Aix*Marseille université
Socialement engagée

NNT/NL : 0000AIXM0000/000ED000

THÈSE DE DOCTORAT

Soutenue à INRAE GAFL Avignon
le 13 septembre 2024 par

Kyoka Kuroiwa

Mise en place d'allèles fonctionnels de résistance aux virus à partir des facteurs de sensibilité eIF4E par édition du génome chez la tomate

Design of virus resistance alleles based on the susceptibility factors eIF4E via genome-editing in tomato

Discipline
Biologie Santé

Spécialité
Biologie Végétale

École doctorale
62 Sciences de la Vie et de la Santé

Laboratoire/Partenaires de recherche
INRAE – Génétique et Amélioration des Fruits et Légumes (GAFL), Avignon

Composition du jury

Laurence ALBAR IRD – UMR PHIM, Montpellier	Rapporteuse
Nathalie GONZALEZ INRAE – UMR BFP, Bordeaux	Rapporteuse
Fabien NOGUÉ INRAE – IJPB, Versailles	Examinateur
Christophe ROBAGLIA AMU/CNRS – BIAM – LGBP, Marseille	Président du jury
Jean-Luc GALLOIS INRAE – GAFL, Avignon	Directeur de thèse
Marianne MAZIER INRAE – GAFL, Avignon	Co-directrice de thèse

INRAE

Affidavit

I, undersigned, Kyoka Kuroiwa, hereby declare that the work presented in this manuscript is my own work, carried out under the scientific direction of Jean-Luc Gallois et Marianne Mazier, in accordance with the principles of honesty, integrity and responsibility inherent to the research mission. The research work and the writing of this manuscript have been carried out in compliance with both the French national charter for Research Integrity and the Aix-Marseille University charter on the fight against plagiarism.

This work has not been submitted previously either in this country or in another country in the same or in a similar version to any other examination body.

Place Yachiyo-shi, Japan, date 28th June 2024



Cette œuvre est mise à disposition selon les termes de la Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International.

Liste de publications et participation aux conférences

Liste des publications réalisées dans le cadre du projet de thèse :

1. Kuroiwa K, Thenault C, Nogué F, Perrot L, Mazier M, Gallois J-L. "CRISPR-based knock-out of eIF4E2 in a cherry tomato background successfully recapitulates resistance to pepper veinal mottle virus". Plant Sci. 2022;316: 111160.
2. Kuroiwa, K., Danilo, B., Perrot, L., Thenault, C., Veillet, F., Delacote, F., Duchateau, P., Nogué, F., Mazier, M., and Gallois, J.-L. (2023). An iterative gene-editing strategy broadens eIF4E1 genetic diversity in Solanum lycopersicum and generates resistance to multiple potyvirus isolates. Plant Biotechnol. J. Advance Access published January 30, 2023, doi:10.1111/pbi.14003.

Participation aux conférences et écoles d'été au cours de la période de thèse :

Oral presentations:

"An iterative gene editing strategy broadens eIF4E1 genetic diversity in Solanum lycopersicum, triggering resistance to multiple potyvirus isolates"

- Plant Biology Europe 2023 (July 3rd-6th 2023) in Marseille, France
- COLLOQUE EDSVS 2023 (May 30th 2023) in Marseille, France ; Oral presentation award 2nd place

"An iterative gene editing strategy broadens *eIF4E1* genetic diversity in *Solanum lycopersicum*, triggering resistance to several potyvirus isolates"

- 19th Rencontres de Virologie Végétale (RVV2023) in Aussois, France (15-19th janvier 2023) ; Best oral presentation award
- 228th Symposium of Hokkaido Branch of the Phytopathological Society of Japan "Front-line researches on viruses and host resistances" in Sapporo, Japan (October 20th, 2022)
- The annual 3rd PlantEd Conference in Dusseldorf, Germany (September 5-7th, 2022)

"Strategies for developing genetic resistances against potyviruses in tomato with CRISPR systems"

- For INRAE meeting « Les Fous de la tomate » in Bordeaux, France (November 25-26th, 2021)

Poster presentations:

"An iterative gene editing strategy broadens eIF4E1 genetic diversity in Solanum lycopersicum, triggering resistance to multiple potyvirus isolates"

- The 12th International Congress of Plant Pathology 2023 (August 20-25th 2023) in Lyon, France
- Mendel Early Career Symposium 16 – 17 May 2024 in Vienna, Austria

Résumé

Les résistances génétiques aux agents pathogènes constituent un levier important pour protéger la production agricole tout en maintenant un faible impact sur l'environnement. Jusqu'à présent, la diversité naturelle des allèles fonctionnels du facteur d'initiation de la traduction eucaryote (*eIF4E*) a fourni des ressources génétiques pour la sélection de la résistance aux potyvirus dans de nombreuses cultures. Cependant, la redondance génique et l'importance physiologique des facteurs eIF4E rendent difficile la mise en œuvre de résistances liées à ces gènes par mutagénèse chez d'autres espèces. Dans cette thèse, j'ai évalué différentes approches d'édition du génome chez la tomate pour générer des allèles de gènes *eIF4E* afin de conférer à la plante une résistance aux potyvirus par perte de sensibilité.

Dans la première partie de ce travail, nous avons inactivé le gène *eIF4E2* par une approche CRISPR/Cas9. Ce travail a confirmé les limites de la stratégie d'inactivation. En effet, mon travail montre que l'inactivation de *eIF4E2* permet de copier une résistance aux potyvirus, mais avec un spectre de résistance très étroit. Dans la deuxième partie, nous avons modifié en deux endroits le gène *eIF4E1* de tomate par une approche de remplacement via l'utilisation de TALEN et par l'édition de base par éditeur de base cytosine, respectivement. Ceci a permis d'introduire des mutations non-synonymes dans *eIF4E1* permettant d'imiter des mutations telles qu'habituellement présentes dans des allèles de résistance naturelle. Je montre que ces modifications précises confèrent une résistance à plusieurs isolats de potyvirus sans compromettre la fonction physiologique de la protéine eIF4E1.

Afin d'évaluer le potentiel de l'utilisation pratique de cette approche de modification génique en agriculture, j'ai poursuivi mon travail par l'évaluation des conséquences phénotypiques potentielles dues aux stratégies d'édition du génome adoptées. Nous avons évalué la croissance des plants de tomates cerises édités dans le gène *eIF4E1* dans des conditions de culture en serre. Nous avons également évalué la durabilité de la résistance aux potyvirus conférée par l'allèle édité *eIF4E1*. Enfin, nous avons également évalué l'impact potentiel de l'édition du gène de *ACETOLACTATE SYNTHASE* (*ALS*), qui est utilisé comme rapporteur en vue de faciliter la sélection des plantes éditées, ceci dans le but de proposer des consignes pour les futures stratégies de coédition. Dans l'ensemble, ce travail met en évidence le potentiel de l'édition génomique

des facteurs de susceptibilité des plantes permettant de concilier résistance aux virus et croissance de la plante.

Mots clés : résistance aux virus, potyvirus, facteurs d'initiation de la traduction (eIF4E), CRISPR/Cas9, l’édition de génome, l'amélioration génétique pour la résistance

Abstract

Genetic resistances against pathogens are an important leverage for protecting crop production while maintaining low environmental impacts. Thus far, the natural diversity of functional *eukaryotic translation initiation factor* (*eIF*) *4E* alleles has provided genetic resources for breeding potyvirus resistance in many crops, but both the redundancy and the physiological importance of eIF4E have made it difficult to implement similar resistance through mutagenesis in other species. In this thesis, I assessed different approaches to generate resistance to potyvirus in tomato by genome-editing of *eIF4E* genes to provide resistance to potyviruses by loss-of-susceptibility.

In the first part, we targeted the *eIF4E2* for gene knockout by CRISPR/Cas9 and confirmed the limitations of this strategy. I show that this allows indeed to copy potyvirus resistance, but with a very narrow resistance spectrum. In the second part, we used gene knock-in by TALEN and base-editing by cytosine base editor in order to introduce nonsynonymous mutations in the *eIF4E1* to mimic mutations as they usually occur in natural resistance alleles. I show that these precise modifications confer resistance to multiple potyvirus isolates without compromising the protein function.

To assess the practical use of this approach in agriculture, I carried on with the evaluation of potential phenotypic consequences due to the genome-editing strategies taken. We evaluated the *eIF4E1*-edited cherry tomato plants for their growth in greenhouse-culture conditions. We also assessed the durability of the resistance to potyvirus conferred by the edition in *eIF4E1*. Finally, we also evaluated the potential impact of editing the reporter *acetolactate synthase* (*ALS*) gene, which we target in order to facilitate the edited plant selection, with the aim of providing a guideline for future co-edition strategies. Overall, this work highlights the potential of genome-editing of plant susceptibility factors to balance resistance and growth.

Keywords: virus resistance, potyvirus, eukaryotic translation initiation factor (eIF) 4E, CRISPR/Cas9, genome-editing, resistance breeding

Acknowledgement

I wish to thank the French Ministry for Higher Education and Research for financing my PhD contract and the INRAE GAFL lab and ReDD team for funding the materials and experiments.

Firstly, I would like to acknowledge my PhD jury members—Laurence Albar, Nathalie Gonzalez, Fabien Nogué, and Christophe Robaglia for accepting to review this thesis manuscript.

I would also like to thank my thesis committee members—Kristiina Mäkinen, Christian Meyer, Christophe Robaglia again as well as the fellow jury Clément Barre-Villeneuve and Manal Boulaiz, who had shown warm support during and also outside of the thesis committees.

I would like to express my deepest gratitude to my PhD supervisors Jean-Luc and Marianne for your guidance both at professional/scientific and personal level. Thank you for running ahead of me, with me, and sometimes behind me to push my back through the journey of three years of PhD plus the six months of Master 2 internship.

This thesis was not realizable without the huge support from my former and current colleagues from but not limited to LBM—Anaë, Anne, Caro, Céline, Claire, Emmanuel and Marion, Flavie, Geoffroy, Naima, Nasradin, Renaud, Vincent, William; GAFL—Alexandre, Emmanuel (the Chef), Fabien, Marie-Noëlle; and members of A2M, especially Nathalie and Mara, as well as Isabelle (Image3A platform) and Raphael (Université d'Avignon). I would like to thank the admin team and especially Astrid for always reassuring and securing my missions and last minute commande. I would like to thank Robin, my first M1 internship student for the work and stimulus. A big thanks to Anaë again for working with me through the last couple months and being probably the best apprenti I can ever collaborate.

Thank you also to my historical office mates: Geoffroy, Mamoudou, Mohamed, Lucas, and Octave for not complaing the constand smell of coffee, biscuit crunching, sleepy eyes, and the pile of papers and plushies…

My special thanks goes to Manu, who has taken such a great care of me and my tomatoes. Thank you for always being there and encouraging me. You were my go-to place for replenishing the energy and confort.

I am also grateful for my colleagues who have hung out with me outside of the lab quite often, Céline, Naima, and William. My life in Avignon could not have been so fun without you. A special shoutout goes to my best buddy from Master 2 internship time: Sarah, who has made my early period in Avignon and onwards such a wonderful experience.

The number of people that I would like to thank is almost countless. But I would like to end my exhaustive list by thanking all my friends and mentors that I have met outside of the lab as well. This also includes many of the families of my colleagues who have welcomed me to their home and allowed me to spend great time with them. I am also thankful to the LGBP members in Luminy who have been always welcoming. And towards last, but not least, thank you Delyan who is basically all of the above, and have initiated me on the seemingly endless cap-binding exeperiments, discussed with me day and night, and driven me through the PhD and after. Finally, I am grateful for my family who is always supportive of my choices and rooting for me.

Table of contents

Affidavit
Liste de publications et participation aux conférences
Résumé
Abstract
Acknowledgement
Table of contents
General Introduction **1**
1. Introduction to plant breeding ... 1
   a. History of plant breeding ... 1
   b. Four types of plant breeding ... 2
   c. Summary on the breeding techniques ... 5
2. Genome-editing accelerated by the CRISPR/Cas9 engineering ... 6
   a. Brief summary of targeted mutagenesis ... 6
   b. History of CRISPR/Cas discovery ... 7
   c. Structure and mode of action of the original and the engineered CRISPR/Cas9 ... 8
   d. Engineering the versatility of the CRISPR/Cas for genome-editing ... 10
   e. From targeted deletion to designing alleles by CRISPR/Cas ... 12
   f. Beyond the technological improvements: strategies to overcome the technical challenges for boosting the CRISPR editing ... 14
3. Genomics enhances genome-editing for trait improvements with precision ... 17
   a. Role of genomics study in crop breeding ... 17
   b. Types of allelic variants and functional validation ... 19
   c. Enhancing agronomic traits by genome-editing ... 22
4. Plant pathogens and humans ... 25
   a. Impact of plant pathogen ... 26
   b. How do pathogens spread and how can we prevent? ... 26
   c. Emerging risks ... 28
5. Plant pathogen resistance ... 30
   a. First layer of defense: host-pathogen compatibility test for entrance ... 30
   b. Second layer of defense A: major immune response for non-viral pathogen ... 31
   c. Second layer of defense B: immune response for viruses ... 32
   d. Third layer of defense: host-pathogen compatibility test for sustenance ... 36
   e. Plant genetic resistance for breeding ... 37
6. Loss-of-susceptibility case file: eIF4E and potyvirus ... 38
   a. Discovery of eIF4E-based resistance against potyvirus ... 39
   b. Biological roles of plant eIF4E ... 42
   c. Role of eIF4E in virus infection – molecular aspect ... 46
   d. Implications of eIF4E-based potyvirus resistance for plant breeding ... 49
7. Thesis objectives ... 52
   a. How can we generate eIF4E-based resistance using genome-editing? ... 52
   b. Are the base-edited tomatoes free from any phenotypic consequences? ... 54

***Chapter 1*: Manuscript 1** **56**
CRISPR-based knock-out of eIF4E2 in a cherry tomato background successfully recapitulates resistance to pepper veinal mottle virus **57**
Abstract ... 57

Introduction 58
Results 61
Discussion 64
Materials and methods 68
References 71
***Chapter 2*: Manuscript 2 75**
An iterative gene editing strategy broadens eIF4E1 genetic diversity in Solanum lycopersicum and generates resistance to multiple potyvirus isolates 76
Abstract 76
Introduction 77
Results 79
Discussion 86
Material and methods 89
References 93
Preface to *Chapter 3* and *4* 97
Chapter 3: *eIF4E1* precision editing provides a solution to the trade-off between the plant development and potyvirus resistance albeit with a low durability. 98
Introduction 98
Results and discussion 100
Conclusion 109
Materials and methods 110
References 113
Chapter 4: Assessment the impact of co-editing the *Acetolactate synthase* reporter gene on *Solanum lycopersicum* development 116
Introduction 116
Results and Discussion 118
Conclusion 122
Materials and methods 124
References 127
General Discussion 130
Perspectives 136
References 138
Annexes 160

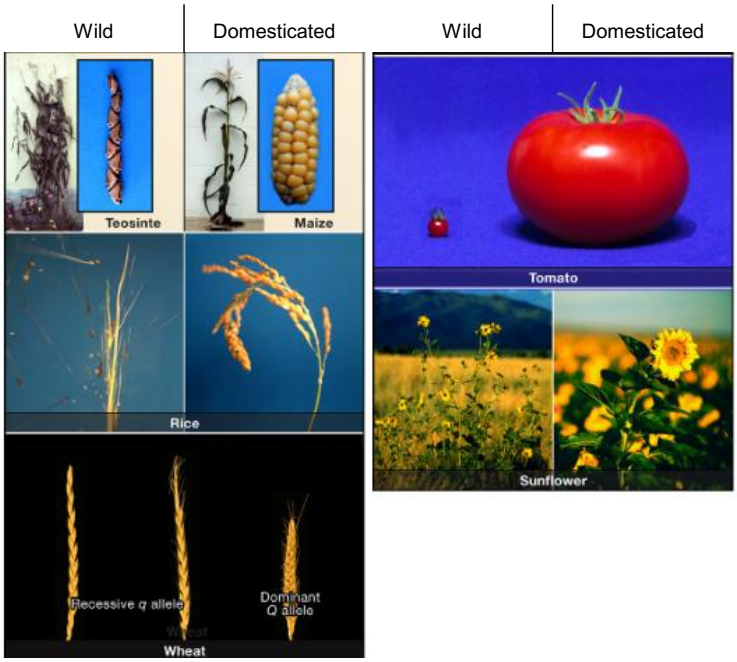


Figure i1-1 Comparison of ancestral, wild crop relatives and modern crop species.

Wild crop relatives (left) and their cultivated crop counterparts (right) are shown. From top left, a teosinte with multiple stalks and long branches and a cultivated maize with its single stalk. Wild rice with a panicle that shatters, and cultivated rice with a solid panicle of grain. A wild wheat with slender, fragile spikes (left) and a cultivated wheat with the causal recessive allele for the same slender, spike trait (middle), and another cultivated wheat with a condensed and tough spike (right). Top right: Wild tomato and the cultivated tomato of increased fruit size. A wild sunflower plant with many small heads borne on multiple slender stalks and a cultivated sunflower plant with a single large head borne on a thick stalk. Figure taken and legend adapted from Doebley, et al., 2006.

Table i1-1 List of major domestication and improvement traits.

Generated based on Meyer and Purugganan, 2013; Traits listed in both columns were continuously selected in both phases.

Domestication	Improvement
• Increased fruit/seed size • Growth habit: determinacy, influorescence architecture • Shorter life cycle • Lack of seed dormancy • Taste e.g. lack of bitterness	• Increased fruit/seed size • Increased fruit/seed number • Loss of seed shattering • Synchronized flowering and fruit ripening • Photoperiod sensitivity • Abiotic stress resistance • Biotic stress/disease resistance • Taste <u>Modern improvement</u> • Abiotic stress resistance • Biotic stress/disease resistance • Increased nutritional value • Added nutritional value

General Introduction

1. Introduction to plant breeding

a. History of plant breeding

Around 10,000-12,000 years ago, our Neolithic ancestors have begun "domestication" of plants, the first step of plant breeding (Doebley et al., 2006; Diamond, 2017; Wallace et al., 2018). Ancestral species of modern crop have been selected—chosen and propagated—for traits that facilitated the cultivation of edible parts of the plants and enhanced the yield (**Figure i1-1**). The domesticated ancestral species then underwent more rounds of selection, today considered as "improvement" steps, which eventually resulted in modern crop species.

i. Human selections in association with favorable phenotypes

The crop domestication and improvement were influenced by two types of human selection: conscious and unconscious selection. Conscious selection is a deliberate selection and preservation of plant individuals with traits of interest by humans; these individuals were reproduced to be used in the cultivation (Zohary, 2004). Such traits that were valued and consciously selected include larger fruit/seed size and number, fruit shape, growth habit (e.g. determinacy), lack of bitter taste and toxicity, and stress resistance (**Figure i1-1, Table i1-1**) (Meyer and Purugganan, 2013). This type of selection has continued during the improvement phases along with human migration and is still in process as in modern breeding program.

Unconscious selection is an unintentional, often automatic, selection of individuals and seeds due to the change in their environment of growth (Zohary, 2004). As the wild species were domesticated, the plants were displaced from their natural habitat into a new growth environment managed by humans and subjected to agricultural practices. This shift induced the selection of the plants that possessed suitable phenotypes for such human agricultural environment and cycle and therefore automatically counter-selected those that were not fit. The unconsciously selected traits are now identified: these are the synchronized flowering and fruit ripening, the photoperiod sensitivity, and the loss of seed shattering, for example (**Table i1-1**) (Meyer and Purugganan, 2013). The

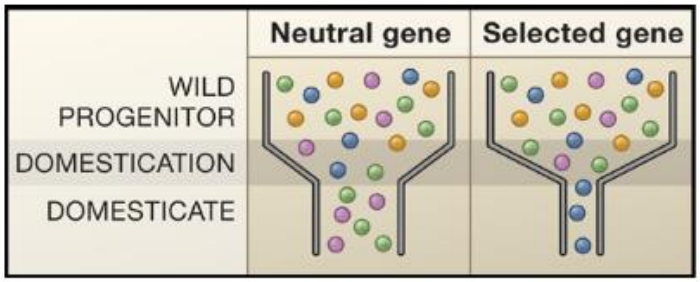


Figure i1-2 Bottleneck effects and selection reduce the genetic diversity of population.

