The Optimisation of CRISPR-Cas9 Transformation in *Brassica* Species Using the *eIEF4E* Resistance Gene as a Model Target

by

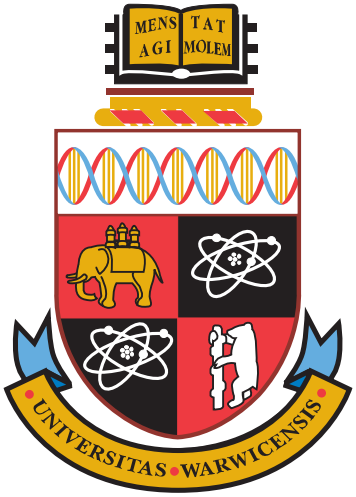
Luca James Champley Illing

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy



School of Life Sciences

September 2019

Supervisors: Guy C. Barker and John Walsh

School of Life Sciences, University of Warwick, CV4 7AL

Table of Contents

Acknowledgements

First and foremost, I would like to express my gratitude to my supervisors Dr. Guy Baker and Dr. John Walsh. I thank you both for your guidance, advice, and continuous support throughout this project.

Secondly, I would like to thank Richard Stark for his invaluable help in bioinformatic matters, as well as all of the staff and students in Life Sciences who have helped me throughout my project, especially those in my MIBTP cohort, Labs C46 & C30 and at the Wellesbourne campus. Additionally, I would like to thank all the friends I have made during my time at the University of Warwick and Leamington RUFC, for keeping me going throughout my project.

I would also like to acknowledge the financial support of my funding bodies, BBSRC and AHDB, without which, this research would not be possible.

Similarly, thank you to my brothers Hugo, Piers, and George and extended family who have offered reassurance and support throughout the years.

Finally, I would like to thank my parents, Glen and Lorraine, for their unwavering belief, support and encouragement throughout not only my PhD but my entire academic career and without which this would not have been achievable.

Declaration

I hereby declare that the work contained in this thesis is the original work of the author, except where specific reference is made to other sources, with the nature and extent of the author's contribution indicated (as appropriate) where work was based on collaborative research. The work was undertaken at the School of Life Sciences, University of Warwick between October 2016 and September 2019 and has not been submitted, in whole or in part, for any other degree, diploma or other qualification.

Luca James Champley Illing

September 2019

Abstract

Genome editing in plants has provided many new breakthroughs due to the development of different methods, such as CRISPR/Cas9, by which this can now be carried out. Beneficial traits should now be more quickly integrated into breeding lines. The Potyvirus TuMV, affects a wide variety of crop species including Brassicas, causing losses in yield and increasing production costs, therefore broad-spectrum resistance to TUMV is needed. Previous work identified a *B. rapa* variety, RLR22, which had resistance to TuMV due to a single mutation in an isoform of *eIF4E*, eIF(iso)4E, truncating the protein. This stops the virus utilising the eIF4F translation initiation complex, which consists of three subunits eIF4E, eIF4A, and eIF4G. This represents an ideal target for developing CRISPR/Cas9 knock outs as equivalent resistance has not yet been found in *B. oleracea*. Further investigation into the genetic makeup of the eIF4F complex enabled the targeting of specific paralogs that might confer resistance. Bioinformatic analysis of the genes involved identified several novel paralogs of *eIF(iso)4e, eIF4A, eIF(iso)4g*. This allowed for the design of paralog specific PCR primers and guide RNAs. RNAseq analysis also revealed alternative splicing taking place in all the gene families involved. CRISPR modification of *B. oleracea* proved difficult, resulting in only one heterozygous edited plant. This is potentially due to lethality of the knock-out of the target gene. In *B. rapa*, novel transformation techniques were employed to overcome the obstacles of recalcitrance to being transformed in this species. Protoplast protocols were optimised for use in Brassica and these increased transformation rates significantly. However, the use of embryogenesis transcription factors to increase regeneration of plants, and the use of Carbon Nanodots to aid in transformation were unsuccessful. CRISPR modification in Brassicas remains difficult, however, these complications stress the obstacles that must be overcome for industry to implement such technologies into their breeding programs.

Abbreviations

Common Abbreviations

TuMV	*Turnip Mosaic Virus*
RNA	*Ribonucleic Acid*
AS	*Alternative Splicing*
ASS	*Alternative Splice Sites*
Bbm	*Baby Boom*
bp	*Base Pair*
Cas	*CRISPR Associated System*
CD	*Carbon Nano-Dots*
cDNA	*Complementary DNA*
CNT	*Carbon Nano-Tube*
CRISPR	*Clustered Regularly Interspaced Short Palindromic Repeats*
DEX	*Dexamethasone*
DNA	*Dioxyribonucleic Acid*
dNTPs	*Deoxyribonucleotide Triphosphates*
EB	*Extraction Buffer*
eIF4A	*Eukaryotic Translation Initiation Factor 4A*
eIF4E	*Eukaryotic Translation Initiation Factor 4E*
eIF4F	*Eukaryotic Translation Initiation Factor 4F*
eIF4G	*Eukaryotic Translation Initiation Factor 4G*
ELISA	*Enzyme Linked Immunosorbent Assay*
EU	*European Union*
gDNA	*Genomic DNA*
gRNA	*Guide RNA*
H_2O	*Water*
HR	*Homologous Recombination*
IBA	*Indole-3-butyric acid*

InDel	*Insertion and Deletion*
iso	*Isoform*
JIC	*John Innes Center*
KO	*Knock Out*
MEE	*Mutually Exclusive Exons*
mOsm	*Milliosmole*
mRNA	*Messenger RNA*
NAA	*1-Naphthaleneacetic acid*
nCBP	*Novel Cap Binding Protein*
NHEJ	*Non-Homologous End Joining*
PABP	*Poly-A Binding Protein*
PAM	*Protospacer Adjacent Motif*
PBS-T	*Phosphate Buffer Saline with Tween*
PCR	*Polymerase Chain Reaction*
qPCR	*Quantitative Polymerase Chain Reaction*
QTL	*Qualitative Trait Locus*
RCF	*Relative Centrifugal Force*
RI	*Retained Intron*
RNAi	*RNA Interference*
RSTV	*Rice Tungro Spherical Virus*
RT-PCR	*Reverse Transcription PCR*
SE	*Skipped Exons*
T-DNA	*Transfer DNA*
Tm	*Primer Melting Temperature*
VIR genes	*Virulence genes*
VPg	*Viral Protein Genome-linked*
Wus2	*Wuschel*
YFP	*Green Fluorescent Protein*

List of Tables

List of Figures

Chapter 1

General Introduction

1.1 *Brassica*

The *Brassica* genus is a highly important crop family which supplies the highest diversity of products from a single genus worldwide, including root, flower and leafy vegetables as well as edible and fuel oils (Russo 2008). In 2013 approximately 7,439,100 tonnes of various *Brassica* crops were produced worldwide, worth roughly 14.6 billion dollars (FAOSTAT). Three diploid species *Brassica rapa* (AA), *Brassica oleracea* (CC) and *Brassica nigra* (BB) have combined in separate speciation events to create three allotetraploid species namely; *Brassica napus* (AACC), *Brassica juncea* (AABB), and *Brassica carinata* (BBCC) (Koh et al. 2017). Together these six species make up the classical Triangle of U (U 1935) (Figure 1.1).

In addition to this complexity there has been a whole genome triplication of the *Brassica* genome since the divergence from *Arabidopsis thaliana.* This event was then followed by significant reshuffling and gene loss within the genome (Liu et al. 2014) making the genomic make-up of the diploid Brassicas extremely complex.

1.1.1 *Brassica rapa*

Brassica rapa (AA Genome) is an important vegetable crop worldwide. There are several *B. rapa* vegetable types including turnip (va*r. rapa*), field mustard (*var. campestris*), pak choi (*var. chinensis*), mizuna (*var. japonica*), tatsoi (*var. narinosa*), yu choy sum (*var. parachinenesis*), and chinese cabbage (*var. pekinensis*) (Dixon 2007). Chinese cabbage is one of the most widely grown vegetable crops; it is a leaf vegetable native to China used many Asian cuisines. It is thought to have arisen from

a natural crossing of turnip (*B. rapa var. rapa*) and pak choi (*B. rapa var. japonica*) (Dixon 2007).

1.1.2 *Brassica oleracea*

Brassica oleracea (*B. oleracea,* CC) is an important vegetable crop worldwide, although it originated in Europe as a leafy vegetable crop and was cross-bred with many closely related *Brassica* species making a large amount of different cultivars over many years of domestication (Dixon 2007). This made the taxonomy of these crops difficult to deduce. The main types of *B. oleracea* include; kale (*var. viridis*), cabbage (*var. capitata*), brussels sprout (*var. gemmifera*), cauliflower (*var. botrytis*), broccoli (*var. italica*), and kohlrabi (*var. gongylodes*) (Dixon 2007).

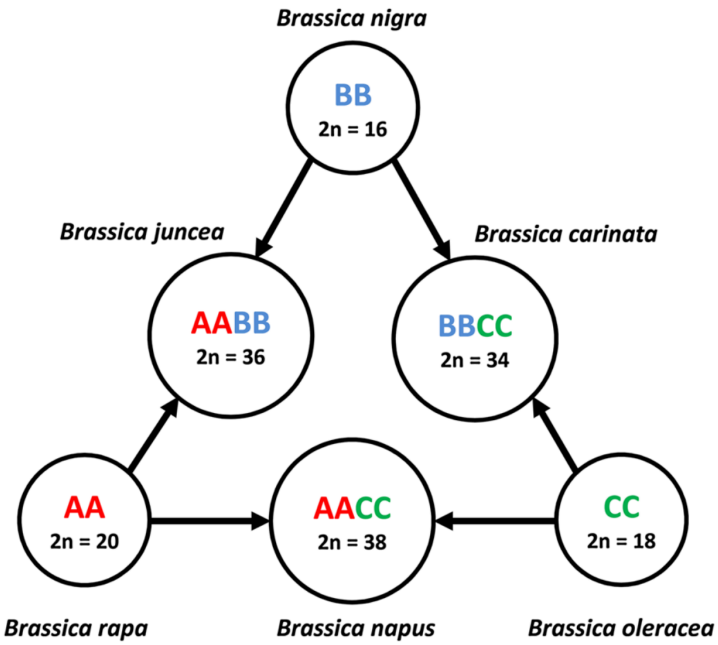


Figure 1.1 - The classic triangle of U

Depicting the relationship between the three diploid species of *B. rapa* (AA genome), *B. nigra* (BB genome), *B. oleracea* (CC genome) with the three allotetrpoloid species *B. juncea* (AABB genome), *B. carinata* (BBCC genome), *B. napus* (AACC genome). (Adapted from Koh et.al, 2017)

1.2 Turnip Mosaic Virus (TuMV)

Turnip Mosaic Virus (TuMV) is a member of the largest virus family *Potyviridae* containing 36% of all known plant viruses (Ward et al. 1991). This family causes significant losses in *Brassica* production worldwide by affecting both yield and quality of produce (Walsh et al. 2002). This results in a significant economic loss in Europe, Asia and North America. There are multiple pathotypes of TuMV that can infect a range of species causing different symptoms and severities of infection. TuMV has a positive sense single stranded ribonucleic acid (RNA) genome of roughly 9830 nucleotides in length encased within a coat protein forming a long filamentous rod-shaped particle (Figure 1.2) (Ward et al. 1991).

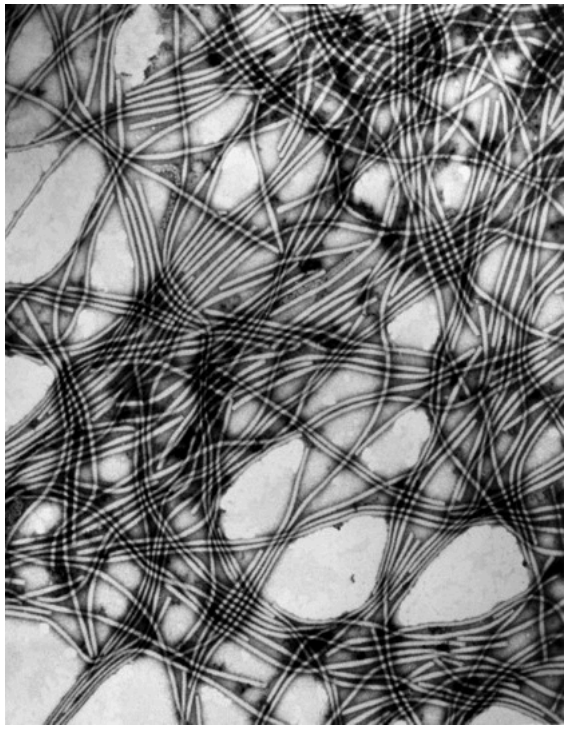


Figure 1.2- Electron Micrograph of TuMV

The long filamentous rod-shaped viral particles can be seen (Colin Clay)

1.2.1 Symptoms

Depending on the TuMV isolate and the *Brassica* crop infected there are many different symptoms and severity of infections. These include severe chlorotic lesions, black necrotic lesions, stunted growth, leaf curling and mosaic patterning in the leaves, as well as cigar burn in storage cabbages (Hunter et al. 2002). Additionally TuMV has been shown to reduce the size of individual seeds and seed having reduced yield as well as seed viability in *B. napus* (Walsh et al. 1985). These symptoms result in the end produce being undesirable and/or unmarketable in both leafy vegetable crops and oilseed crops.

1.2.2 Transmission and Life Cycle

The virus is transmitted by aphid vectors in a stylet-born non-persistent manner. At least 90 species of aphid can transmit TuMV due to the coat protein and helper components allowing the binding of the virus to the aphid stylet (Wang et al. 1999). The aphid will only retain the virus for a short period of time, 3-5 hours (Shattuck 2010), in which it must either probe or feed on a susceptible plant in order to spread the virus. Once a primary infection is established spread of the virus can take place rapidly depending on aphid number which are highly seasonal and are influenced greatly by temperature and weather conditions.

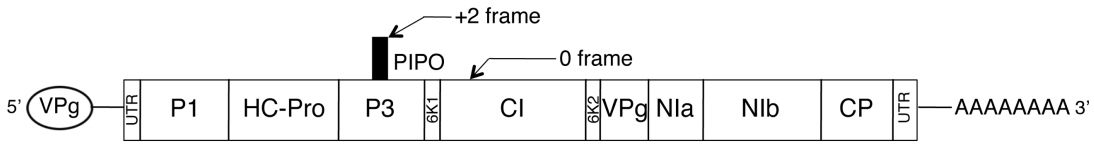


Figure 1.3 – Diagram Representing the TuMV Genome

Showing the ten genes encoding the ten mature proteins. Also included is the second open reading frame present in P3 of unknown function. Adapted from (Chung et al. 2008)

Once the virus is inside the plant cell the RNA genome is uncoated. Viruses are unable replicate on their own so hijack the host plants molecular machinery in order to complete their life cycle. Using a Viral Protein Genome-linked (VPg) protein encoded

by the virus, the viral RNA is able to bind to the translation initiation complex eIF4F. This is possible due to the versatile nature of the VPg which has a intrinsically disordered structure allowing it to bind to multiple proteins (Léonard et al. 2004, Jiang et al. 2011). Once bound to the host's translation machinery the open reading frame of the virus is translated to produce a large polyprotein which is then cleaved into ten mature proteins (Figure 1.3) (Walsh et al. 2002). A further reading frame is present in protein 3 and has been shown to form a fusion protein with protein 3 (Chung et al. 2008). Genome replication also occurs during the uncoated phase. After assembly of the new virus partials the virus can spread through the plant using pre-existing pathways in order to systemically infect the host plant (Carrington et al. 1996).

1.2.3 Control of TuMV

In order to control the spread of TuMV the use of pesticides to remove the aphid vector has been found to be inefficient, as the viruliferous aphids are able to migrate and transfer the virus before the insecticide takes effect. Additionally, since a European Union (EU) ruling in April 2018, the neonicotinoid pesticides used to control aphid numbers has had three of the main active ingredients, imidacloprid, clothianidin, thiamethoxam, banned from use in the field with several other pesticides also in line to be banned in the near future. Therefore the avoidance of infected plants with the use of resistant cultivars appears to be the cheapest and easiest long term way to control the transmission of TuMV and thereby maintaining crop yields and quality (Shattuck 2010). One method of preventing infection could be through blocking the binding of the virus to the host translation machinery, such as the eIF4F complex.

1.3 Eukaryotic Translation in Plants

Eukaryotic translation in plants is a complex process by which multiple proteins are needed in order to translate messenger RNA (mRNA) into polypeptide chains. This processes is initiated by one of two process, cap-dependent or cap-independent, translation initiation. Cap-dependent initiation is most common and involves the interaction of the 7-methyl-guansoine-containing cap structure on the 5' end of the

mRNA binding to an initiation factor complex. This then recruits further translation machinery in order to complete the translation of the mRNA.

1.3.1 Eukaryotic Translation Initiation Factor 4F (eIF4F) Complex

The eIF4F complex is a cap biding initiation factor complex formed of three sub-units. Firstly eukaryotic initiation factor 4E (eIF4E), which binds to the 5' cap nucleotide structure of the mRNA due to a number of tryptophan residues located in its binding pocket at amino acid positions 56 & 102 (Browning 1996). eIF4E is bound to eukaryotic initiation factor 4G (eIF4G), of which its exact function is unknown, however, it is thought to be an essential scaffold protein in the formation of the eIF4F complex (Macovei et al. 2018). The third subunit is eukaryotic initiation factor 4A (eIF4A) is a DEAD box RNA helicase (Browning 1996). Although eIF4E is the subunit that binds directly to the 5' cap it is has been shown that eIF4A and eIF4G also play a role in the binding of the mRNA to the eIF4F complex (Gallie 2001). The complex then binds to the Poly-A binding protein (PABP) causing the mRNA to loop. It then recruits the eIF3 protein, binding to the eIF4G subunit, which in turn recruits the 40S ribosomal subunit in order to begin translation (Browning 1996) (Figure 1.4).

It is known that there are multiple paralogs of genes producing the subunits of the eIF4F complex. The *eIF4E* gene family is known to contain 3 isoforms in *A. thailana*; *eIF4E*, *eIF(iso)4E* and *novel cap binding protein (nCBP*) (Browning 1996, Ruud et al. 1998). In Diploid *Brassicas* there is thought to be 3 paralogs of *eIF4E* and *eIF(iso)4E* and they share roughly 50% sequence identify (Browning 1996, Jenner et al. 2010), however little is published on *nCBP* in brassicas. The *eIF4G* gene family is known to have a single isoform, *eIF(iso)4G*, however this is much smaller than eIF4G, being 86kDa compared to eIF4G 180 kDa (Gallie et al. 2001). In *Brassica* the exact number of *eIF4G* and *eIF4A* paralogs are not known despite genome sequencing.

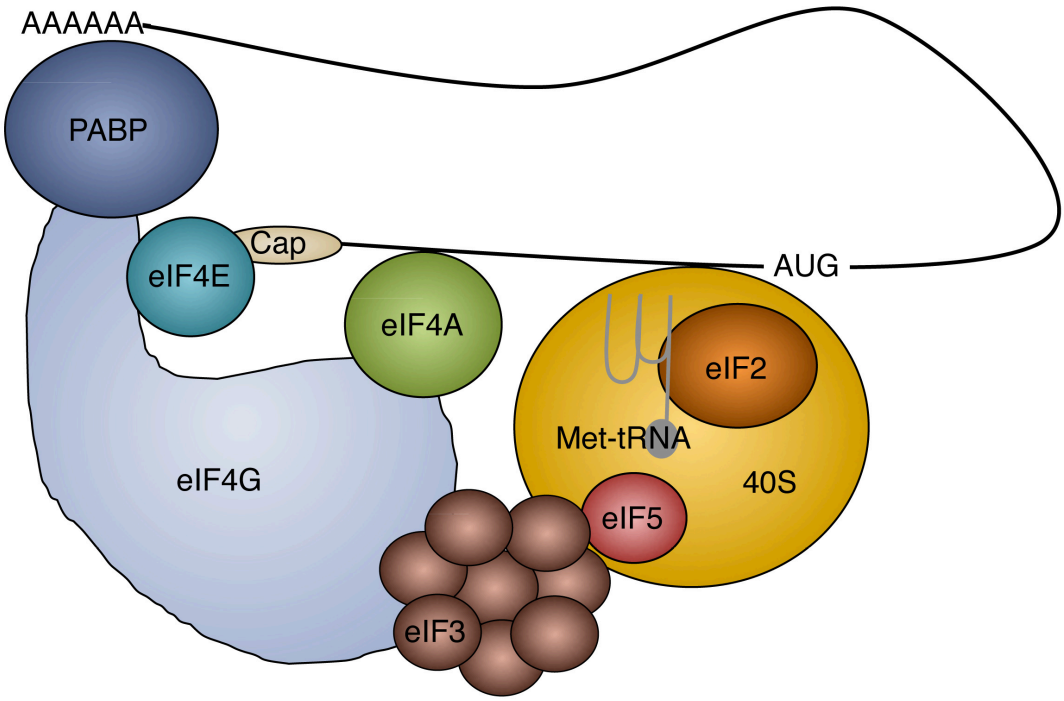


Figure 1.4 - Diagram of The Eukaryotic Initiation Factor complex

The complex is made up of eIF4E, eIF4G and eIF4A, bound to the mRNA molecule, PABP and the 40s ribosomal subunit via eIF3. Adapted from Robaglia et al. (2006).

1.4 TuMV Resistance in *Brassica*

The first resistance to TuMV was observed in *B. napus* although this was only to a small subset of viral isolates. On investigation it was found that six, so called, *TuRB0* genes conferred a dominant strain-specific resistance (Walsh et al. 1999, Hughes et al. 2002, Jenner et al. 2003).

The first broad-spectrum TuMV resistance in *Brassica* was found in *B. rapa* ssp. *pekinensis* (Chinese cabbage) in the plant line RLR22. The resistance seen gave resistance to all isolates of TuMV known at the time (Walsh et al. 2002). Using qualitative trait locus (QTL) analysis by crossing RLR22 with the known susceptible line R-o-18, it was found that the resistance had two loci, one recessive locus located on the upper section of chromosome A4, known as *recessive TuMV resistance 01 (retr01).* A second dominant locus was found on chromosome A8 called *conditional TuMV resistance 01 (ConTR01)*. It was found that *retr01* was epistatic to *ConTR01*

meaning it relied on the presence of the other locus to confer resistance (Rusholme Pilcher et al. 2007). This type of recessive resistance is thought to be more durable than a dominant resistance gene, and has been seen multiple times in initiation factors in many other plant species (Truniger et al. 2009). The *retr01* locus was found to be at the location of the A4 paralog of *eIF(iso)4E* which had a mutation resulting in a truncated protein (Jenner et al. 2010, Nellist et al. 2014). In addition resistance has been seen in *Capsicum* with a missing tryptophan residue in the binding pocket of the eIF(iso)4E protein which is thought to prevent the binding of the viral VPg to the eIF4F complex and can therefore no longer use the host plant machinery to replicate (Charron et al. 2008). This resistance mechanism has been seen before in other plant species in which a mutation in some aspect of eIF4F complex confers virus resistance (Table 1.1) (Robaglia et al. 2006).

Table 1.1 – Table of TuMV Resistance Genes

Resistance gene	Plant Species	Gene Responsible
Retro1	*B. rapa*	*eIF(iso)4E*
Contro1	*B. rapa*	*Unkown*
pvr2	*Capsicum spp.*	*eIF4E*
Turbo1	*B. napus*	*uknown*
mo1	*Lactuca spp.*	*eIF4E*
sbm1	*Pisum sativium*	*eIF4E*
pot1	*Lycopesicon spp.*	*eIF4E*
lsp1	*A. thaliana*	*eIF(iso)4E*

1.4.1 Engineering the eIF4E resistance mechanism

Due to the fact the eIF4E resistance mechanism is well known and has been seen in multiple plant species, attempts have been made to introduce it into a range of species and varieties. One method used a TILLING population of *Solanum lycopersicum* (Tomato) to mutate the *eIF4E* gene. They found that by mutating only one of the two paralogs of the gene they did not get resistance and needed to combine mutations in both paralogs of the gene to get the resistance sought (Gauffier et al. 2016). Another, used RNA interference (RNAi) in *Cucumis melo* (melon) to knockdown *eIF4E*

expression giving resistance to four different potyvirus species (Rodriguez-Hernandez et al. 2012). A third method used in-vitro site directed mutagenesis of the *eIF(iso)4E* gene which was then over-expressed in a susceptible *B. rapa* line, conferring increased resilience to multiple TuMV strains, however they were not completely resistant (Kim et al. 2014).

1.5 CRISPR/Cas9

Genetic modification of crop species is a key tool to help understanding as well as improve yield and other favourable traits. Clustered regularly interspaced short palindromic repeats (CRISPR)/ CRISPR –associated (Cas) systems were discovered in prokaryotes several decades ago. In bacteria and archaea it is used as an adaptive immune system against viruses and plasmids by using RNA to guide the Cas machinery to the viral genes (Qi et al. 2016). Some Cas systems such as Cas9 are able to introduce double stranded breaks into the target deoxyribonucleic acid (DNA). The double stranded break can then be repaired in two ways using the cell's native repair mechanisms. These are either homologous recombination (HR) or non-homologous end joining (NHEJ). NHEJ can be error prone resulting in random sized insertions and deletion (InDel) mutations (Britt 1999). A simplified version of the CRISPR/Cas9 system can be altered by changing the guide-RNA (gRNA) to introduce double stranded breaks to almost any target DNA (Jinek et al. 2012). The gRNA is a 20-nucleotide region of the RNA that binds to the Cas9 protein that shares homology with the target sequence (Figure 1.5) (Razzaq et al. 2018). The main requirement for the gRNA is that it contains a protospacer adjacent motif (PAM) site at the end of the gRNA, most commonly NGG, which is fundamental for Cas9 activity (Farboud et al. 2015). The InDel mutations that occur as a result of NHEJ can lead to a frame shift which may result in premature stop codons, effectively knocking out the target genes (Andersson et al. 2017). This relatively new system allows for specific genome editing of targeted genes as well as the potential to introduce desired mutations into a gene of interest.

