs could be used to replace these proteins and maintain native translation. This resistant allele could be a transgene from a plant known to be naturally resistant to the viruses-of-interest, or can be synthetically engineered, as has been successfully done before in *Arabidopsis* (Bastet et al., 2018). This resistant allele would, however, need to be introduced to the mutant plants transgenically, which would nullify the previous efforts made to create the mutant plants in a nontransgenic manner. Base editing with a CRISPR/dCas9 RNP introduced to protoplasts provides a nontransgenic method of engineering a resistant allele. Resistance to another potyvirus (Tobacco Etch Virus) has been achieved in tomato through expression of a *Capsicum eIF4E* with a single glycine-to-arginine amino acid substitution (Yeam et al., 2007). This substitution, unfortunately, also inhibited the native function of the eIF4E protein. It does, however, provide proof that a single substitution can confer viral resistance, as would be the goal of base editing experiments.

Before such work is performed *in vivo,* it would be prudent to examine protein-interaction studies with mutated eIF4E isoforms to confirm that the protein maintains its native function while inhibiting VPg binding. Base editing holds great potential for engineering viral resistance, and it will be interesting to see how it is utilized in crop species in the upcoming years.

5.3 References

Bastet A, Lederer B, Giovinazzo N, Arnoux X, German-Retana S, Reinbold C, Brault V, Garcia D, Djennane S, Gersch S, Lemaire O, Robaglia C, Gallois JL (2018) Trans-species synthetic gene design allows resistance pyramiding and broad-spectrum engineering of virus resistance in plants. Plant Biotechnol J **16:** 1569-1581

Nicolia A, Proux-Wéra E, Åhman I, Onkokesung N, Andersson M, Andreasson E, Zhu L (2015) Targeted gene mutation in tetraploid potato through transient TALEN expression in protoplasts. J Biotechnol **204:** 17-24

Yeam I, Cavatorta JR, Ripoll DR, Kan B, Jahn MM (2007) Functional dissection of naturally occurring amino acid substitutions in eIF4E that confers recessive potyvirus resistance in plants. Plant Cell **19**: 2913-2928