(Left) Random demographic changes such as domestication bottleneck and dispersal bottleneck (see texts) reduce the population size and the genetic diversity of neutral genes as well as beneficial and detrimental genes. (Right) Selection only keeps an interesting variant of a gene (blue) that code for the trait of interest, decreasing the diversity of the gene. Modified from Doebley, et al., 2006

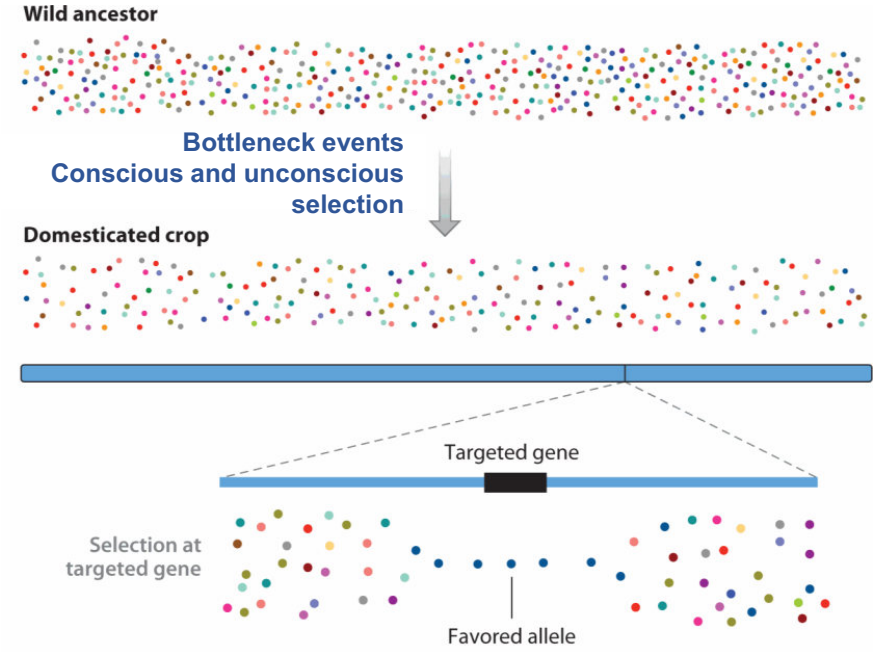


Figure i1-3 Representative image of the shift in diversity of neutral and favored alleles.

Colored dots represent the neutral allelic diversity at different positions of a chromosome (blue rectangle). Bottleneck events and repeated conscious and unconscious selection by humans reduce the allelic diversity in a domesticated crop compared to the wild ancestor at the top. At a targeted gene for conscious selection, the allelic diversity is even lower and often contains only the favored allele, represented by the blue dots. (Modified from Olsen and Wendel, 2013)

unconscious selection has continued, as the humans migrated to a new land with a different climate from where the wild species were first domesticated; only the plants that were adapted to these new climates could survive and be continuously cultivated and used for further breeding (Zohary, 2004). In both types of selection, plants were selected due to particular phenotypes which were attributed to the natural genetic variations, and as a result, these favorable variations were maintained in the population.

ii. Random demographic changes outside selection

In addition, a demographic event such as population bottleneck has played a role in shaping modern crop phenotypes and diversity. Bottleneck effect is defined by the decrease and shift in genetic diversity caused by a decrease in the population size due to different reasons including sheer chances, for instance, the initial random choice of individuals (domestication bottleneck) and the migration of the human populations (dispersal bottleneck) (**Figure i1-2** left) (Doebley et al., 2006; van de Wouw et al., 2010). These events necessarily determined the first set of the wild species to be domesticated and of the cultivated accessions to be improved in a new environment. Consequently, the genetic diversity in the gene pool of a crop species was also determined.

The directed selection of the domestication/improvement traits as well as the bottleneck effects have driven the modification of genetic structures of the crop species compared to their progenitors (van de Wouw et al., 2010) (**Figure i1-2,3**). Modern four plant breeding techniques are an extension of these historical selections with an aim to accumulate the advantageous traits in the crop cultivars in order to meet today's demand.

b. Four types of plant breeding

There are four major techniques of plant breeding that have been developed during the modern crop "improvement" phase: cross breeding, mutation breeding, transgenic breeding, and genome-editing (Chen et al., 2019; Gao, 2021).

i. From conventional breeding to chemical breeding

Cross breeding is a method, where an elite variety of a crop species is crossed with a donor variety so that the desirable trait in the latter is introduced into the elite variety via sexual recombination. It is the most conventional and major method of crop improvement that has selected and accumulated the desirable traits and the associated alleles and quantitative trait loci (QTL) with polymorphisms that have naturally occurred

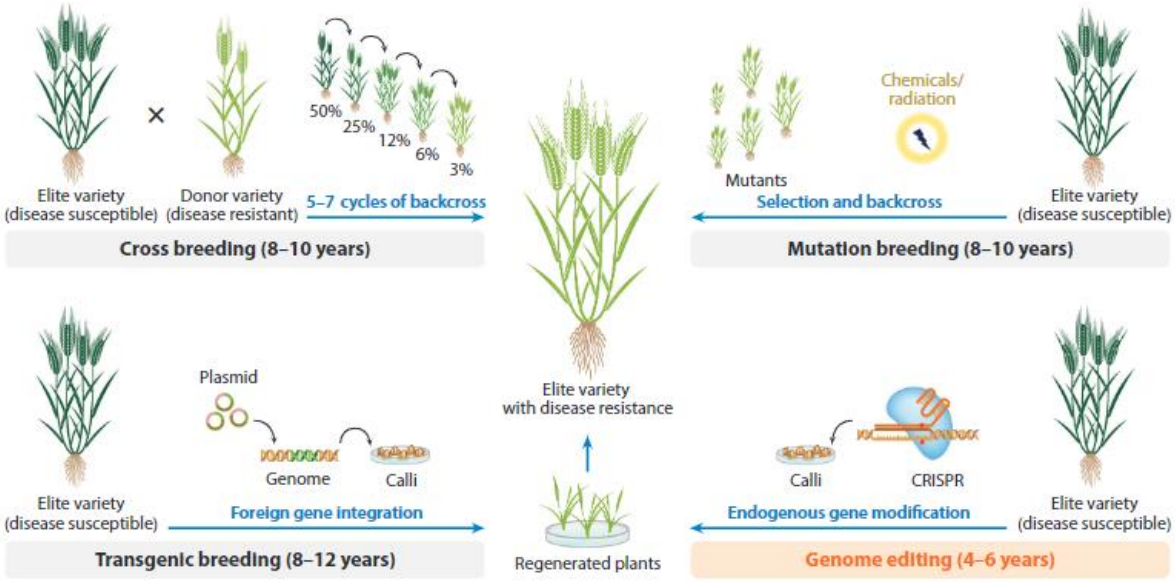


Figure i1-4 Summary of four breeding methods in modern agriculture.

Cross breeding: an elite cultivar is improved through crossing with a donor line and selecting progeny with the desired trait. Mutation breeding: improving a trait using chemical or physical mutagens to treat plant materials (such as seeds) and generate mutants via random mutagenesis. Two methods require several rounds of backcrossing with the recipient elite line to eliminate unexpected linked traits from the donor line or unwanted mutations. Transgenic breeding: improving a trait by transforming the foreign genes into elite varieties. Evaluation for product launch may take many years. Genome editing: improving a trait by directly modifying the genome of elite varieties. Taken from Chen et al., 2019.

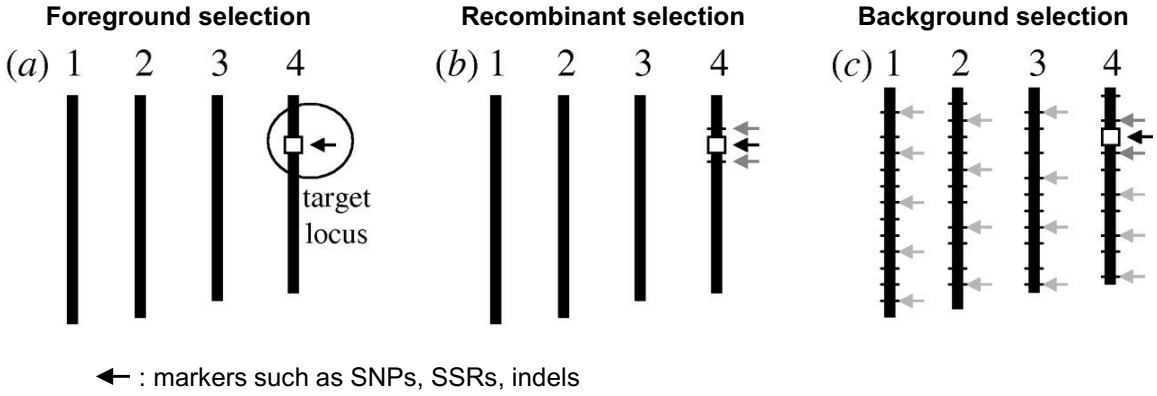


Figure i1-5 Steps of marker-assisted backcrossing (MAB).

The three levels of marker-assisted backcrossing (MAB) based on genetic markers to increase the efficiency of selection during cross breeding. (a) Foreground selection: DNA markers are used to select for the target gene or quantitative trait loci (QTL), facilitating the identification of desirable traits early in the seedling stage and enabling the selection of recessive alleles. (b) Recombinant selection: this step involves selecting backcross progeny that carry the target gene along with recombination events between the target locus and linked flanking markers. This process reduces the size of the donor chromosome segment, minimizing linkage drag by using markers flanking the target gene to decrease the donor fragment size. (c) Background selection: markers unlinked to the target locus are used to select backcross progeny with the highest proportion of the recurrent parent genome. This accelerates the recovery of the RP genome, potentially achieving complete line conversion by BC4, BC3, or even BC2, saving two to four backcross generations. Arrow represents the location of markers. SNP: single nucleotide polymorphism, SSR: simple seqence repeat, indel: insertion and deletion. Figure adapted from Collard and Mckill, 2008

in the genome of a given crop species (**Figure i1-4** top left). The second method is mutation breeding which was first demonstrated in maize in 1928 (Stadler, 1928; For review, Holme et al., 2019). In this method, chemical mutagenesis via ethylmethanosulfonate (EMS) and radiation are used to induce random mutations in the genome. This allowed to generate a wide variety of mutant alleles that did not occur naturally, thereby adding the new alleles and traits to the gene pool (**Figure i1-4** top right).

These two methods require multiple rounds of backcrossing and selection to retain all the desirable traits while removing the harmful ones. This process depends on the natural recombination of the loci in the genome, hence in a stochastic manner. Selection was originally based on the phenotypic observation which was time-consuming and difficult especially for recessive traits. Development of different genomic methods since 1980s has greatly contributed to accelerated selection. Traits of interest were associated with genetic markers such as single nucleotide polymorphism (SNP) and simple sequence repeat (SSR), as well as insertion and deletion (indel), and this association allowed to detect the causal genes and QTLs of the traits linked to the markers (For review, Collard and Mackill, 2008). Marker-assisted selection (MAS) or marker-assisted backcrossing (MAB) based on these genetic markers linked to QTL has been implemented to facilitate the selection step (**Figure i1-5**).

However, even with these methods, the cross breeding and mutation breeding remain labor-intensive and time-consuming, taking 8-10 years until a desirable new variety is obtained (Chen et al., 2019) (**Figure i1-4**). The pace of crop improvement by the above methods is considered too slow to the meet the demand for increase in crop production (Scheben et al., 2017; Hickey et al., 2019). Furthermore, because these methods depend on the genetic recombination during the crossing, segregation of only the beneficial traits can be difficult or even impossible when the causal genes and other undesirable genes are tightly linked due to their close locations on chromosome. It is also limited by the interspecific barriers. The alternative methods that are not restricted by these drawbacks have been sought after and developed based on the genetic studies.

ii. Transgenic breeding

Around the same time as the genomic study has developed, in 1983, first examples of transgenic plant breeding were demonstrated (For review, Raman, 2017). In transgenic breeding, a gene of interest from an organism is transferred into the genome of the elite

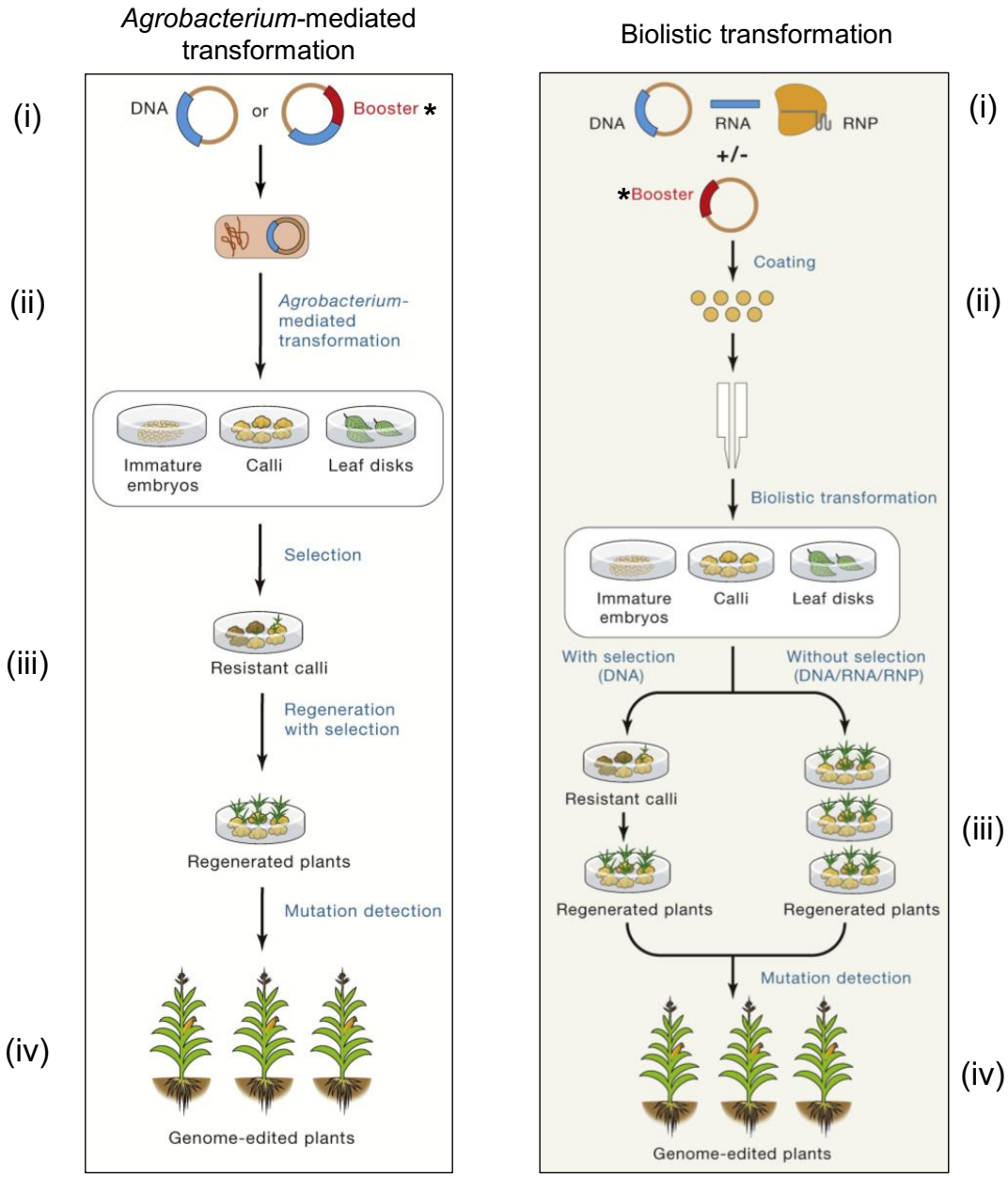


Figure i1-6 Plant Transformation methods: *Agrobacterium*-mediated transformation (left) and biolistic transformation (right) and regeneration.

Methods are applicable to both generating transgenic plants and genome-edited plants. (i) To make a transgenic plant, first, a gene of interest is cloned from a genome of a donor organism. In *Agrobacterium*-mediated transformation, the cloned gene is inserted into the transfer DNA (T-DNA) of an *Agrobacterium* tumor-inducing (Ti) plasmid. In biolistic transformation, the cloned gene is used to generate a template DNA (or RNA transcript). For genome-editing, this gene of interest is replaced with a gene cassette that contains the genome-editing machineries and the target recognition domain. The T-DNA or DNA template usually contains a selectable marker gene. Ti plasmid is used for *Agrobacterium tumefaciens* transformation. (ii) The cells of tissue explants of the elite variety are transformed by *Agrobacterium* carrying the above Ti plasmid or by a biolistic that physically shoots the template in the cells ("transformation"). (iii) The explants containing transformed cells are selected by survival on the selection media that contains antibiotics or herbicide to which the selectable marker gene confers resistance. Surviving explants are transferred to the callus induction media and then to the shoot induction media ("regeneration"). The regenerated shoots are transferred to the root induction media to develop as seedlings (iv). (*Booster mainly applies to the plant developmental regulators used to boost genome-edited plant regeneration; cf. *Introduction 2.f.i*). Figure adapted from Gao, 2021.

background variety (**Figure i1-4** bottom left). It is performed mostly *in vitro*: a gene of interest is cloned from a genome of a donor organism either to insert into the transfer DNA (T-DNA) of an *Agrobacterium* tumor-inducing (Ti) plasmid or to produce a template DNA (or RNA transcript). The T-DNA or template contains a selectable marker gene. Then, the cells of tissue explants of the elite variety are transformed by *Agrobacterium tumefaciens* carrying the above Ti plasmid or by a biolistic that physically shoots the template into the cells ("Transformation") (**Figure i1-6**). After the transformation, the explants containing transformed cells are selected by survival on the selection media that contains antibiotics or herbicide to which the selectable marker gene confers resistance. Surviving explants are transferred to the callus induction media and then to the shoot induction media ("regeneration"). The regenerated shoots are transferred to the root induction media to develop as seedlings.

This method allows more rapid and flexible use of genetic variety to improve specific traits of a given plant and overcomes the interspecific barriers. However, the nature of this technique—random insertion of a foreign DNA into the plant genome—provokes the public concerns for safety and limits its commercial use in many countries (Raman, 2017; Turnbull et al., 2021). This method is also limited to species that can be transformed; and some plant species including pepper (*Capsicum* spp.) are said to be "recalcitrant to transformation". For example, *C. annuum* plants are difficultly transformed by agrobacterium-mediated method due to the low regeneration capacity and the poor overlap between tissues with the transformation competence and the regeneration capacity (Heidmann et al., 2011).

iii. Post-transgenic breeding: genome-editing

The fourth method: plant breeding by genome-editing has appeared in 1990s when it was discovered that the site-specific nucleases (SSNs) are engineerable (For review, Iqbal et al., 2020). Genome-editing technologies developed based on SSNs include homing endonucleases (HEs), zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and the recently developed, type II clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated protein (Cas) (CRISPR/Cas) system (Hua et al., 2019) (**Table i1-2**). Unlike the other methods, the SSNs induce mutations at target sites in the genome of a given plant ("targeted mutagenesis") (**Figure i1-4** bottom right). These mutations can be either random or specified by a template.

Table i1-2 Comparison of different site-specific nucleases used in genome-editing.

	Homing nucleases	ZFN	TALEN	CRISPR/Cas
Core components	Two protein domains: DNA binding, DNA cleavage	Zinc finger protein domain, FokI nuclease	TALE-DNA binding protein domain, FokI nuclease	sgRNA, Cas protein
Sequence availability	Limited by existing enzymes	Limited by triplet recognition	Limited by first nucleotides	Virtually unlimited
Target sequence length	14-40 bp	24-36 bp	24-59 bp	20-22 bp
Target recognition efficiency	Low to moderate	Moderate	Moderate	High
Multiplexing	Not possible	Difficult	Difficult	Easy

Generated based on Belfort et al., 2014 and Iqbal et al., 2020.

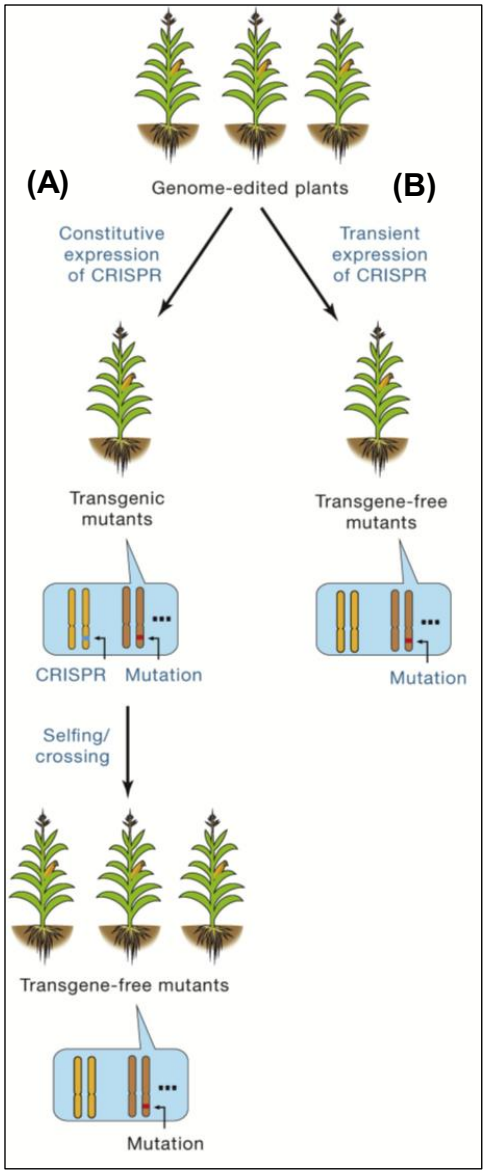


Figure i1-7 Generation of transgene-free genome-edited plants after stable and transient expression.
A Plants that are stable transformed with genome-editing machinery (here CRISPR as an example) contain mutations as well as the transgene. These plants can be selfed or crossed with the parental genotype in order to remove the transgene in the next generation. **B** Plants that are transiently transformed with genome-editing machinery will not retain the transgene in the plant genome; the plants may be directly transgene-free. Figure adapted from Gao, 2021.