1.5.1 CRISPR and eIF4E Resistance

The eIF4E broad-spectrum resistance mechanism has previously been shown to be replicable using other molecular techniques either knocking down or knocking out the part of the gene family. Therefore it is a natural target for CRISPR/Cas9 mutagenesis techniques (Gal-On et al. 2017). First seen was a CRISPR knock out of *eIF4E* in *Cucumis sativus* (cucumber) which, after selfing, gave T3 generation plants that did not contain the Cas9 transgene and were homozygous for the knock out. These plants were resistant to several potyviruses that infect cucumber (Chandrasekaran et al. 2016). Another study in *A. thaliana* showed introducing an InDel mutation via CRISPR in to the *eIF(iso)4E* isoform resulted in a functional knockout line that was resistant to potyviruses, including TuMV (Pyott et al. 2016). More recently, in *Manihot esculenta* (cassava), a dual CRISPR knock out of the two paralogs of the NCBP isoform has been shown to give greatly reduced symptoms in the plant when it was infected with potyvirus. This had not been seen before as breeding the multiple resistance loci into one plant line had proven difficult (Gomez et al. 2019). Additionally, as the eIF4F complex is extremely important to the plant it has been thought that instead of a knock out, the precision editing that CRISPR/Cas9 allows means it is possible to mutate the binding pocket of the eIF4E gene so that VPg can

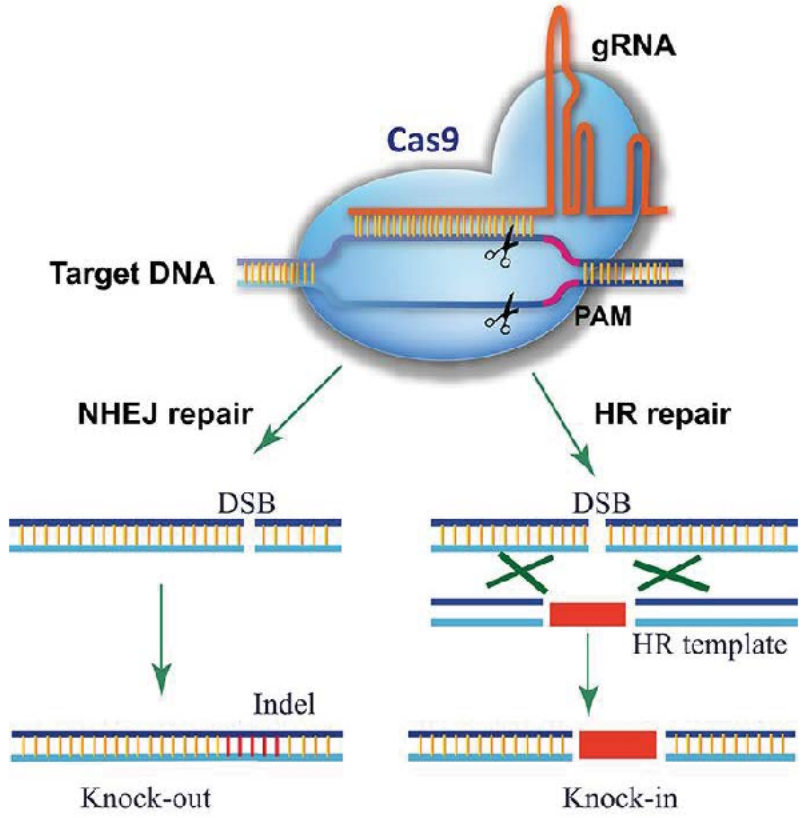


Figure 1.5 - Diagram of the Cas9 machinery

Showing the Cas9/gRNA complex bound to the host DNA just upstream of the PAM site. The Cas9 protein forms double stranded breaks (DSB) which are then repaired by the hosts native systems in either homologous or non-homologous recombination. Adapted from (Razzaq et al. 2018)

no longer bind but does not produce any adverse phenotypic affect to the plant. This has been achieved in *A. thaliana,* and when combined with loss of function *eIF(iso)4E,* it gives resistance to a larger range of potyviruses including TuMV (Bastet et al. 2019).

In addition to the modification of *eIF4E* it has also been shown that modification to the other subunits on the eIF4F complex can confer resistance. In *Oryza sativa var. indica* (rice) it was shown that CRISPR-induced mutations in *eIF4G* gave resistance to Rice tungro spherical virus (RTSV) (Macovei et al. 2018).

1.5.2 CRISPR in *Brassica*

The CRISPR/Cas9 system has been shown to work in multiple plant species, however, even though there is a large diversity of *Brassica* it has only been achieved in two species and only in a restricted number of lines of these species. Firstly, in *B. napus* (*var. westar*) in which relatively high mutation rates were seen of between 30-40% (Yang et al. 2017, Sun et al. 2018). Secondly, in *B. oleracea* (AG DH1012) in which one study transformed and mutated a double haploid line with mutation rates of 10%, this line was chosen for its favourable transformation rates (Lawrenson et al. 2015). Another study in *B. oleracea var. abloglraba* had a mutation rate of 76% but only a transformation rate of 1.5% (Sun et al. 2018). Both gave rise to heritable mutations. More recently mutations have been successfully introduced in *B. oleracea var capitata,* and shown to be heritable to the T1 generation, although again mutation efficiencies of 13% or less were seen across four separate genes (Ma et al. 2019). The small number of successful studies showing the use of CRISPR/Cas9 system in *Brassica* indicates that the process is not straightforward.

1.6 Transformation of *Brassica*

In order to deliver the CRISPR/Cas9 machinery into a plant of interest some method of transformation must be used. There are many methods used, including older methods such as *Agrobacterium* mediated transformation and protoplast transformation, as well as novel methods such as functionalized carbon particles.

1.6.1 *Agrobacterium* Transformation

Agrobacterium tumefaciens has long been used as the backbone of plant transformation due to its natural ability to alter the plant's genetic make-up. Using a modified plasmid containing the DNA to be transferred (T-DNA) flanked by border sequences as well as virulence genes (*VIR* genes) necessary for the transformation process. The T-DNA can be introduced randomly into the target plant genome giving heritable transformed plants (Gelvin 2003). Previous studies have shown that *Agrobacterium* transformation of multiple brassica species is possible including, *B. napus, B. juncea, B. rapa and B. oleracea* (De Block et al. 1989, Moloney et al. 1989,

Barfield et al. 1991, Radke et al. 1992). Although these transformations worked, the transformation rates were relatively low (< 10%) in most species, except in *B. napus* (30-40%). *B. rapa* in particular has proven especially difficult to transform with no readily transformable lines known. This has been attributed to the genotype restrictions in many *Brassica* species, with many genotypes remaining recalcitrant. In *B. oleracea,* a study was carried out at the John Innes Centre (JIC) where an existing double haploid genetic mapping population was screened for transformation efficiency, one line was found (DH 1012) which gave transformation rates of 15% (Sparrow et al. 2004). This line has become a model *B. oleracea* genotype for transformations and has been transformed several times with the CRISPR/Cas9 machinery, resulting in heritable mutations (Lawrenson et al. 2015, Sparrow et al. 2015).

Another approach to improving recalcitrance in *Brassica* may be by transforming with embryogenesis promoting transcription factors. This allows the formation of transgenic calli. One example of this is the use of Baby Boom (*Bbm*) and Wuschel (*Wus2*) genes to generate transgenic calli in a maize inbred line (Lowe et al. 2016). Additionally, transcription factor *RKD4* has also been shown to induce embryogenesis (Waki et al. 2011), and has been shown to produce transgenic calli in a previously un-transformable wheat line (unpublished).

1.6.2 Protoplast Transformation

Protoplasts are an important system for studying organisms with cell walls. There are protocols available tailored to multiple different plant species in both monocots and dicots. Protocols generally involve the enzymatic degradation of the cell wall in multiple tissues. This process varies across species and tissue type as factors such as thickness of cell walls, temperature, duration of digestion, pH and osmolarity all effect protoplast release (Sinha et al. 2003). The removal of the cell wall allows for many experimental procedures to take place which would normally be impeded by the cell wall. This includes studying the cellular membrane and easier transformation of cells with isolated DNA resulting in stable nuclear transformations (Davey et al. 2005). An

adapted *A. thaliana* protocol has previously been shown to enable isolation of *Brassica* protoplasts (Yoo et al. 2007). Additionally, it has been shown that protoplasts can be used as an ideal platform in which to test gRNAs used in the CRISPR/Cas9 system in *B. oleracea* (DH 1012). This is due to the relatively short process (3-4 days) that is used to isolate, transform and analyse protoplast DNA, whereas transformation and growth of a plant can take 2-3 months until analysis is possible (Lin et al. 2018).

1.6.3 Nanoparticle transformation

Although agrobacterium and protoplast transformations have been well used, they remain difficult to implement in recalcitrant plant varieties. Novel technologies are offering new methods of DNA transformation which show the promise of easy transformation of recalcitrant plant varieties. Nanoparticle delivery of DNA has several potential benefits including avoiding immune responses and carrying larger payloads. Several particles have been used for DNA transformation including functionalised gold particles, graphene, and quantum dots (Riley et al. 2017). Although useful, these particles had several problems including expense and toxicity to the plant. Another nanoparticle that has recently be shown to successfully overcome the plant cell wall barrier without mechanical aid is carbon nano-tubes (CNTs). These tubes have been chemically functionalised to bind to DNA and passively cross plant membranes resulting in 85% transformation efficiencies (Demirer et al. 2019). One drawback of CNTs however is the intricate protocol for creating and functionalising the particles as well as DNA binding. One method which may overcome this is carbon nanodots (CDs), which have similar attributes to CNTs but with the benefit of being formed and functionalised using a one-step microwave assisted protocol (Liu et al. 2012). The recent improvements in the functionalisation of the CDs (Swift et al. 2018) has allowed for larger plasmids, such as CRISPR/Cas9- gRNA plasmids, to be delivered into plant leaf tissue in a previously recalcitrant wheat line (unpublished). The CDs are currently in the patent application process, as such the exact method of functionalisation and DNA binding has not been published. CDs give the potential for plasmid DNA delivery in recalcitrant *Brassica* without transgene integration. The CDs also have the added benefit of being taken up by the plant into several tissue types (unpublished), meaning it could potentially be taken into the germline. There could

therefore be potential for a floral dip method to be applied similar to that used in *A. thaliana*.

1.7 Aims and Objectives

There were three main aims of this study and were as follow; firstly, to further the understanding of the genetic make-up of the genes making up the eIF4F complex in diploid *Brassicas*. Secondly to mimic the eIF4E TuMV broad-spectrum resistance mechanism found in *B. rapa* line, RLR22, in *B. olerecea* using CRISPR/Cas9 gene-editing techniques. Finally, to access, utilise, and optimise novel transformation techniques to allow the use of CRISPR/Cas9 transformation techniques in recalcitrant *Brassica* lines

Chapter 2

Materials and Methods

2.1 Molecular Biology Techniques

2.1.1 Genomic DNA Extraction

Roughly 100mg of plant leaf tissue was extracted into 2ml 96 well plates (Thermo-Scientific AB-1127). Two tungsten carbide beads were placed into each well. The whole plate was then frozen at -80°C for 20 minutes. An oscillatory grinder (Qiagen TissueLyser) then shook the plates at 22 oscillations/sec for 1 min or until tissue was ground to a fine powder. The plate was centrifuged at 4500g for 1 min then, to each well, 300µl extraction buffer (EB) (100mM Tris pH 8.0, 50mM EDTA pH 8.0, 500mM NaCl) with 2µl/ 10ml EB of β-mercaptoethanol added just before use. After adding 20µl of 20% SDS the sample was incubated at 65°C for 25 minutes. The plate was then placed on ice for 10 minutes and 100µl of 5M potassium acetate is added and incubated on ice for a further 30 minutes. The plate was centrifuged at 3007g for 20 minutes. 125µl of supernatant was then transferred to a clean plate and 200µl isopropanol was added and incubated at -20°C for 1 hour or overnight. The plate was then spun at 3007g for 20 minutes. The isopropanol was then tipped off and 400µl of 70% ethanol was added to clean the pellet, and spun at 3007g for 10 minutes. The ethanol was then tipped off and the pellet was left to air dry at 37°C for 30 minutes or until all the ethanol had evaporated. The dried pellet was resuspended in 50µl Tris-EDTA (TE) buffer (10mM Tris, 1mM EDTA pH 8.0) containing 20mg/ml RNase A (Thermo-Fisher R1253). DNA was stored at -20°C. When smaller numbers were required the method was scaled to use 2ml Eppendorf tubes instead of the 96 well plate.

2.1.2 Primer Design

Primers were designed manually in order to amplify individual gene copies. Primer characteristics were checked using the OligoEvaluator (Sigma-Aldrich). The characteristics of primers were approximately 50% G/C content, Tm 60°C, and low or no secondary structure. Primers were then ordered dry using desalt purification at a yield of 0.025 μM (Sigma).

2.1.3 Polymerase Chain Reaction

Polymerase chain reactions (PCR) was used for analysis of both cDNA and gDNA. Two PCR reaction mixes were used. For simple PCR reactions the mix was comprised of 12.5 μl RedTaq PCR Master Mix (Sigma-Aldrich), 1 μl forward primer (10μm), 1μl reverse primer (10μm), 10ng template DNA and H_2O to 25 μl reaction. The second reaction mix was used when sequencing of the PCR product was required, comprised 10μl 5X Phusion HF Buffer (NEB), 1μl 10mM dNTPs, 2.5 μl forward primer (10μM), 2.5 μl reverse primer (10μM), 10 ng template DNA, 0.5 μl Phusion DNA Polymerase (NEB). Reactions were carried out in the following conditions; 95°C for 5 minutes, 30 cycles of 95°C for 30 seconds, 55 - 65°C for 30 seconds to 1 minute depending on the primer pair, 72°C for 1 – 3 minutes depending on length of expected product. When necessary, PCR products were purified using Qiagen PCR Purification kit, ethanol precipitation or Sephadex purification.

2.1.4 Sephadex Purification

Sephadex G-50 (Sigma G5080) was poured into a Multiscreen column loader (Millipore) and excess scrapped off. A Multiscreen filter plate (Millipore) was placed upside down onto the loader and inverted. 300μl of sterile H_2O was put into each well and incubated at 4°C for 1 hour. The filter plate was then placed and aligned on top of a used round bottom 96 well plate and spun at 910g for 5 minutes. PCR reactions in a 96 well plate were then pipetted into the corresponding well of the filter plate onto

the middle of the sephadex. The filter plate was taped onto a clean collection plate, and spun at 910g for 5 minutes. Purified DNA was stored at -20°C

2.1.5 Quantitative Polymerase Chain Reaction

Each Quantitative Polymerase Chain Reaction (qPCR) reaction was carried out using 10µl of 2x LightCycler 480 SYBR Green I Master Mix (Roche), 3µl H_20, 1µl forward primer (5mM), 1µl reverse primer(5mM) and 5µl of cDNA per well. For relative qPCR a 1/5 dilution series of cDNA was made and dilutions 1/5, 1/25, 1/125, 1/625 were used. The genes *β-tublin* (XM_009153140) and *TIP41* (XM_009116214) were used as relative expression controls. Samples were randomised on the 96 well plate. The reactions were run on the Agilent Technologies Stratagene Mx3005P (Agilent), and data was processed using the Agilent MxPro software.

2.1.6 Sequencing

Sequencing was carried out using the Eurofins GATC sequencing LightRun Tube Service. 5µl of 60-80ng of DNA and 5µl of 5µM primer were submitted. Analysis was carried out using SnapGene and DNASTAR Suite.

2.1.7 RNA Extraction

Approximately 100mg of plant leaf tissue was placed into an Eppendorf tube and snap frozen in liquid nitrogen. Tissue was then ground to a fine powder using a pestle and mortar or using the Dremel attachment. TRIzol reagent (Thermo-Fisher) (1ml) was added and the tube vortexed, then incubated for 5 minutes at room temperature. Chloroform (200µl) was added and tube shaken by hand for 15s and then incubated for 3 minutes at room temperature. Samples were then centrifuged at 12,000g for 15 minutes at 4°C. The Aqueous RNA layer was transferred to a fresh tube and 500µl isopropanol was added. Samples were incubated for 10 minutes at room temperature and then centrifuged at 12,000g for 15 minutes at 4°C. The supernatant was removed and the pellet was washed with 1ml 75% Ethanol. Samples were then centrifuged at

7500g for 5 minutes at 4°C, the supernatant was removed and the pellets air dried for 15-20 minuets. The pellets were dissolved in 40µl RNase free H_2O and RNA was stored at -80°C.

2.1.8 Reverse Transcription PCR

Reverse Transcription PCR (RT-PCR) was used to create cDNA for subsequent PCR and qPCRs. RNA (2.5µg) was diluted into a total volume of 11.5µl of H_2O, this was then incubated for 5 minutes at 60°C. A mix of 1µl RNaseOut (Thermo-Fisher), 10µl 5x RT Buffer (Thermo-Fisher), 100mM DTT (Thermo-Fisher), 1.6µl dNTPs 20mM each (NEB), 3µl oligo dT primer (Thermo-Fisher), 0.5µl SuperScirpt II (Thermo-Fisher) and 17.4µl H_2O. This mix was added to the RNA and incubated for 1 hour at 42°C. H_2O (200µl) was then added for use in PCRs.

2.1.9 Agarose Gel Electrophoresis

DNA fragments were separated for analysis on 1% agarose gels using electrophoresis. Agarose powder (GeneFlow LE Agarose) was dissolved in 1x Tris-Borate-EDTA (TBE) buffer at 1%, with 5 µl/100ml GelRed (Biotium Inc). Samples were loaded with die included in master mix or with the addition of 6x DNA Loading Buffer IV (NEB), additionally a 100bp Ladder (NEB) or 1kb Ladder (Invitrogen) was loaded depending on expected size of product. Gels were run at between 80-120V for between 30 minutes and 2 hours depending on product size. When necessary, DNA bands were excised with a clean scalpel and purified using Qiagen Gel Purification kit.

2.1.10 Plasmid Assembly and Amplification

Several plasmids were used during the project. The plasmid used for eYFP expression in protoplasts was pEarlyGate104 -eYFP. The plasmid used for CRISPR/Cas9 expression was L2 Dicot Dual guide plasmid, designed by Tom Lawrenson at JIC. In order to insert gRNAs into the plasmid two cloning reactions were performed to insert each gRNA. The gRNA forward and complement strands containing an overhang necessary to bind to the respective restriction site sticky end were created. The first

gRNA was inserted at the *Bsa1* site which flanked a *LacZ* gene allowing for selection once insertion is completed. The reaction was set up as follows: 200ng dual accepter vector, 1µl of each forward and reverse gRNA oligo pair from 2µM stocks, 1µl of 10x T4 ligase buffer, water to 8.5µL, 0.5µl (10 units) *Bsa1*(NEB), 1µl (400 units) T4 ligase. The reaction was then cycled as follows: 1x 20 seconds at 37ºC, 26x 37ºC for 3 minutes/16ºC for 4 minutes, 1x 50ºC for 5 minutes, 1x 80ºC for 5 minutes. The construct was then cloned and amplified ready to ligate in the second gRNA. This process was repeated for the second gRNA insertion at the *Esp31* site which flanks an RFP region which allowed for a similar colony selection. Plasmid assembly was confirmed by sequencing. Plasmids were then amplified to concentrations used for transformations with Qiagen Kits.

2.1.11 Cloning

Purified plasmid (2µl) was added to thawed 50µl One Shot™ TOP10 Chemically Competent *E. coli* (Thermo-Fisher) and incubated on ice for 30 minutes. The cells were then heat shocked at 42°C for 40 seconds and placed back onto ice. SOC media (250µl of 2% Tryptone, 0.5% Yeast Extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl2, 10 mM MgSO4, 20 mM glucose) was placed with the cells and incubated at 37°C for 1 hour. The cell suspension (100µl) was spread on an agar plate using appropriate antibiotic selection, and incubated overnight at 37°C. Colonies were then picked, checked by PCR and grown overnight on a shaker at 37°C in LB broth with appropriate antibiotics. Plasmids were then purified using the Qiagen Kit.

2.1.12 gRNA Design

gRNAs were designed using the CRISPR-P 2.0 website (http://crispr.hzau.edu.cn/CRISPR2/) and then checked manually using NCBI BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi). They were then ordered as isolated nucleotides (Sigma-Aldrich), both forward and complement strands containing an overhang necessary to clone into the L2 Dual Guide Plasmid.

2.1.13 Bioinformatics

Blast work was carried out using NCBI BlastN (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Alignments were completed using the SeqMan Pro software (DNAstar). qPCR data was analysed using the Agilent MxPro software.

2.2 RNA Sequencing

The RNAseq library generation and sequencing was carried out by the Warwick University School of Life Sciences in-house genomics lab for 3 *B. rapa* R-o-18 and 3 RLR22 RNA samples. RNA quality was analysed using an Aligent Bioanalyser. Libraries were then prepared using Illumina index adapter sequences. Sequencing was then carried out on an Illumina NextSeq as a 150bp, single-read, high-output run.

2.2.1 RNA Sequencing Analysis

The raw BCL files were converted to Fastq files using 'bcl2fastq' (Illumina). Quality control was carried out using 'fastqc'(Wingett et al. 2018) and 'multiqc' (Ewels et al. 2016). Genome and annotation files of *B. rapa* were obtained from Assembly: GCA900416815.2 (http://www.genoscope.cns.fr/externe/plants/chromosomes.html). Alignment and counting were carried out using the STAR software package (Dobin et al. 2012). Differential expression analysis was carried out using the DEseq2 (Love et al. 2014) R package. Raw alignment files were sorted and indexed for viewing on the IGV viewer (Robinson et al. 2011) using the Samtools software package (Li et al. 2009). The RNAseq analysis script in its entirety can be found in the appendix.

Alternative splicing analysis was carried out with the help of Richard Stark. The Trinity pipeline (Grabherr et al. 2011)was used to create a super-transcript annotation file from the original RNAseq reads. The reads were then aligned to this annotation file using STAR and individual exon usage was counted. The DEseq2 R package was then used to analyse the differential exon usage between R-o-18 and RLR22.

2.3 Plant Transformation Techniques

2.3.1 Protoplast Isolation

Leaves were harvested from 1-2 month old plants. Leaves were sliced into 1mm strips and placed into a 60mm petri dish. Enzyme solution (6ml of 20mM MES ph5.7, 0.4M Mannitol, 20mM KCL, 1.5% Cellulase R10 (Duchefa), 0.4% Macerozyme R10 (Duchefa) was added to leaf strips and vacuum infiltrated twice for 5 minutes at -0.8bar. The petri dishes were then incubated on a slow shaker at 25°C for 4 hours. Protoplasts were filtered through a 70 μm nylon cell strainer (Fisher) and rinsed with 6ml cold W5 buffer (2mM MES ph 5.7, 154mM NaCl, 125mM $CaCl_2$, 5mM KCL). Protoplasts were then centrifuged at 100g for 2 min and the supernatant removed, then washed in 6ml cold W5 buffer and re-pelleted. Protoplasts were then diluted in 1ml MMG solution. Depending on the experiment either 0.4M mannitol MMG (4mM MES pH 5.7, 0.4M mannitol, 15 mM $MgCl_2$) or 0.6M mannitol MMG (4mM MES pH 5.7, 0.6M mannitol, 15 mM $MgCl_2$) was used and protoplasts counted on a Fuchs-Rosenthal haemocytometer using a microscope.

2.3.2 Protoplast Transformation

In a 2ml round bottomed tube, 80,000 protoplast cells were suspended in 100 μl MMG solution. Desired plasmid DNA (8 μg) was added, followed by 150 μl PEG solution, then incubated for 10 minutes at room temperature. If 0.4M Mannitol MMG was used, a 0.2M Mannitol PEG solution (40% w/v PEG4000, 0.2 M Mannitol, 100mM $CaCl_2$) was used for transformation. If 0.6M mannitol MMG was used a 0.6M PEG solution (40% w/v PEG4000, 0.2 M Mannitol, 100mM $CaCl_2$) was used for transformation. 1ml W5 ($CaCl_2$ 100mM, NaCl 154mM, KCl 5mM, MES 2mM, pH5.7) was added to stop the reaction. It was centrifuged at 100g for 1 min, the supernatant removed, and the pellet resuspended in 500 μl WI solution (4 mM MES pH 5.7, 0.5 M mannitol, 20 mM KCL). The tube was sealed with micropore tape and incubated for 24/48 hours in 16h daylight at 23°C depending on the experiment. The protoplasts were then imaged with a florescent microscope or DNA extracted for analysis.