The delivery of the DNA coding for the genome-editing machineries and the template is performed by *Agrobacterium*-mediated transformation or biolistic transformation like the transgenic breeding (**Figure i1-6**). Once the gene is edited as desired, these foreign genetic materials can be removed by a simple crossing for sexually propagated species (**Figure i1-7A**). Additional delivery methods that are DNA-free, i.e. templates in the form of RNA, protein, or ribonucleoprotein (RNP), are also available for the genome-editing techniques using biolistic and expands the usage of the SSNs in the asexually/vegetatively propagated species for which outcrossing the foreign DNA is laborious (Ran et al., 2017) (**Figure i1-6**).

Thus, this fourth method enables the targeted and direct manipulation of the genetic materials in the elite cultivars, devoid of the linkage drag, interspecific barriers, and random insertion of the foreign gene in the genome in many species. Although there are legislative and technological issues similar to the transgenic approach, this method holds promising potential to accelerate the plant breeding programs. This thesis is based on the genome-editing via two SSN technologies: TALEN and CRISPR/Cas system, with a particular focus on the CRISPR/Cas. The targeted mutagenesis and the development of the CRISPR/Cas systems will be reviewed in later sections.

c. Summary on the breeding techniques

Above four major modes of plant breeding have been developed owing to the advances in molecular and cellular biology. Among the four, the cross breeding and the mutation breeding remain the major methods for generation of new varieties. They will continue to be supported by the genetic and genomic approaches that identify the new alleles and QTLs of interesting traits.

However, genomic studies also point out that the effort to accumulate desirable traits in a few cultivars has led to the decrease in genetic diversity and the fixation of a large proportion of the crop genome. Genetic diversity of the ancestral and intermediate crop species is the major source of the plant breeding, and the loss of diversity would mean the potential loss of traits that are yet to be discovered and beneficial (van de Wouw et al., 2010). Also, certain alleles are introduced in multiple cultivars of a crop species, for instance, to confer pathogen resistance, and this wide application of few specific genes may risk an eventual, massive resistance breakdown in the crop. Another concern is that the genome of major crop species has been loaded with a large number of major-effect

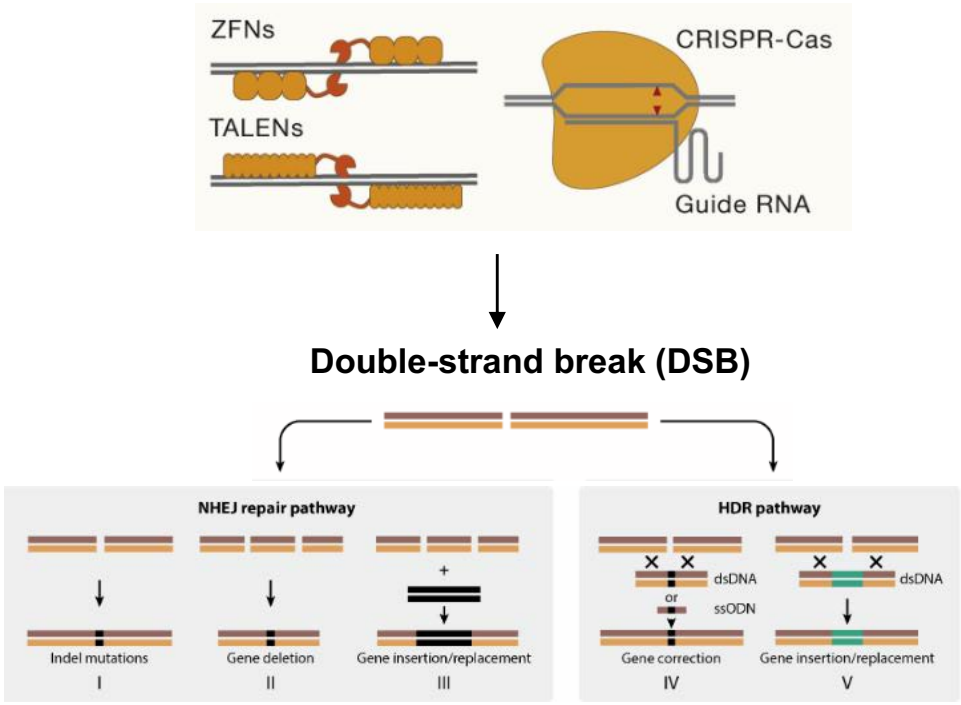


Figure i2-1 Single-site nucleases (SSNs) generate double-strand break (DSB) and triggers different repair pathways.

SSNs: ZFNs, TALENs, and CRISPR/Cas searches the target sequence and generates a DSB. The DSB is repaired via non-homologous end-joining (NHEJ) pathway or the homology-directed repair (HDR) pathway in a presence of a template. NHEJ repair can result in three major modification events. HDR occurs when sufficient amount of a template double-stranded DNA (dsDNA) or single-stranded oligodeoxynucleotide (ssODN) is available. HDR repair can have two outcomes depending on the template. Figures are adapted from Chen et al., 2019 and Gao, 2021.

genes, QTLs, and their linked genes which can be beneficial, harmless, or neutral. For example, the modern tomato cultivar contains nearly 25,6% of the assembled genome occupied by the genes/traits that have been selected and the linked loci in the course of domestication and improvement (Lin et al., 2014). An increasing fixation of crop genome as seen in tomato (*Solanum lycopersicum*) would limit the future improvement via crossings.

The alarmingly small diversity and room are left for improvement of some crop species by conventional breeding, calling for alternative methods such as targeted mutagenesis by genome-editing techniques that can directly generate new alleles in the genome of the elite variety. Among the different genome-editing tools, the recent, CRISPR/Cas system demonstrates a great versatility which enable the designing of new alleles and traits. In the next section, the overview on the targeted mutagenesis, with a special focus on the CRISPR/Cas systems and the state-of-art improvements of CRISPR-based editing will be provided.

2. Genome-editing accelerated by the CRISPR/Cas9 engineering

a. Brief summary of targeted mutagenesis

The new breeding technique—genome-editing—triggers targeted mutagenesis in the genome of a crop of interest. In this technique, the plant cells are generally transformed with the SSN machinery and the optional template in the same way as the transgenic plant generation. Once expressed in the nucleus, the SSN recognizes the target site in the genomic DNA and creates a double-strand break (DSB). This is followed by an innate DNA repair mechanism either via the non-homologous end-joining (NHEJ) pathway or the homology-directed repair (HDR) pathway if a template is provided, and the edition occurs during this phase (**Figure i2-1**) (For review, Xue and Greene, 2021). NHEJ results in random insertions or deletions of one or more nucleotides in the DSB site, often leading to frameshift mutations that result in loss of function of the target gene. HDR occurs when a suitable DNA template, which harbors homologous sequences to the flanking regions of DSB site, is available; this results in incorporation of custom mutations based on the template at the DSB site ("gene-targeting"). Gene-targeting can generate

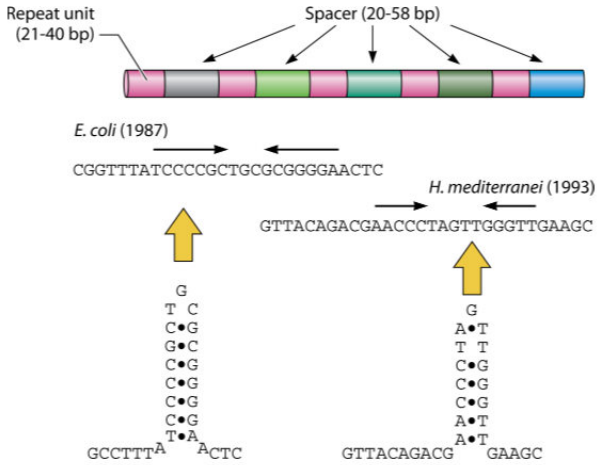


Figure i2-2 The structure of CRISPR sequence found in bacteria and archea.

The repeat units (pink) and spacers (other, non-pink colors) are alternated. The palindromic repeat unit forms a secondary structure. The example sequences from *E. coli* and *H. mediterranei* are shown. Taken from Ishino et al., 2018.

specific indels of several nucleotides as well as insertion of an entire gene, according to the template. In plants, however, the NHEJ occurs preferentially to the HDR, making it difficult to implement the desired modifications via gene targeting (Puchta, 2005).

Among the first SSN tools developed, the ZFN and the TALEN have played the important role in the advance of plant breeding in early 2000s based on targeted mutagenesis. Both consist of two protein domains fused artificially: a specific DNA-binding domain and the non-specific cleavage domain of FokI endonuclease (**Table i1-2**) (For review, Iqbal et al., 2020). In the ZFN and the TALEN, the DNA-binding domain is respectively an array of zinc fingers and of TALEs, specifically designed for the target DNA sequence. Guided by the DNA-binding domain, the FokI domain can generate DSB in the genome that is followed by the repair mechanism.

However, a game changer, the CRISPR/Cas system was developed in 2012 (Jinek et al., 2012) and has revolutionized the genome-editing technologies since the first demonstrations in 2013 in mammalian cells and then plant cells by engineered CRISPR/Cas (Cong et al., 2013; Mali et al., 2013; Shan et al., 2013). The engineered CRISPR/Cas system consists of a single guide RNA (sgRNA) that binds specifically to the target DNA sequence and a Cas protein, the endonuclease (**Figure i2-1**). The decisive difference between these technologies lie in the modes of target site recognition: the ZFN and TALENs recognize the target DNA sequence via engineered domain of protein, whereas the CRISPR/Cas system does via complementary sequence of sgRNA (Gao, 2021). This has democratized the genome-editing by reducing greatly the cost and complexity of designing a specific gene-editing machinery which were the difficulties encountered for using ZFNs and TALENs (Chen et al., 2019).

Since its discovery, the CRISPR/Cas system has been rapidly developed and engineered for various usages. This thesis work is founded upon two of such CRISPR/Cas-based genome-editing technologies. Thus, following sections will be dedicated to the discovery and development of the engineered CRISPR/Cas system.

b. History of CRISPR/Cas discovery

The CRISPR/Cas system was originally discovered as bacterial and archeal immunity systems. The following part will be written based on a historical review (Ishino et al., 2018), written by one of the first discoverers of the CRISPR. The identification of CRISPR/Cas system started by the discovery of the CRISPR, the then mysterious, unusual

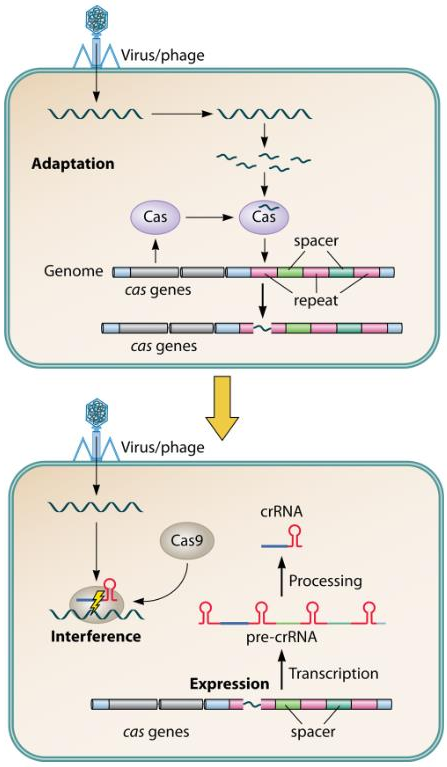


Figure i2-3 The CRISPR/Cas-mediated bacterial immune system.

Cas effectors are involved in different steps: adaptation (purple Cas), expression (not depicted, involved in pre-crRNA processing), and interference (gold Cas9). During "adaptation", Cas recognizes the foreign DNA from from bacteriophage, cleaves it, and incorporates it into the spacer region of CRISPR sequence. During the next attack by bacteriophage of same species, the CRISPR sequence is transcribed (pre-crRNA) and processed into crRNA by another Cas ("Expression"). The crRNA is then bound by Cas endonuclease (e.g. Cas9) and this complex can cleave the homologous, foreign DNA ("Interference"). Taken from Ishino et al., 2018.

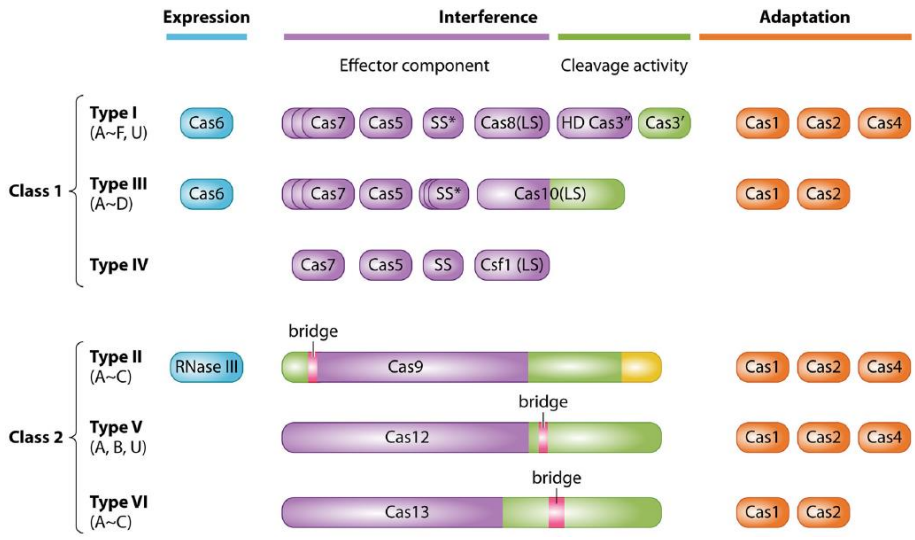


Figure i2-4 CRISPR/Cas classification based on the sequence analyses and gene organization.

The functions of Cas proteins in bacterial immune system (as described in Figure i2-2) are shown above and color-coded. Class 1 and 2 contains multiple effectors and single effector, respecitvely. Types are defined by the presence of signature Cas proteins. Taken from Ishino et al., 2018.

repeated sequences in bacteria *Escherichia coli* in 1987 (Ishino et al., 1987). In 1993, a similar genomic structure was reported in archaea *Haloferax mediterranei* (Mojica et al., 1993). This sequence structure comprised of two types of nucleotide sequences: spacer sequences (20-58 bp) and repeat unit (21-40 bp), which is palindromic and thus forms secondary DNA structure, and the spacer sequences are located between multiple repeat units (**Figure i2-2**). The sequence was hence named clustered regularly interspaced short palindromic repeat (CRISPR) sequence in 2002 (Jansen et al., 2002).

The function of the CRISPR sequence was unknown until 2005 when the spacer sequences were found to be similar to sequences in bacteriophage, archaeal viruses, and plasmids, suggesting their role in the bacterial and archaeal immunity system (Mojica et al., 2005; Pourcel et al., 2005; Bolotin et al., 2005). Around the same time, genomic analyses revealed protein coding sequences that are strictly CRISPR-associated (hence named Cas). Then, these Cas proteins in association with CRISPR were shown to be the prokaryotic acquired immune system (Makarova et al., 2006; Barrangou et al., 2007). Subsequently, Cas proteins with functions at different stages of the immune response were identified (**Figure i2-3**).

The CRISPR/Cas immune systems are variant between the strains and species, while some are conserved across bacterial species. They are divided into two classes based on the number of *Cas*-encoded effector proteins: the class 1 has multiple proteins that form an effector complex, whereas the class 2 has a single protein effector. Each class is divided into three types: class I into type I, III, and IV, and class II into type II, V, VI, which are further divided into subtypes (**Figure i2-4**). Based on their simple effector structure, the class 2 systems, namely the Cas9 (type II), the Cas12 (type V), and the Cas13 (type VI) genes, have been intensively studied for the potential engineering into genome-editing tools.

c. Structure and mode of action of the original and the engineered CRISPR/Cas9

Among class 2 immune systems, the type II, Cas9 effector is the most extensively studied. The Cas9 effector requires a CRISPR RNA (crRNA) and a *trans*-activating CRISPR RNA (tracrRNA) for target recognition and DNA cleavage (Jinek et al., 2012). The crRNA forms a duplex with the target site on the complementary strand of DNA, thereby serving as a guide, whereas the tracrRNA activates the Cas9 to cleave and process the pre-crRNA transcribed by CRISPR sequence. Importantly, the Cas9 as well as other Cas proteins

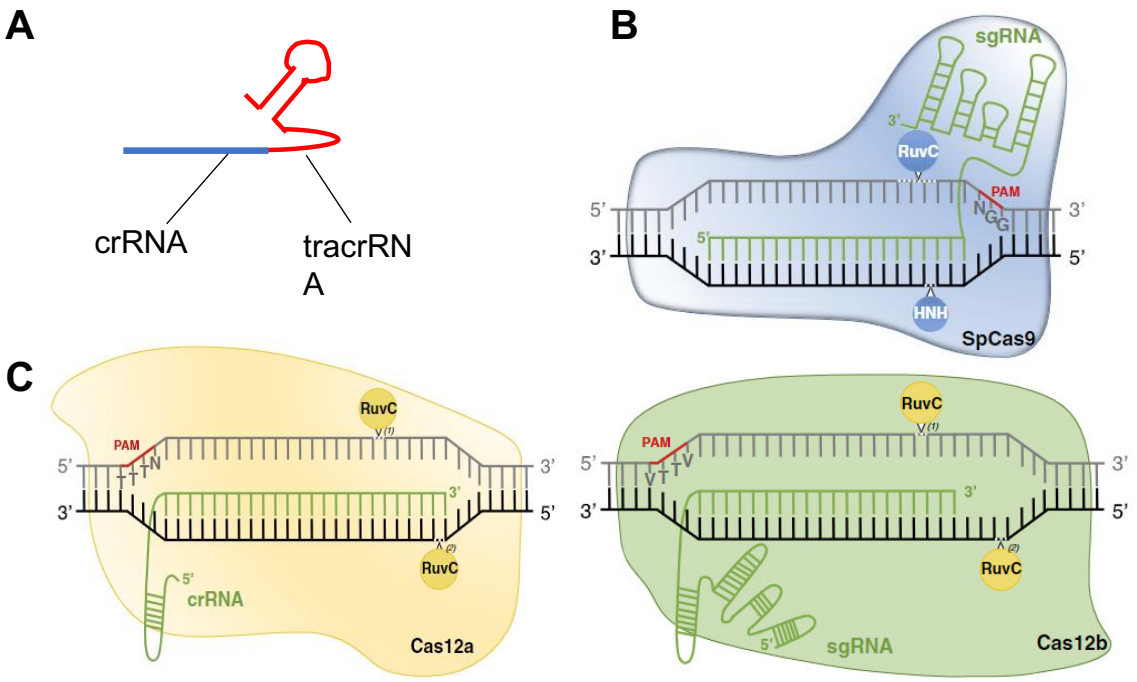


Figure i2-5 The engineered sgRNA and CRISPR/Cas structure schematics.

A sgRNA is a single RNA scaffold, made from fusion between the crRNA and the tracrRNA. **B** The engineered CRISPR/SpCas9 is composed of the sgRNA (green) and the SpCas9 effector protein (light blue). The approximative cleavage sites by the distinct domains RuvC and HNH are shown by the white arrows, at 3 bp upstream of PAM (5'-NGG-3'). **C** The engineered CRISPR/Cas12a is composed of crRNA (green) and Cas12a effector protein (light yellow) and the CRISPR/Cas12b is composed of sgRNA (green) and Cas12b effector protein (light green). The approximative cleavage sites by RuvC domains downstream of each PAM site are shown by white arrow. Taken from Veillet et al., 2020.

requires that the genomic sequence have a specific sequence, called a protospacer adjacent motif (PAM), at a suitable location in relation to the target sequence in order to cleave the DNA. For cleavage function, the Cas9 itself contains two nuclease domains, RuvC and HNH, that cuts the non-target and target strands of DNA, respectively, when activated (Jinek et al., 2012).

i. Engineering the CRISPR/Cas9 system into a genome-editing tool

At the same time as they revealed the detailed mode of function of CRISPR/Cas9, Jinek et al. (2012) demonstrated that a chimeric single RNA made from fusion of crRNA and tracrRNA, and a *Streptococcus pyogenes* Cas9 (SpCas9) can be engineered as a programmable genome-editing tool *in vitro*. This was soon followed by further engineering and applications in mammalian cells and then plant cells (Cong et al., 2013; Mali et al., 2013; Shan et al., 2013). Today, the widely used engineered CRISPR/Cas9 system consists of two parts: the sgRNA derived from the original crRNA-tracrRNA chimera and the SpCas9 (**Figure i2-5A, B**). The sgRNA can be designed to target any genomic region of interest as long as the target DNA sequence has the PAM sequence 5′-NGG-3′ (N: any of the four nucleotides) directly downstream of the target sequence. Interestingly, the PAM requirements and position in the genome in respect to the target sequence differ between the Cas effectors; for example, other Cas proteins such as Cas12a and Cas12b recognize 5′-TTTN-3′ and 5′-VTTV-3′, respectively (V: G, C, or A), opening a possibility to edit genomic sequences that are not accessible by the Cas9 (**Figure i2-5C, D**).

The mechanism of genome-editing by the CRISPR/Cas9 is as follows: briefly, the sgRNA binds the Cas9 to activate and guides the Cas9-sgRNA complex to the target DNA sequence; the activated Cas9 probes and recognizes the PAM sequence in the proximal genomic region; and finally, the target region is cleaved by RuvC and HNH domains to make the DSB at about 3 bp upstream of the PAM. The DSB is followed by DNA repair mechanisms via the NHEJ or the HDR pathway, introducing one or more nucleotide indels and insertions or deletions of a larger fragment or an entire gene (**Figure i2-1**) (Xue and Greene, 2021). Mutations can be found at homozygous, heterozygous, and biallelic state. In plants, the CRISPR/Cas9 has been widely and efficiently used to generate a knock-out (KO) mutants of a gene of interest due to the preference of the NHEJ to the HDR (Puchta, 2005).

Table i2-1 Examples of engineered cas9 variants

Name	PAM recognition (5′-3′)	Methods	Reference
SpCas9-VQR	NGAN	bacterial selection/structure-guided	Kleinstiver et al., 2015
SpCas9-VRER	NGCG	bacterial selection/structure-guided	Kleinstiver et al., 2015
SpCas9-NG	NG	Structure-guided	Nishimasu et al., 2018
xCas9	NG, GAA, and GAT	phage-assisted continuous evolution	Hu et al., 2018
SpG	NGN	Structure-guided	Walton et al., 2020
SpRY	NRN and NYN	Structure-guided	Walton et al.,2020

R: purine, Y: pyrimidine

d. Engineering the versatility of the CRISPR/Cas for genome-editing

The engineered CRISPR/Cas9 is a two-component system made of the sgRNA and the Cas9 that have independent and uncoupled functions. This makes the system amenable to improvements and customization of different parts and opens numerous ways of exploiting and further expanding the potential of the system into a versatile tool. Here, some of major examples of expansion of CRISPR/Cas9 usage will be provided.

i. Broadening the Cas9 target

One of the limitations of the classic CRISPR/Cas9 system is the SpCas9-specific PAM sequence NGG that restricts its realm of target to the GC-rich genomic region where PAM is available at proximal site to the target locus. One way to overcome this limitation is to use the Cas12a (PAM: TTTN) or Cas12b (PAM: VTTV) that have different PAM requirements. Another way is to broaden the target of the CRISPR/Cas9 by using natural variants or engineered variants that recognize non-NGG PAM (For review, Cebrian-Serrano and Davies, 2017).

ii. Natural Cas9 variants

Among the natural Cas9 effector orthologs mined from bacterial and archaeal species, for example, *Staphylococcus aureus*-derived Cas9 (SaCas9) with NNGRRT (R: A or G) PAM was detected and applied for genome-editing in mammalian systems and some in plant systems (Ran et al., 2015; Qin et al., 2019). Other examples include but not limited to *Streptococcus thermophilus*-derived Cas9 (PAM 5′-3′: NNAGAAW or NGGNG; W: A or T) and *Neisseria meningitidis*-derived Cas9 (PAM 5′-3′: NNNNGATT) (Cebrian-Serrano and Davies, 2017).

iii. Engineered Cas9 variants

Cas9 variants have also been artificially generated using experimental evolution-based methods and structure-guided methods (**Table i2-1**). For instance, a bacterial selection-based directed evolution[2] has generated several variants of SpCas9 with nonsynonymous mutations that can efficiently edit the genome in zebrafish and human

[2]The system selects bacteria based on the survival that is enabled by Cas9-mediated cleavage of a selection plasmid encoding an inducible toxic gene with a PAM of interest. At the beginning, bacteria contain a Cas9 coding sequence with mutagenized PAM-interacting site (based on structural information). Only the bacteria that contained Cas9 that recognized the PAM of interest and cleaved the toxic gene can survive, thereby selecting the PAM-recognition variants.

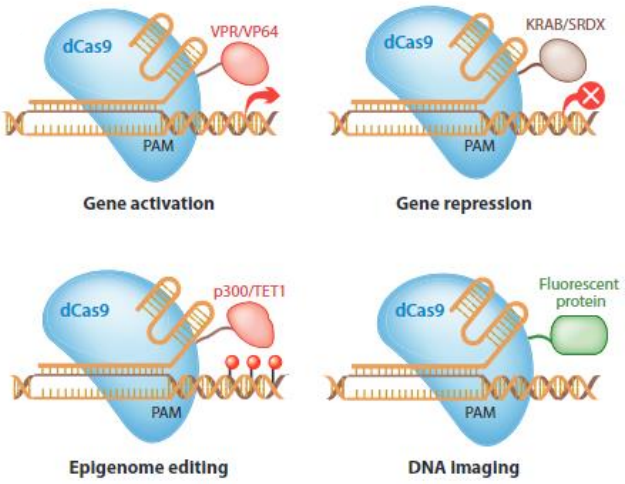


Figure i2-6 The examples of dCas9-based gene manipulation systems. The catalytically inactivated dCas9 is fused with proteins including transcriptional activators (e.g., VP64, p65AD, and VPR) or transcriptional or repressors (e.g., the KRAB domain and SRDX domain), epigenetic effectors (e.g. histone acetyltransferase p300, and ten-eleven translocation 1 (TET1)), and fluorescent proteins such as GFP. Taken from Chen et al., 2019.

cells (Kleinstiver et al., 2015). Cas9 variants with mutations D1135V/R1335Q/T1337R (VQR) and D1135V/G1218R/R1335E/T1337R (VRER) that recognized NGAN and NGCG, respectively, were obtained. These variants described in Kleinstiver et al. (2015) were successfully applied to genome-editing in rice with same PAM requirements, increasing the number of sites accessible for CRISPR/SpCas9 by approximately two-fold in the rice genome compared with what was possible by the wild-type SpCas9 (Hu et al., 2016). Phage-assisted continuous or non-continuous evolution (PACE or PANCE)[3] has also been utilized to generate SpCas9 variants for animal systems that can recognize multiple PAMs with low off-target activity (Hu et al., 2018) and non-G PAMs (Miller et al., 2020).

Engineering guided by the structural information of the Cas9 has also given a rise to SpCas9 variant that recognizes NG as a PAM (SpCas9-NG) (Nishimasu et al., 2018). Development of the SpG (PAM 5′-3′: NGN) and SpRY (NRN and NYN) (Y: T or C) has basically achieved PAM-less genome-editing by SpCas9 though it is accompanied by a slightly increased off-target activity in the absence of fidelity-enhancing substitutions (Walton et al., 2020). The application of SpRY variant was then demonstrated in rice and the Dahurian larch (a coniferous tree, gymnosperm) (Ren et al., 2021). Together, natural and engineered Cas9 variants broaden the possibility of genome-editing in nearly any genomic regions.

iv. Customizing the Cas9

The function of a basic SpCas9 is to cleave both target and non-target strands of the DNA with distinct RuvC and HNH domains, which necessarily makes a DSB. Inactivation of these domains has been attempted to render the CRISPR/Cas system more versatile. As a result, inactivation of one of the domains generated a DNA nickase (nCas9) that cleaves only one strand of the DNA at target site, while that of two domains generated the dead Cas9 (dCas9), which can function as a scaffold for other effector proteins at the DNA target site. For instance, the fusion of dCas9 to the transcriptional activator or repressor, epigenetic effector, and fluorescent protein have been achieved to control gene expression, to change the epigenetic state, and to label a specific gene, respectively (Reviewed in Chen et al., 2019) (**Figure i2-6**). The nCas9 is extensively studied and used

[3]A method for the rapid, directed evolution of proteins based on the activity of the evolved protein coded by bacteriophage that determines the fitness of bacteriophage. PANCE is an analogous method that allows propagation of phages coding for modestly beneficial mutants of the protein.

in several advanced genome-editing technologies such as base-editor and prime-editor, and these will be explained later in the section.

v. Multiplex genome-editing possible with sgRNAs

The second component of the CRISPR/Cas9 system, sgRNA, can also bolster the genome-editing. High efficiency of CRISPR/Cas9 system allows simultaneous mutations of multiple loci in the genome. This is possible by designing a common sgRNA for targeting multiple homologs in the genome, and it is highly useful for simultaneous editing of homologous genes in the polyploid species such as wheat (Reviewed in Gao, 2021). It is also possible by designing multiple sgRNAs that target different genes of interest due to the facility of sgRNA designing compared to the specific DNA-binding domain of the other SSNs like ZFN or TALEN (cf. *Introduction 2.a*). Multiplex genome-editing accelerates the trait stacking in the elite cultivars (For review, Manghwar et al., 2019). (The word "multiplex editing" in literatures refers more often to the second concept). In plants, simultaneous editing of as many as eight genes in *Nicotiana benthamiana* and of twelve genes in *Arabidopsis thaliana* have been reported using distinct sgRNAs to target each gene and an intron-optimized Cas9, zCas9i (Stuttmann et al., 2021).

Indeed, high editing efficiency of the CRISPR/Cas9 has raised concern for the potential off-target effect in the plant genome. To minimize such side-effects, A number of *in silico* off-target prediction tools have been developed to assist optimal design of sgRNA, which is key to preventing off-target effect (Chen et al., 2023).

e. From targeted deletion to designing alleles by CRISPR/Cas

In plants, the CRISPR/Cas9 system has been very efficient at generating loss-of-function alleles of targeted genes via NHEJ, but less successful at generating loss- or gain-of-function alleles that required specific indels/substitution/gene provided by the DNA template inserted via HDR (Puchta, 2005; For review, Scheben et al., 2017). To achieve such precise editing beyond gene KO, several technologies base on the nCas9—a Cas9 modified to retain only one active cleavage HNH domain—has been developed.

i. Base editors

One of such precision editing technologies is the base editors that can convert one or more bases into other base within the target region directly, without a donor template.

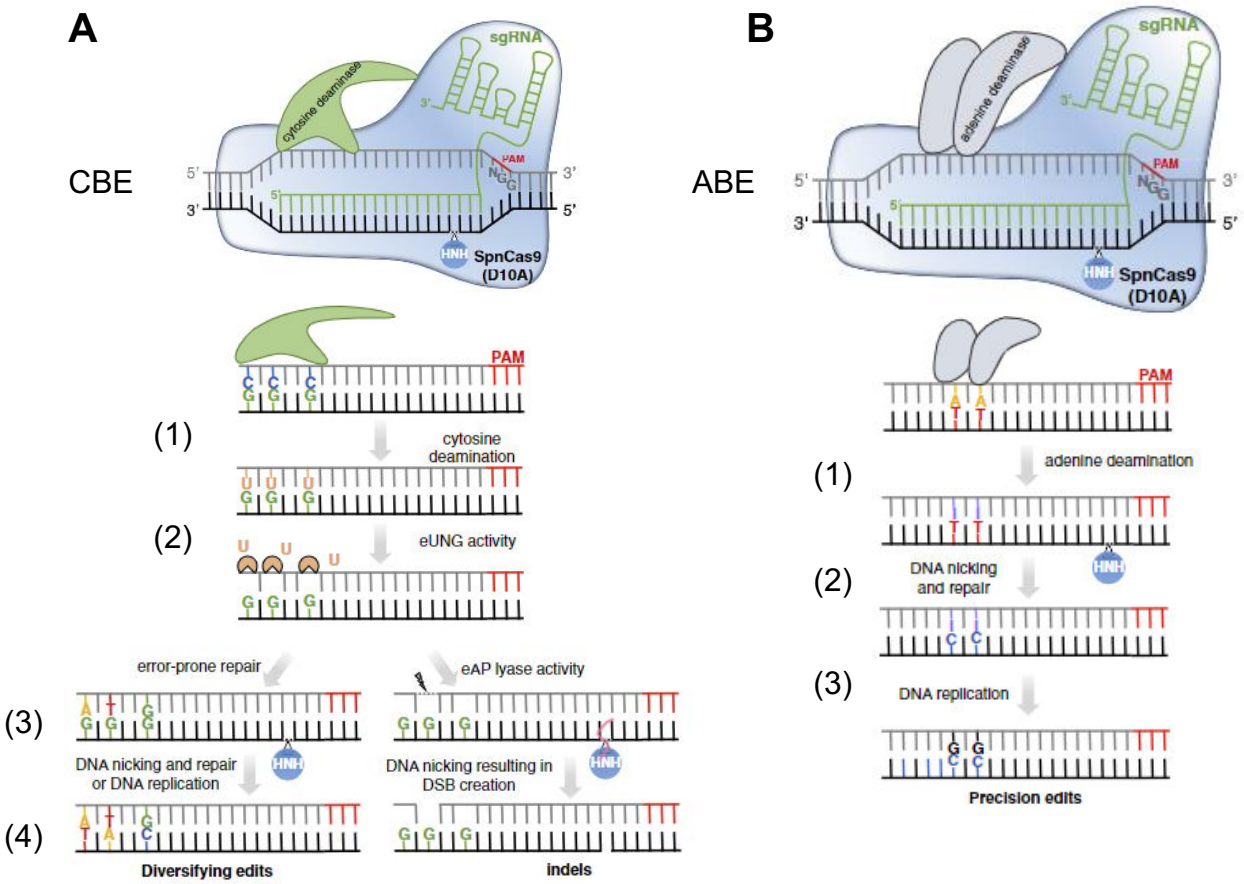


Figure i2-7 Base-editor and nucleotide substitution schematics. A Cytosine base editor (CBE) is composed of sgRNA, SpnCas9, and cytosine deaminase. (1) Cytosines in the editing window targeted by sgRNA and PAM are deaminated into uracils. (2) Endogenous uracil DNA glycosylase (eUNG) detects and removes the U, leading to an abasic site. (3-4) As the HNH domain cleaves the target strand, such site is repaired by error-free (U back to C), error-prone (U to non-C), or indels due to the endogenous AP (eAP) lyase activity which introduces a cleavage in non-target strand. **B** Adenine base editor (ABE) is composed of sgRNA, SpnCas9, and adenine deaminase. (1) Adenines in the editing window targeted by sgRNA and PAM are deaminated into inosine intermediates. (2-3) As the HNH domain cleaves the target strand, the edited sites are repaired and replicated to produce precise editing of adenine to guanine. Taken from Veillet et al., 2020.

Table i2-2 List of genes targeted for editing by base editors CBE and ABE.
Taken from Mishra et al., 2020.

Crop name	Targeted genes	Type of base editor used	Functions	References
Oryza sativa	*NRT1.1B and SLR1*	CBE	Enhance nitrogen use efficiency	Lu and Zhu (2017)
	C287	CBE	Herbicide resistant	Shimatani *et al.* (2017)
	OsPDS, OsSBEIIb	CBE	Nutritional improvement	Li *et al.* (2017)
	OsCDC48	CBE	Regulate senescence and death	Zong *et al.* (2017)
	OsSPL14	CBE	Herbicide resistance	Tian *et al.* (2018)
	OsMPK6	ABE	Pathogen-responsive gene	Yan *et al.* (2018)
	OsACC-T1	ABE	Herbicide resistance	Li *et al.* (2018)
	SLR1	ABE	Della protein for plant height	Hua *et al.* (2018)
	OsSPL14	ABE	Plant architecture and grain yield	Hua *et al.* (2018)
	OsRLCK185, OsCERK1	CBE	Defence response	Ren *et al.* (2018)
	Pi-d2	CBE	Blast resistance	Ren *et al.* (2018)
	Wx	ABE	Rice amylose synthesis	Hao *et al.* (2019)
	GL2/OsGRF4, OsGRF3	ABE	Grain size and yield	Hao *et al.* (2019)
	ALS	CBE	Herbicide resistance	Veillet *et al.* (2019)
Triticum aestivum	*TaLOX2*	CBE	Lipid metabolism	Zong *et al.* (2017)
	TaDEP1, TaGW2	ABE	Panicle length and grain weight	Li *et al.* (2018)
Zea mays	*ZmCENH3*	CBE	Chromosomal segregation	Zong *et al.* (2017)
Solanum tuberosum	*StALS, StGBSS*	CBE	Herbicide resistance, Starch synthesis	Zong *et al.* (2018)
Solanum lycopersicum and *S. tuberosum*	*SLALS1*	CBE	Herbicide resistance	Veillet *et al.* (2019)
Citrullus lanatus	*ALS*	CBE	Herbicide resistance	Tian *et al.* (2018)

These are the cytosine base editors (CBE) and the adenine base editor (ABE) that can induce substitution of cytosine and adenine, respectively, into other nucleotides (Komor et al., 2016; Gaudelli et al., 2017). This can achieve targeted modification of a given gene without knocking out the gene.

The CBE is composed of an sgRNA and an nCas9 protein fused to a cytosine deaminase (**Figure i2-7**). In ABE, the cytosine deaminase is replaced by the adenine deaminase. Like the classic CRISPR/Cas9 system, the base editor is guided by the sgRNA to the complementary target DNA adjacent to the PAM. In the case of CBE, the cytosine deaminase catalyzes the deamination of one or more cytosines to uracil within the small editing window at the 5′ end of non-target strand, while nCas9 cleaves the target strand with HNH domain. The base excision repair (BER) pathway is then triggered by detection of the uracils to induce error-free (U to C) or error-prone (U to T or A or G) repair, resulting in diverse nucleotide substitutions, or indels (Komor et al., 2016; For review, Hess et al., 2017; Veillet et al., 2020b) (**Figure i2-7A**). In the case of ABE, the adenine deaminase catalyzes the deamination of one or more adenines to inosine intermediate, which does not trigger BER, and mediates efficient A-to-G conversion (Gaudelli et al., 2017; Fo review, Veillet et al., 2020b) (**Figure i2-7B**).

Both the CBE and ABE have been used to modify various phenotypes of plants—herbicide resistance, pathogen resistance, yield, plant architecture—that are conferred by induced polymorphisms in specific genes and not gene KO (**Table i2-2**). They can also generate the targeted genetic variations in a genomic region and therefore may help identify key amino acid(s) or domain in a protein of interest contributing to the phenotypic variations. Several downsides such as the wide size of editing window and the stochastic outcomes of modifications achieved by CBE have been pointed out (For review, Veillet et al., 2020b). For this, high-precision CBE have been developed, which show to prevent the unwanted additional mutations in the target region (Tan et al., 2019). The base editors are also limited by the choice of target nucleotide (C or A) that do not allow a completely flexible genome-editing. Nevertheless, the base-editing technologies open many possibilities for gene function studies as well as the direct improvement of traits.

ii. Prime editors

To achieve the precise incorporation of multiple nucleotide substitutions, insertions, and deletions, in a target site of the genome, a search-and-replace technology called prime editing has been developed. Prime editing was demonstrated in mammals

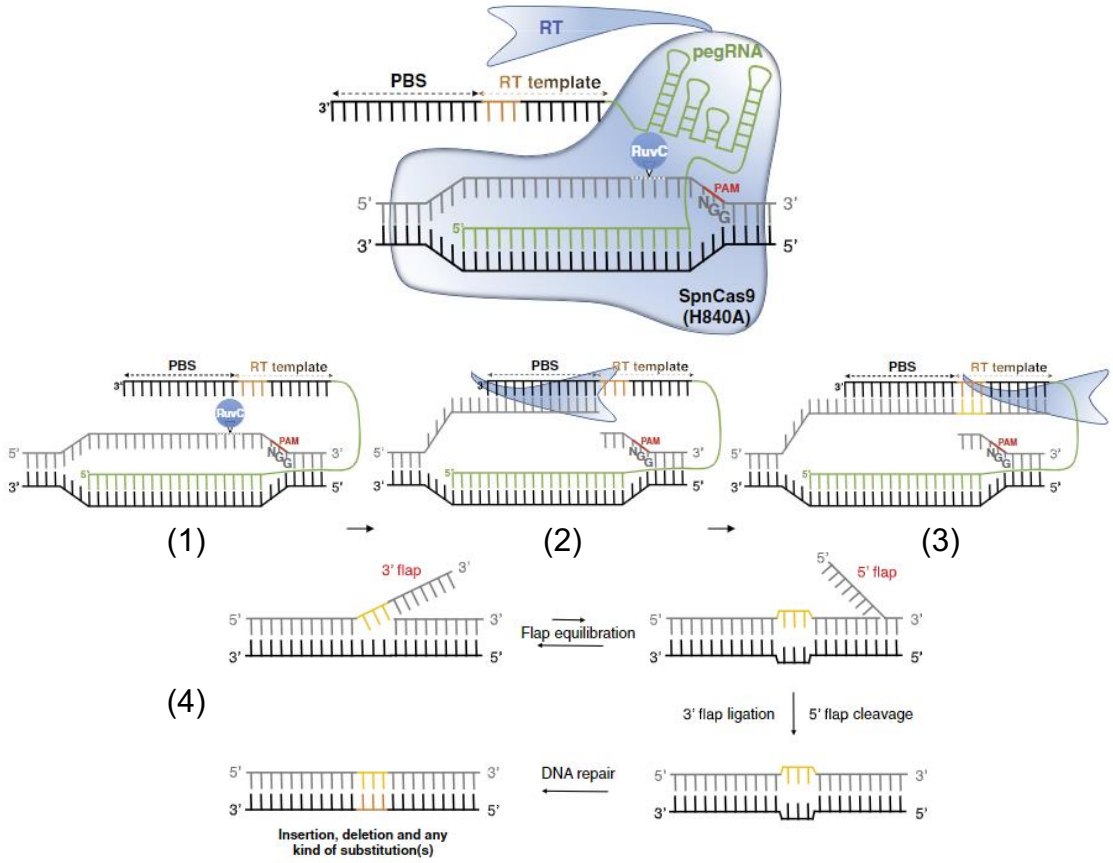


Figure i2-8 The prime editor (PE) and editing process schematics. The PE is composed of prime editing RNA (pegRNA) and SpnCas9 tethered with reverse transcriptase (RT). The pegRNA consists of a classical DNA targeting sgRNA fused to a customizable 3' extension that includes a primer binding sequence (PBS) and an RT template. RT template can be designed to introduce the desired polymorphism. (1) The PE is guided by the pegRNA to cleave the non-target DNA strand by RuvC domain of SpnCas9. (2) The 3'-end of cleaved part is bound by the PBS where the priming of reverse transcription takes place. (3) The primed RT does reverse transcription of the RT template and write new DNA sequence. (4) The 3' flap ligation is resolved and the the DNA repair mechanism integrates the new DNA sequence. Taken from Veillet et al., 2020.

for the first time (Anzalone et al., 2019) and then in cereal crops by several groups (Lin et al., 2020; Reviewed in Vu et al., 2023). The prime editor (PE) composed of a reverse transcriptase (RT) tethered to nCas9 (**Figure i2-8**). The prime editing sgRNA (pegRNA) consists of a classical sgRNA fused to a customizable 3′ extension that includes a primer binding sequence (PBS) and an RT template bearing the desired polymorphism (Veillet et al., 2020b). The PE is guided by the pegRNA to cleave the non-target DNA strand by HNH-impaired nCas9, and the 3′-end of cleaved part is bound by the PBS where the reverse transcription takes place, based on the RT template to introduce the desired indels and substitutions of different sizes. This technology can theoretically realize any type of modifications in the genome. Thus far, the editing efficiency of PEs remain low in many plant species, but several ways of improvement have been suggested, including modifying the pegRNA structure, using a paired pegRNA, adding nonsynonymous mutations in the RT template, for example (Lin et al., 2021; For review, Vu et al., 2023; Perroud et al., 2023).

f. Beyond the technological improvements: strategies to overcome the technical challenges for boosting the CRISPR editing

Another important challenges in the plant genome-editing besides the CRISPR/Cas9 technology are in the experimental step of the plant transformation/regeneration and the selection of the edited plants. In this section, different strategies to overcome the above bottlenecks will be presented.

i. Improving plant regeneration

Transformation[4] and regeneration are indeed huge bottlenecks for genome-editing for many plant species because the plant's ability to be transformed and to regenerate are genotype-dependent, but even more so for species that are generally recalcitrant like pepper (*Capsicum* spp.) and sorghum (*Sorghum bicolor*) (Ramu et al., 2017; Zhang et al., 2019a). It is also the bottleneck for orphan crop species that are less studied and therefore lacking research and protocol optimization for different steps of genome-editing (Venezia and Creasey Krainer, 2021). Regeneration is often considered a bigger bottleneck than the other steps of the CRISPR/Cas9-editing (Altpeter et al., 2016). To overcome this, several strategies based on the plant developmental regulators have

[4] Transformation sensu stricto is introduction of the transgene into the plant cell and to the plant genome.