2.3.3 Brassica Transformation

Brassica plants were transformed as described by Hundleby et al. (2015). Seeds were surface sterilised and 20 germinated in 90mm petri dish on germination media (4.3 g/L MS salts, 30 g/L sucrose, pH 5.7, 8 g/L Phytagar, 1mg/L myoinositol, 10 mg/L thiamine-HCL, 1 mg/L pyridoxine, 1mg/L nicotinic acid) and incubated at 23°C in 16h day length for 4 days. *Agrobacterium tumefaciens* (strain AGL1, containing an appropriate plasmid) were grown in LB medium at 28°C for 48h then 50 µl of the bacterial suspension was added to LB and grown overnight at 28°C. This was then pelleted at 1000g for 5 min and the pellet resuspended in MS media (4.3 g/L MS salts, 30 g/L sucrose, pH 5.7) at OD_{650} = 0.1. Cotyledons were excised from seedlings by slicing though the petiole with a sharp scalpel and placed onto co-cultivation media (4.3 g/L MS salts, 30 g/L sucrose, pH 5.7, 8 g/L Phytagar, 2 mg/L BAP). The excised cotyledons were then inoculated by dipping the cut petiole into an agrobacterium suspension. Cotyledons were returned to the co-cultivation plates, sealed with micropore tape and incubated at 23°C in 16h day length for 72 hours. Cotyledons were then transferred to plates containing selection medium (4.3 g/L MS salts, 30 g/L sucrose, pH 5.7, 8 g/L Phytagar, 2 mg/L BAP, 160 mg/L Timentin, 15 mg/L Kanamycin). When necessary, calli were placed on to DEX induction plates (identical to co-cultivation plates with 20µM DEX).

2.3.4 Chemically Competent Agrobacterium

The AGL1 *Agrobacterium* strain was grown overnight on a shaker at 28°C in 10ml LB broth with 100µg/ml carbenicillin and 150µg/ml rifampicin. The overnight culture (2ml) was then added to 50ml LB broth containing antibiotics in a 250ml flask. This was then incubated on shaker at 28°C until the culture grew to an OD_{600} of 0.5-1. The culture was then chilled on ice and spun at 3000g for 5 minutes at 4°C. The supernatant was discarded and cells resuspended in 1ml of ice-cold sterile 20mM $CaCl_2$. Aliquots (100µl) were made and cells snap frozen in liquid nitrogen and stored -80°C. To transform the *Agrobacterium*, 1µg DNA was added and cells heat shocked at 37°C for 5 minutes. LB broth (1ml) was then added and cells incubated at 28°C for 3 hours.

Then cells were spread on LB plates containing the appropriate antibiotics and incubated at 28°C for 2-3 days.

2.3.5 Carbon Nano-Dot Application

Carbon Nano-Dots were made by H. Whitney, University of Bristol. To bind plasmid DNA to the CDs, 85µl TE Buffer, 85µl plasmid DNA (200ng/ul) and 85µl CDs (2ug/ul) are placed into a 50ml falcon tube. Agitated by hand and incubated at room temperature for 5 minutes. 20ml H_2O is then added to the tube and placed into a spray bottle. Plants were grown to the 4 true-leaf stage and then were sprayed once a day for 4 consecutive days. The treatments were either CD+DNA (85µl TE buffer, 85µl DNA, 85µl CND in 20ml distilled water) or only DNA -ve DNA (85µl TE buffer, 85µl_DNA in 20ml distilled water). Plants were then visualised on a confocal microscope (Zeiss LSM 880), looking for YFP expression (527nm) and using Chlorophyll A (735 nm), 4 days after the last spray.

2.4 Plant Growing Conditions

Plants were grown either in a growth cabinet (Panasonic MLR-352-PE), for protoplast and transformation work, or in a glasshouse, for everything else. The growth cabinet was maintained at 23°C under 16-h day length of 70 µmol/m 2 /s and watered every other day. Plants were grown in P7 (7mm) pots in Levington Advance M2 potting and bedding media. The glasshouse was maintained at 18C +/- 2°C under 16-h supplementary lighting and plants watered every day.

2.5 TuMV Inoculation

Plants to be infected with TuMV at 4-5 true leaf stage were sprayed with carborundum powder. Infected tissue from a donor plant was ground with a pestle and mortar with a small amount of inoculation buffer (K_2HPO_4, 10g/L; Na_2SO_3 ,1g/L). This mixture was then wiped on the leaves with a piece of muslin fabric. Inoculated leaves were

marked with a punch in order to identify systemic infections. All inoculations were carried out using TuMV isolate GBR 6 (pathotype 4). Leaves from TuMV challenged plants were imaged on a light box.

2.6 ELISA Assay

Enzyme-Linked Immunosorbent assays (ELISA) were carried out as follows; leaf samples were ground through electric rollers and the sap collected in Eppendorf tubes. Each sample (100µl) was then pipetted into two wells of a 96 well plate (Thermofisher) containing 100µl of coating buffer (15mM Na_2CO_3, 35mM Na_2HCO_3) and incubated overnight at 4°C. The samples were flicked out of the plate and wells washed by filling with 1X phosphate-buffered saline pH 7.3 with 0.05% Tween 20 (PBS-T) and soaked at room temperature for 3 minutes. The PBS-T was tapped out and the wash procedure carried out two additional times. PBS-T with 0.2% bovine serum albumin was then used to dilute the TuMV antibody, EMA67 (Jenner et al. 1999), to a concentration of 1/500. Diluted antibody (150µl) was then added to each well, the plate covered and incubated for 2 hours at room temperature. The plate was then washed as before 3 times with PBS-T. The secondary antibody, goat anti-mouse alkaline phosphatase conjugate (Sigma A-3562), was diluted to 1/2000 using PBS-T with 0.2% bovine serum albumin. Diluted secondary antibody (150µl) was then added to each well, the plate covered and incubated for 2 hours at room temperature. The plate was then washed as before 3 times with PBS-T. Substrate solution (150µl of 9.7% diethanolamie pH 9.8, 1 alkaline phosphatase substrate tablet/5ml) was then added to each well. The plate was then incubated at room temperature until reaction had progressed sufficiently to allow visualization. The plate was then read on an ELISA plate reader (Anthos, model no. 37506) at wavelength 405nm.

Chapter 3

Paralogs of the eIF4E Translation Initiation Complex

3.1 Introduction

Previous work found that alternative splicing of a single gene, *eIF(iso)4E*, was able to confer recessive resistance in the *B. rapa* line RLR22 (Jenner et al. 2010). This gene is part of the *eIF4E* gene family of which there are three members, two isoforms of *eIF4E* and a *Novel Cap Binding Protein* (*nCBP*) (Ruud et al. 1998). Nellist et al. (2014) found there were three copies of both *eIF4E* and *eIF(Iso)4E* in the *B. rapa* genome, however did not investigate *nCBP* in *B. rapa.* Within the eIF4F translation initiation complex there are several other proteins present, including eIF4G and eIF4A, which are also involved indirectly with cap binding of mRNA and which are highjacked by the potyvirus VPg. Knockouts of *eIF(iso)4G* copies in *A. thaliana* resulted in TuMV resistant lines (Nicaise et al. 2007). Additionally, eIF4E-dependent translation initiation has been described which involves the use of eIF4A and eIF4G (Gallie 2001). Since this work, the genome of *B. rapa* (Chiifu-401) has been published (Wang et al. 2011) as well as the genome of *B. oleracea (Liu et al. 2014)* allowing for the reinvestigation of the genetic make-up of the eIF4F translation initiation complex.

3.2 Results

3.2.1 BLAST Analysis of eIF4E Components

Six genes of interest involved in the eIF4F translation initiation complex were identified from the literature and previous work. They included *eIF4E, eIF4(iso)4E, eIF4G, eF(iso)4G, eIF4A,* and *nCBP*. For each of the genes, the *A. thaliana* paralogous sequence was obtained from NCBI. These were then compared against the *B. rapa* and *B. oleracea* genomes using BLASTn (NCBI). Sequences with e-values of less than 0.0005 were identified as orthologs of the *A. thaliana* sequences. In *B. rapa* 5 copies of *eIF4E*, 3 copies of *eIF(iso)4E*, 1 copy of *eIF4G*, 5 copies of *eIF(iso)4G*, 11 copies of *eIF4A*, and 3 copies of *nCBP* were identified (Table 3.1). In *B. oleracea* 5 copies of *eIF4E*, 3 copies of *eIF(iso)4E*, 1 copy of *eIF4G*, 3 copies of *eIF(iso)4G*, 10 copies of *eIF4A*, and 3 copies of *nCBP* were identified (Table 3.1). These sequences were then aligned and the most probable orthologs, based on sequence identity, were identified between *B. rapa* and *B. oleracea* (Table 3.1). A large amount of sequence conservation was seen between the *A. thaliana* and both *Brassicas* of 60-70% across the genes of interest. A further alignment was carried out between all the paralogous sequences from *B. rapa* and *B. oleracea* respectively. Between the *Brassicas* sequence conservation of 80-90% was seen across the genes of interest. The non-consensus regions of each alignment were then used to design primers able to amplify specific paralog copies within their respective genomes, both in genomic DNA and cDNA (Appendix Table 1).

3.2.2 Genomic DNA Analysis

Genomic DNA analysis was carried out across all paralogs in all six genes of interest, in order to compare variation across the *Brassica* genus. The sequences were amplified in eight *Brassica* lines, including two *B. juncea* (TGM, MSD), two *B. rapa* (R-o-18, RLR22) and four *B. oleracea* (A12, BRO5058, BRO5060, BRO5058). These were then sequenced and aligned against the *B. rapa* (Chiifu-401) reference gDNA and cDNA. In all gene alignments only small variations were seen, such as point mutations and small 5-10bp indel mutations in the intronic regions (Figure 3.1).

Table 3.1 – Orthologs of *Arabidopsis thaliana* Genes of interest in *Brassica oleracea* and *Brassica rapa*

Species	*A. thaliana*	*B. rapa*	*B. oleracea*
Gene Name	TAIR Gene ID	Location[1]	Location
eIF4E	AT1G29590	A9	C5
		A7-104	C7-168
		A3	C7-408
	AT4G18040	A8	C8
	AT1G29550	A1	C1
		XA7- 837	XC6-461
			XC02-910
			XC02-969
eIF(iso)4E	AT5G35620	A04	C4
		A8	C8
		A5	C6
eIF4G	AT3G60240	A4	C4
eIF(iso)4G	AT5G57870	A2	C2
		A3	C7
	AT2G24050	A1-330	C1
		A8-135	
	AT4G30680	A1-326	
eIF4A		A8-607	C3-641
	AT3G13920	A3-168	C3-255
		A1	C1
	AT3G19760	A8-138	C6-140
		A7	C6-326
	AT1G54270	A6	C6-832
	AT1G72730	A2	C2-218
	AT1G51380	A5	C5
		A3-647	C3-836
		A8-153	C2-379
		A4	
nCBP	AT5G18110	A10	C9
		A3	C3
		A2	C2

[1] Location indicates which chromosome the ortholog is present on, if two orthologs are on the same chromosome the first 3 numbers of its specific location are included. Locations marked with X were later deemed to not be true orthologs

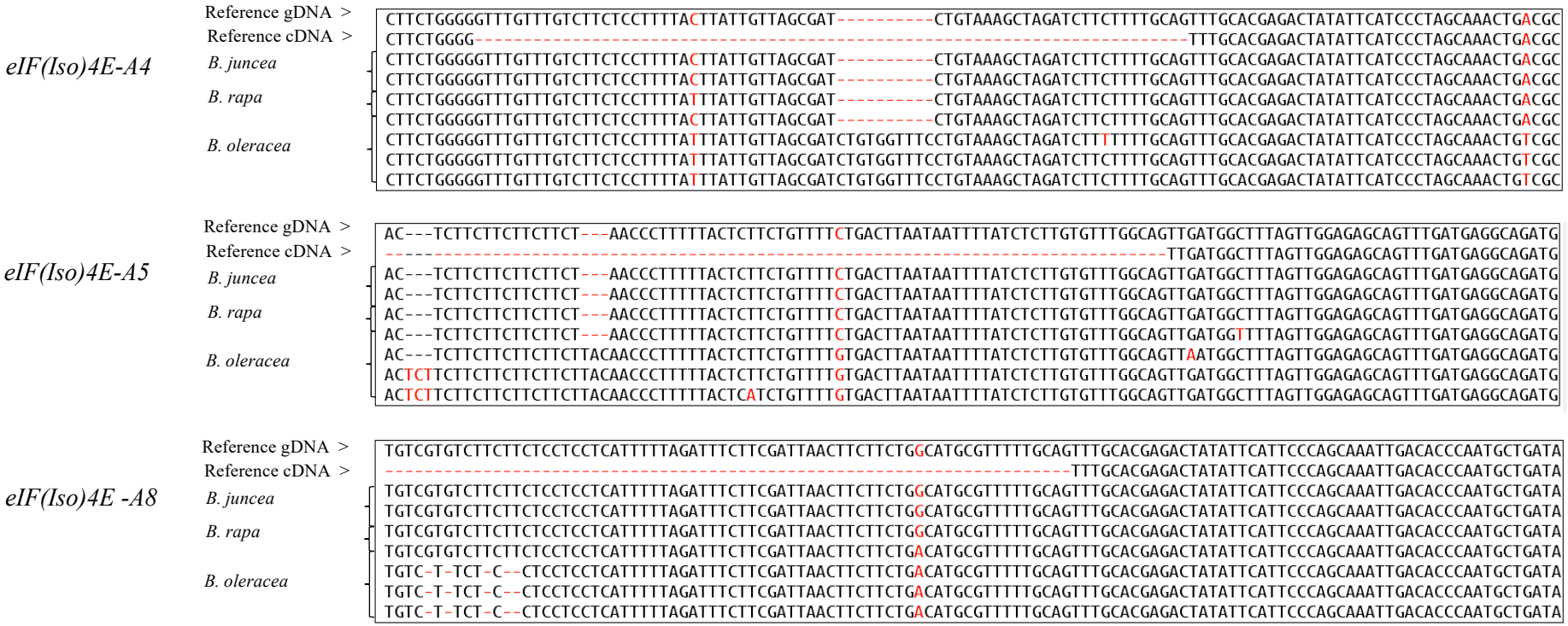


Figure 3.1- Example Alignment of Three *EIF(iso)4E Paralogs*

The three *eIF(iso)4E* paralogs after gDNA amplification and sequencing with paralog specific primers. Point mutations are highlighted in red. Small insertions in the intronic regions are seen in *eIF(iso)4-A4* and *eIF(iso)4E-A5* in the *B. oleracea* sequences. Small deletions in the intronic regions are seen in *eIF(iso)4E-A8 B. oleracea* sequences. The Reference gDNA and cDNA is *B. rapa* (Chifu-401).

3.2.3 qPCR Analysis

The primers designed for cDNA amplification (Appendix Table 1) were used in qPCR to investigate if there were any differences in expression of the genes of interest between TuMV-inoculated and mock-inoculated plants. In both *B. rapa* and *B. oleracea* no significant differences in relative quantity of transcripts was seen in any of the genes of interest (Figure 3.2).

The qPCR analysis required some optimisation of primers. To do this, primers were designed to amplify the entire cDNA of several genes in order to confirm their sequences. When the total cDNA of *eIF4E-A9* was amplified and analysed on an agarose gel several bands were present (Figure 3.3). Following cloning of the PCR products and sequencing, the larger band was found to be a splice variant of *eIF4E-A9* in which the intron regions between exons 3 and 4 and exons 5 and 6 had been retained. Further analysis of the predicted protein sequence produced by this transcript showed that a premature stop codon was introduced in the included intron region between exons 3 and 4 (Figure 3.4).

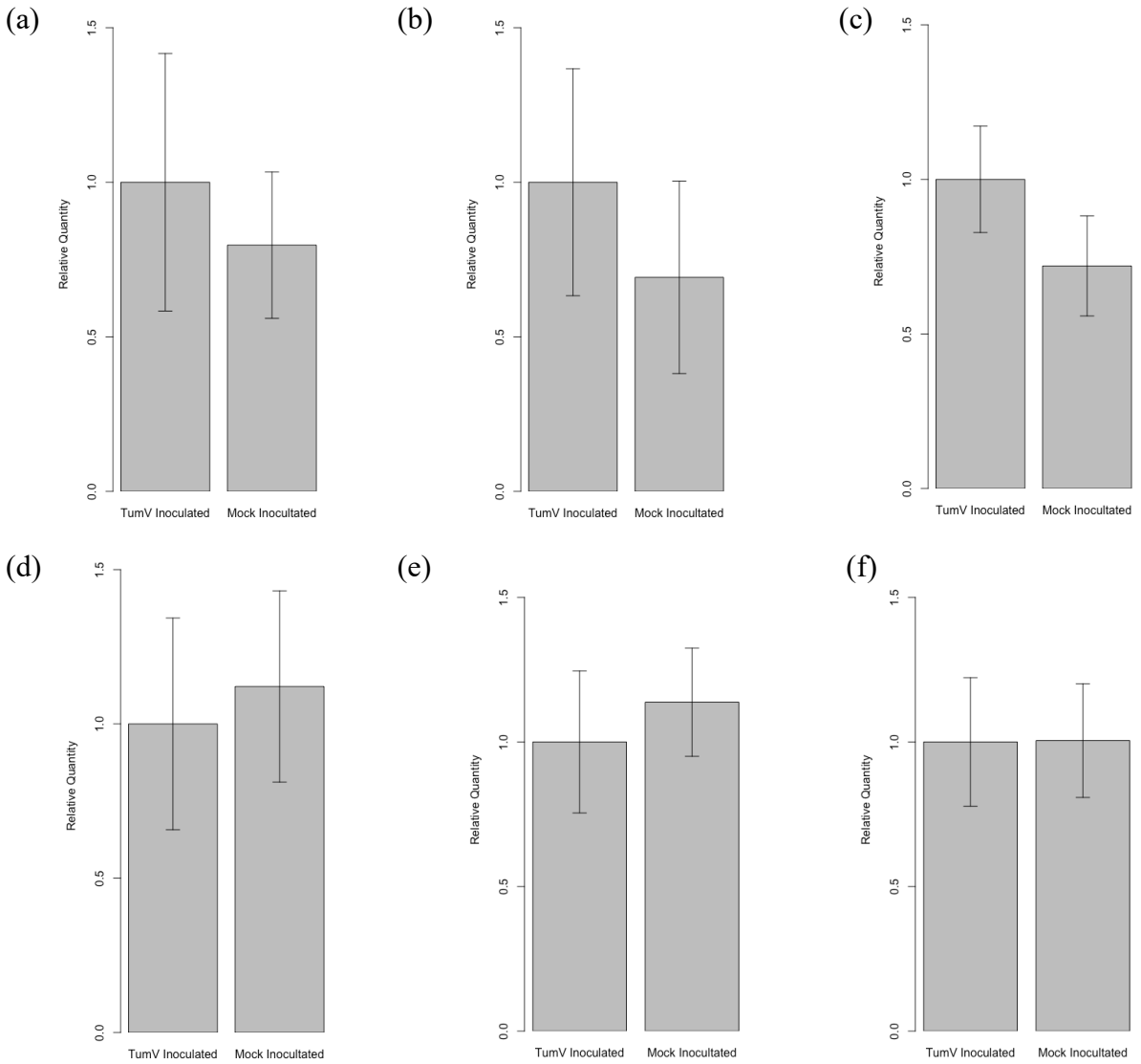


Figure 3.2- Example plots of relative quantity of transcripts of the three *eIF(Iso)4E* paralogs.

(a) *eIF(Iso)4E-A4* in *B. rapa* (R-o-18). (b) *eIF(Iso)4E-A5* in *B. rapa* (R-o-18). (c) *eIF(Iso)4E-A8* in *B. rapa* (R-o-18). (d) *eIF(Iso)4E-A4* in *B. oleracea* (A12). (e) *eIF(Iso)4E-A5* in *B. oleracea* (A12).(f) *eIF(Iso)4E-A8* in *B. oleracea* (A12). All samples were normalised using the housekeeping genes β-*tublin, TIP41* and *Actin-7*. PCR products were verified by sequencing.

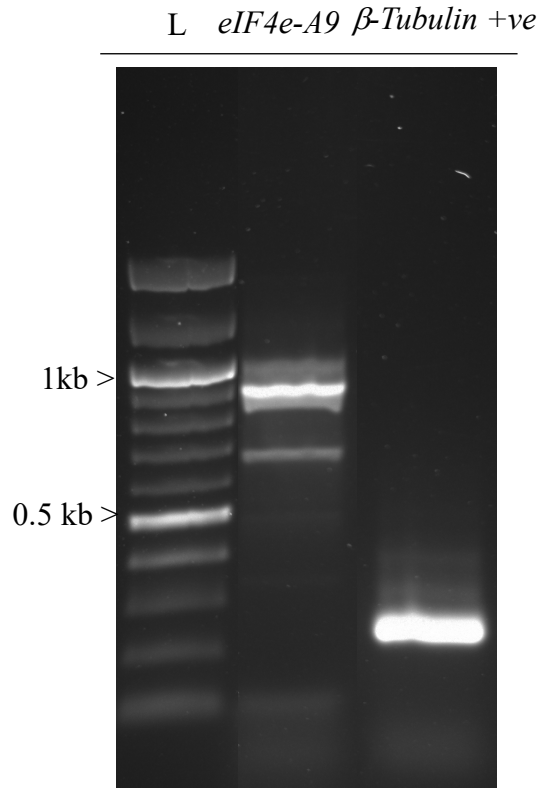


Figure 3.3- PCR Amplification of *eIF4E-A9* in *B.rapa* (R-o-18)

PCR amplification of *eIF4E-A9* with total length cDNA primers (F-LI79, R- LI104). Several bands are visible, both larger and smaller than the expected band size of 858bp. 1kb ladder and *β-tublin* positive (248bp) are also shown.

Figure 3.4 - Alignment of cloned *eIF4E-A9* PCR product from *B. rapa* (R-o-18)

Alignment of cloned *eIF4E-A9* PCR product to Reference gDNA and cDNA. Clone 1 includes the intron region between 687bp and 793bp, as well as between 1,132bp and 1,221bp. The introduced premature stop codon in *eIF4E-A9*-clone 1 is boxed in red. The Reference gDNA and cDNA is *B. rapa* (Chifu-401).

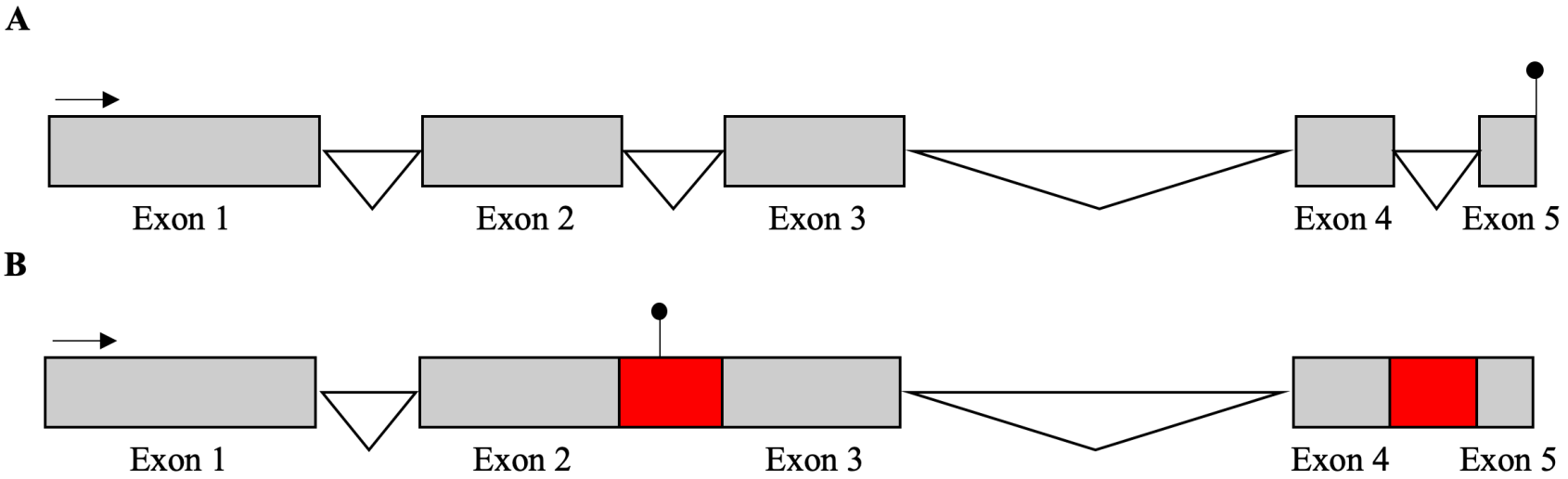


Figure 3.5 – Diagram of *eIF(iso)4E-A9* mRNA Structure

(A) Diagram of the native *eIF(iso)4E-A9* mRNA as seen in the diploid *Brassicas*, with 5 exons and the stop codon at the end of exon 5. (B) Diagram of the *eIF(iso)4E-A9* alternative splice-form seen when amplifying the gene in *B. rapa* R-o-18, with two intronic regions included and a stop codon introduced in the region between exon 2 and exon 3.

3.2.4 *B. rapa* RNAseq analysis

The gene expression profiles of known TuMV resistant, RLR22, and non-resistant, R-o-18, *B. rapa* lines were compared using RNAseq. Both alternative splicing and differential expression were analysed. Principal component analysis of the data shows two distinct principal components made up of the RLR22 data and the R-o-18 data (Figure 3.6).