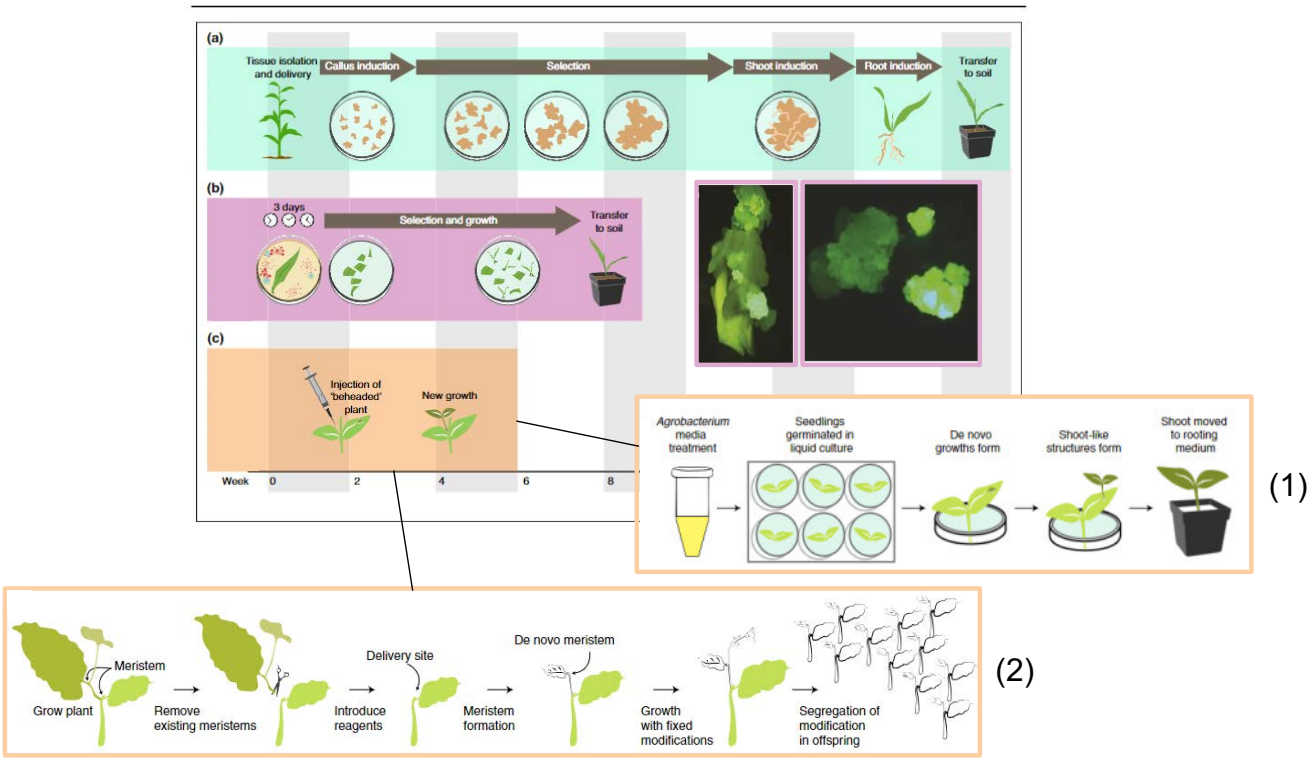


Figure i2-9 Methods utilizing developmental regulators in comparison to the classic regeneration methods. (a) Traditional regeneration method. (b) Induction of somatic embryogenesis: expression casettes of developmental regulators *BABYBOOM* (*BBM*) and *WUSCEL2* (*WUS2*) can be introduced by *Agrobacterium*-mediated transformation. The overexpression of two genes promote somatic embryogenesis in the tissue, and the callus formation was observed on mature embryo slice (left) and leaf tissue (right) in maize (Lowe et al. 2016). (c) Induction of de novo meristem: expression casettes of developmental regulators *WUS2* and *isopentenyl transferase* (*ipt*) can be introduced by *Agrobacterium*-mediated transformation on the co-culture media (1) or perfusion into a soil-grown plant (2) Modified from Maher et al., 2020 and Atkins et al., 2020.

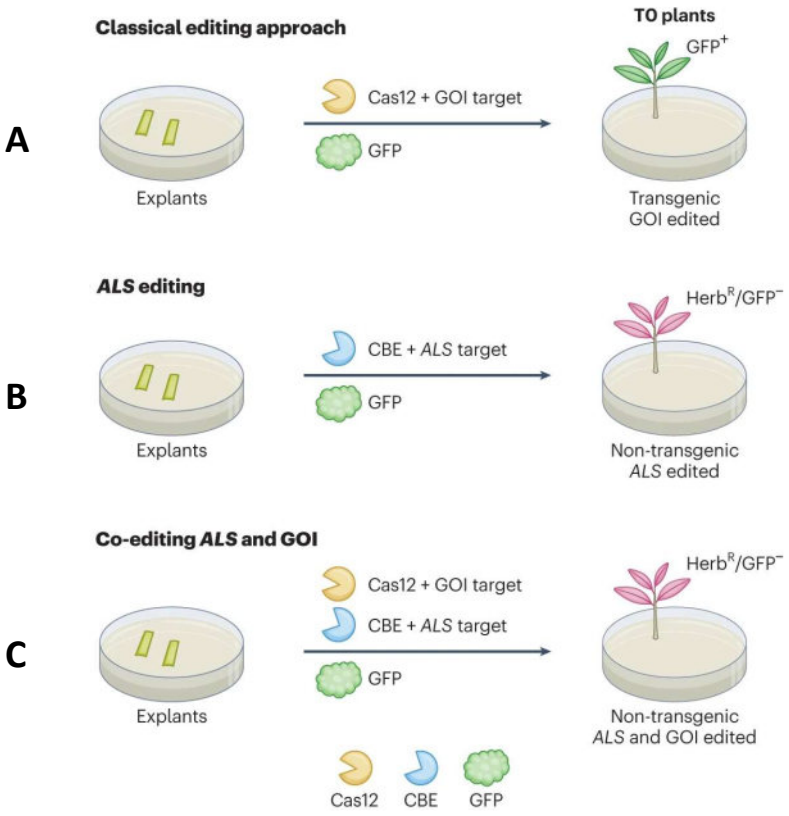
been demonstrated (For review, Atkins and Voytas, 2020) (**Figure i2-9**). Two different strategies will be introduced below.

One strategy is boosting the somatic embryogenesis by overexpression of *BABYBOOM* (*BBM*) and *WUSCHEL2* (*WUS2*) genes. Introducing overexpression cassettes of *BBM* and *WUS2* genes enhanced the recovery of transgenic shoots in multiple inbred lines of maize including some recalcitrant ones as well as sorghum, sugarcane (*Saccharum officinarum*), and indica rice (*Oryza sativa* ssp. *indica*) (Lowe et al., 2016). More recently, this method was successfully used in combination with CRISPR/Cas9 genome-editing in sorghum in a genotype-independent manner (Che et al., 2022).

Another strategy is *de novo* meristem induction by exogenous expression of *WUS2* and the cytokinin biosynthesis gene *isopentenyl transferase* (*ipt*) in leaf tissue. Leaf tissue of Cas9-expressing *N. benthamiana* was transformed with plasmid(s) carrying *WUS2* and *ipt*, and sgRNA targeting *phytoene desaturase* (*PDS*); new shoots were directly regenerated from the starting tissue and some of them were edited in *PDS* gene, without the transgene (Maher et al., 2020). The authors also demonstrated meristem induction in a tissue culture-free method, on a soil-grown *N. benthamiana* plant. These methods based on developmental regulators are promising in expanding the use of CRISPR/Cas-based genome-editing in the plant species and accessions that have been difficult to transform or neglected and in reducing the cost and time required to produce edited plants.

ii. Selection of genome-edited plants

Selection of genome-edited plant is another major obstacle to smooth production of the edited plants because it is costly and often requires the expression of a selectable reporter from the transgene. Conventionally, after the transformation step, the transformed plants are selected based on the presence of transgene, i.e. T-DNA, containing genome-editing machinery and antibiotic or herbicide resistance gene, that can be verified by survival on the selection media. The selected transformants then undergo another round of selection based on the individual editing events in the gene of interest by genotyping. Finally, the edited plants can be selfed or crossed to segregate away the transgene. This method is not only costly and time-consuming, but also undesirable for asexually propagated crop like potato (*S. tuberosum*) and cassava (*Manihot esculenta*) as well as tree species because they cannot be easily crossed to remove the inserted transgene without risking inbreeding depression or loss of desirable traits (Ramu et al., 2017; Zhang et al., 2019a). Therefore, development of the efficient, transgene-free



*Alternatively, antibiotic resistance gene like *NptII* (kanamycin resistance) is used as a reporter and selected on the antibiotic-containing media.

Figure i2-10 Recovery and selection methods of genome-edited plants, based on the strategy used from Huang et al., 2023.

A In classical editing approach, for example, a casette expressing Cas12 and GOI target (sgRNA targeting gene of interest (GOI)) along with a reporter gene GFP is delivered to the cells of explants. The transformants integrating the transgene (GFP-positive (+) plants) are selected and regenerated. Then, these plants are analyzed for the GOI editing. The frequently used reporter genes include antibiotic resistance genes which can be selected by the corresponding antibiotic-containing media. **B** In *ALS*-editing approach, a casette expressing CBE and sgRNA targeting *ALS* gene along with the reporter GFP is delivered to the cells. The *ALS*-edited T0 plants are selected on the herbicide (chlorsulfuron)-containing media ($Herb^R$). Meanwhile, GFP^+ plants can be counter-selected as transgenic plants (thus GFP^- is selected). Likewise, in **C**, co-editing *ALS* and GOI, a casette expressing Cas12 + GOI target, CBE + sgRNA targeting *ALS* gene, along with a reporter gene GFP is delivered to the cells. Using the chlorsulfuron media, only the *ALS*-edited, GFP^- (thus transgene-free) plants can be selected and regenerated, and these plants can be analyzed for the GOI editing. *ALS*-edited plants are often found to be edited in the GOI by Cas12. Taken from Gallois and Nogué, 2023.

selection method would be interesting for implementation of genome-editing in the breeding program.

iii. Transient expression

One way to bypass the transgene integration into the plant genome is the transient expression of the CRISPR/Cas9 system without a selectable marker in the plant cells (**Figure i1-7B**). It is followed by a large-scale screening of all the treated tissues for the edition. Thus, this strategy would skip the transformant selection step and directly detects the edited plants by sequence validation of all the treated plants (Zhang et al., 2016). While the transient expression method indeed generates transgene-free, homozygous mutants in T0 generation, the screening by genomic analyses can be laborious and very costly, especially when simultaneously targeting multiple genes or plants with polyploidy.

iv. Endogenous reporter editing

Another way to avoid transgene integration while achieving selection efficiency is to edit endogenous gene as a reporter and perform selection based on the generated phenotype. *Phytoene desaturase* (*PDS*) gene whose KO mutant displays albino phenotype and *acetolactate synthase gene* (*ALS*) gene whose site-of-action mutant displays resistance to different herbicides are among such genes that can be targeted as reporter. *PDS* gene was indeed among the first genes targeted for gene KO by the CRISPR/Cas9 (Shan et al., 2013). *ALS* was also among the first genes targeted for base-editing by the CBE to introduce substitutions within the exon that conferred resistance to herbicides, which could be selected by herbicide-containing media (Tian et al., 2018; Zong et al., 2018). While the albino phenotype of the *pds* KO would not be agronomically useful, herbicide resistance phenotype of the functional *ALS* mutant had the potential to be used as a reporter, without the need of transgene.

v. Co-editing of reporter and gene of interest by transient expression

Combination of these strategies together, the transient expression of CBE, delivered by *Agrobacterium*-mediated and biolistic transformation was successfully shown to base-edit *ALS* gene in tomato and potato (Veillet et al., 2019) and wheat (Zhang et al., 2019b), respectively, allowing selection on the sulfonylurea-containing media (**Figure i2-10B**); and some of the T0 mutants were directly transgene-free. These studies have opened the avenue to the co-editing of the endogenous *ALS* gene as a selectable marker and the gene of interest in plant species to realize a cost-effective, transgene-free

genome-editing. The cherry tomato mutants that were studied in the *Chapter 2* of this thesis were partially generated via the co-editing of *SlALS* and *SleIF4E1* using CBE. The co-editing of *ALS* by the CBE and the gene(s) of interest by CRISPR/Cas12a in tomato, potato, tobacco, and citrus (*Citrus maxima* cv. *Siamese Sweet*) was also successfully demonstrated, underscoring the potential of *ALS* co-editing strategy (Huang et al., 2023). In this paper, authors have additionally used *green fluorescent protein* (GFP) gene as a reporter gene to counter-select the transgene, further increasing the efficient selection of transgene-free mutants (**Figure i2-10C**).

The emergence of CRISPR/Cas technologies in the genome-editing toolbox for plant improvement was groundbreaking. Combined with the improvement in plant transformation methods, the power of the CRISPR/Cas system and the base editors is now readily harnessed to overcome the weaknesses of previous techniques and improve various traits in the elite varieties of crops as well as in the wild crop relatives.

3. Genomics enhances genome-editing for trait improvements with precision

a. Role of genomics study in crop breeding

Agronomic traits of modern crop varieties we see today have been selected throughout the domestication of ancestral accessions and the improvement that followed. These traits are coded by one or more genes or often in a quantitative manner by QTLs, and the phenotype-based selection has allowed to accumulate these loci at higher frequency than other loci within the population. Studying the evolutionary process that has shaped the modern crop traits, e.g. fruit size and growth habit, and identifying the causal genotypes based on the genomic sequences have been an important aspect of recent plant breeding. This has revealed many plant growth and development-related genes and QTLs in different species (Doebley et al., 2006). In parallel, the exploration of natural gene reservoirs among the landraces and wild relative species has been carried out in order to introgress more beneficial traits such as abiotic and biotic resistance into the elite cultivar (For review, Mickelbart et al., 2015; Dempewolf et al., 2017).

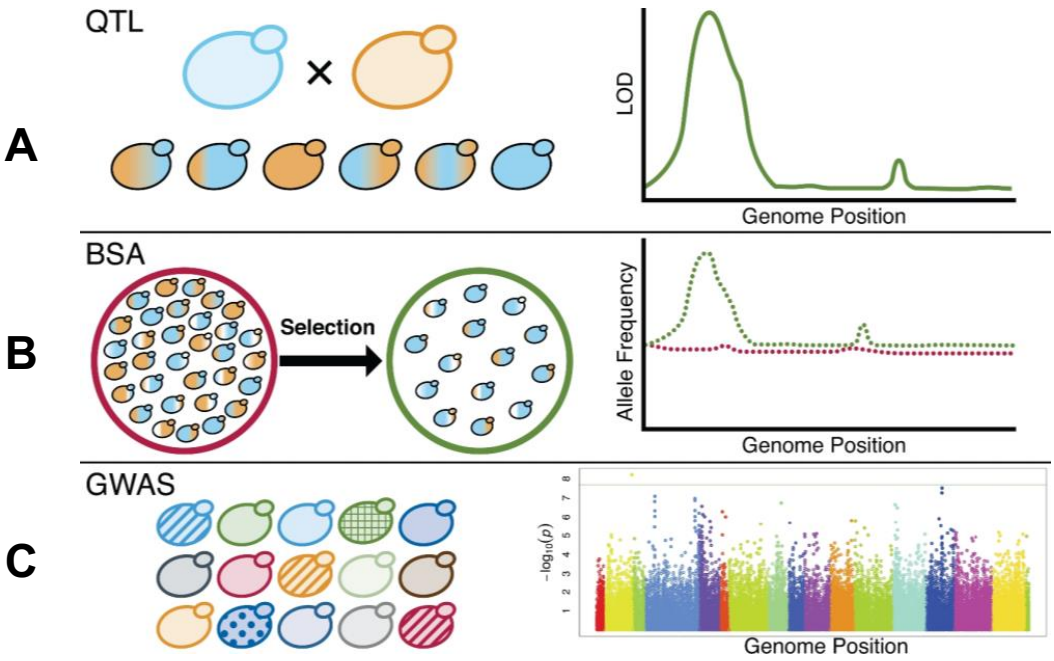


Figure i3-1 Concept figures of different genomic approaches.
A QTL mapping: two parental lines are crossed (and backcrossed and/or selfed multiple times) to generate a population of progenies. Genotype and phenotype of the individuals are analyzed. The accumulated data for the population is used to estimate the correlation levels (log-odds (LOD) score) between genotype-phenotype, and the loci of a high LOD is predicted as the QTL of interest. **B** BSA: two parental lines with extreme phenotypes are crossed to generate a population of a large number of progenies. Progenies are pooled according to the two opposing phenotypes. The pooled populations, or the bulks, are sequenced and the frequency of genomic regions are analyzed; a high allele or loci frequency in one bulk population (green circle) is predicted as causal genomic region for the associated phenotype in a population. **C** GWAS: a population of varieties with known phenotypes are subjected to the whole-genome sequencing. Correlations unlikely to occur by chance between the genotype and the phenotype across many varieties are detected (–log of the p-value) and the genome position of implicated loci are detected. Figure taken from Sardi and Gasch, 2017.

i. Genomics approach to identify causal alleles and QTLs

The detection of genes and QTLs of interest in the crop cultivars and their wild relatives was made possible with development of genomic approaches. Namely, the candidate gene approach and the QTL mapping and cloning have been among the first methods that localized the genetic markers and associated traits on the crop genome (For review, Collard and Mackill, 2008; Meyer and Purugganan, 2013). The more recent methods based on the whole-genome (re)sequencing (WGS), such as bulk-segregant analysis sequencing (BSA-seq) and WGS-based genome-wide association studies (GWAS) are able to detect the SNPs and eventually the loci associated with traits at high precision (**Figure i3-1**) (For review, Meyer and Purugganan, 2013; Chat et al., 2022; Majeed et al., 2022).

As the WGS has become more accessible in terms of cost and computational resources, the number of plant species sequenced and assembled expanded. Today, one or more genome assembly data of high quality are available for many crop species and their wild relatives: 798 land plant species of which 135 domesticated, 127 cultivated, 120 natural commodities[5], and 12 feral[6] species, plus 404 wild species including 77 that are wild relatives of crops (Marks et al., 2021). A growing collection of genome assembly data will help find the genes/QTLs for a wide range of species including minor or orphan crops whose plant breeding program has been less studied and invested (Tadele, 2019; Venezia and Creasey Krainer, 2021).

ii. Genomics and plant breeding

Mapping of alleles and QTLs of interesting agronomic traits has greatly contributed to the modern crop breeding because the identified genetic markers of these loci improved the accuracy and speed of selection during cross breeding and subsequent backcrossing (i.e. the marker-assisted selection and the marker-assisted backcrossing) (Collard and Mackill, 2008). Furthermore, cloning of loci provided the basis for the functional validation as well as transgenic breeding.

Today, the genomic data of high resolution provided by an increasing number of whole-genome assemblies has the great potential to boost the crop improvement mediated by the CRISPR/Cas9 systems. The precise information on the structure and

[5] plants that are harvested with little cultivation (Marks et al. (2021))
[6] plants that are not economically important but have still been influenced by human selection (Marks et al. (2021))

Table i3-1 Intragenic mutation types and sites found in different genes and QTLs
TE: transposable element

Mutation types/site	Exon (Coding region)	Intron	5'- or 3'-UTR	5' or 3' flanking regions
SNP	Nonsense mutation > Loss of function Missense mutation > Loss or Gain of function	In splice site or other splicing-related *cis*-elements > Alternative splicing > Loss or Gain of function	In *cis*-regulatory elements > Change in gene expression level/localization > Loss or Gain of function	
Indels	Frameshifting > Loss of function In-frame modifications > Loss or Gain of function			
TE insertion	Induction of truncated or aberrant transcripts > Loss or Gain of function		In or near *cis*-regulatory elements > Change in gene expression level/localization > Loss or Gain of function	

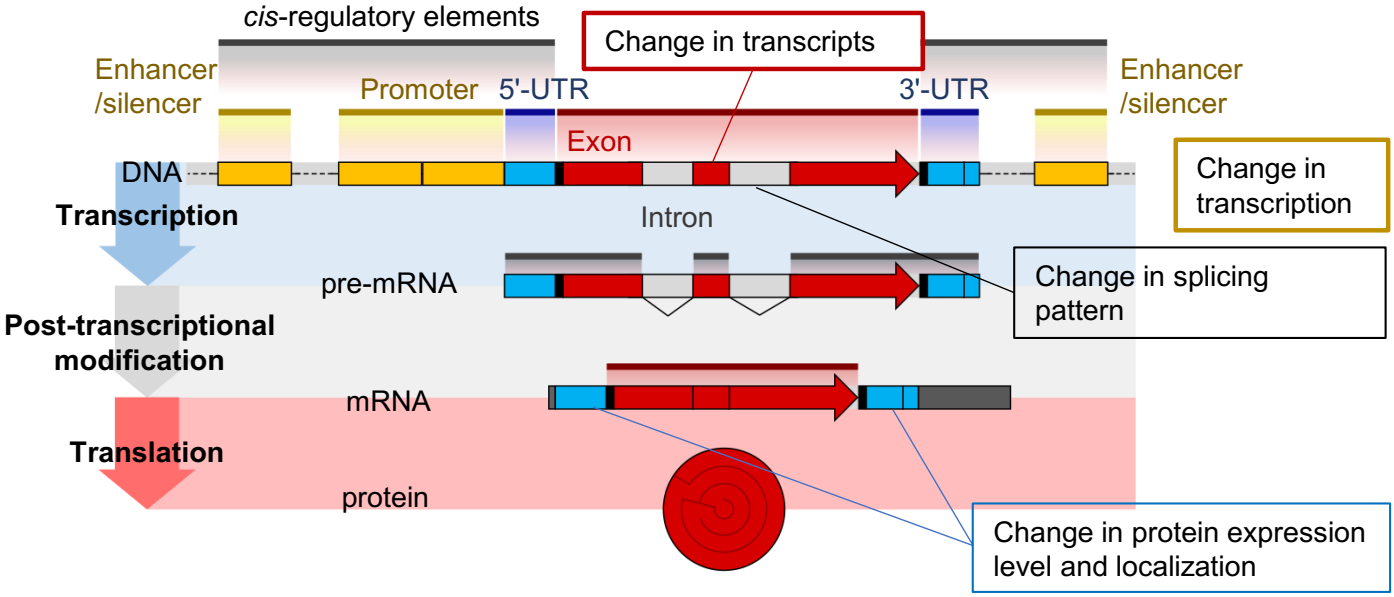


Figure i3-2 Possible sites of mutations in a gene and their effects. Boxes indicate different elements of a gene and its regulatory elements. Red is the coding region, blue is the untranslated region, yellow is the different *cis*-regulatory elements at transcriptional level. Genomic region that harbors the *cis*-regulatory elements is indicated by black bar on the top. The effects of mutation in each site are indicated by texts in the colored box. Figure modified from Shafee and Lowe, 2016.

Table i3-2 Genomic variations found at a larger scale than a gene

Mutation types	Effects
Genome duplication	Increase in ploidy level > Gain of function
Gene duplication	Increase in copy number, neofunctionalization, subfunctionalization > Gain of function
Chromosomal rearrangement: inversion, translocation, segmental duplication	Presence or absence of one or more genes, variation in gene copy number > Loss and Gain of function

molecular function of the identified gene or QTL of interest offer hints for modifications and designing of the elite cultivar alleles using the CRISPR/Cas9-derived genome-editing tools. This would allow to confer one or more desirable phenotypes, without repeated backcrossings and interspecific barrier. The following sections will present different features of genes and QTLs revealed with genomic approaches and functional validation; and discuss how such information can be used to fine-tune the agronomic traits by genome-editing.

b. Types of allelic variants and functional validation

Before the emergence of SSNs, especially the CRISPR/Cas9 technologies, the mapping of the loci contributing to the phenotypes of interest in the donor genome and the loci itself provided the most important resource for the breeding programs. Now the CRISPR/Cas9 system can directly access and modify the target loci in an elite cultivar, with a guide of the genomic sequence. Hence, the information on the loci, i.e. the types of variations and their effect on the gene expression, contributing the beneficial traits in the donor have become the most essential resource, which can be transferred to the elite cultivars.

i. Types of allelic variants

In order to apply new traits into a crop of interest using the CRISPR/Cas9 system, it is important to understand the nature of the causal alleles and QTL*s*. The causal alleles and QTL*s* of a trait can be either loss-of-function or gain-of-function variants with different levels of effect. These variants are due to different types of mutations such as SNPs, indels, transposable element (TE) insertions, and genomic structural variations: gene or genome duplications, and large-scale chromosomal rearrangements (**Table i3-1,2**) (For review, Olsen and Wendel, 2013; Meyer and Purugganan, 2013). The SNPs, indels, and TE insertions can occur in the coding region of a gene, or exon, as well as non-coding region: introns, 5′-untranslated region (UTR), 3′-UTR, and the 5′ and 3′ flanking regions, which contain different *cis*-regulatory elements of transcription and translation. The *cis*-regulatory element mutation in the 5′ and 3′ flanking regions influence the gene transcription, whereas that in the exon, intron, and 5′- or 3′-UTRs induce changes at post-transcriptional level, i.e. mRNA translation and localization (**Figure i3-2**). The genomic structural variations, the mutations in the genome at a larger scale, are summarized in **Table i3-3** and will not be discussed in detail here (For review, Flagel and Wendel, 2009; Olsen and Wendel, 2013).

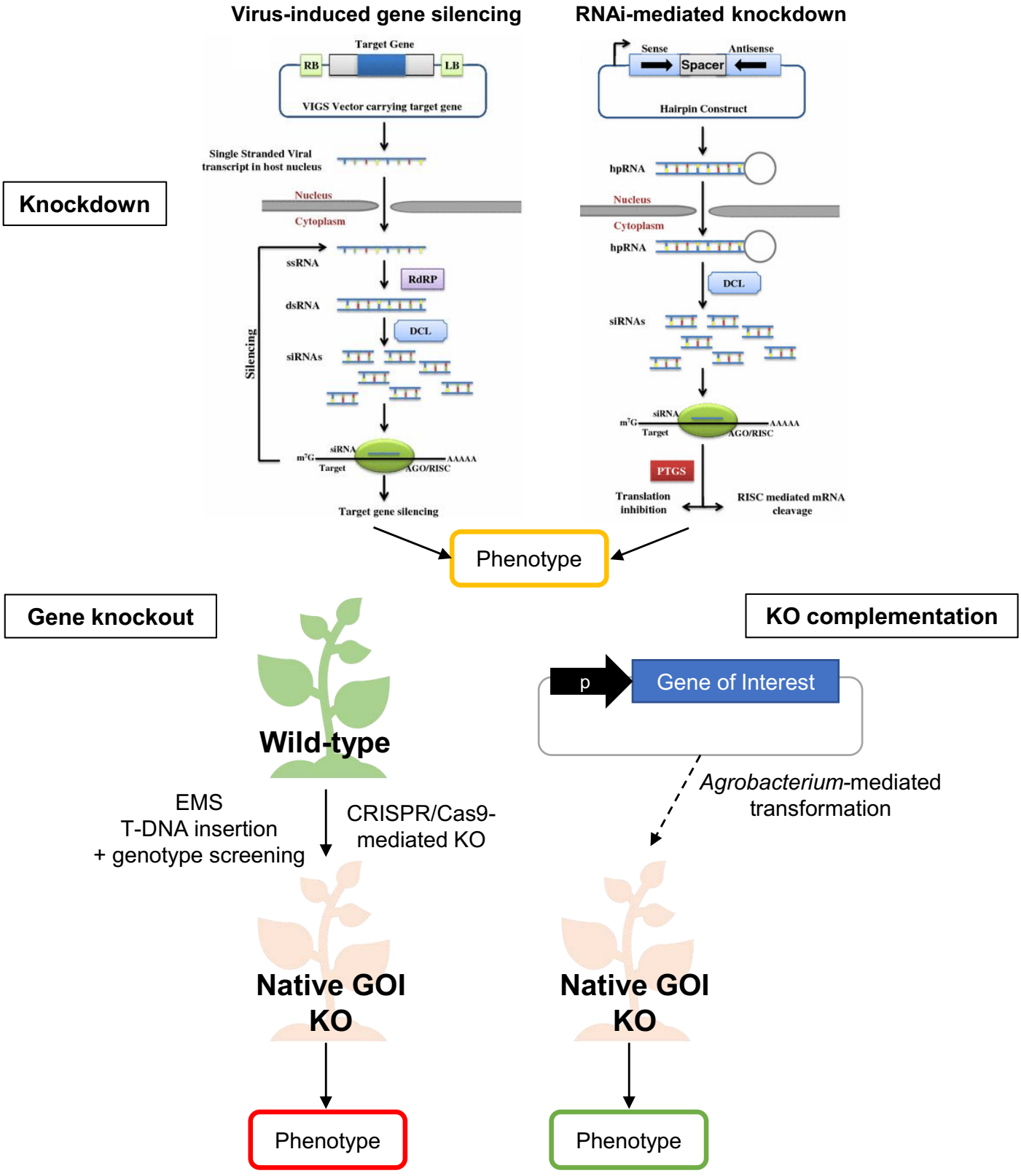


Figure i3-3 Functional validation by gene knockdown, gene knockout, and complementation.

Gene knockdown by virus-induced gene silencing (VIGS) and RNA interference (RNAi)-mediated knockdown are shown on top. Gene knockout and complementation assay by (over)expression of a gene of interest (GOI) on bottom. In VIGS, the VIGS vector carrying a host target GOI is transformed into plant cell. In RNAi, a typical hairpin RNA construct harboring the target GOI driven under the control of strong promoter is transformed into the plant cell. In both methods, the expressed RNA induces host RNA silencing which results in silencing the host GOI (knockdown) and mimicking gene KO phenotype. Figure adapted from Tiwari et al., 2014. Gene KO is done by chemical mutagenesis like EMS or T-DNA insertion (for Arabidopsis). The phenotype of gene KO for GOI allows validation of the function of the GOI. KO mutants can also be used for complementation assay of the variants of GOI by expressing the GOI from an overexpression or native promoter (p).

The effects of mutations in a gene depend on both the type and site of mutations and are summarized in **Table i3-2**. Briefly, SNPs in the exon that introduce premature termination codon (PTC) (nonsense mutation) and indels in the exon that induce frameshifting and PTC result in loss of function the gene (For review, Xu and Guo, 2020). These are considered the most frequent causal mutations identified in the domestication genes (Meyer and Purugganan, 2013). SNPs in the exon that introduce nonsynonymous mutations (missense mutation) or in-frame modifications may alter the structure and the function of the protein, leading to either loss-of-function or gain-of-function variant.

Both SNPs and indels in the intronic region can result in either loss or gain of function of the gene due to the altered splicing. TE insertion in the exonic or intronic region of a gene leads to the production of truncated or aberrant transcripts (For review, Dubin et al., 2018). On the other hand, the SNPs, indels, and TE insertion in or near the *cis*-regulatory element: the 5′- or 3′-UTR and the 5′ or 3′ flanking regions, often alter the gene expression level. Such alleles often show quantitative and varying effect on phenotype.

Altogether, various types of mutations modify the protein expression, structure, localization, and/or function in either loss- or gain-of-function manner, and the effect is specifically determined by the mutations and their site.

ii. Functional validation of genes and QTLs

Once the candidate mutations in the allele/QTL are found, it is also interesting to perform the functional validation of the mutations in order to understand the gene functions and determine the effects of different mutations on the traits.

Conventionally, the functional validation of a gene or QTL, of especially loss-of-function variant, is performed *in planta* by gene knockdown or knockout methods in an attempt to reproduce a signature phenotype (**Figure i3-3**). RNA interference (RNAi)-mediated knockdown or virus-induced gene silencing (VIGS)[7] are major knockdown techniques, which are performed in the native background, or often in *A. thaliana* (hereafter referred to as Arabidopsis) (**Figure i3-3** Knockdown). Gene knockout plants have been obtained by targeting induced local lesions in genomes (TILLING) with EMS in crop species (**Figure i3-3** Knockout). The TILLING mutants have additional advantage as

[7] VIGS: an RNA-mediated reverse genetics technology for analyzing the function of genes. It downregulates the endogenous gene of interest in plants by triggering the posttranscriptional gene silencing via infection of a viral vector carrying a part of the target gene.

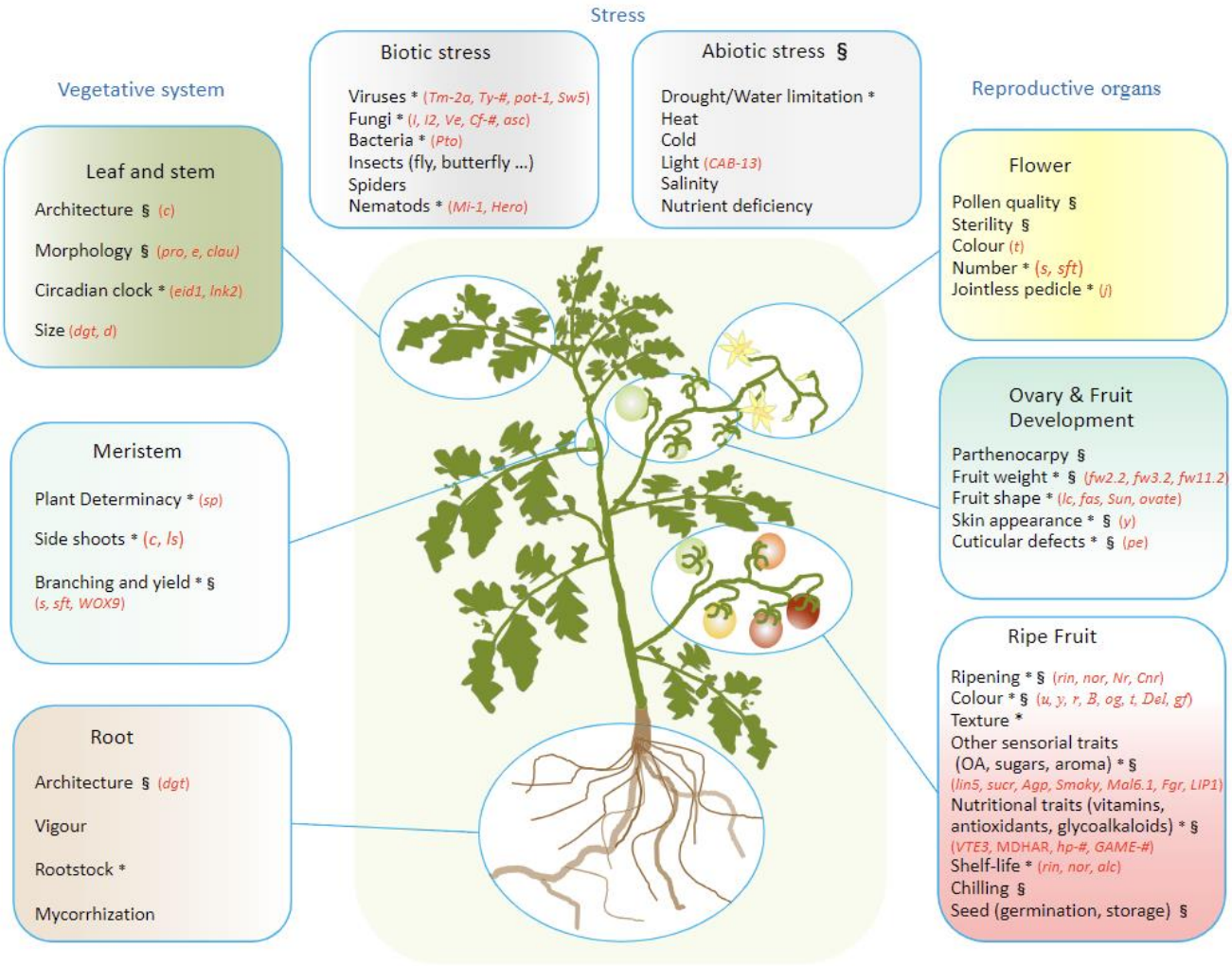


Figure i3-4 List of extensively studied tomato phenotypic traits and identified causal genes

* : traits selected during tomato domestication and improvement. §: traits highly susceptible to variations in environmental conditions. In red, non-exhaustive list of major genes and QTLs identified in natural genetic diversity. Figure taken and legend adapted from Rothan et al., 2019.

they can be generated in any genetic backgrounds and also used a source of plant breeding though it requires a large-scale screening as well as multiple rounds of backcrossings to isolate such mutants (Scheben et al., 2017). The KO mutants can also be used in a complementation assay to verify the gain-of-function variants by transgenic (over)expression (**Figure i3-3** KO complementation).

On another note, function of the putative alleles of interest are often extensively studied in the model plant Arabidopsis because there are larger collections of genomic data, natural variants and mutants, and functional analysis tools available for Arabidopsis (Scheben et al., 2017). The crop gene of interest could be expressed in Arabidopsis in which its ortholog is knocked out, as to check the complementation of the gene. Conversely, Arabidopsis ortholog could be expressed in the KO line of crop though this is a less popular option because it is more difficult to transform crop species than Arabidopsis plants (Yaschenko et al., 2024). The discoveries made in Arabidopsis serve as strong basis of crop trait improvement and have been often translated to different crop species (Yaschenko et al., 2024). While these methods were strongly indicative of the allele function, there were concerns that the results were not reflecting the true behavior of the alleles due to the experimental settings, i.e. use of overexpressing promoter, incomplete suppression of gene expression, and non-native environment, could influence the effect of allele.

Today, with the development of the CRISPR/Cas technologies and of plant transformation techniques, *in situ* functional validation has become much simpler and more accessible (**Figure i3-3**). The alleles can be directly targeted in a species of interest for gene knockout or other type of modifications using the CRISPR/Cas to induce DSB followed by NHEJ repair pathway that randomly introduces indels. For instance, targeting exonic region(s) of a gene likely leads to a classic knockout of the gene, while targeting *cis*-regulatory elements of a gene can generate quantitative phenotypes (Gao, 2021). This would accelerate the reverse genetics study to understand the function of the genes as well as QTLs. With the rise of base-editing, it is also possibly to introduce substitutions in the genomic sequence that modifies the gene function without knocking out. Such strategy will allow to pinpoint the functional domains and motifs that are important for the structure and function of the proteins, and conversely these functionally important regions can be screened by using sgRNA library (Guyon-Debast et al., 2021; Gaillochet et al., 2021).