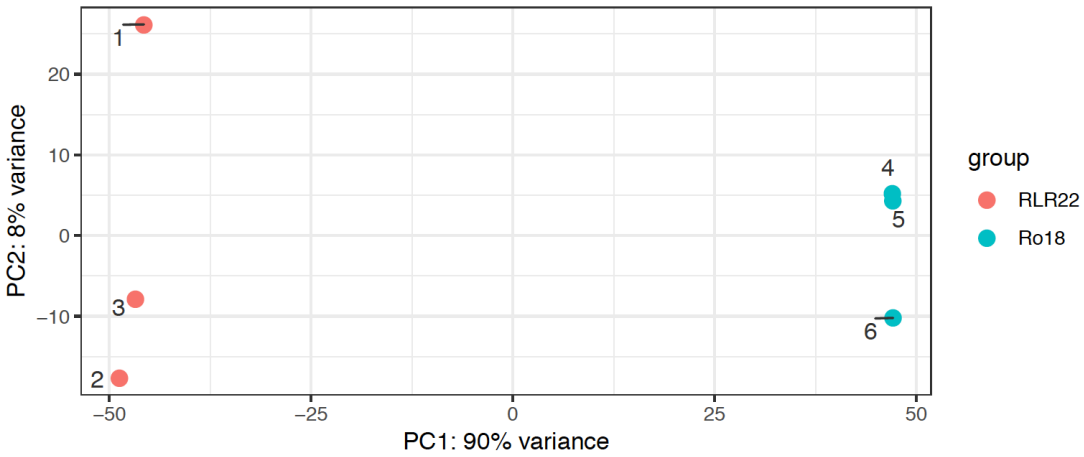


Figure 3.6 – Principal component analysis of the RNAseq Data

Principal component analysis of the RNAseq transcript count data taken, shows two distinct groups of data allowing for further analysis. Each point represents one RNA leaf tissue sample taken a six weeks from each of their respective *B. rapa* lines (R-o-18 or RLR22) which has then been sequenced in triplicate.

A volcano plot of differential gene expression shows a large number of genes are significantly differentially expressed between the R-o-18 and RLR22 across all 6 replicates (Figure 3.7). This plot of significance Log_{10} vs fold change allowed identification of genes with the largest fold change that are also statistically significant. The XY plot shows the spread of expression of genes between the replicate average of R-o-18 and RLR22, the top differentially expressed genes are highlighted in red where the significance *p*-value is greater than 0.0001 (Figure 3.8).

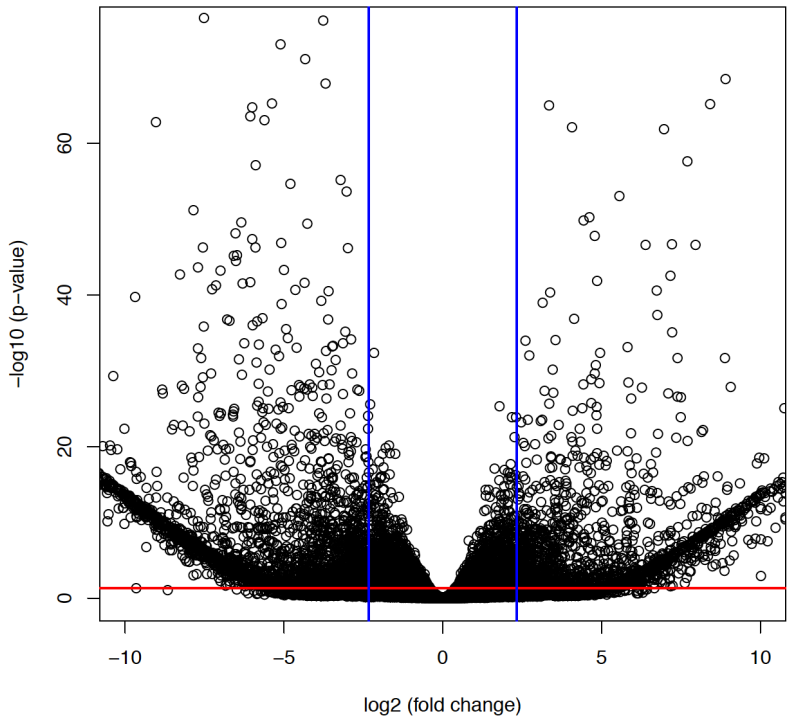


Figure 3.7 - Volcano plot of differentially expressed genes

Y axis is the significance of differential expression between RLR22 and R-o-18 with the red line defining a significance cut off of 0.05. X axis is the $\log_{10}$ of fold change between RLR22 and R-o-18 with the blue lines defining a significance cut off of 0.05.

Differential expression of the genes of interest was analysed from the data. Six genes showed differential expression in RLR22 compared to R-0-18; *eIF4E-A7-104* (1.29 Log2FoldChange), *eIF4E-A3* (-6.41 Log2FoldChange), *eIF(Iso)4E-A4* (-0.89 Log2FoldChange), *eIF(Iso)4E-A8* (1.11 Log2FoldChange), *nCBP-A10* (2.73 Log2FoldChange) *and nCBP-A3* (0.96 Log2FoldChange). *nCBP-A10* showed the most significate difference between the two samples (adjusted p-value of 9.45E-12) with a 2.72 fold change (Table 3.2).

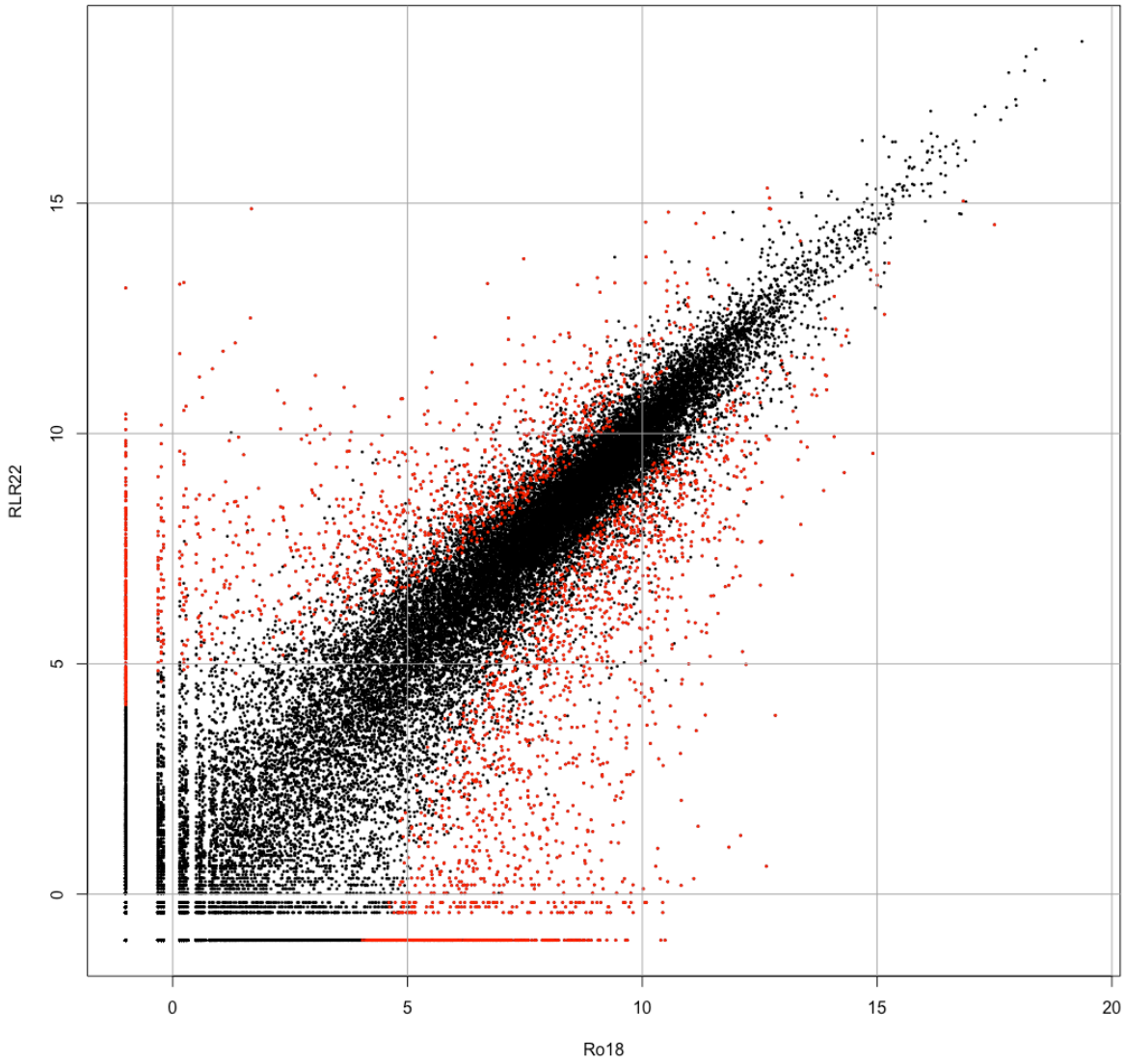


Figure 3.8 - XY plot of the $\log_2$ gene expression values of RLR22 Vs R-o-18
Genes with a differential expression significance of lower than 0.0001 are highlighted in red.

Table 3.2 - Differential expression of the genes of interest

Differential expression of the genes of interest within R-o-18 and RLR22 comparison of transcript data. Six genes showed statistically significant differential expression *eIF4E-A7-104, eIF4E-A3, eIF(Iso)4E-A4, eIF(Iso)4E-A8, nCBP-A10 and nCBP-A3*

Gene Name	Chromosome	Geneoscope ID	baseMean[1]	log2FoldChange[2]	lfcSE[3]	Stat[4]	Pvalue[5]	Padj[6]	DE[7]
eIF4E	A9	BraA09t39200Z	3.51	0.93	1.36	0.69	0.49	0.68	No
	A7-104	BraA07t28852Z	44.81	1.29	0.40	3.24	0.00	0.01	Yes
	A3	BraA03t14185Z	7.43	-6.41	1.53	-4.18	0.00	0.00	Yes
	A8	BraA08t33323Z	1152.82	-0.43	0.28	-1.56	0.12	0.27	No
	A1	BraA01t00953Z	412.14	0.27	0.23	1.17	0.24	0.43	No
eIF(iso)4E	A4	BraA04t16641Z	266.53	-0.89	0.28	-3.13	0.00	0.01	Yes
	A8	BraA08t32975Z	547.46	1.11	0.33	3.34	0.00	0.00	Yes
	A5	BraA05t20553Z	N/A		Due to no transcript in annotation file				No
eIF4G	A4	BraA04t15901Z	2450.07	-0.01	0.24	-0.04	0.97	0.98	No
eIF(iso)4G	A2	BraA02t05888Z	3151.30	0.26	0.16	1.65	0.10	0.23	No
	A3	BraA03t15064Z	440.26	-0.03	0.33	-0.11	0.92	0.96	No
	A1-330	BraA01t00698Z	1136.93	0.47	0.43	1.08	0.28	0.48	No
	A8-135	BraA08t33872Z	29.01	0.34	0.61	0.56	0.57	0.74	No
	A1-326	BraA01t00691Z	49.85	1.17	0.48	2.42	0.02	0.05	No
eIF4A	A8-607	BraA08t32161Z	1385.91	-0.12	0.33	-0.37	0.71	0.84	No
	A3-168	BraA03t13003Z	1101.36	0.02	0.19	0.12	0.90	0.95	No
	A1	BraA01t04011Z	4389.07	0.07	0.27	0.26	0.80	0.89	No
	A8-138	BraA05t20661Z	4133.94	0.28	0.23	1.24	0.21	0.40	No
	A7	BraA07t31441Z	3.47	-3.33	1.51	-2.21	0.03	0.09	No
	A6	BraA06t23256Z	N/A						No
	A2	BraA02t06862Z	N/A						No
	A5	BraA05t22359Z	2526.00	0.17	0.28	0.60	0.55	0.73	No
	A3-647	BraA06t23256Z	N/A						No
	A8-153	BraA08t34210Z	178.76	0.03	0.32	0.10	0.92	0.96	No
	A4	BraA01t02053Z	10.65	2.11	0.88	2.40	0.02	0.06	No
nCBP	A10	BraA10t44202Z	75.44	2.73	0.37	7.30	0.00	0.00	Yes
	A3	BraA03t10188Z	119.41	0.96	0.33	2.87	0.00	0.02	Yes
	A2	BraA02t05433Z	584.09	0.19	0.23	0.83	0.41	0.61	No

[1] baseMean refers to the mean of normalized counts of all samples, normalizing for sequencing depth.

[2] Log2FoldChange refers to the log-ratio of a gene's expression values in the two different *B. rapa* lines

[3] lfcSE refers to the Standard error of the Log2FoldChange

[4] Stat refers to the z-statistc after running a Wald test for significance where the Log2FoldChange is divided by its standard error

[5]P-value is the significance value given by the Wald tests

[6] Padj is the p-value adjusted using the Benjamini and Hochberg procedure which accounts for false positives when doing multiple comparisons

3.2.5 Alternative Splice-form Analysis

Alternative splicing (AS) was identified by chance in the *eIF4E-A9* paralog in *B. rapa* (R-o-18) during testing of primers for qPCR. Analysis of the RNAseq data was carried out in order to further investigate the AS events in the eIF4F complex and in *B. rapa* as a whole. The raw RNAseq data was used in the Trinity software pipeline to create a novel super-transcript annotation file, allowing for the understanding of the transcriptional complexity of the gene. DESeq2 software package was then used to compare alternative splicing between the two *B. rapa* lines analysed, R-o-18 and RLR22. It was found that, of the transcripts amplified, 14311 of 74434 genes where alternatively spliced with a significance value of less than 0.01. This is 19% of all genes amplified in this experiment showing some form of alternative splicing. This includes retained introns (RI), skipped exons (SE), mutually exclusive exons (MEE) or alternative splice sites (ASS).

Alternative splicing of the genes of interest was also analysed by looking at the read alignments in IGV genome viewer. Of the genes of interest 15 showed alternative splicing. This included the *eIF4E-A9* paralog in *B. rapa* R-o-18 showing that alternative splicing is taking place (Table 3.3). In addition, confirmation of the alternative splicing of the RLR22 *eIF(iso)4E-A4* resistance paralog was also seen.

Table 3.3 – Alternative Splicing of the Genes of Interest in *Brassica rapa*

Gene Name[1]	Chromosome	Geneoscope_ID	Alternative Splicing	AS Event
eIF4E	A9	BraA09t39200Z	Yes	RI
	A7-104	BraA07t28852Z	Yes	RI/ASS
	A3	BraA03t14185Z	No	
	A8	BraA08t33323Z	No	
	A1	BraA01t00953Z	Yes	RI/SE
eIF(iso)4E	A4	BraA04t16641Z	Yes	RI
	A8	BraA08t32975Z	Yes	RI
	A5	BraA05t20553Z	N/A	
eIF4G	A4	BraA04t15901Z	Yes	RI/SE
eIF(iso)4G	A2	BraA02t05888Z	Yes	RI/ASS
	A3	BraA03t15064Z	Yes	SE
	A1-330	BraA01t00698Z	No	
	A8-135	BraA08t33872Z	No	
	A1-326	BraA01t00691Z	Yes	SE
eIF4A	A8-607	BraA08t32161Z	Yes	RI
	A3-168	BraA03t13003Z	Yes	RI/ASS
	A1	BraA01t04011Z	Yes	SE
	A8-138	BraA05t20661Z	No	
	A7	BraA07t31441Z	No	
	A6	BraA06t23256Z	N/A	
	A2	BraA02t06862Z	N/A	
	A5	BraA05t22359Z	No	
	A3-647	BraA06t23256Z	No	
	A8-153	BraA08t34210Z	Yes	RI
	A4	BraA01t02053Z	No	
nCBP	A10	BraA10t44202Z	Yes	RI/SE
	A3	BraA03t10188Z	No	
	A2	BraA02t05433Z	Yes	RI

[1] The AS event refers to the type of splicing observed, retained introns (RI), skipped exons (SE), mutually exclusive exons (MEE) or alternative splice sites (ASS).

3.3 Discussion

The analysis of the genetic make-up of the genes that contribute to the eIF4F initiation complex, using the published genome of *B. rapa* (Chiifu-401) (Wang et al. 2011) confirmed the presence of several extra paralogs to *eIF4E, eIF4A* and *eIF(iso)4G* (Table 3.1) . This was a useful starting point from which to further investigate the eIF4F translation initiation complex, as comparison of the paralogs allowed identification of orthologs between *B. rapa* and *B. oleracea.* In addition, this identified regions of non-consensus sequence between the paralogs which made it possible to design primers which were able to amplify individual paralogs both in cDNA and genomic DNA (Appendix Table 1).

Amplification of the genes of interest in the genomic DNA across *B. juncea, B. oleracea and B. rapa* revealed that these genes are highly conserved and the only differences identified were point mutations and some InDel mutations (Figure 3.1). This was to be expected, however, as all of these genes are expressed through all stages of the plant life and are involved in the essential process of translation. The InDel mutations were interesting to some extent, however, as they also seemed to be conserved across several lines which may infer that they may play some role in the genes' regulation, expression or processing.

The qPCR analysis of the expression between TuMV-infected and mock-inoculated samples in R-o-18 did not show any significant differences in expression (Figure 3.1). Although this does not aid in the identification of novel targets for resistance, it did confirm that all the genes were being expressed to some extent. It is therefore interesting that in the resistant *B. rapa* line (RLR22) a mutation in only a single paralog of the *eIF(iso)4E* gene confers resistance, as all other paralogs are being expressed and share close sequence similarity, one might expect the virus to be able to utilise more than the single paralog.

PCR amplification of the *eIF4E-A9* full cDNA sequence in R-o-18 gave several bands on the agarose gel both larger and smaller than the expected product size (Figure 3.3). Cloning of these PCR products and subsequent sequencing revealed the larger band to be a splice variant of *eIF4E-A9* in which intronic regions had been included in the mRNA. Despite numerous attempts, the smaller band was unable to be cloned. Analysis of the splice variant sequence revealed that a stop codon had been introduced between exons 3 and 4. This is similar to the mutation that was introduced in the *B. rapa* resistant line (RLR22) in which a stop codon was introduced between exons 1 and 2 (Nellist et al. 2014).

RNAseq analysis was carried out comparing *B. rapa* TuMV resistant (RLR22) and non-resistant (R-o-18) lines in order to see differential expression between the two lines as well as look for further splice variants in the genes of interest. There was a large number of genes seen with differential expression between the two samples (Table 3.2 & Table 3.3). To some extent the fact that R-o-18 and RLR22 are two different subspecies, *ssp trilocularis* and *ssp pekinensis* respectively, might explain some differences, however it may also identify some potential genes of interest in the context of resistance. Additionally, of the genes being investigated, the fact that *nCBP-A10* showed a marked difference in expression is interesting because previous work has shown that modification of this gene in *cassava* resulted in reduced susceptibility.

The identification of alternative splicing of many of the genes of interest has not been reported before. There is the potential that these splice variants would be removed by the plant's native proof-reading machinery. However, it should be noted that many of the splice variants had retained intron events which are similar to that seen in the *eIF(iso)4E-A4* paralog in RLR22 that confers resistance.

In addition, many genes were alternatively spliced when comparing between R-o-18 and RLR22. This again could be explained by the difference in subspecies, however, the reason for alternative splice forms is unknown and would be an interesting vein of research.

Chapter 4

CRISPR/Cas9 in *B. oleracea*

4.1 Introduction

Genetic modification of crop species is a key tool to help understanding of gene function as well as to improve yield and other favourable traits. The CRISPR/Cas9 system has been shown as an effective way of introducing knock-out modifications within plant genomes (Lawrenson et al. 2015, Chandrasekaran et al. 2016, Yang et al. 2017, Sun et al. 2018, Bastet et al. 2019, Gomez et al. 2019). In order to guide the Cas9 machinery to a target gene guide RNAs (gRNAs) are designed that are complementary to the target sequence, the Cas9 protein then introduces a double-stranded break which is repaired by non-homologous end-joining which will introduce an indel mutation often resulting in a knock-out of the target gene (Jinek et al. 2012). The CRISPR system has also been proven to function in one *B. oleracea* line (DH1012) and produce heritable mutations which are passed to subsequent generations (Lawrenson et al. 2015).

The *B. rapa* (RLR22) resistance to TuMV previously identified has been shown to be the result of a truncation of the *eIF(Iso)4E-A4* gene (Nellist et al. 2014). However, to date, no equivalent resistance to TuMV has been found in *B. oleracea* using traditional screening methods or in the breeding population.

Using the CRISPR/Cas9 system it has been shown that knocking out the *eIF(iso)4E-A4* ortholog in *A. thaliana* and *eIF(iso)4E Cucumis sativus* (cucumber) confers resistance to specific potyviruses (Chandrasekaran et al. 2016, Pyott et al. 2016). The CRISPR/Cas9 system was used to investigate whether a similar resistance could be conferred in *B. oleracea* by knocking out the *eIF(iso)4E-A4* ortholog *eIF(iso)4E-C4.*

4.2 Results

4.2.1 DH1012 Susceptibility to TuMV

To demonstrate DH1012 was susceptible to TuMV infection and therefore could be used as a model for a potential resistance phenotype after a CRISPR mutation was introduced, the CRISPR/Cas9 amenable line was challenged with TuMV. Examination of plant material inoculated with TuMV (GBR 6) showed the characteristic mosaic pattern in systemic leaf tissue 3 weeks after infection (Figure 4.1). In addition, ELISA analysis was carried out to further verify infection. All plants in ELISAs that had been inoculated with TuMV were shown to have greater absorbance than the non-infected control by more than plus 2 standard deviations (Table 4.1). This included the sample in which the systemic and infected tissues were analysed separately.

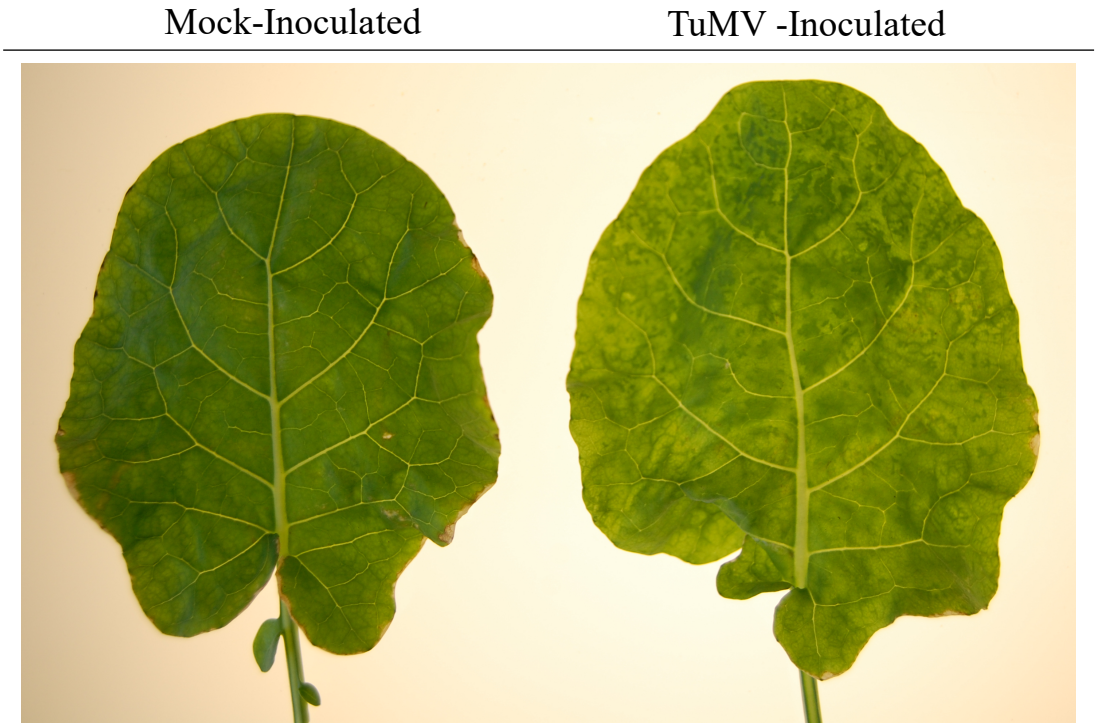


Figure 4.1 – DH1012 Leaf TuMV Infection

Leaves from Mock-inoculated (left) and TuMV-inoculated (right) plants on light box. Mosaic pattern is visible in the TuMV-inoculated leaf.

Table 4.1 – ELISA Results of DH1012 TuMV Inoculation

Sample #	Treatment	Average Absorbance	Within > Control +2 SD
1	TuMV -Inoculated Plant	1.4875	No
2	TuMV -Inoculated Plant	0.5105	No
3	TuMV -Inoculated Plant	0.4610	No
4	TuMV -Inoculated Plant	0.4735	No
5	TuMV - Inoculated Leaf [1]	0.1945	No
5	TuMV-Uninoculated Systemic Leaf	1.1565	No
6	Mock-Inoculated	0.0525	Yes
	Susceptible Positive	0.6263	
	Susceptible Negative	0.0515	
	Empty wells	0.0145	
	Standard Deviation of Susceptible Positive	0.0007	

[1] All TuMV-inoculated samples showed an absorbance greater than 2 standard deviations above the known susceptible negative. Samples 1-4 are replicates of each other with two systemic leaves taken from each plant then processed for ELISA. Sample 5 the inoculated leaves and the uninoculated systemic leaves were tested separately. Sample 6 two leaves were taken from a mock inoculated plant. *B. juncea* line TGM was used as a susceptible control. All samples were run in triplicate on the ELISA plate.

4.2.2 Guide RNA Design

Sequencing of the DH1012 *eIF(iso)4E-C4* paralog was carried out using the previously designed specific *B. oleracea* primers (Appendix Table 1). This, in conjunction with the previously collected sequencing data, was used to design guide RNAs specific for the *eIF(iso)4E-C4* paralog. Four guides were designed by Tom Lawrenson at JIC using the sequencing I provided and four guides were designed by myself using the CRISPR-P2.0 webtool and then manually checked using NCBI-BLAST (Table 4.2) (Figure 4.2).

Table 4.2 – Guide RNAs Designed to target *eIF(iso)4E-C4*

[1]Name	Sequence (5'-3')
gRNA-C4-A6	GAGCCAAGCCTTGTCTAAAGC
gRNA-C4-B6	GTGGCAGTTGATGGCTCTTGT
gRNA-D4-C6	GCAGAACAGCTTCATTAGAT
gRNA-D4-D6	GACTTCTTCCCAATACCCATC
LI-gRNA-2	GCGAAGGGAGGCTCCCCAGG
LI-gRNA-6	ACAATAATGGGAGATCGCAC
LI-gRNA-3	AGTACCGGCAACAGAGACGA
LI-gRNA-7	TTACGTGGCTTACTAAAATG

[1] The first four guide RNAs were designed by Tom Lawrenson at JIC. The second four guide RNAs were designed by myself following no edits being found in the T0 plantlets generated using the first four gRNAs. The four LI gRNAs were used in the protoplast work.

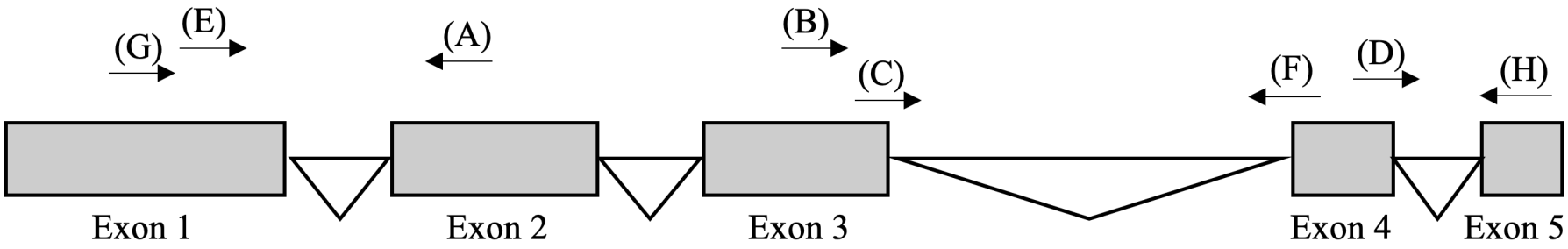


Figure 4.2- Diagram of location of Guide RNAs in *eIF(iso)4E-C4*

Location of 20bp guides is indicated by arrows with their direction depicting if on the forward or reverse strand. (A) gRNA-C4-A6 (B) gRNA-C4-B6 (C) gRNA-D4-C6 (D) gRNA-D4-D6 (E) LI-gRNA-2 (F) LI-gRNA-6 (G) LI-gRNA-3 (H) LI-gRNA-7

4.2.3 CRISPR Constructs

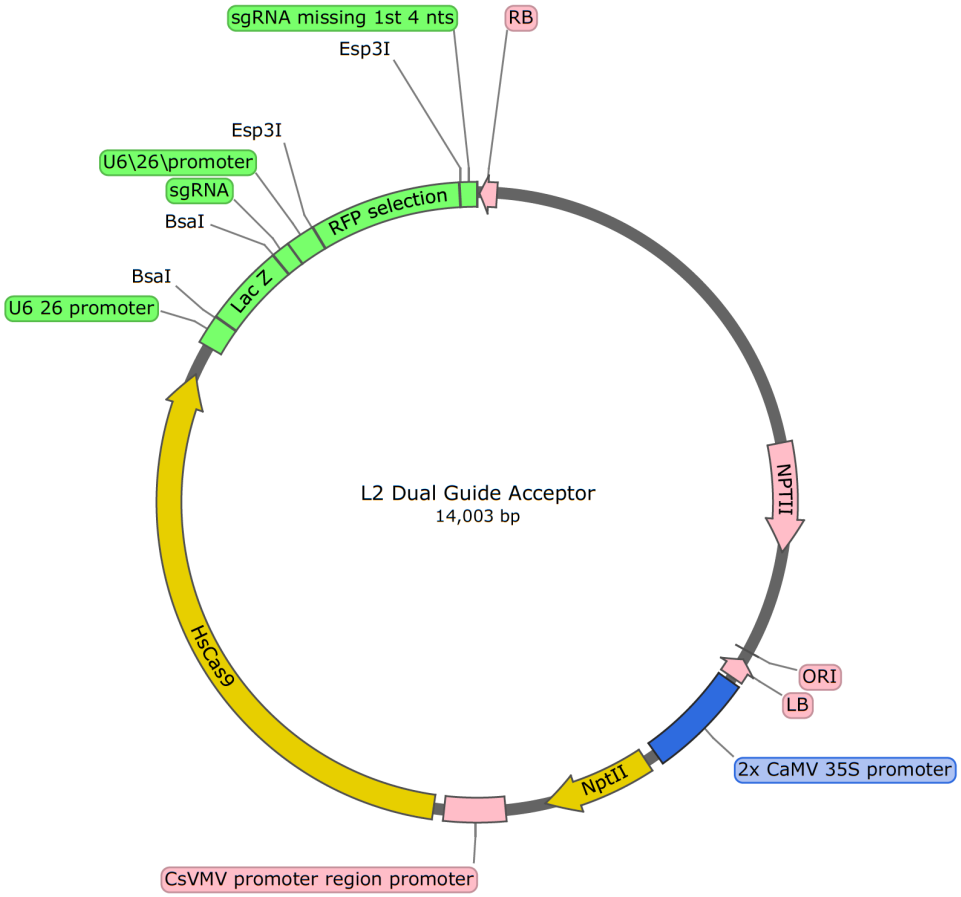
The plasmid used for the editing of the *eIF(iso)4e-C4* paralog was designed by Tom Lawrenson at JIC. The major components included the Cas9 protein, left and right borders for *Agrobacterium* transformation, kanamycin selection, as well as two restriction sites, *BsaI* and *EspBI*, containing LacZ and RFP genes into which the guides could be ligated and reverse selected for (Figure 4.3). A second plasmid was also created later for protoplast transformation work, this contained the eYFP protein which would allow for confirmation of transformation into the protoplasts by florescence microscopy.

4.2.4 Agrobacterium Transformation and Genetic Analysis

In collaboration with the BRACT project at JIC, two plasmids containing the first four guide RNAs were transformed into the *B. oleracea* DH1012 line using the *Agrobacterium* method outlined by Hundleby et al. (2015). Two plasmids were transformed each containing two guides, C4 and D4. Cotyledons were transformed for each gRNA pair (C4 and D4). Shoots (100) were generated for the C4 plasmid and 52 shoots were created for the D4 plasmid. They were confirmed to contain the *Cas9* transgene via qPCR. However, PCR and sequencing analysis using the previously designed specific *eIF(iso)4E-C4* primers (LI-83 and LI-99), showed no mutation of the target gene in the T0 plantlets. The plantlets were then incubated at higher temperatures for several days to try and increase the Cas9 activity. Further PCR and sequencing analysis showed no edits in the target gene. The plantlets were then taken through to plants; of the original 100 C4 plantlets and 52 D4 plantlets only 62 plants and 20 plants respectively survived.

These plants were then grown to seed. Of the C4 plants only 23 produced seed and of the D4 plants only 15 produced seed. The seed production was low with plants producing between 1 and 20 seeds and a total of 306 seeds was obtained from all the plants. These seeds were then planted and the T1 generation grown. Of the seeds

planted 267 germinated and 39 did not. Leaf samples were taken and genomic DNA extracted (Table 4.3). PCR and sequencing analysis of the target gene was then performed as previously described. Of the 267 plants analysed only 1 showed signs of Cas9 activity (plant LI-295), this was the next generation (T1) of a T0 D4-transformed line. Sequencing showed a clear chimeric expression of the *eIF(iso)4E-C4* directly following the gRNA-D4-D6 PAM site (Figure 4.4).



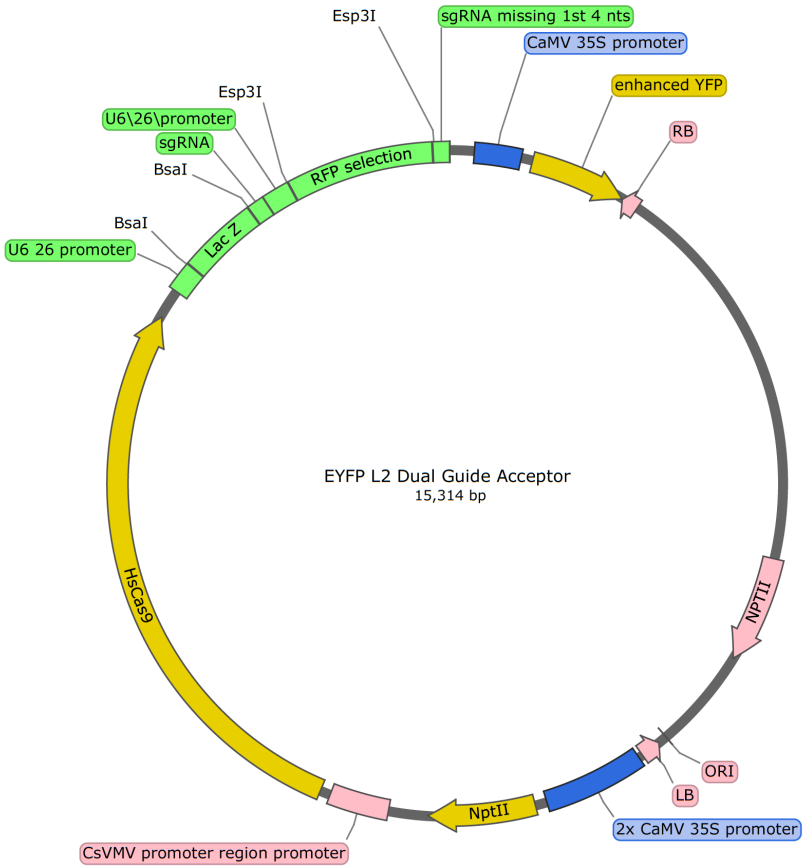


Figure 4.3 - Plasmid maps of CRISPR/Cas9 constructs used

Constructs used while attempting to edit *eIF(iso)4E-C4*. The L2 Dual Guide Acceptor vector was used for the collaboration with JIC and was developed by JIC (unpublished). In addition, it was used to test gRNAs in the protoplast assays. The EYFP L2 Dual Guide Acceptor vector was developed in order to track the plasmid during transformation into protoplasts (GenScript).

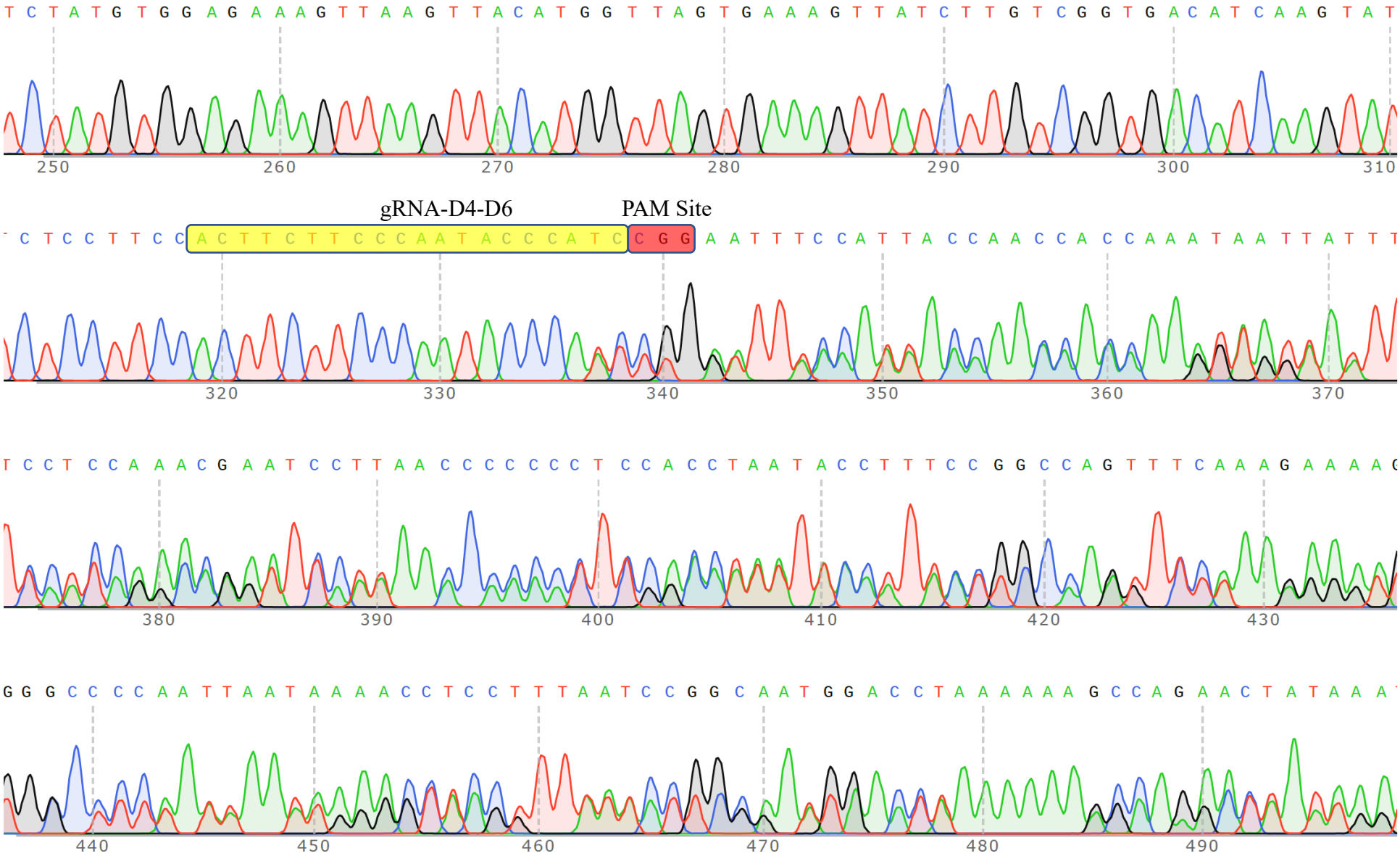
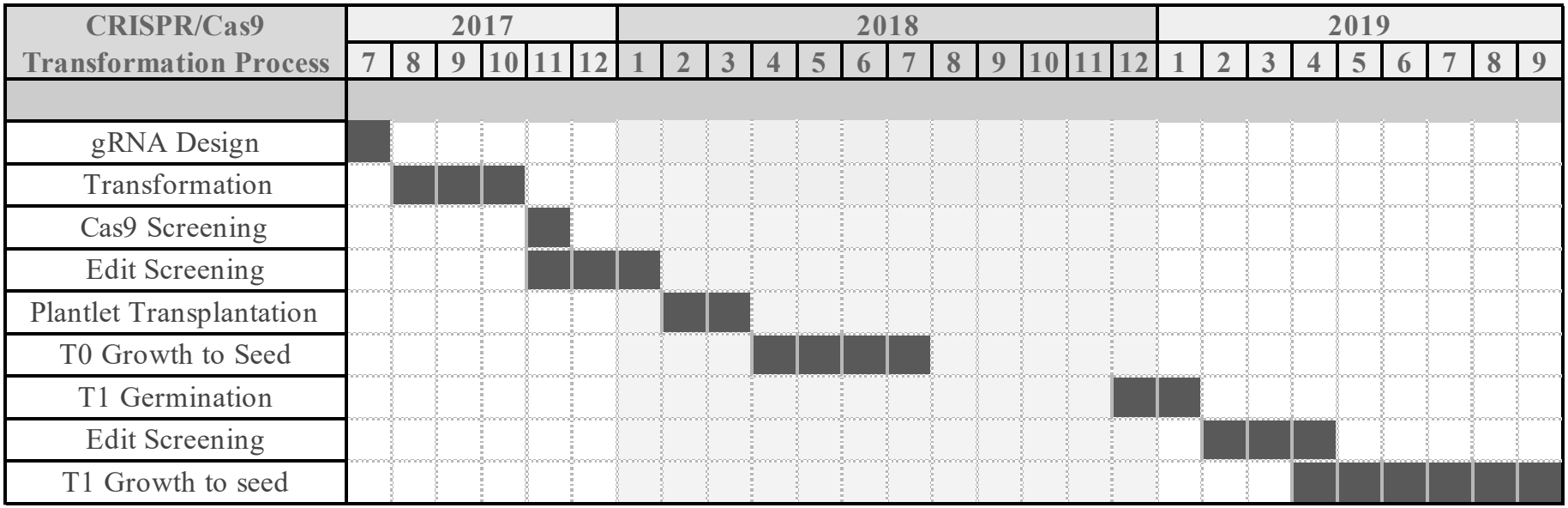


Figure 4.4 - Sequencing of *eIF(iso)4E-C4* in T1 *Brassica oleracea* plant LI-295

Sequencing of the reverse strand of the *eIF(iso)4E-C4* in T1 plant LI-295. Guide RNA gRNA-D4-D6 is highlighted in yellow and its respective PAM site is highlighted in red. Double peaks are seen in the chromatogram starting at the end of the gRNA.

Table 4.3 – Gant Chart of the CRISPR/Cas9 transformation process

[1]



[1] The timeline of events in the process of CRISPR/Cas9 design, transformation, plant tissue culture, screening and seed collection over two generations of *Brassica olerecea* (DH1012).

4.3 Discussion

The DH1012 line was shown to be susceptible to TuMV infection. The characteristic mosaic pattern of infection could be seen in the inoculated leaf tissue (Figure 4.1). The infection was confirmed by ELISA which showed systemic infection of the virus throughout the plant. This showed that the DH1012 line could to be used for phenotyping any resistance after a CRISPR/Cas9 edit was achieved.

In collaboration with the JIC BRACT project, four guides were designed and inserted into two plasmids, C4 and D4. These plasmids were transformed into the DH1012 cotyledons and plantlets were produced. Unfortunately, no edits were seen in the T0 generation at any stage, because of this, four more guides were designed. Due to the large amount of time needed to transform and culture the plant tissue it was thought best to test the guides in protoplasts first. This work is detailed in Chapter 5.

The plants transformed with the C4 and D4 CRISPR plasmids were seeded and carried on to T1 generation. Of the 267 plants only one showed signs of an edit, LI-295. Sequencing showed double peaks in the chromatogram starting at the location of the PAM site of gRNA-D4-D6 (Figure 4.4). This is characteristic of a chimeric edit where only part of the tissue has been edited or of a homogenous heterozygote mutation, both of which are screened for normally during the T0 stage. The extremely low frequency of edits found could be explained by several hypotheses, firstly it could be that the guide RNAs are sub-standard and are not guiding the Cas9 protein to the intended site correctly. Secondly, it could be that the edit is being made, however it is causing lethality in the plant tissue when homozygous. Thirdly it could be an unaccounted-for external factor such as homology-directed repair.

Chapter 5

Novel Technologies for Plant Transformation

5.1 Introduction

The CRISPR/Cas9 system is normally delivered to a plant via *Agrobacterium* mediated transformation. However, commercial genotypes of many crop species are often non-transformable using this *Agrobacterium* method as the resulting material does not develop transgenic calli. Recent studies, however, have shown that transforming with embryogenesis promoting transcription factors can allow the formation of transgenic calli (Lowe et al. 2016). The transcription factor *RKD4* has been shown to induce embryogenesis (Waki et al. 2011), and has been shown to produce transgenic calli in a previously un-transformable wheat line . This could be used to allow regeneration in recalcitrant *Brassica* lines such as the rapid cycling *B. rapa* line R-o-18, which is often used as a model *B. rapa* species.

Another method in which non-regenerative lines could be transformed is by avoiding the regeneration step altogether using novel technologies. Carbon nanodots (CDs) are nano sized carbon structures which can be functionalised by the addition of PEG groups which allows them to bind to DNA. They have previously been shown to transform multiple different tissue types in several plant lines including recalcitrant wheat (unpublished). In addition, it has been shown that CRISPR activity can be achieved when transforming using CDs, with the added benefit of being transiently expressed. In theory a floral dip method would allow for transformation of the germplasm, which could then be grown normally meaning no regeneration would be necessary.

Alternatively, protoplasts could be utilised. Protoplasts are plant or bacteria cells whose cell walls have been removed. They are an important system for studying organisms with cell walls. There are protocols available tailored to multiple different plant species which allows the removal of the cell wall. This permits experimental procedures to take place which would normally be impeded by the cell wall. Uses include studying the cellular membrane and transformation of cells with isolated DNA resulting in stable nuclear transformations (Davey et al. 2005). An adapted *A. thaliana* protocol has previously been shown to enable isolation of *Brassica* protoplasts (Yoo et al. 2007).

In the previous chapter it was seen that the attempt to edit *eIF(iso)4E-C4* was initially thought to have not worked at the T0 stage, due to the relatively common problem of the guides targeting the gene incorrectly. Four new guides were designed and inserted into the L2 Dual Guide plasmid (gRNA 2&6, gRNA 3&7). Instead of going through the *Agrobacterium* transformation protocol to test these, which can take up to 5 months before results are known, it was thought best to develop a relatively quick protoplast assay which would allow for quick screening of guide RNA functionality. It has been previously shown that protoplasts can be used as an ideal platform in which to test gRNAs used in the CRISPR/Cas9 system. This is due to the relatively short process (3-4 days) that is used to isolate, transform and analyse protoplasts, (Lin et al. 2018)

5.2 Results

5.2.1 Protoplast Optimisation

Optimisation of protoplast isolation and transformation was carried out in *B. oleracea* DH1012 and A12 lines. This allowed for the relatively rapid transformation and functionality testing of gRNAs in *B. oleracea.*

First protoplast isolation was optimised from the Schafer et. al (unpublished) protocol previously adapted from Yoo et al. (2007) protocol which was developed for *A.*

thaliana isolations. Using the same enzyme solution but altering the length of time incubated in solution, it was found that 4 hours of incubation gave the most protoplasts which also looked the healthiest as after longer periods of time protoplasts were thought to begin apoptosis (Table 5.1).

Table 5.1 – Effect of Digestion Incubation Time On Protoplast Isolation

Hours Incubated	Number of protoplasts/ml [1]
2	3.5 million
3	6.5 million
4	12.5 million
5	9.5 million
6	8 million
7	5 million
8	5.5 million
Over Night	3 million

When performing the protoplast isolation, it was noted that leaves from different areas of the plant gave drastically different numbers and quality of protoplasts. It was found that the first two true leaves of the plant gave the largest number of protoplasts that were the healthiest with the little to no debris in the protoplast solution (Table 5.3). The younger leaves produced lower numbers or, most notably, were very unhealthy and had a lot of debris present in solution.

[1] Average number of protoplasts isolated from DH1012 leaves after incubation over a range of hours at room temperature.

During the initial protoplast transformations, an mCherry expression plasmid was used as a method of observing transformation efficiency. The mCherry florescence was observed using a mercury lamp florescence microscope. The initial transformation rates were quite low at between 5% and 10%. In order to improve this, it was thought the osmolarity of some of the buffers may be too low for *Brassica* protoplasts, which are larger than *A. thaliana* protoplasts, causing them to burst. The osmolarity of all the buffers used were measured and it was found that the transformation buffer (0.4M mannitol MMG) and overnight incubation buffer (0.4M mannitol WI) had the lowest osmolarities, 510 mOsm and 632 mOsm respectively (Table 5.2). They were then modified by increasing the concentration of mannitol to 0.6M or 0.7M, which increased the osmolarity to 832 mOsm and 831 mOsm respectively for 0.6M, and 1027 mOsm and 1025 mOsm respectively for 0.7M (Table 5.2). When protoplasts were isolated and then transformed with these solutions no transformation was seen, however a lot more protoplasts survived and appeared healthier. It was thought this was due to the PEG solution, with 0.2M mannitol not being able to give a big enough osmotic shock when mixed with the increased osmolarity MMG solutions. An alternate PEG solution was found in the literature with 0.6M mannitol, and on transformation with this and 0.6M MMG, transformation rates of 10-15% were seen.

Table 5.2 – Osmolarity of Protoplast Solutions

Solution[1]	osmolarity (mOsm)
Enzyme Solution	630
W5 Buffer	694
MMG (0.4 mannitol)	510
MMG w/PEG & W5	746
MMG w/PEG (0.2M mannitol)	N/A
WI solution	632
MMG (0.6M mannitol)	831
WI (0.6M mannitol)	830
MMG (0.7M mannitol)	1027
WI (0.7M mannitol)	1025
PEG (0.6M mannitol)	N/A

These transformation rates were still relatively low, from the literature it was noted that optimising the length of incubation of the transformation can have an effect on transformation rates. An experiment was carried out using the 3 different pairings of solutions; 0.4M MMG and WI with 0.2M PEG, 0.6M MMG and WI with 0.6M PEG, and 0.7M MMG and WI with 0.7M PEG. They were incubated during transformation for 5, 10, 20, and 30 minutes. It was found that incubating the transformations for 20 minutes gave the highest transformation rates in the 0.4M MMG and WI with 0.2M PEG and 0.6M MMG and WI with 0.6M PEG solutions (Table 5.4). Although these two solution pairs had similar rates, the total number of transformed protoplasts was much higher in the 0.6M transformation as more protoplasts survived and looked much healthier than the original solutions (Figure 5.1). This transformation protocol

[1] Solutions used during isolation and transformation of *B. oleracea* protoplasts. The solutions containing large amounts of PEG were unable to be measured as the solutions were too viscous for the osmometer.

was also shown to work well in *B. juncea* (*var. tsunga* GRU 004270) giving a 60% transformation rate, which was also higher than previously seen (Figure 5.2).