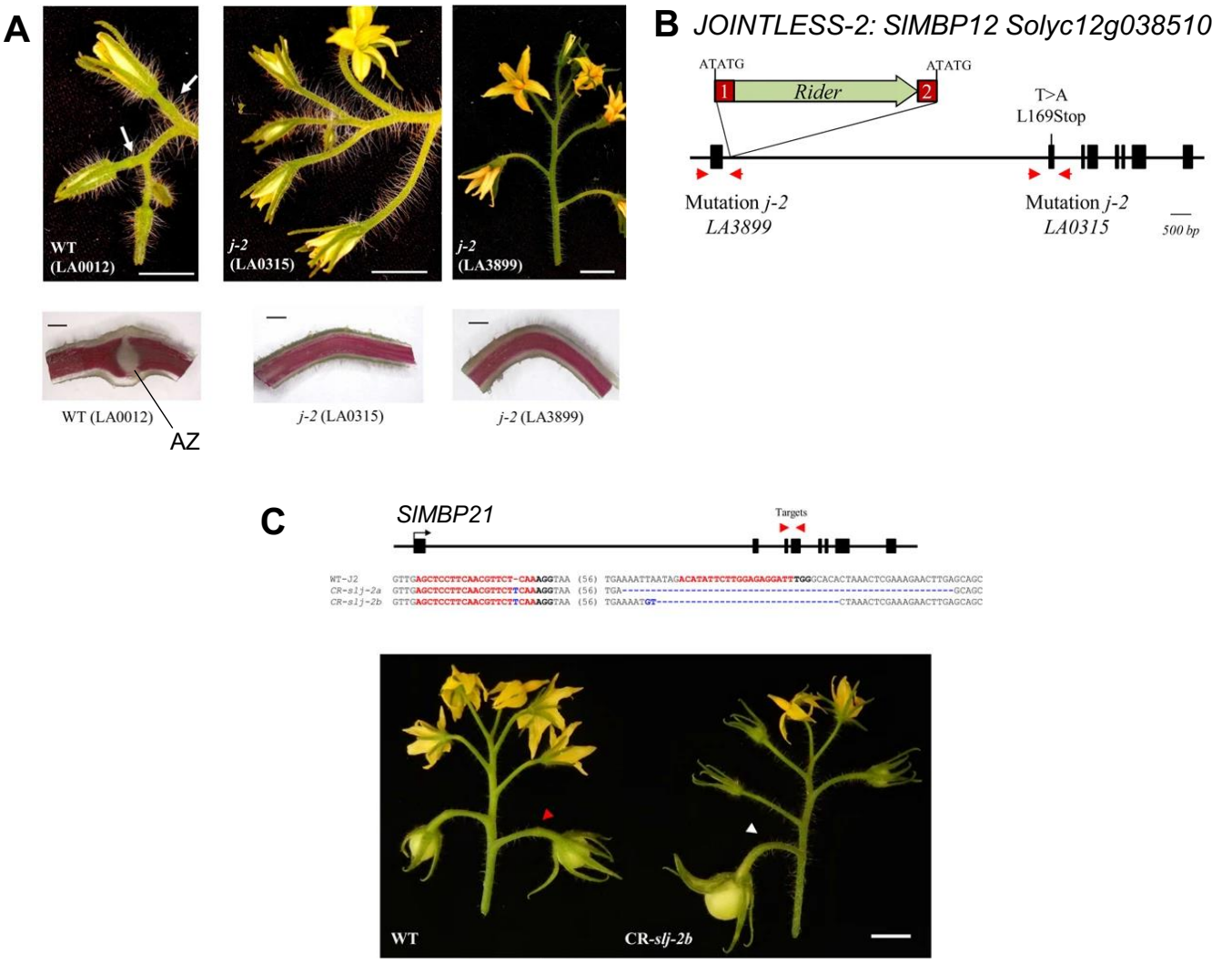


Figure i3-5 The jointless pedicel phenotype of tomato is due to *JOINTLESS-2* (*J-2*) allele and can be engineered by CRISPR/Cas9-mediated gene KO.

A Photo of Wild-type variety Pearson and *j-2* mutant tomato accessions LA0315 and LA3899. Top: *j-2* tomatoes lack the joints on the pedicel compared to the WT (indicated by white arrows). Bottom: the longitudinal sections of flower pedicels stained with phloroglucinol show the absence of the abscission zone (AZ) in the *j-2* mutants. Scale = 2 mm. **B** Diagram of *JOINTLESS-2* (*J-2*) gene. Distinct mutations found in loss-of-function j-2 alleles LA0315 and LA3899 accessions are shown. *Rider* in green arrow is a retrotransposable element after exon1 in LA3899-derived *j-2* and a PTC in exon 2 in LA0315-derived *j-2*. Modified and Roldan et al., 2017 and Mao et al., 2000. **C** The jointless pedicel phenotype in tomato conferred by *j-2* allele was reproduced by targeted deletions in *SlMBP21* (*J-2*) gene. Top: the targeted site of *SlMBP21*. Two sgRNAs targeting exon 3 and 4 were used for CRISPR/Cas9 editing. The alignement shows the two edited alleles, each with an insertion and a deletion (in blue). Bottom: photos of WT (var. M82) and the *SlMBP21* edited tomato (*CR-slj-2b*). *CR-slj-2b* shows jointless phenotype. Taken from Roldan et al., 2017.

c. Enhancing agronomic traits by genome-editing

Improvement of techniques for gene/QTL fine mapping has allowed identification of numerous causal genes and loci along with their genetic markers for important phenotypes for many crop species at high resolution. Many alleles associated with agronomically interesting traits were discovered among the naturally occurring mutations in the wild relatives, land laces, and cultivated accessions as well as artificially induced mutants. These discoveries were complemented with reverse genetics studies of the function of causal alleles using the model species *A. thaliana* and the crop species of interest, parallelly. Propelled by these accumulated genomic and genetic knowledge, the CRISPR/Cas9 system has readily generated beneficial phenotypes in the crop species.

i. The CRISPR/Cas9 systems can generate loss-of-function gene and fine-tuned QTL

Great examples of the CRISPR/Cas9 system recreating the phenotypes found in the natural variations have been demonstrated in one of major crop species, tomato by (Soyk et al., 2017; Roldan et al., 2017; Rodríguez-Leal et al., 2017). Here, a few agronomic phenotypes are featured, but the list of phenotypes that have been studied and can be improved is exhaustive; such list for tomato is provided in **Figure i3-4**.

ii. Copying a trait by gene KO

Soyk et al. (2017) and Roldan et al. (2017) copied the jointless pedicel phenotype that is found from certain tomato accessions to a cultivated tomato variety M82. Jointless phenotype is important for the post-harvest fruit quality and the shelf-life; one of the causal loci *j-2* has been characterized from the wild relative *S. cheesmaniae* (LA0166) and a commercial variety (**Figure i3-5A**) (Mao et al., 2000). From the *j-2* locus, two loss-of-function variants of *SlMBP21* have been identified: one was due to an insertion of a *Rider* retrotransposon, and the other was due to a nonsense mutation in the exon 2, leading to a truncated form of protein (**Figure i3-5B**) (Soyk et al., 2017; Roldan et al., 2017). Based on this information, the two teams targeted the exon 3 and 4 of *SlMBP21* using two sgRNAs and the CRISPR/Cas9 in a cultivated tomato variety M82; this induced an insertion and a large deletion in the targeted genomic region and mimicked the jointless pedicel trait of the natural *j-2* loss-of-function alleles of *SlMBP21* (**Figure i3-5C**: example from Roldan et al., 2017). Roldan et al. (2017) also noted that the CRISPR KO *SlMBP21* plants did not show some morphological characters found in the natural *j-2* introgressed

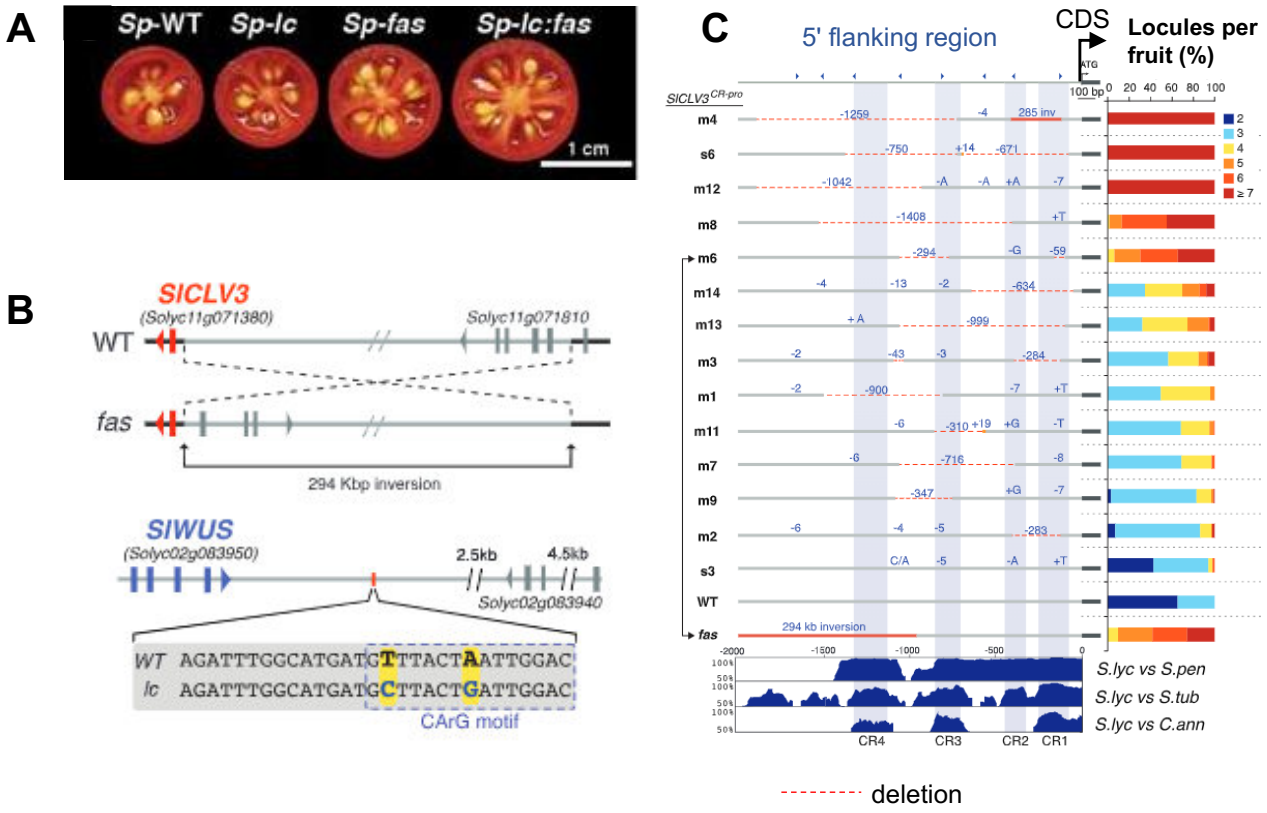


Figure i3-6 The locule numbers in tomato are controlled by *CLV3* and *WUS* alleles and variations can be engineered by CRISPR/Cas9-mediated gene KO of promoter region.
A Photo comparing the tomato fruit of wild ancestor *S. pimpinellifolium* (*Sp*-WT) and mutants of locule number/fruit size QTLs (*lc*, *fas*, and double mutants *lc:fas*) in the *Sp* background. Locule number is increased in *Sp-fas* and *Sp-lc*:*fas* fruits. **B** Diagram of QTLs *fas* and *lc* in comparison to their wild type alleles, *SlCLV3* and *SlWUS*, respectively. and Rodriguez-Léal et al., 2017. **C** Promoter region of *SlCLV3* was targeted for editing by CRISPR/Cas9. Alleles with various deletion profiles was obtained and the locule number per fruit for each edited lines was analyzed. The percentage of locule numbers are shown by colors. CR1-4 are the Conserved Region in the promoter sequence between the cultivated tomato and wild relative (*S. penelli*) other Solanaceae crops (*S. tuberosum* and *Capsicum annuum*). CDS: coding sequence. Figures taken from Sommsich et al., 2016; Rodríguez-Leal et al., 2017.

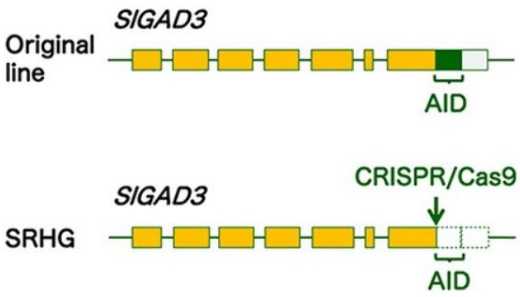


Figure i3-7 Examples of CRISPR/Cas genome-editing for biofortification.

Comparison of *SlGAD3* alleles in the parental tomato line and the edited variety “Sicilian Rouge High GABA (SRHG)”. Exon7 containing the autoinhibitory domain of the GABA synthesis gene *SlGAD3* was targeted for deletion by CRISRP/Cas9. Taken from Ezura et al., 2021.

variety, speculating that this was likely due to the absence of linkage drag in the CRISPR line.

iii. Generating quantitative variations by QTL editing

In contrast, Rodríguez-Leal et al. (2017) has generated tomato plants with varying locule numbers by inducing various mutations at QTL. In tomato, the fruit size is controlled by many genes and QTL, which have been found as the domestication and improvement genes and QTLs (Tanksley, 2004). Among them, naturally occurring *fascinated* (*fas*) QTL and *locule number* (*lc*) are the major contributors to the increased locule numbers compared to two locules in wild relatives, hence the large fruit size, in the large-fruited cultivars (**Figure i3-6A, B**) (Lippman and Tanksley, 2001). Of the two, *fas* is a partial loss-of-function QTL due to the inversion of 295kbp genomic region in the upstream of *CLV3* gene including *CLV3* promoter (Huang and van der Knaap, 2011; Xu et al., 2015) (**Figure i3-6B** top) (Muños et al., 2011; Somssich et al., 2016; Rodríguez-Leal et al., 2017). Based on the genomic structure of *fas* QTL, Rodríguez-Leal et al. (2017) has targeted the promoter region of *SlCLV3* to introduce random indels by CRISPR/Cas9. The results have validated the dosage effects of the *fas* QTL and demonstrated the possibility to modulate the level of a QTL-coded phenotype, i.e. generate tomatoes with range of locule numbers (**Figure i3-6C**). Targeted mutagenesis by the CRISPR/Cas allows to modify or mimic and observe the behavior of alleles in their native conditions and leads to accurate understanding of the gene function.

iv. From functional validation to biofortification by gene KO

Additionally, there is an interesting example of how functional validation of the metabolic pathway gene in tomato has led to development of the first CRISPR/Cas-edited crop to go on a market. Nonaka et al. (2017) has generated a Micro-tom variety of tomato that shows increased accumulation of γ-aminobutyric acid (GABA), a non-proteinogenic amino acid known to lower hypertension in humans. The group targeted two homologous genes coding for *glutamate decarboxylase* (*SlGAD*), a key enzyme in GABA biosynthesis in tomato for gene knockout of autoinhibitory domain by the CRISPR/Cas9 and successfully generated GABA-enriched tomatoes (**Figure i3-7**). This research later became a foundation of "Sicilian Rouge High GABA", the first CRISPR-edited tomato commercial variety to be sold on the market in the world (Ezura, 2022).

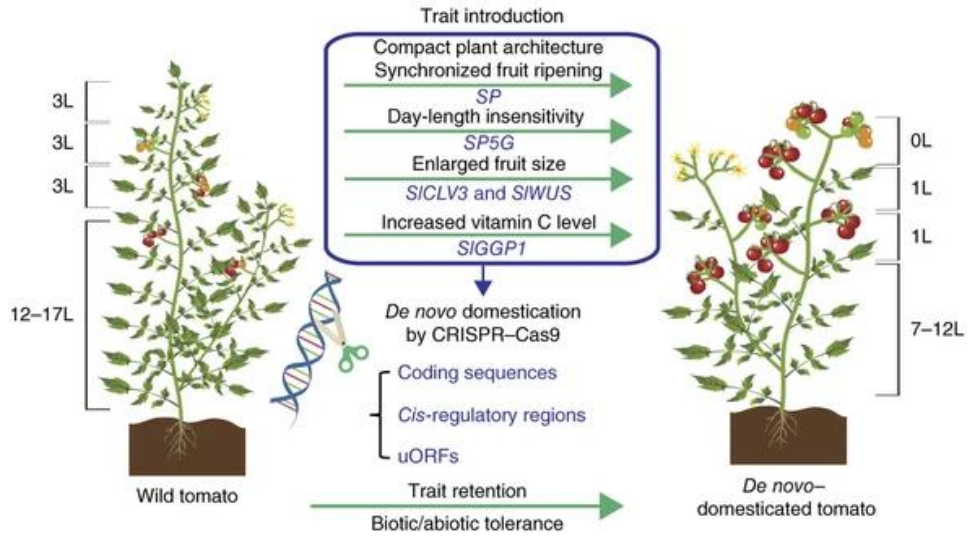


Figure i3-8 Concept of trait induction to achieve *de novo* domestication of wild tomato relative.

The genes underlying major domestications traits (in blue box) were targeted by CRISPR/Cas9 while retaining the beneficial traits of wild relative such as biotic and abiotic stress resistance. These genes were edited in coding sequence, CRE, or upstream ORFs (uORFs). The de novo-domestcated tomato is expected to have less foliar growth and large fruit traits, for example. L: leaf. Taken from Li et al., 2018.

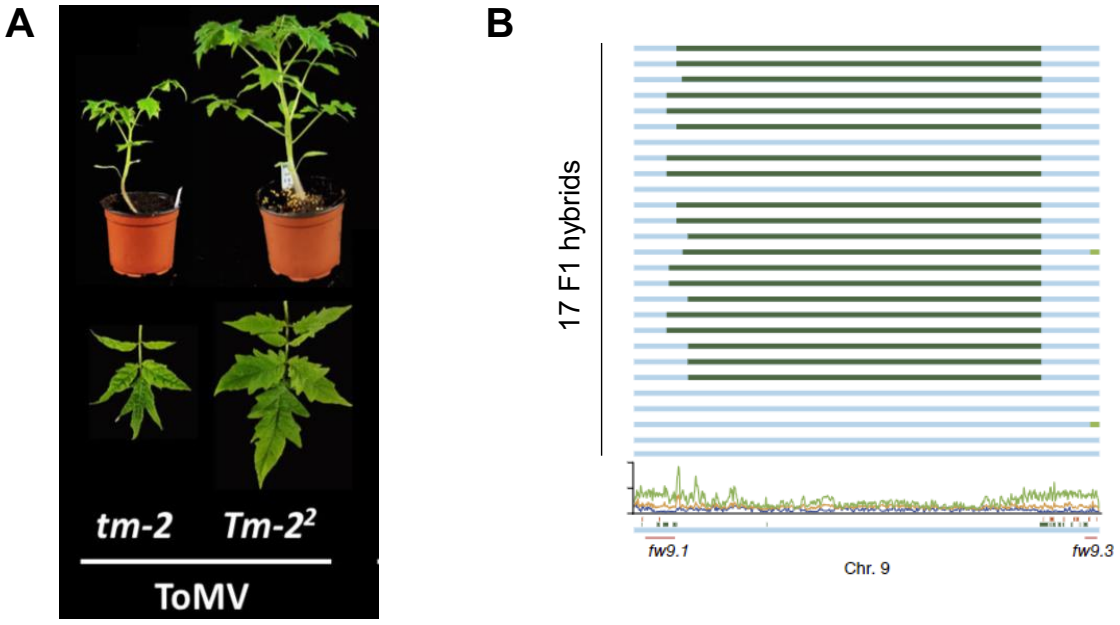


Figure i3-9 *Tm-2²*: natural resistance allele introgressed in the modern tomato cultivars.

A *Tm-2²* (also known as *Tm-2ᵃ*) introgression line shows resistance against tomato mosaic virus (ToMV). ToMV-infected susceptible wild-type tomato (*tm-2*) shows growth inhibition symptom, while Tm-2a introgression line does not. Taken from Hak and Spiegelman, 2021. **B** *Tm-2²*-containing introgression spans a huge part of chromosome 9 in many hybrid tomato cultivars. Each bar is a hybrid cultivar out of 17 F1 hybrids analyzed in Li et al., 2014. Green bar is the introgression from *S. peruvianum* to confer *Tm-2²* resistance allele.

These examples display a promising potential to utilize the CRISPR/Cas9-based genome-editing to transfer or generate beneficial traits and expand variations in the elite cultivar of crop species. Recently, an interesting alternative has been proposed to this classic flow of plant breeding of "making the elite cultivar better" due to the versatility of the CRISPR/Cas9.

v. Alternative plant breeding: de novo domestication

Accessibility to multiplex editing by the CRISPR/Cas system has inspired an alternative plant breeding strategy: *de novo* domestication of wild species and minor or orphan crops. Instead of recuperating the useful traits such as various stress resistance and nutritional contents among the ancestral or minor accessions, this strategy proposes to accelerate the domestication of these accessions by modifying their domestication and improvement genes. The concept was proposed for the first time in 2017 and since then multiple examples of de novo domestication of wild species and orphan crop species have been demonstrated (Østerberg et al., 2017; Zhang et al., 2023). Two groups independently showed a multiplex CRISPR/Cas9 approach to target different sets of multiple domestication genes in *S. pimpinellifolium* and to improve traits like fruit size (**Figure i3-8**) (Zsögön et al., 2018; Li et al., 2018). In this work, Li et al. (2018) verified also that the salt stress and bacterial disease resistance traits were retained in the genome-edited *S. pimpinellifolium* plants.

Around the same time, rapid domestication of a *Solanaceae* crop 'groundcherry' (*Physalis pruinosa*) was demonstrated by genome-editing the domestication and improvement genes related to plant and fruit architecture (Lemmon et al., 2018). This work has also provided an experimental framework for *de novo* domestication strategy: the group has started with the whole-genome sequencing to obtain the genome assembly and then identified the orthologous genes of the major domestication and improvement genes to target by comparing with a major *Solanaceae* species, tomato, for which extensive data on phenotypes and genomics are available.

The possibility to manipulate the genes and QTLs with the CRISPR/Cas9 has implied two advantages in plant breeding: one is that the elite varieties can be directly improved based on the genetic information of the beneficial traits without going through intensive rounds of crossing and backcrossing and fearing undesirable linkage drag; and another is that the wild crop relatives or orphan crops can now be improved to develop new crops based on the knowledge from major crop relatives.