The most efficient transformation techniques were used to transform the four new guides in two new plasmids into the protoplasts. DNA was then extracted and PCR analysed using previously designed primers (LI83-LI99). Unfortunately, after multiple repetitions, no edits were seen in the target gene. Therefore, a new plasmid was developed, eYFP L2 Dual guide, which contained an *eYFP* expression gene under a constitutive promoter. The idea being that the transformation of the plasmid could be tracked as previously with the mCherry expression plasmids. Unfortunately, when transforming with this plasmid no EYFP expression was seen, nor edits in the target gene.

However, when collaborating with the Gutierrez-Marcos (SLS, University of Warwick) group in which these transformation methods were used to transform *B. oleracea* A12 with their own binary vector (unpublished), edits in their target gene were seen.

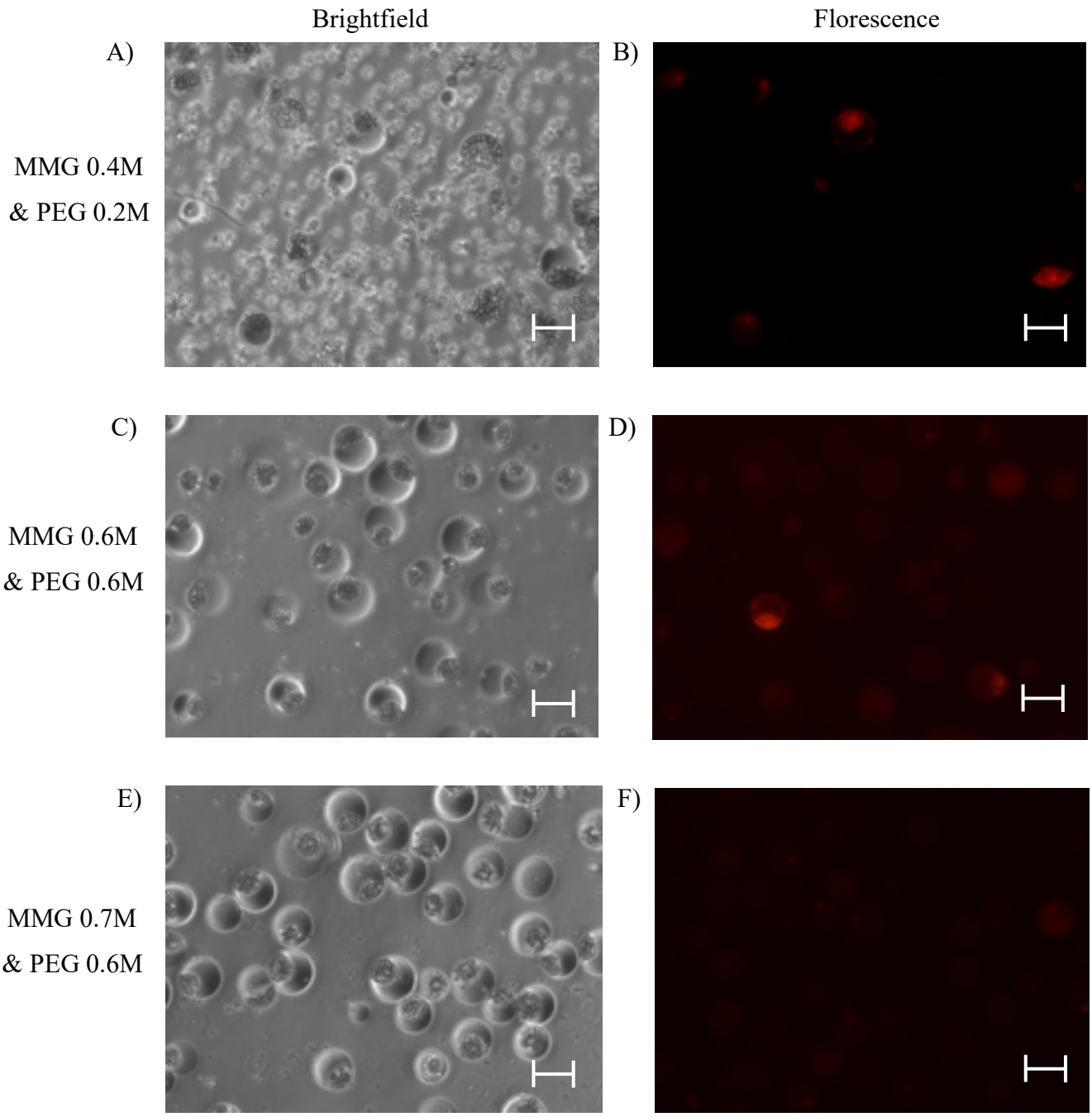


Figure 5.1 - *Brassica oleracea* protoplast transformation with mCherry expression plasmid

A&B) 0.4M mannitol MMG and WI with 0.2M mannitol PEG Solution. C&D) 0.6M mannitol MMG and WI with 0.6M mannitol PEG Solution. E&F) 0.7M mannitol MMG and WI with 0.6M mannitol PEG Solution. Brightfield images are on the left and florescence images on the right. Scale bars 25µm

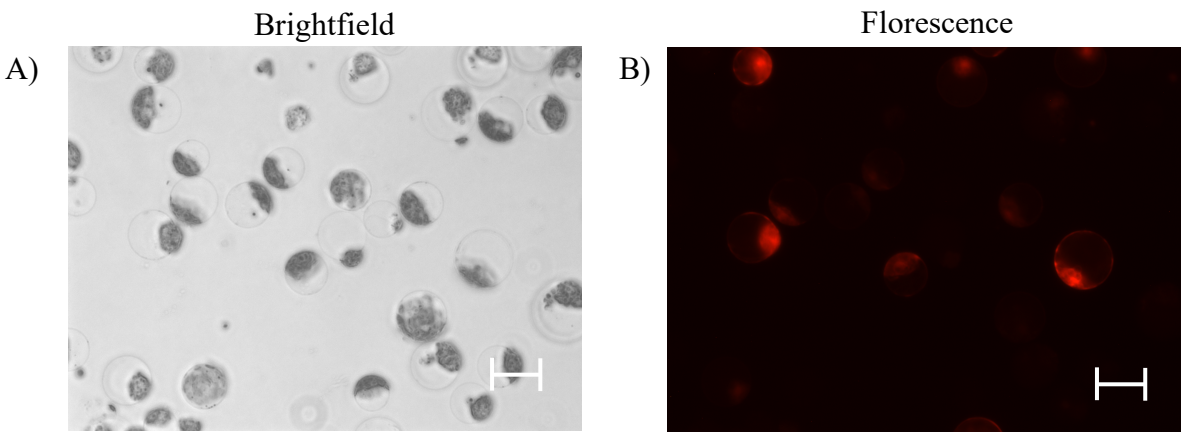


Figure 5.2 - *Brassica juncea* protoplast transformation with mCherry expression plasmid 0.6M mannitol MMG and WI with 0.6M mannitol PEG Solution. A) Brightfield image. B) florescence image. Scale bars 25µm

Table 5.3 – Effects of Leaf Type on Protoplast Isolation

Leaf[1]	Size	Area of plant	Comment	# of Protoplasts/ml	Protoplast Apperance
1	Small	Very top	Newest Leaf	6 million	Not healthy/ Debris in Solution
2	Medium	Middle/Top	Normal/Waxy	2 million	Low number but healthy
3	Medium	Middle	Red in colour	13 million	Not healthy/ Debris in Solution
4	Large	Middle	Normal/Waxy	5 million	Not healthy/ Debris in Solution
5	Large	Bottom	First two true leaf, Old and Red	18 million	Very Healthy
6	Large	Bottom	First two true leaf, Normal	16 million	Very Healthy

[1] Leaves were chosen from the top, middle and bottom of *B. oleracea* from three A12 plants in varying stages of growth. All produced protoplasts however of varying number and appearance.

Table 5.4 – Effect of Solution Osmolarity on Protoplast Transformation

Transformation Solutions Mannitol Concentration[1]			Transformation Incubation Time (Minutes)				
MMG & WI	PEG	Negative	5	10	20	30	Observations
0.4M	0.2M	none	none	~3%	~17%	~20%	Transformed Protoplast were healthy but Large majority had died
0.6M	0.6M	none	none	~5%	~25%	~20%	Larger number of overall healthy protoplasts
0.7M	0.6M	none	none	none	none	none	Healthy but no transformation

[1] Transformation rates of *B. oleracea* protoplasts using three different transformation solution mannitol concentrations. Transformation percentages were estimated from the microscope field of view. Transformations were incubated over four time periods as well as a negative transformation control. Observations of the health of the protoplast were taken during counting.

5.2.2 *RKD4* Assisted Regeneration

Multiple commercial genotypes of many crop species are often non-transformable with the standard *Agrobacterium* method as the transformed tissue is non-regenerative. This is the case for *B. rapa* line R-o-18. Recent studies however have shown that transforming with embryogenesis promoting transcription factors, such as *RKD4,* allows the formation of transgenic calli in some previously non-regenerative species (Waki et al. 2011). Therefore, it was hypothesised that the transformation of *B. rapa* with a plasmid expressing the *A. thaliana RKD4* gene would allow for regeneration from the calli stage. An RKD4 expression plasmid was sourced from the Gutierrez-Marcos (SLS, University of Warwick) group. In this plasmid the *RKD4* gene is controlled by a DEX inducible promoter.

Attempts were made to transform 120 R-o-18 cotyledons using the method laid out by Hundleby et. al (Sparrow et al. 2015). Calli (100) were generated and grown on selection plates. Several rounds of dissection and replacing on the selection plates were carried out in order to ensure the cells were all transformed. Following this, the *RKD4* expression was induced using plates containing DEX. However, after several months, only two of the R-o-18 calli had regenerated shoots and no calli had regenerated roots. The two shooting plantlets were placed on root-inducing media containing IBA. After a month they regenerated roots and were grown to seed. DNA was extracted from these two plants and analysed for the presence of the *RKD4* gene, which was not found. All other calli were placed into a different regeneration medium, containing NAA and IBA. They were left for 6 months, however they showed no signs of regenerating and just continued growing relatively slowly in size (Figure 5.3).

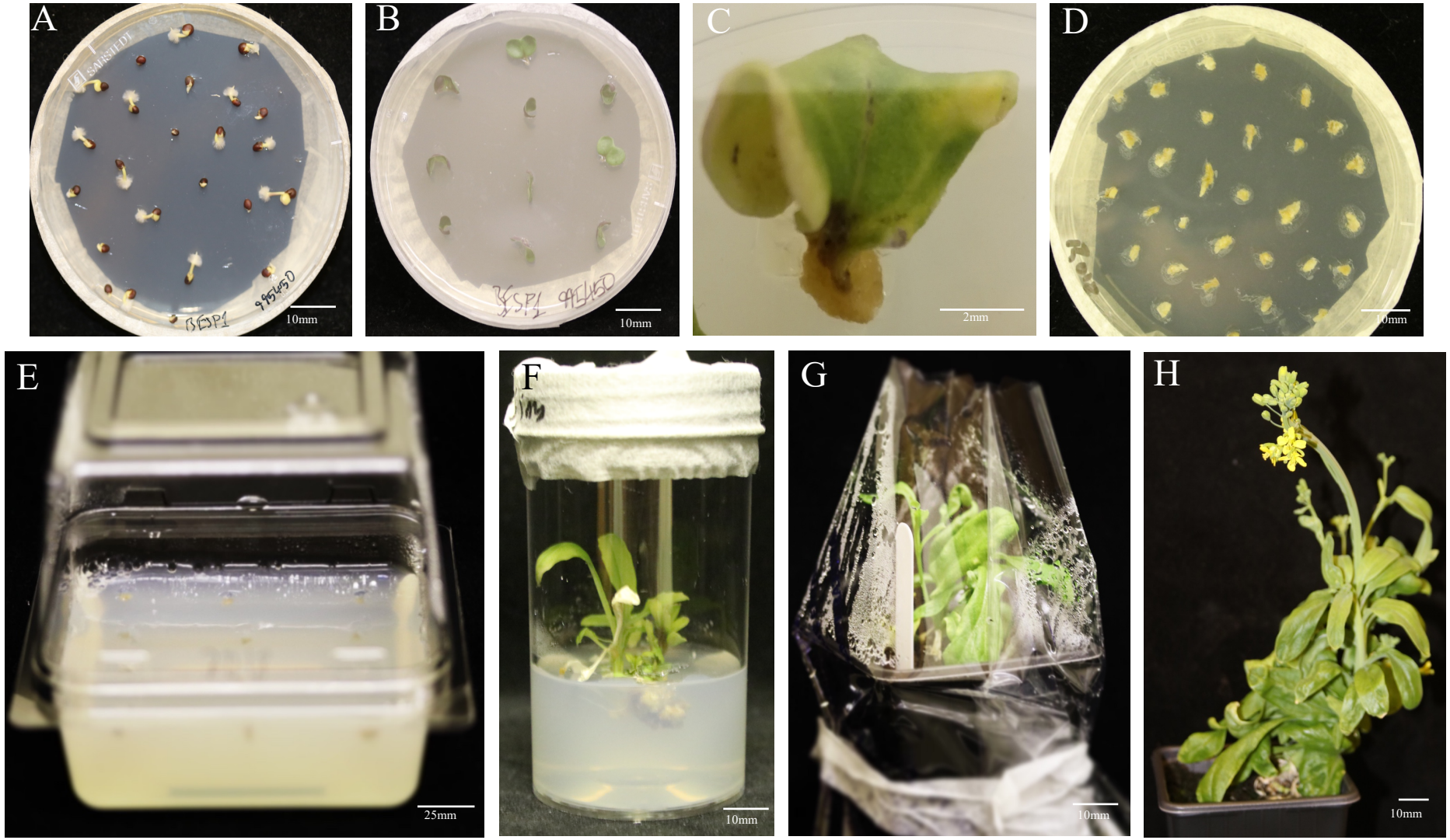


Figure 5.3 – Process of Regeneration in *Brassica rapa* (R-0-18)

A) Sprouting seeds on germination media. B) Individual cotyledons on co-cultivation plate after excision and dipping into agrobacterium suspension. C) Calli forming at the bottom of a cotyledon after 2-3 weeks on selection plate. D) Calli on selection plate after division to allow for internal cells to be exposed to selection E) Calli in regeneration media. F) Shoot and callus in rooting media after shoot has developed from callus. G) Plant transferred to soil after root development, kept in plastic bag to keep humidity high. H) Flowering plant after removal of plastic bag.

5.2.3 Carbon Nanodot Transformation

The novel nanodot technology has been shown to allow transformation and transient expression of CRISPR machinery in difficult to transform wheat lines (scientist unpublished communication). In order to show the potential for the use of nanodots in *Brassica* transformation, a YFP expression plasmid was inserted into both *B. rapa* (R-o-18) and *B. oleracea* (DH1012). This work was carried out in collaboration with the H. Whitney group at the University of Bristol, who supplied nanodots which had been functionalised with a PEG group. The Whitney protocol was followed for binding of the DNA and application to plant material. Three initial attempts at spraying the leaves of plants with the CDs resulted in no expression of YFP. It was observed that due to the leaves of both *B. oleracea* and *B. rapa* being very waxy, the solution containing nanodots was repelled off of the leaves. This was attempted to be rectified in two ways, firstly the surfactant, Silwet, was used in the solution, this is commonly used in *A. thaliana* floral dips to reduce the effect of the hydrophobic nature of the plant. Secondly carborundum powder was used to generate small legions in the waxy layer, similar to its use in virus inoculation, to allow the solution to penetrate the waxy layer. After several attempts no florescence expression was seen with the use of Silwet, however there was some florescence seen with the use of carborundum (Figure 5.4).

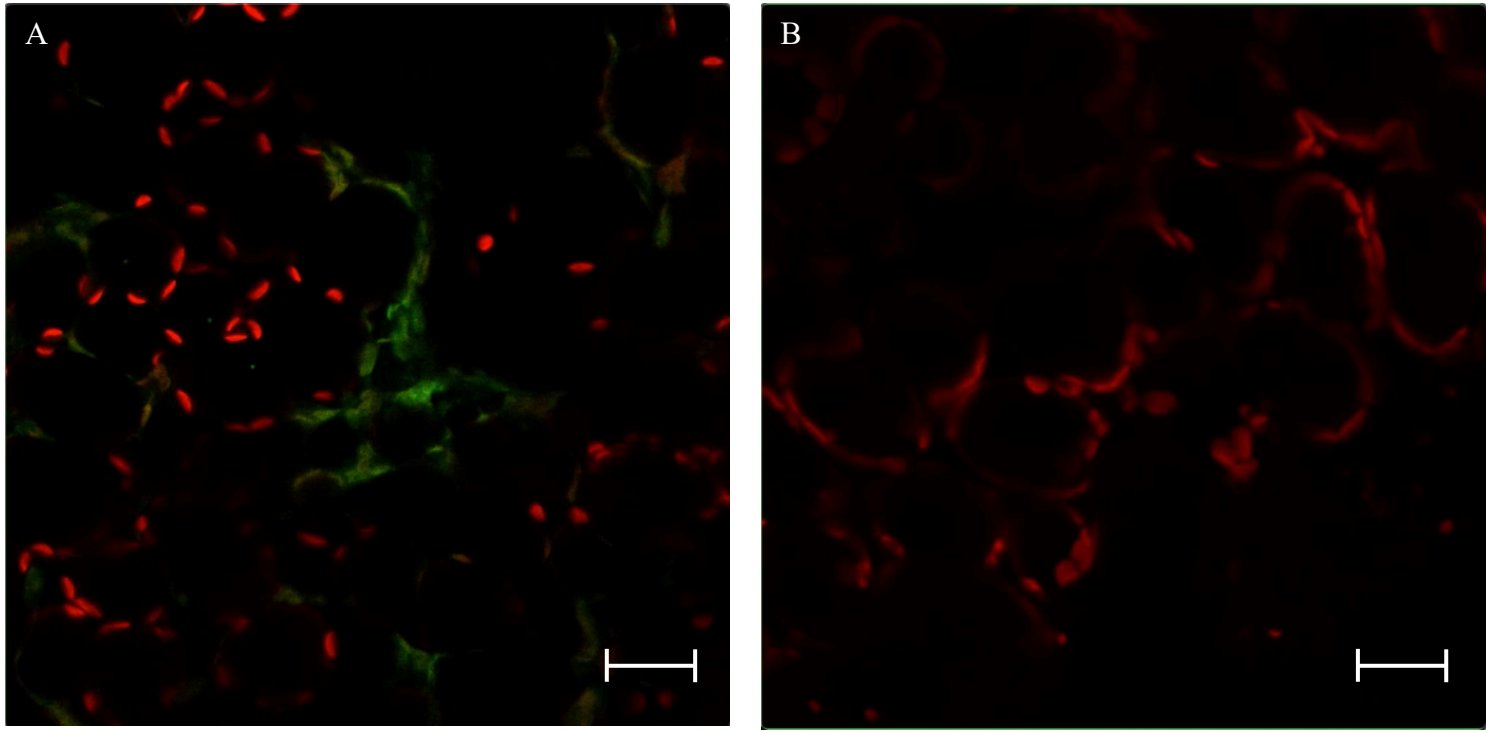


Figure 5.4 – Confocal Microscopy Image of Carbon Nanodot Transformation in *Brassica rapa*

B. rapa leaf tissue imaged using a fluorescent microscope under YFP and Chlorophyll A wavelengths. A) Leaf treated with carborundum powder and carbon dots. B) Leaf treated with plasmid DNA only. Size bars equate to 100nm. Plants were then visualised on a confocal microscope (Zeiss LSM 880), looking for YFP expression (green 527nm) and using Chlorophyll A (red 735 nm)

5.3 Discussion

5.3.1 Protoplast

Through the work carried out on the optimisation of protoplast transformation in *B. oleracea* lines DH1012 and A12 it was found that the optimal results were obtained using the first two true leaves and incubation of the digestion solution for 4 hours. For transformation the highest efficiencies were seen using 0.6M mannitol MMG and WI solutions in conjunction with 0.6M PEG solutions. This was mainly due to the improved overall health of the protoplasts that these conditions solutions conferred.

No edits were seen using the LI guide RNA plasmids. There are several potential reasons for this. Firstly, this could be due to the guide RNAs not being suitable, however it is unlikely that of all 8 guides tried, not one works. Secondly, it could be that the L2 Dual Guide plasmid is too large. It is known that the larger plasmids generally have lower transformation efficiency (Bart et al. 2006). This is supported by the fact that the eYFP L2 Dual Guide plasmid did not appear to be transform into the protoplasts, as this is an even larger plasmid. The successful edits seen when transforming with our collaborators with much smaller binary vectors also supports this theory. This showed that the optimised transformation protocol can be effective as a method for testing the efficiency of gRNAs.

5.3.2 *RKD4*

Many crop species and commercial plant varieties are difficult to transform for the principal reason they are non-regenerative. One method which has been shown to improve regeneration efficiency is the use of embryogenesis promoting transcription factors. The transcription factor *RKD4* has been shown to allow regeneration of recalcitrant wheat varieties (unpublished). The agrobacterium transformation using this transcription factor was attempted in *B. rapa* line R-o-18. Although there were many calli generated that were able to grow on kanamycin selection plates, only two calli regenerated; *RKD4* was not found to be integrated into the genome of these calli.

Although this does not show that *RKD4* does not work, it may show that the initial transformation with agrobacterium is not working correctly and would need to be optimised in future work.

5.3.3 Carbon Nanodots

In order to overcome the problems with agrobacterium transformation it was decided to use the relatively novel carbon nanodot technology developed by the Whitney group at Bristol University. Previous work showed they were able to transform wheat with both florescent reporter constructs and CRISPR/Cas9 constructs. Attempted transformation of both *B. oleracea and B. rapa* using the standard method resulted in no florescence expression being seen. Using carborundum powder resulted in some florescence being seen, however this was later thought to be the result of auto-florescence from the apoptosis of the leaf tissue due to the carborundum powder. The fact that no florescence was seen, may be due to the fact that the DNA is not binding to the carbon nanodots, this may be the case as previous tests have only been carried out with completely freshly functionalised CDs in Bristol. It is thought that there might be some problem with the functional PEG group, that allows the DNA to bind to the nanodot, detaching over time or in transit from Bristol.

Chapter 6

General Discussion

The *Brassica* genus contains many key crop species, including root and leafy vegetables as well as oilseed crops, and is of significant economic importance around the world. The *Potyvirus,* Turnip mosaic virus (TuMV), causes substantial losses in *Brassica* production worldwide by affecting both yield and quality of produce (Walsh et al. 2002). Previous work identified a variety of *B. rapa* (RLR22) which had broad-spectrum recessive resistance to TuMV as a result of a mutation in an *eIF(iso)4E* gene (Jenner et al. 2010, Nellist et al. 2014). It was shown in *A. thaliana* and other species, that exploiting this resistance mechanism through molecular techniques, such as CRISPR/Cas9, conferred resistance to *Potyviradae* (Rodriguez-Hernandez et al. 2012, Kim et al. 2014, Chandrasekaran et al. 2016, Bastet et al. 2019, Gomez et al. 2019). This project aimed to utilise the eIF4E resistance mechanism as a model target to demonstrate applicability of CRISPR/Cas9 in *Brassica.* In addition, the project was designed to further the understanding of the genetic make-up of the eIF4F translation initiation complex in *B. rapa* and *B. oleracea,* with the aim of identifying new targets for inducing resistance.

6.1 Bioinformatics

Analysis of the genetic make-up of the subunits of the eIF4F translation initiation complex identified several novel paralogs of *eIF4E, eIF4G* and *eIF4A* in *Brassica*. This fundamental knowledge was necessary for the design of specific primers for PCR amplification as well as gRNAs for the explicit targeting of the *eIF(iso)4E* paralogs. The amplification of the genes of interest across several *Brassica* species and varieties showed that most of the genes making up the eIF4F complex are highly conserved. In addition, qPCR analysis of the expression in TuMV-infected and uninfected *B. rapa* plants showed there was little difference in expression of these genes. This raises the question as to why resistance is seen in the RLR22 line when only one paralog of

eIF(iso)4E is mutated and all other paralogs are being expressed. It has previously been suggested that there are mechanisms in place that means the VPg of the TuMV either preferentially binds to one particular paralog or cannot bind to the other paralogs and isoforms of *eIF4E*, however this has not yet been demonstrated (Bastet et al. 2017). This has also been seen in an unsuccessful complementation study in which the virus was able to overcome resistance by utilising the other isoforms in *A. thaliana (Jenner et al. 2010).* In addition, the potential for TuMV to overcome resistance may be higher than previously thought, as we have shown that there are more paralogs for potential use.

Amplification of the *eIF4E-A9* paralog also revealed previously unreported alternative splicing taking place in the *eIF4E* paralogs, this was supported by RNAseq alternative splicing analysis. Alternative splicing taking place may initially seem unremarkable as it has been well documented as a level of transcriptional regulation in plants (Reddy et al. 2013). However, prior to this study there had been no reports of alternative splicing taking place in many of the subunits of the eIF4F complex. It was confirmed by RNAseq that alternative splicing is happening across a number of paralogs in *B. rapa*. This was supported by a study of alternative splicing carried out in *B. oleracea* during the course of this project by *Xu et al. (2019)*. This increases the complexity of the eIF4F subunits and their respective gene families, raising the question of why is alternative splicing taking place. In addition, there were differences seen in the alternative splicing between R-o-18 and RLR22 and many of the splicing events looked similar to the intron inclusion of the RLR22 resistance mechanism. This poses the question, might the alternative splicing of other paralogs be involved in the resistance, and, if not, what role do they play?

6.1.1 Future work

Following on from the RNAseq analysis more work could be done looking into genes that have been differentially expressed between RLR22 and R-o-18, it is especially interesting that one of the NCBP paralogs was differentially expressed and more work need to be done in order to understand its role in translation initiation and resistance

in brassicas. In addition, the role of the alternative splice forms in the plant, if any, on translation initiation and resistance mechanisms needs to be investigated.

6.2 Targeting *eIF(iso)4E-C4* in *B. oleracea*

The only edit seen of the *eIF(iso)4E- C4* paralog in *B. oleracea* (DH1012) was found in the T1 generation after over 267 plants had been screened. This is a very low number of edits bearing in mind other targets that were edited as part of the BRACT project at JIC generated at least 10% mutation rates in the T0 generation. The low mutation rate seen could be due to a number of factors, firstly it could be that the guide RNAs are substandard and are not guiding the Cas9 protein to the intended site. Although this is one of the most common problems with using the CRISPR/Cas9 system (Farboud et al. 2015), the fact that there was an edit present in T1 shows at least one of the gRNAs was working correctly and therefore it would be expected that more edits would have been seen.

Secondly, it could be that the edit is being made however it is causing some lethality in the plant cells, and therefore no edits can be detected. This might be expected, due to the important role the *eIF(iso)4E* protein plays in translation. It may be that *B. oleracea* relies more heavily on the C4 paralog than the other A4 paralog found in resistant *B. rapa* line RLR22. This idea is supported by the fact that the only edit found was heterozygous, as this plant would have an alternative allele it could utilize. In addition, no native recessive TuMV resistance has been found in any *B. oleracea* variety despite extensive searching, this would support the possibility in the *eIF(iso)4E-C4* allele is lethal.

6.2.1 Future Work

The seed collected from the one *B. oleracea* DH1012 edited plant that was found needs to be screened for homozygous edited plants in the T2 generation. If these can be found, homozygous plants would need to be selfed and the T3 seed grown and

phenotyped for TuMV resistance. If only heterozygous plants are found this would further support the idea that a homozygous edit is lethal. However, it would still be of benefit to phenotype the heterozygous plants as a single allele knock out might provide some reduced symptom phenotypes similar to that seen in a tomato TILLING population and *B .rapa* (Nellist et al. 2014, Gauffier et al. 2016) . In addition to this, it would be beneficial to investigate if a KO in paralog *eIF(iso)4E- C4* is actually lethal in *B. oleracea.* This could be done using a controlled mutation method such as temperature sensitive mutants or inducible Cas9 protein expression or silencing of the gene.

6.3 Novel techniques

Protoplasts have been used for many years as a transformation tool and have been employed more recently as a method to quickly test gRNA functionality. The optimisation of protoplast isolation and transformation in different *Brassica* species worked well and allowed for of an increased number of protoplasts to be transformed. Although there was no evidence of editing using the L2 Dual Guide CRISPR plasmid, edits were made using a collaborator's binary vector system, showing the transformation process is sound. The reason for the L2 Dual Guide transformation not working could be due to the size of vector, ~14Kb, which is a known cause of reduced transformation efficiency (Simon et al. 1986).

The use of the transcription factor *RKD4* to increase the regeneration efficiency of recalcitrant lines did not seem to work, however this was due to the fact that the agrobacterium transformation of the *B. rapa* R-o-18 line was unsuccessful following the same protocol used for *B. oleracea* CRISPR/Cas9 transformation. This technique would need to be optimised for *B. rapa*.

The use of the carbon nanodots to transform recalcitrant lines looks promising based on the unpublished work done in wheat. Unfortunately, no concrete evidence of transformation was seen in *B. oleracea* or *B. rapa.* This is believed to be because the

new nature of the technology means little is known about the transportation and shelf-life of the carbon nano-dots; this has hindered the attempts made at Warwick.

6.3.1 Future Work

With the protoplast work, the next step would be to attempt to transform using the smaller binary vector in order to validate the gRNAs. However this may also be a problem as stated previously there is the possibility that the knock-out of *eIF(iso)4E* in *B. olerecea* may be lethal therefore a known editable target could be used as a positive control or the binary vectors could be validated in *B. rapa* where this knock out is known to not be lethal. Once this is established, it would provide a rapid week-long protocol to establish gRNA viability, meaning when transforming with agrobacterium both the guides would be known to function correctly, increasing the likelihood of a desired edit.

With regards to the *RKD4* work, the first step would be to identify if the agrobacterium transformation method is working in the recalcitrant R-o-18 line and DNA is actually being introduced into the plant genome. This could be done by transforming with a plasmid carrying a fluorescent marker gene and then check for expression at the calli stage using both microscopy and DNA analysis.

With the carbon nanodots, the best experiment would be to transform the *Brassica* lines at Bristol and transform the plants there with the freshest CDs. In addition, it would be beneficial to test the binding of the DNA plasmid to the CDs using Dynamic Light Scattering (DLS). The carbon nanodot method, although in its infancy and with multiple problems, still has potential to allow easy transformation of many recalcitrant species and therefore would still be extremely beneficial to pursue.

6.4 Summary

This project showed that there is more intricacy in the genetic make-up of the eIF4F complex than previously thought, demonstrated by the previously unseen paralogs and alternative splicing of several of the genes of interest. This emphasises the need to better understand the complex as a whole. In addition, this project highlighted the difficulties with using the CRISPR/Cas9 system in brassicas, finding only one genome edited plant after two generations. This could be due to there being some lethality with knocking out the *eIF(iso)4E-C4* ortholog in *B. oleracea*, potentially explaining why there has been no recessive TuMV resistance found in *B. oleracea* as of yet. Furthermore, the use of novel techniques to overcome recalcitrance in *B. rapa,* showed some promise, however more work is necessary before they are viable transformation procedures. This project required for many more techniques to be understood, optimised and utilised than originally anticipated. However, the number of complications encountered stresses the difficulties that must be overcome in order for industry to routinely implement these technologies into their breeding programs.

Bibliography

Andersson, M., H. Turesson, A. Nicolia, A.-S. Fält, M. Samuelsson and P. Hofvander (2017). "Efficient targeted multiallelic mutagenesis in tetraploid potato (Solanum tuberosum) by transient CRISPR-Cas9 expression in protoplasts." Plant Cell Reports **36**(1): 117-128.

Barfield, D. G. and E.-C. Pua (1991). "Gene transfer in plants of Brassica juncea using Agrobacterium tumefaciens-mediated transformation." Plant Cell Reports **10**(6): 308-314.

Bart, R., M. Chern, C.-J. Park, L. Bartley and P. C. Ronald (2006). "A novel system for gene silencing using siRNAs in rice leaf and stem-derived protoplasts." Plant Methods **2**(1): 13.

Bastet, A., C. Robaglia and J.-L. Gallois (2017). "eIF4E Resistance: Natural Variation Should Guide Gene Editing." Trends in Plant Science **22**(5): 411-419.

Bastet, A., D. Zafirov, N. Giovinazzo, A. Guyon-Debast, F. Nogué, C. Robaglia and J. L. Gallois (2019). "Mimicking natural polymorphism in eIF4E by CRISPR-Cas9 base editing is associated with resistance to potyviruses." Plant Biotechnology Journal.

Britt, A. B. (1999). "Molecular genetics of DNA repair in higher plants." Trends in Plant Science **4**(1): 20-25.

Browning, K. S. (1996). "The plant translational apparatus." Plant Molecular Biology **32**(1): 107-144.

Carrington, J. C., K. D. Kasschau, S. K. Mahajan and M. C. Schaad (1996). "Cell-to-Cell and Long-Distance Transport of Viruses in Plants." The Plant cell **8**(10): 1669-1681.

Chandrasekaran, J., M. Brumin, D. Wolf, D. Leibman, C. Klap, M. Pearlsman, A. Sherman, T. Arazi and A. Gal-On (2016). "Development of broad virus resistance in non-transgenic cucumber using CRISPR/Cas9 technology." Molecular Plant Pathology: n/a-n/a.

Charron, C., M. Nicolaï, J.-L. Gallois, C. Robaglia, B. Moury, A. Palloix and C. Caranta (2008). "Natural variation and functional analyses provide evidence for co-evolution between plant eIF4E and potyviral VPg." The Plant Journal **54**(1): 56-68.

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

Davey, M. R., P. Anthony, J. B. Power and K. C. Lowe (2005). "Plant protoplasts: status and biotechnological perspectives." Biotechnology Advances **23**(2): 131-171.

De Block, M., D. De Brouwer and P. Tenning (1989). "Transformation of Brassica napus and Brassica oleracea using Agrobacterium tumefaciens and the expression of the bar and neo genes in the transgenic plants." Plant Physiology **91**(2): 694-701.

Demirer, G. S., H. Zhang, J. L. Matos, N. S. Goh, F. J. Cunningham, Y. Sung, R. Chang, A. J. Aditham, L. Chio, M.-J. Cho, B. Staskawicz and M. P. Landry (2019). "High aspect ratio nanomaterials enable delivery of functional genetic material without DNA integration in mature plants." Nature Nanotechnology **14**(5): 456-464.

Dixon, G. R. (2007). Vegetable brassicas and related crucifers, Crop Production Science in Horticulture. Oxforshire, UK, CAB International.

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson and T. R. Gingeras (2012). "STAR: ultrafast universal RNA-seq aligner." Bioinformatics **29**(1): 15-21.

Ewels, P., M. Magnusson, S. Lundin and M. Käller (2016). "MultiQC: summarize analysis results for multiple tools and samples in a single report." Bioinformatics **32**(19): 3047-3048.

FAOSTAT. (2013). from http://faostat3.fao.org.

Farboud, B. and B. J. Meyer (2015). "Dramatic Enhancement of Genome Editing by CRISPR/Cas9 Through Improved Guide RNA Design." Genetics **199**(4): 959.

Gal-On, A., M. Fuchs and S. Gray (2017). "Generation of novel resistance genes using mutation and targeted gene editing." Current Opinion in Virology **26**: 98-103.

Gallie, D. R. (2001). "Cap-Independent Translation Conferred by the 5′ Leader of Tobacco Etch Virus Is Eukaryotic Initiation Factor 4G Dependent." Journal of Virology **75**(24): 12141-12152.

Gallie, D. R. and K. S. Browning (2001). "eIF4G Functionally Differs from eIFiso4G in Promoting Internal Initiation, Cap-independent Translation, and Translation of Structured mRNAs." Journal of Biological Chemistry **276**(40): 36951-36960.

Gauffier, C., C. Lebaron, A. Moretti, C. Constant, F. Moquet, G. Bonnet, C. Caranta and J. L. Gallois (2016). "A TILLING approach to generate broad-spectrum resistance to potyviruses in tomato is hampered by eIF4E gene redundancy." Plant Journal **85**(6): 717-729.

Gelvin, S. B. (2003). "*Agrobacterium*-Mediated Plant Transformation: the Biology behind the "Gene-Jockeying" Tool." Microbiology and Molecular Biology Reviews **67**(1): 16.

Gomez, M. A., Z. D. Lin, T. Moll, R. D. Chauhan, L. Hayden, K. Renninger, G. Beyene, N. J. Taylor, J. C. Carrington, B. J. Staskawicz and R. S. Bart (2019). "Simultaneous CRISPR/Cas9-mediated editing of cassava eIF4E isoforms nCBP-1 and nCBP-2 reduces cassava brown streak disease symptom severity and incidence." Plant Biotechnology Journal **17**(2): 421-434.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman and A. Regev (2011). "Full-length transcriptome assembly from RNA-Seq data without a reference genome." Nature Biotechnology **29**(7): 644-652.

Hughes, S. L., S. K. Green, D. J. Lydiate and J. A. Walsh (2002). "Resistance to Turnip mosaic virus in Brassica rapa and B. napus and the analysis of genetic inheritance in selected lines." Plant Pathology **51**(5): 567-573.

Hunter, P. J., J. E. Jones and J. A. Walsh (2002). "Involvement of Beet western yellows virus, Cauliflower mosaic virus, and Turnip mosaic virus in Internal Disorders of Stored White Cabbage." Phytopathology™ **92**(8): 816-826.

Jenner, Keane, Jones and Walsh (1999). "Serotypic variation in turnip mosaic virus." Plant Pathology **48**(1): 101-108.

Jenner, C. E., C. F. Nellist, G. C. Barker and J. A. Walsh (2010). "Turnip mosaic virus (TuMV) Is Able to Use Alleles of Both eIF4E and eIF(iso)4E from Multiple Loci of the Diploid Brassica rapa." Molecular Plant-Microbe Interactions **23**(11): 1498-1505.

Jenner, C. E., X. Wang, K. Tomimura, K. Ohshima, F. Ponz and J. A. Walsh (2003). "The Dual Role of the Potyvirus P3 Protein of Turnip mosaic virus as a Symptom and Avirulence Determinant in Brassicas." Molecular Plant-Microbe Interactions **16**(9): 777-784.

Jiang, J. and J.-F. Laliberté (2011). "The genome-linked protein VPg of plant viruses—a protein with many partners." Current Opinion in Virology **1**(5): 347-354.

Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna and E. Charpentier (2012). "A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity." Science **337**(6096): 816-821.

Kim, J., W.-H. Kang, J. Hwang, H.-B. Yang, K. Dosun, C.-S. Oh and B.-C. Kang (2014). "Transgenic Brassica rapa plants over-expressing eIF(iso)4E variants show broad-spectrum Turnip mosaic virus (TuMV) resistance." Molecular Plant Pathology **15**(6): 615-626.

Koh, J. C. O., D. M. Barbulescu, S. Norton, B. Redden, P. A. Salisbury, S. Kaur, N. Cogan and A. T. Slater (2017). "A multiplex PCR for rapid identification of Brassica species in the triangle of U." Plant Methods **13**(1): 49.

Lawrenson, T., O. Shorinola, N. Stacey, C. Li, L. Østergaard, N. Patron, C. Uauy and W. Harwood (2015). "Induction of targeted, heritable mutations in barley and Brassica oleracea using RNA-guided Cas9 nuclease." Genome Biology **16**(1): 258.

Léonard, S., C. Viel, C. Beauchemin, N. Daigneault, M. G. Fortin and J.-F. Laliberté (2004). "Interaction of VPg-Pro of Turnip mosaic virus with the translation initiation factor 4E and the poly(A)-binding protein in planta." Journal of General Virology **85**(4): 1055-1063.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin and S. Genome Project Data Processing (2009). "The Sequence Alignment/Map format and SAMtools." Bioinformatics **25**(16): 2078-2079.

Lin, C. S., C. T. Hsu, L. H. Yang, L. Y. Lee, J. Y. Fu, Q. W. Cheng, F. H. Wu, H. C. Hsiao, Y. Zhang, R. Zhang, W. J. Chang, C. T. Yu, W. Wang, L. J. Liao, S. B. Gelvin and M. C. Shih (2018). "Application of protoplast technology to CRISPR/Cas9 mutagenesis: from single-cell mutation detection to mutant plant regeneration." Plant Biotechnol J **16**(7): 1295-1310.

Liu, C., P. Zhang, X. Zhai, F. Tian, W. Li, J. Yang, Y. Liu, H. Wang, W. Wang and W. Liu (2012). "Nano-carrier for gene delivery and bioimaging based on carbon dots with PEI-passivation enhanced fluorescence." Biomaterials **33**(13): 3604-3613.

Liu, S., Y. Liu, X. Yang, C. Tong, D. Edwards, I. A. P. Parkin, M. Zhao, J. Ma, J. Yu, S. Huang, X. Wang, J. Wang, K. Lu, Z. Fang, I. Bancroft, T.-J. Yang, Q. Hu, X. Wang, Z. Yue, H. Li, L. Yang, J. Wu, Q. Zhou, W. Wang, G. J. King, J. C. Pires, C. Lu, Z. Wu, P. Sampath, Z. Wang, H. Guo, S. Pan, L. Yang, J. Min, D. Zhang, D. Jin, W. Li, H. Belcram, J. Tu, M. Guan, C. Qi, D. Du, J. Li, L. Jiang, J. Batley, A. G. Sharpe, B.-S. Park, P. Ruperao, F. Cheng, N. E. Waminal, Y. Huang, C. Dong, L. Wang, J. Li, Z. Hu, M. Zhuang, Y. Huang, J. Huang, J. Shi, D. Mei, J. Liu, T.-H. Lee, J. Wang, H. Jin, Z. Li, X. Li, J. Zhang, L. Xiao, Y. Zhou, Z. Liu, X. Liu, R. Qin, X. Tang, W. Liu, Y. Wang, Y. Zhang, J. Lee, H. H. Kim, F. Denoeud, X. Xu, X. Liang, W. Hua, X. Wang, J. Wang, B. Chalhoub and A. H. Paterson (2014). "The Brassica oleracea genome reveals the asymmetrical evolution of polyploid genomes." Nat Commun **5**.

Love, M. I., W. Huber and S. Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." Genome Biology **15**(12): 550.

Lowe, K., E. Wu, N. Wang, G. Hoerster, C. Hastings, M.-J. Cho, C. Scelonge, B. Lenderts, M. Chamberlin, J. Cushatt, L. Wang, L. Ryan, T. Khan, J. Chow-Yiu, W. Hua, M. Yu, J. Banh, Z. Bao, K. Brink, E. Igo, B. Rudrappa, P. M. Shamseer, W. Bruce, L. Newman, B. Shen, P. Zheng, D. Bidney, C. Falco, J. Register, Z.-Y. Zhao, D. Xu, T. Jones and W. Gordon-Kamm (2016). "Morphogenic Regulators Baby boom and Wuschel Improve Monocot Transformation." The Plant Cell **28**(9): 1998.

Ma, C., C. Zhu, M. Zheng, M. Liu, D. Zhang, B. Liu, Q. Li, J. Si, X. Ren and H. Song (2019). "CRISPR/Cas9-mediated multiple gene editing in Brassica oleracea var. capitata using the endogenous tRNA-processing system." Horticulture research **6**: 20-20.

Macovei, A., N. R. Sevilla, C. Cantos, G. B. Jonson, I. Slamet-Loedin, T. Čermák, D. F. Voytas, I.-R. Choi and P. Chadha-Mohanty (2018). "Novel alleles of rice eIF4G generated by CRISPR/Cas9-targeted mutagenesis confer resistance to Rice tungro spherical virus." Plant Biotechnology Journal **16**(11): 1918-1927.

Moloney, M. M., J. M. Walker and K. K. Sharma (1989). "High efficiency transformation ofBrassica napus usingAgrobacterium vectors." Plant Cell Reports **8**(4): 238-242.

Nellist, C. F., W. Qian, C. E. Jenner, J. D. Moore, S. Zhang, X. Wang, W. H. Briggs, G. C. Barker, R. Sun and J. A. Walsh (2014). "Multiple copies of eukaryotic translation initiation factors in Brassica rapa facilitate redundancy, enabling diversification through variation in splicing and broad-spectrum virus resistance." The Plant Journal **77**(2): 261-268.

Nicaise, V., J.-L. Gallois, F. Chafiai, L. M. Allen, V. Schurdi-Levraud, K. S. Browning, T. Candresse, C. Caranta, O. Le Gall and S. German-Retana (2007). "Coordinated and selective recruitment of eIF4E and eIF4G factors for potyvirus infection in Arabidopsis thaliana." FEBS Letters **581**(5): 1041-1046.

Pyott, D. E., E. Sheehan and A. Molnar (2016). "Engineering of CRISPR/Cas9-mediated potyvirus resistance in transgene-free Arabidopsis plants." Mol Plant Pathol.

Qi, Y., L. Lowder and A. Malzahn (2016). "Rapid evolution of manifold CRISPR systems for plant genome editing." Frontiers in Plant Science **7**(1683).

Radke, S. E., J. C. Turner and D. Facciotti (1992). "Transformation and regeneration of Brassica rapa using Agrobacterium tumefaciens." Plant Cell Reports **11**(10): 499-505.

Razzaq, A. and A. Masood (2018). "CRISPR/Cas9 System: A Breakthrough in Genome Editing." Molecular Biology **07**.

Reddy, A. S. N., Y. Marquez, M. Kalyna and A. Barta (2013). "Complexity of the Alternative Splicing Landscape in Plants." The Plant Cell **25**(10): 3657.

Riley, M. K. and W. Vermerris (2017). "Recent Advances in Nanomaterials for Gene Delivery-A Review." Nanomaterials (Basel, Switzerland) **7**(5): 94.

Robaglia, C. and C. Caranta (2006). "Translation initiation factors: a weak link in plant RNA virus infection." Trends in Plant Science **11**(1): 40-45.

Robinson, J. T., H. Thorvaldsdóttir, W. Winckler, M. Guttman, E. S. Lander, G. Getz and J. P. Mesirov (2011). "Integrative genomics viewer." Nature biotechnology **29**(1): 24-26.

Rodriguez-Hernandez, A. M., B. Gosalvez, R. N. Sempere, L. Burgos, M. A. Aranda and V. Truniger (2012). "Melon RNA interference (RNAi) lines silenced for Cm-eIF4E show broad virus resistance." Molecular Plant Pathology **13**(7): 755-763.

Rusholme Pilcher, R., E. E Higgins, J. Walsh and D. Lydiate (2007). "Genetic control of broad-spectrum resistance to Turnip mosaic virus in Brassica rapa (Chinese cabbage)." The Journal of general virology **88**: 3177-3186.

Russo, V. M. (2008). "Vegetable Brassicas and Related Crucifers. Crop Production Science in Horticulture 14." International Journal of Vegetable Science **14**(1): 93-93.

Ruud, K. A., C. Kuhlow, D. J. Goss and K. S. Browning (1998). "Identification and Characterization of a Novel Cap-binding Protein from Arabidopsis thaliana." Journal of Biological Chemistry **273**(17): 10325-10330.

Shattuck, V. I. (2010). The Biology, Epidemiology, and Control of Turnip Mosaic Virus. Horticultural Reviews, John Wiley & Sons, Inc.**:** 199-238.

Simon, D., A. Rouault and M. C. Chopin (1986). "High-efficiency transformation of Streptococcus lactis protoplasts by plasmid DNA." Applied and Environmental Microbiology **52**(2): 394.

Sinha, A., A. C. Wetten and P. D. S. Caligari (2003). "Effect of biotic factors on the isolation of *Lupinus albus* protoplasts." Australian Journal of Botany **51**(1): 103-109.

Sparrow, P. A. C., P. J. Dale and J. A. Irwin (2004). "The use of phenotypic markers to identify Brassica oleracea genotypes for routine high-throughput Agrobacterium-mediated transformation." Plant Cell Reports **23**(1): 64-70.

Sparrow, P. A. C. H. n. and J. A. Irwin (2015). Brassica oleracea and B. napus. Agrobacterium Protocols: Volume 1. K. Wang. New York, NY, Springer New York**:** 287-297.

Sun, B., A. Zheng, M. Jiang, S. Xue, Q. Yuan, L. Jiang, Q. Chen, M. Li, Y. Wang, Y. Zhang, Y. Luo, X. Wang, F. Zhang and H. Tang (2018). "CRISPR/Cas9-mediated mutagenesis of homologous genes in Chinese kale." Scientific Reports **8**(1): 16786.

Sun, Q., L. Lin, D. Liu, D. Wu, Y. Fang, J. Wu and Y. Wang (2018). "CRISPR/Cas9-Mediated Multiplex Genome Editing of the BnWRKY11 and BnWRKY70 Genes in Brassica napus L." International Journal of Molecular Sciences **19**(9).

Swift, T. A., M. Duchi, S. A. Hill, D. Benito-Alifonso, R. L. Harniman, S. Sheikh, S. A. Davis, A. M. Seddon, H. M. Whitney, M. C. Galan and T. A. A. Oliver (2018). "Surface functionalisation significantly changes the physical and electronic properties of carbon nano-dots." Nanoscale **10**(29): 13908-13912.

Truniger, V. and M. A. Aranda (2009). Chapter 4 - Recessive Resistance to Plant Viruses. Advances in Virus Research. G. Loebenstein and J. P. Carr, Academic Press. **75:** 119-231.

U, N. (1935). "Genome analysis in Brassica with special reference to the experimental formation of B. napus and peculiar mode of fertilization." Japanese Journal of Botany **7**: 389-452.

Waki, T., T. Hiki, R. Watanabe, T. Hashimoto and K. Nakajima (2011). "The Arabidopsis RWP-RK Protein RKD4 Triggers Gene Expression and Pattern Formation in Early Embryogenesis." Current Biology **21**(15): 1277-1281.

Walsh, J. A. and C. E. Jenner (2002). "Turnip mosaic virus and the quest for durable resistance." Molecular Plant Pathology **3**(5): 289-300.

Walsh, J. A., R. L. Rusholme, S. L. Hughes, C. E. Jenner, J. M. Bambridge, D. J. Lydiate and S. K. Green (2002). "Different Classes of Resistance to Turnip Mosaic Virus in Brassica rapa." European Journal of Plant Pathology **108**(1): 15-20.

Walsh, J. A., A. G. Sharpe, C. E. Jenner and D. J. Lydiate (1999). "Characterisation of resistance to turnip mosaic virus in oilseed rape (Brassica napus) and genetic mapping of TuRB01." Theoretical and Applied Genetics **99**(7): 1149-1154.

Walsh, J. A. and J. A. Tomlinson (1985). "Viruses infecting winter oilseed rape (Brassica napus ssp. oleifera)." Annals of Applied Biology **107**(3): 485-495.

Wang, R. Y. and T. P. Pirone (1999). "Purification and Characterization of Turnip Mosaic Virus Helper Component Protein." Phytopathology **89**(7): 564-567.

Wang, X., H. Wang, J. Wang, R. Sun, J. Wu, S. Liu, Y. Bai, J.-H. Mun, I. Bancroft, F. Cheng, S. Huang, X. Li, W. Hua, J. Wang, X. Wang, M. Freeling, J. C. Pires, A. H. Paterson, B. Chalhoub, B. Wang, A. Hayward, A. G. Sharpe, B.-S. Park, B. Weisshaar, B. Liu, B. Li, B. Liu, C. Tong, C. Song, C. Duran, C. Peng, C. Geng, C. Koh, C. Lin, D. Edwards, D. Mu, D. Shen, E. Soumpourou, F. Li, F. Fraser, G. Conant, G. Lassalle, G. J. King, G. Bonnema, H. Tang, H. Wang, H. Belcram, H. Zhou, H. Hirakawa, H. Abe, H. Guo, H. Wang, H. Jin, I. A. P. Parkin, J. Batley, J.-S. Kim, J. Just, J. Li, J. Xu, J. Deng, J. A. Kim, J. Li, J. Yu, J. Meng, J. Wang, J. Min, J. Poulain, J. Wang, K. Hatakeyama, K. Wu, L. Wang, L. Fang, M. Trick, M. G. Links, M. Zhao, M. Jin, N. Ramchiary, N. Drou, P. J. Berkman, Q. Cai, Q. Huang, R. Li, S. Tabata, S. Cheng, S. Zhang, S. Zhang, S. Huang, S. Sato, S. Sun, S.-J. Kwon, S.-R. Choi, T.-H. Lee, W. Fan, X. Zhao, X. Tan, X. Xu, Y. Wang, Y. Qiu, Y. Yin, Y. Li, Y. Du, Y. Liao, Y. Lim, Y. Narusaka, Y. Wang, Z. Wang, Z. Li, Z. Wang, Z. Xiong and Z. Zhang (2011). "The genome of the mesopolyploid crop species Brassica rapa." Nature Genetics **43**: 1035.

Ward, C. W. and D. D. Shukla (1991). "Taxonomy of Potyviruses: Current Problems and Some Solutions." Intervirology **32**(5): 269-296.

Wingett, S. and S. Andrews (2018). "FastQ Screen: A tool for multi-genome mapping and quality control [version 2; peer review: 4 approved]." F1000Research **7**(1338).

Xu, Y., A. Zeng, L. Song, J. Li and J. Yan (2019). "Comparative transcriptomics analysis uncovers alternative splicing events and molecular markers in cabbage (Brassica oleracea L.)." Planta.

Yang, H., J.-J. Wu, T. Tang, K.-D. Liu and C. Dai (2017). "CRISPR/Cas9-mediated genome editing efficiently creates specific mutations at multiple loci using one sgRNA in Brassica napus." Scientific Reports **7**(1): 7489.

Yoo, S.-D., Y.-H. Cho and J. Sheen (2007). "Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis." Nature Protocols **2**: 1565.

Appendix

Primers

Appendix Table 1 – Primers designed from paralog sequence alignment for amplification of specific paralog copies in *Brassica rapa* and *Brassica oleracea.*

Primer	DNA Type	Sequence (5'-3')	Direction	Gene	Specific Target
LI1	cDNA	CATCGAGGAGTTCTGGAGTCTTTACAATAAC	Forward	eIF4E	C5,A7,A9,C7-168
LI2	cDNA	CATAGTCCACTTTCCTCCATTAGCACATAC	Reverse	eIF4E	C5
LI3	cDNA, gDNA	TTCCAGAGTAGCCTTAGGGAACATCATAGT	Reverse	eIF4E	A7
LI4	cDNA, gDNA	CAGTGTAGCCGTAGGGAACATCATAGTC	Reverse	eIF4E	A9
LI5	cDNA, gDNA	GTAAGCGCTCTTTGCGCCTCTG	Reverse	eIF4E	A1
LI6	cDNA	AAGTCTTGGCTTTACACCTTGCTTGC	Forward	eIF4E	A3,C7-408
LI7	cDNA, gDNA	TAAGCGTCTTTCGCGTTCCTATCC	Reverse	eIF4E	A3
LI8	cDNA	TCGAGGAGTTCTGGAGTCTGTTCAATAAC	Forward	eIF4E	A8
LI9	cDNA, gDNA	CAACAGCTCCGCATATCTCATCTCC	Reverse	eIF4E	A8
LI10	cDNA	TAAGCGCTCTTCGCGTTCCTATC	Reverse	eIF4E	C7-408
LI11	cDNA	CCACTTTCCTCCATGAGCACAAAC	Reverse	eIF4E	C7-168
LI12	cDNA	TTACACCTTGCTTGCATTGATTGGAG	Forward	eIF4E	C1,A1
LI13	cDNA	GTAAGTGTTCTTTGCGCCCCTGTC	Reverse	eIF4E	C1

LI14	cDNA	AAGATTTCTGGGGTTTGCACGAGAC	Forward	eIF(Iso)4E	A5,C8,A8,A4,C4,C6
LI15	cDNA	CACTAGCAACAACACCACAAATCTCATCT	Reverse	eIF(Iso)4E	A5
LI16	cDNA	CTAGCAACTACACCACAAATCTCATCTGC	Reverse	eIF(Iso)4E	C8
LI17	cDNA	GATCTCATCTGCCTCATCAAACTGCTC	Reverse	eIF(Iso)4E	A8
LI18	cDNA	AACCACCCCACAGATCTCATCAGC	Reverse	eIF(Iso)4E	A4
LI19	cDNA	CTAGCAACCACACCACAAATCTCATCA	Reverse	eIF(Iso)4E	C4
LI20	cDNA	TGGCGCACACTAGCAACAAGACC	Reverse	eIF(Iso)4E	C6
LI21	cDNA	CTGATTTGATGAAGATTCCTGCTCG	Forward	eIF4G	A4
LI22	cDNA	GCGTTTGGCTTAGAGGGGACATA	Reverse	eIF4G	A4
LI23	cDNA	CCTGAAGACTGTCAAAGGGATAC	Forward	eIF(Iso)4G	A2,C2
LI24	cDNA	CTCTTCAGGTGGGAAGGTGGGTAG	Reverse	eIF(Iso)4G	A2
LI25	cDNA	CAGAAAGTCGGCTCAAGGACGG	Reverse	eIF(Iso)4G	C2
LI26	cDNA	CAATCTCTCCGACTCACACCATTCC	Forward	eIF(Iso)4G	A3,C7
LI27	cDNA	CAAAATCTCATGACCCTCTCCTCTCC	Reverse	eIF(Iso)4G	A3,C7
LI28	cDNA	GAAAACAAGCCTGCAGCTCAAGTTC	Forward	eIF(Iso)4G	A1,C1
LI29	cDNA	CTTTTAGCTGACCATGGGACCTCTG	Reverse	eIF(Iso)4G	A1,C1

LI30	cDNA	GATCAGAAGCTCAACGAAGTCCTGG	Forward	eIF4A	C3-641
LI31	cDNA	GACCAGAAGCTCAATGAAGTCCTCG	Forward	eIF4A	A8-607
LI32	cDNA	GCACTGGACAAGGGTGAAGTCG	Reverse	eIF4A	C3-641,A8-607
LI33	cDNA	ACGAAGTTCTTGAGGGACAGGACG	Forward	eIF4A	A3-168,C3-255
LI34	cDNA	GCACTGGACGAGGGTGAAGTCT	Reverse	eIF4A	A3-168
LI35	cDNA	TGCCATACTCGGCGAGGAAGG	Forward	eIF4A	A2
LI36	cDNA	GCAACACGCCAGAGCAGAAG	Reverse	eIF4A	A2
LI37	cDNA	GAACGAAGTTCTCGAGGGACAAG	Forward	eIF4A	A1
LI38	cDNA	GCCTTTGCAGAACGGGACG	Reverse	eIF4A	A1,C1,C6-140
LI39	cDNA	GGGTATCTATGCCTATGGTTTTG	Reverse	eIF4A	A8-138
LI40	cDNA	GATCCTCACGGACACTGGTTC	Reverse	eIF4A	A8-138,C6-326
LI41	cDNA	CCATACTCGGTGAGGAAGGACAG	Forward	eIF4A	A7
LI42	cDNA	GGCACTGGAGGAGAGTGTAGTCG	Reverse	eIF4A	A7
LI43	cDNA	GAAGTCCTGGAGGGACAAAATGAG	Forward	eIF4A	A6
LI44	cDNA	GCAAGACACCAGAGCAGAACG	Reverse	eIF4A	A6
LI45	cDNA	CCATGCCATCAAAACTCTCATAAAC	Reverse	eIF4A	A3-647
LI46	cDNA	CTATTCTGTTTCTGCTTGTCATGGC	Forward	eIF4A	A3-647
LI47	cDNA	CTCTTAATCAACCTCAAAGTCATGG	Forward	eIF4A	A3-647
LI48	cDNA	CAAAGGTTCTAAATCATGGCAGGAT	Forward	eIF4A	A3-647
LI49	cDNA	ACGAAGTTCTTGAAGGACAGGATG	Forward	eIF4A	A5
LI50	cDNA	CGGTACCGGACTGAGCCTGC	Reverse	eIF4A	A5

LI51	cDNA	CGAAGTTCTCGAGGGACAGGAC	Forward	eIF4A	C3-255
LI52	cDNA	CAGCAGGAATGGCGTCAGAAGG	Forward	eIF4A	C2-218
LI53	cDNA	GACCCTTGCAGAAGGGAATAATC	Reverse	eIF4A	C2-218
LI54	cDNA	TAACAGCAGTCATGGCAGGAATG	Forward	eIF4A	C2-218
LI55	cDNA	CGAAGTTCTCGAGGGACAAGATG	Forward	eIF4A	C1
LI56	cDNA	GGTTCCACCAACACAAGCATGG	Reverse	eIF4A	C6-140
LI57	cDNA	CCGGATTGGGCTTGCTGGAT	Reverse	eIF4A	C6-326
LI58	cDNA	TGTTCTTGACGAGGCCGACG	Forward	eIF4A	C5
LI59	cDNA	CTCAACGTTGACGTAAAACTGCTT	Reverse	eIF4A	C5
LI60	cDNA	AGGCCGGCGTCCACGTC	Forward	eIF4A	C5
LI61	cDNA	TCATGGATCTGGAGCCTGGAAC	Forward	β-tublin	All
LI62	cDNA	GGAATGGCAAACCTGAAACCC	Reverse	β-tublin	All
LI63	cDNA	CACCGGAAGCCTCTGACTGAT	Forward	TIP41	All
LI64	cDNA	TGATGGTGTGCTTGTGCGGTTGAG	Reverse	TIP41	All
LI65	cDNA	TGATGAACAAGACGAGCAAGAAAC	Forward	SYP61	All
LI66	cDNA	ATAGATGAACTGGGAACGGAGATG	Reverse	SYP61	All
LI67	cDNA	GGAAGATCCTGAGTGTGCTAA	Forward	eIF(Iso)4E	A4
LI68	cDNA	GTGAAGGCTGGGTTTGCTGGTG	Forward	Actin-7	All
LI69	cDNA	CCATGTCATCCCAGTTGC	Reverse	Actin-7	All
LI79	gDNA	ATGGCGGAAGAGGAGAATC	Forward	eIF4E	A9,A7
LI80	gDNA	TGACCATCCCTACAAGAAG	Forward	eIF4E	A3
LI81	gDNA	TTGCGGAAGAAGCGAAC	Forward	eIF4E	A8
LI82	gDNA	CACTCAAGCCTAATGTCG	Forward	eIF4E	A1

LI83	gDNA	GAGGATGTGAACGAAGC	Forward	eIF(Iso)4E	A5,A4,A8,C7
LI84	gDNA	CATTTGATTTAGTCCGTGTC	Reverse	eIF(Iso)4E	A5
LI85	gDNA	CAACACCAAAGGTACATAGTG	Forward	eIF(Iso)4E	A4
LI86	gDNA	CTGCATATACATAGACCACAG	Reverse	eIF(Iso)4E	A8
LI87	gDNA	AAAGCAGCACCGAACGTG	Forward	eIF4G	A4
LI89	gDNA	GATTGATCGTCGCTGAGAAG	Forward	eIF(Iso)4G	A2
LI90	gDNA	GACCAAAACTCCCACAAG	Reverse	eIF(Iso)4G	A2
LI92	gDNA	GCTTACACCAAACTACGGC	Reverse	eIF(Iso)4G	A3
LI94	gDNA	TCATCAAGCATAGACCCG	Reverse	eIF(Iso)4G	A1
LI99	gDNA	GTCCAAAATTTGATGCTTC	Reverse	eIF(Iso)4E	C7
LI100	gDNA	GCATCAAGGCCCAAAG	Forward	eIF(Iso)4E	C6
LI101	gDNA	CGGAAGTAACGGAGAAG	Reverse	eIF(Iso)4E	C6
LI102	cDNA	GCAAGTGGGAACGCTAATC	Reverse	eIF4E	C5
LI103	cDNA	GCCAATCAACACTCGAGAAC	Reverse	eIF4E	C1
LI104	cDNA	TCACCTCAAACACAAACTG	Reverse	eIF4E	A9
LI109	gDNA	CATCCTCTACGGGTTTG	Reverse	nCBP	A10
LI110	gDNA	CAGGGTCAAATCTCAGTAG	Forward	nCBP	A10
LI111	gDNA	CGGGTTCGTCCTCTAAT	Forward	nCBP	A3
LI112	gDNA	GTCCATACGTCTTCCCAG	Reverse	nCBP	A3

LI113	gDNA	GCAAACATCGACAACCTC	Forward	nCBP	A2
LI114	gDNA	GTTGCAGAAGTAGCCAGG	Reverse	nCBP	A2
LI115	gDNA	GCTGGTCTTCATCCTCTAC	Forward	nCBP	C3
LI116	gDNA	GCTACCCAAATCTTTCTCC	Reverse	nCBP	C3
LI117	gDNA	CGAGGTCAAGGATGGTC	Forward	nCBP	C2
LI118	gDNA	CACTTTGAACCTATCCTCTC	Reverse	nCBP	C2
LI119	gDNA	ATGGAGATTACGGAGAGGAG	Forward	nCBP	C9
LI120	gDNA	GCGAAAGATCTTGTGAATAG	Reverse	nCBP	C9
LI125	cDNA	GAAGGGAGCGAGCTAGCTA	Reverse	nCBP	A10
LI126	cDNA	CATAAATGGTCCAATGTCACCT	Reverse	nCBP	A10
LI127	cDNA	CCATAACATATGAATGAGGTAGC	Reverse	nCBP	A3
LI128	cDNA	CACATAAGCATGAGGCAGC	Reverse	nCBP	A2
LI129	cDNA, gDNA	GAAAGTGTCAGTCAAGTCATG	Reverse	eIF4E	A1
LI132	gDNA	GTTCGGAGAAGAGAAGACG	Forward	eIF4E	A1
LI133	gDNA	GATACAGGAGTCTGTCCATG	Reverse	eIF4G	A4
LI134	gDNA	GTGTTGAGCTTGAGGCCTG	Forward	eIF(Iso)4G	A3
LI135	gDNA	CTTACCAATGCCGAAGACG	Forward	eIF(Iso)4G	A1
LI140	gDNA	GCTCCAGTACAGCCTTCTC	Reverse	eIF(Iso)4G	A3
LI141	gDNA	CCTGTTCTGGTGAAAGC	Forward	eIF(Iso)4G	A3
LI142	gDNA	CGCGAATCTAGAGATGC	Forward	eIF(Iso)4G	A1
LI143	gDNA	GCTGGCATCACATCTG	Forward	eIF(Iso)4G	A4

LI149	gDNA	GGACAGAGATTGATTACTTGTAGG	Forward	eIF4G	A1
LI151	gDNA	TTCTGGGAGGATCTGGT	Forward	nCBP	A10,A3, A2

Programming Script

Raw RNASeq Processing and Alignment

#made SampleSheet.csv and put into top directory

#used index adapters from illumina PDF

#convert Raw BCL to Fasta.gz

```
bcl2fastq --input-dir
/home/u1560846/190425_NB501709_0127_AH52CWBGXB/Data/Intensities/Base
Calls/ -p 12 --output-dir
/home/u1560846/190425_NB501709_0127_AH52CWBGXB/fastq_files --no-lane-
splitting
```

```
for i in $(ls *.gz | cut -f1 -d'.' | sort -u); do fastqc ${i}.gz -o
/home/u1560846/190425_NB501709_0127_AH52CWBGXB/multiqc/ -t 6 ; done
```

#combine fastqc results using multiqc

```
multiqc /home/u1560846/190425_NB501709_0127_AH52CWBGXB/multiqc
```

#download genome and annotation files

```
wget  http://www.genoscope.cns.fr/externe/plants/data/Brapa_chromosomes.fasta
```

```
wget http://www.genoscope.cns.fr/externe/plants/data/Brapa_annotation.gff
```

#Make STAR genome index with GFF file included

```
STAR --runMode genomeGenerate --runThreadN 20 --genomeDir
/home/u1560846/B.rapa/star_genome_index/ --genomeFastaFiles
/home/u1560846/B.rapa/star_genome_index/Brapa.fa --sjdbGTFfile
/home/u1560846/B.rapa/star_genome_index/Brapa_annotation.gff --
sjdbGTFfeatureExon CDS --sjdbGTFtagExonParentTranscript Parent --
sjdbGTFtagExonParentGene Name --sjdbOverhang 149 --readFilesCommand zcat
```

Align with STAR R-o-18_1

```
for i in $(ls | cut -f1 -d'.' | sort -u); do STAR --genomeDir
/home/u1560846/B.rapa/star_genome_index/ --runThreadN 20 --readFilesIn
/home/u1560846/190425_NB501709_0127_AH52CWBGXB/fastq_files/Brapa/${i}
.fastq.gz --outFileNamePrefix /home/u1560846/B.rapa/star_out_R/${i}/${i}_ --
readFilesCommand zcat --outFilterScoreMinOverLread 0.3 --
```

```
outFilterMatchNminOverLread 0.3 --alignIntronMax 5000 --quantMode
GeneCounts --alignEndsType EndToEnd; done

                #Change STAR output ready for DESeq2

for i in $(ls | cut -f1 -d'.' | sort -u); do awk 'NR>4 {print $1 "\t" $2}'
/home/u1560846/star_out_R/${i}/${i}_ReadsPerGene.out.tab >
/home/u1560846/star_out_R/${i}/${i}_counts.txt ; done

                #Move to terminal with scp, DiffExpr  analysis done on terminal
following ProcessData.R

                #covert .sam output to .bam

for i in $(ls | cut -f1 -d'.' | sort -u); do samtools view -S -b
/home/u1560846/star_out/${i}/${i}_Aligned.out.sam >
/home/u1560846/B.rapa/star_out/${i}/${i}_Aligned.out.bam ; done

                #Sort .bam file

for i in $(ls | cut -f1 -d'.' | sort -u); do samtools sort
/home/u1560846/star_out/${i}/${i}_Aligned.out.bam
/home/u1560846/B.rapa/star_out/${i}/${i}_Aligned.out.sorted -@20 ; done



                #index bam files

for i in $(ls | cut -f1 -d'.' | sort -u); do samtools index
/home/u1560846/B.rapa/star_out/${i}/${i}_Aligned.out.sorted.bam ; done
```

DEseq2 Analysis

Written with the help of Dr. Sascha Ott

```
# load the required functions
library('DESeq2')
source(file='rnaSeqAnalysisFuncs.r')

##  Read in the file identifying the different samples
sampleTable = read.table(sampleDataFilename,header=TRUE)
head(sampleTable)

##  Read in the results from the LibiNorm analysis (the counts files)
ddsHTSeq<-DESeqDataSetFromHTSeqCount(sampleTable=sampleTable,directory
= htseqDir,design = ~condition)

##  And perform the analysis
dds<-DESeq(ddsHTSeq)



## PCA plots
## Apply rlog transformation to the data.
rld <- rlog(dds)
```

```
PCx = 1
PCy = 2
plotPCAEx(rld,PCx,PCy,labels=TRUE)



reference = 'R-o-18'
treatment = 'RLR22'
Pvalue = 0.05
res = results(dds, contrast=c(factor,treatment,reference))
View(res)

# convert pvalue to -log10(pvalue)
# convert fold change to log2(fold change).DEseq already produces log2 fold
change; log2FoldChange.
# set x limit and y limit depending on fold change and p value.
log10.pval <- -log10(res$padj)
log2.fc    <- res$log2FoldChange
plot(log2.fc,log10.pval,

        xlab="log2 (fold change)",
         ylab="-log10 (p-value)",
        col="black",xlim=c(-10,10),
        ylim=c(0,75))

abline(h = -log10(0.05),col='red',lwd=1.5)
abline(v=-log2(2),col='blue',lwd=1.5)
abline(v=log2(2),col='blue',lwd=1.5)




##  xy plot
##  Identify the genes that meet criteria
topGenes <- res$padj < Pvalue
plotxy(dds,reference,treatment,topGenes)

## Output significant DE genes
DE_genes<-res[which(res$padj < Pvalue),]
write.table(DE_genes,file=paste("result_DE",treatment,"",reference,".txt",sep=""),se
p="\t",col.names=NA,quote=FALSE)
```

TRINITY Annotation

Written with the help of Richard Stark

Trinity De novo:

```
Trinity --seqType fq --single
/home/u1672689/Luca/Brapa/RLR22_01_S4_R1_001.fastq.gz,/home/u1672689/Luc
a/Brapa/R-o-
```

```
18_01_S10_R1_001.fastq.gz,/home/u1672689/Luca/Brapa/RLR22_02_S5_R1_001.f
astq.gz,/home/u1672689/Luca/Brapa/R-o-
18_02_S11_R1_001.fastq.gz,/home/u1672689/Luca/Brapa/RLR22_03_S6_R1_001.f
astq.gz,/home/u1672689/Luca/Brapa/R-o-18_03_S12_R1_001.fastq.gz --CPU 16 --
max_memory 120G --output /home/u1672689/Luca/trinity

### Supertranscripts approach:
### https://github.com/trinityrnaseq/trinityrnaseq/wiki/SuperTranscripts

Trinity_gene_splice_modeler.py \

      --trinity_fasta /home/u1672689/Luca/trinity/Trinity.fasta
for ii in $(find /home/u1672689/Luca/Brapa -name R*); do
echo $ii;
jj=$(basename $ii)
jj=${jj/_R1_001.fastq.gz/}
echo $jj $jj
$ii >> samples.txt
done

dexseq_wrapper.pl

          --genes_fasta SuperTranscripts.fa \

          --genes_gtf SuperTranscripts.gtf \

           --samples_file /home/u1672689/Luca/samples.txt \

          --out_prefix DTU --aligner STAR

align_and_estimate_abundance.pl --transcripts Trinity.fasta --est_method RSEM --
aln_method bowtie --trinity_mode --prep_reference

for ii in $(find /home/u1672689/Luca/Brapa -name R*); do
echo $ii;
jj=$(basename $ii)
jj=${jj/_R1_001.fastq.gz/}
align_and_estimate_abundance.pl --seqType fq  \
    --single $ii \
    --transcripts /home/u1672689/Luca/trinity/Trinity.fasta \
    --est_method RSEM  --aln_method bowtie \
    --thread_count 16 \
    --trinity_mode --output_dir ${jj}.RSEM
done
```

DEseq2 Alternative Splicing Analysis

Written with the help of Richard Stark

```
library(DEXSeq)
samples_info = read.table("DTU.sample_table", header=T, row.names=1)
dxd = DEXSeqDataSetFromHTSeq(as.vector(samples_info$counts_filename),
 sampleData=samples_info, design = ~ sample + exon + condition:exon,
 flattenedfile="trinity_genes.gtf.dexseq.gff")
dxd = estimateSizeFactors( dxd )
```

```
dxd = estimateDispersions( dxd )
plotDispEsts( dxd )
dxd = testForDEU( dxd )
dxd = estimateExonFoldChanges( dxd, fitExpToVar="condition")
dxr1 = DEXSeqResults( dxd )
dxr1.sorted = dxr1[order(dxr1$padj),]
save(list = ls(all=TRUE), file = "DTU.dexseq.Rdata")
write.table(dxr1.sorted, file="DTU.dexseq.results.dat", quote=F, sep=" ")
pdf("DTU.dexseq.pdf")
top_genes = unique(dxr1.sorted$groupID[dxr1.sorted$padj < 0.1 & !
 is.na(dxr1.sorted$padj)])
top_genes = top_genes[1:min(50, length(top_genes))]
message("Top 50 genes: (", paste(top_genes, collapse=','), ")")
for (gene in top_genes) {
    plotDEXSeq( dxr1 , gene, legend=TRUE, cex.axis=1.2, cex=1.3, lwd=2 ,
 expression=FALSE, norCounts=TRUE, splicing=TRUE,
 displayTranscripts=TRUE)
}
```