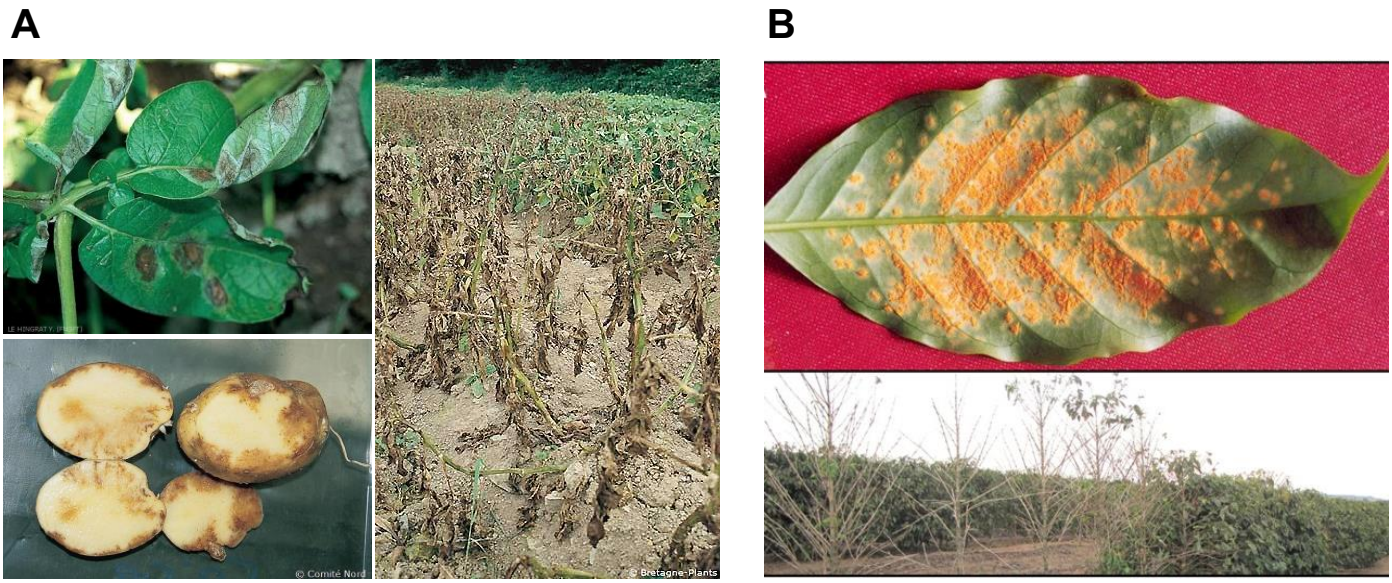


Figure i4-1 Symptom of pathogens that have triggered historical pandemic in crops. A Potato late blight caused by oomycete *Phytophthora infestans* in potato culture. Photos from database Ephytia.inrae.fr. **B** Coffee rust caused by a hemicyclic fungus *Hemileia vastatrix* on coffee culture. Photos from Talhinas et al., 2017.

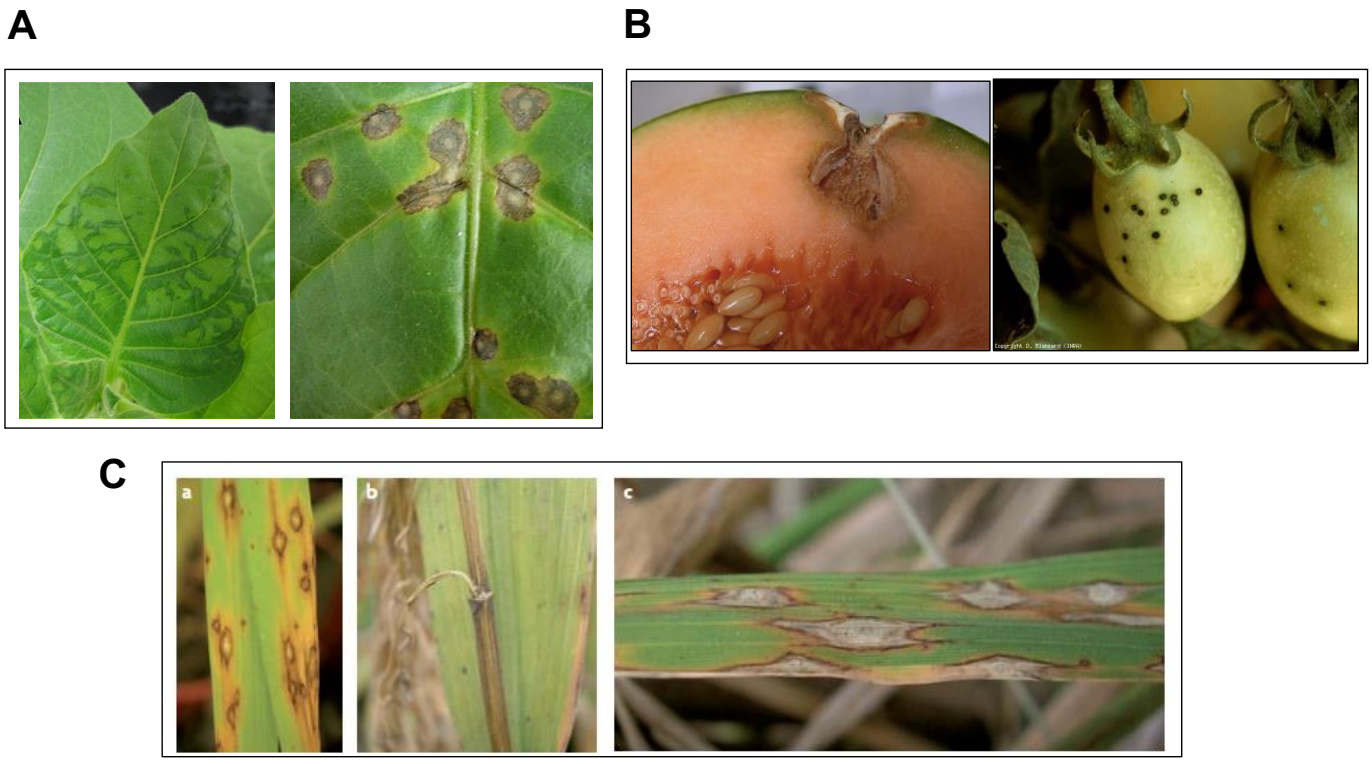


Figure i4-2 Disease symptoms of most important pathogens of virus, bacteria, and fungus in today's agriculture.
A Mosaic symptom (left) and necrotic lesion (right) as a result of hypersensititve reaction (HR) due to tobacco mosaic virus (TMV) infection on tobacco leaves. Photos from https://www.apsnet.org/edcenter/apsnetfeatures/Pages/TMV.aspx . **B** Lesions due to bacteria blight on melon fruit by *Psudomonas syringae* pv. *aptata* (left) and black superficial specks due to bacterial speck on tomato fruit by *P. syringae* pv. *tomato* (right). Photos from database Ephytia.inrae.fr. **C** From left, necrotic lesions with chlorotic margin (a), rot on neck and underdeveloped panicle (b), and large necrotic legions (c) due to rice blast by fungus *Magnaporthe oryzae*. Photos from Wilson and Talbot, 2006.

vi. Genome-editing for revamping the crop resistance to pathogen

Among different agronomic traits, biotic stress resistance has been of great importance to maintain the stable crop yield. Domestication and years of selection for improvement of cultivated crops have incurred the loss of genetic diversity and lack of sufficient resistance traits; consequently, increasing the vulnerability of the whole (mono)culture to pathogen invasion (Dempewolf et al., 2017; Venbrux et al., 2023). The efforts to re-introduce the pathogen resistance traits from ancestral species have been made (Zhang et al., 2017). However, the introgression of beneficial genes often resulted in introgression of non-beneficial or harmful genes due to the linkage drag, even after generations of backcrossing, as seen in the example of tomato commercial hybrid genome whose chromosome 9 is largely occupied by the introgressed virus resistance loci *Tm-2²* and linked loci (**Figure i3-9**) (e.g. Lin et al., 2014). In particular, many pathogen resistance traits, especially against viruses, are inherited in a recessive manner (Truniger and Aranda, 2009), which makes it difficult for breeding of highly heterozygous species like potatoes.

There is also the risk of resistance breaking by evolution of pathogens. Along with considerable time and cost demanded for repeated crossing and selection, the risk of linkage drag and of resistance breaking has hampered us from fully exploiting the biotic stress resistance genes available in the natural genetic diversity. Therefore, it is essential to establish a strategy to recreate natural resistance and eventually *de novo* resistance alleles that bypasses the downside of cross breeding without introducing a transgene in the plant genome. In part of this thesis, a proof of concept of such strategy was attempted using different genome-editing techniques in a model crop tomato with an aim to generate recessive potyvirus resistance. The following sections will cover the key concepts and significance of plant pathogen infection and the plant genetic resistance.

4. Plant pathogens and humans

In this thesis project, the trait of interest to be improved via genome-editing techniques is the resistance against one class of plant pathogens: viruses. In the following section, the overview of the plant pathogen infection on crops and the methods of prevention including use of genetic resistances will be provided.

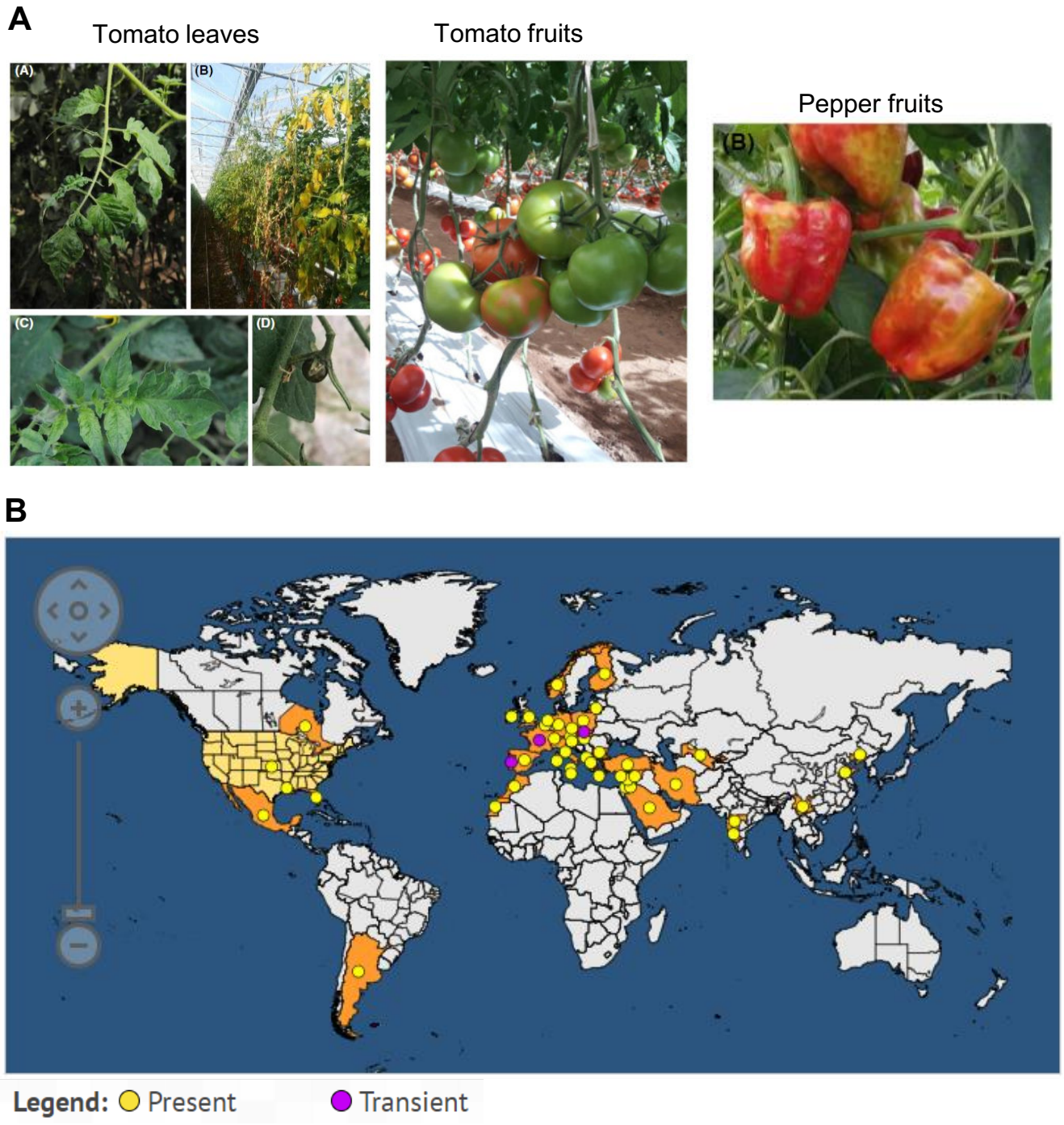


Figure i4-3 Tomato brown rugose fruit virus (ToBRFV) symptoms and affected countries.

A Symptoms of ToBRFV on tomato leaves (left) and fruits (middle) and on pepper fruits (right). Symptoms on leaf include narrowing, blistering of the surface; wilting, followed by yellowing and plant death; mosaic symptoms; and necrosis of the sepals on young tomato fruit. Tomato fruit shows marbling and delyaed ripening. Pepper fruit shows uneven color. Taken from Tomato brown rugose fruit virus (2020) **B** Global distribution of the ToBRFV infection. Yellow dots are the countris and regions where the presence of ToBRFV is observed; purple is transient. Taken from EPPO Global Database https://gd.eppo.int/taxon/TOBRFV/distribution ; 2024/06/10 update.

a. Impact of plant pathogen

Plant pathogens have changed several times the course of human history. The Irish population was in a grave peril in 1845 to 1849 largely due to the spread of potato late blight by the oomycete *Phytophthora infestans* (**Figure i4-1A**). About the same period, Ceylon (now Sri Lanka) was boasting its coffee production until the invasion of coffee rust fungus *Hemileia vastatrix* obliged the plantation owners to switch to tea production in late 1880s (Gail L. Schumann, APS; http://tinyurl.com/ymdvjm4m) (**Figure i4-1B**).

Today, such pathogens—fungi, fungal-like organisms, bacteria, nematodes, phytoplasmas, viruses, viroids—together with invertebrate pests and parasitic plants continue to threaten the crop productions, accounting for 20-40% of the global crop production loss annually (FAO, 2022). **Figure i4-2** shows symptoms on crop plants infected by top pathogens among viruses (tobacco mosaic virus, TMV), bacteria (*Pseudomonas syringae* pathovars), fungi (*Magnaporthe oryzae*) according to Venbrux et al. (2023). Moreover, they constantly hold the risk of triggering global pandemics due to the increased exchanges of people and items (Venbrux et al., 2023). Among such is tomato brown rugose fruit virus (ToBRFV) which emerged ten years ago. ToBRFV can infect 100% of plants in the field and cause 30-70% of tomato yield loss in the plants (Tomato brown rugose fruit virus, 2020). ToBRFV infection causes various symptoms on leaves like narrowing, blisters, mosaic, as well as wilting followed by yellowing and senescence, and on fruits like marbling and delayed ripening (**Figure i4-3A**). Its first outbreak was observed in Israel in 2014 and since then the virus is gradually spreading to different continents despite the phytosanitary measures taken upon importation and exportation of susceptible materials (**Figure i4-3B**). The number of crop loss may surge at any moment with evolution and emergence of powerful pathogens like ToBRFV and climate change that affect the life cycle of vectors and plant immunity (Desaint et al., 2021; Hamann et al., 2021; Zandalinas et al., 2021).

b. How do pathogens spread and how can we prevent?

Diseases of sessile organisms like plants spread due to the abiotic and biotic agents (**Figure i4-4**). Identifying the modes of transmission of different pathogens are essential in order to implement the correct prevention methods against them. Today, this is particularly important for the Integrated Pest Management (IPM), a new concept of disease and pest controls, encouraged for in the agriculture in order to promote sustainable agriculture. Part of the objective of the IPM is to reduce the use of chemicals

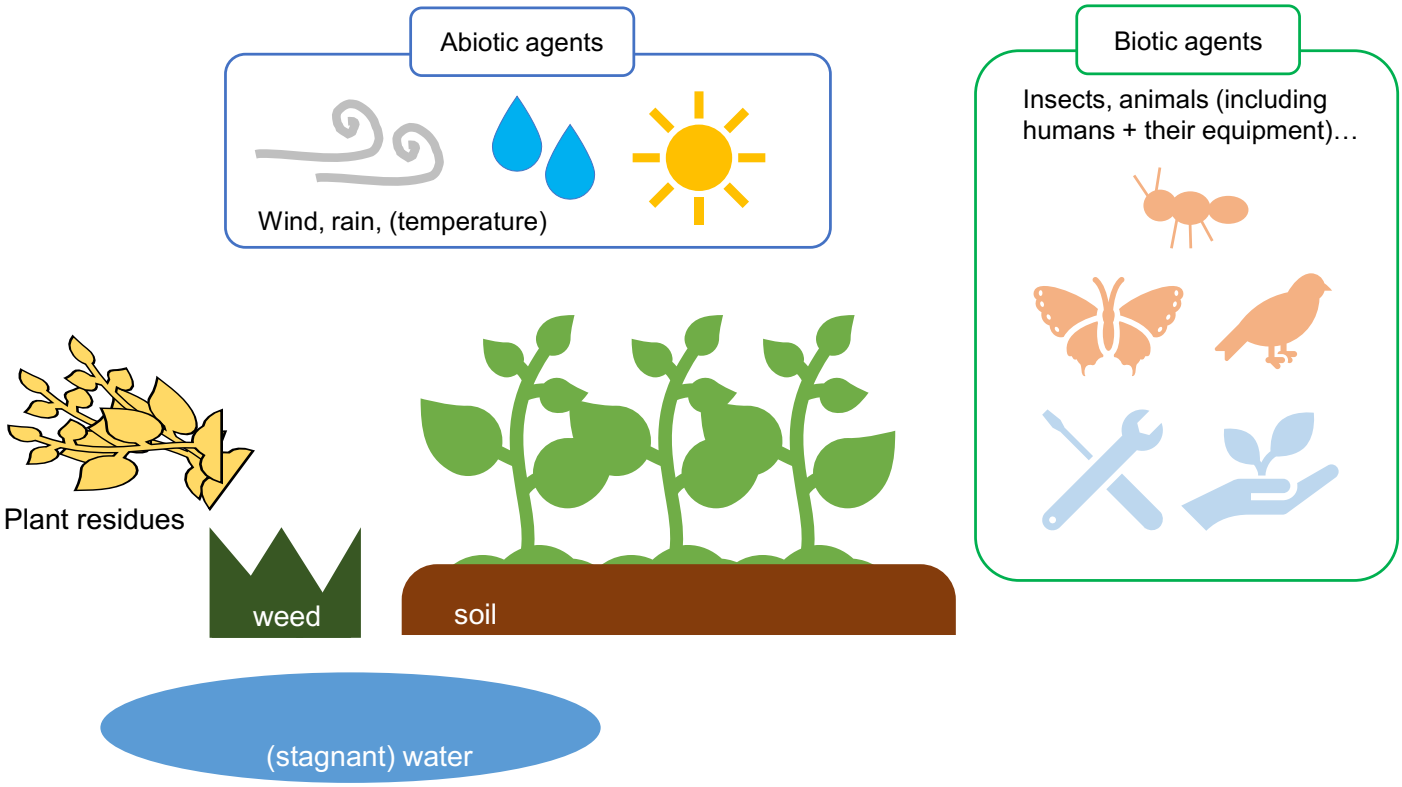


Figure i4-4 Abiotic and biotic agents and the environmenral reservoirs that mediate transmission and spread of diseases.

Abiotic agents like wind and rain carry the bacteria, fungi (spores), and oomycetes. Wind can also cause injuries on plants. Temperature and humidity affect the lifecycle of biotic agents that carry the pathogens and pathogen lifecycle themselves as well as the plant physiology and defense. Biotic agents include insects that are vector and non-vector species and animals. Human activity especially farming practices also play an important role in the spread of crop diseases. For convenience, the farming equipment are included in the biotic agent box. Environmental reservoirs are also shown in this diagram. These include weed species that can be host to diseases and participate in bi-directional spill over between the crop hosts. Water, soil, and plant or crop residues that were previously infected can remain as reservoirs of pathogens for culture.

such as pesticides, which can have huge negative health, environmental, and economic impact. The IPM takes into account the best combination of biological, chemical, physical, and crop-specific management practices that are adapted to each culture (**Figure i4-5**) (FAO website: https://www.fao.org/pest-and-pesticide-management/ipm/integrated-pest-management/en/) (For review, Khoury and Makkouk, 2010; Stenberg, 2017; Karlsson Green et al., 2020). These practices can take place both before and after the pathogen attack, including choice of suitable cultivars and pathogen-free seeds and tubers, disinfecting the soil and material, and use of appropriate chemicals and natural enemies to control pests, to name a few.

i. Preventing abiotic modes of transmission

Major abiotic agents that can spread the plant diseases are the wind and rain that mediate both the migration of fungal spores and bacteria and the accidental wounding of plants, which serve as the entry point for pathogens (**Figure i4-4**). They also promote pathogen entry into plants from the natural openings like stomata (stomata's role in disease reviewed in Xiang et al., 2021). This type of transmission can be prevented by physically protecting the crop plants from the wind and rain. It is also important to remove the reservoir of pathogens like (stagnant) water, soil, and neighboring weed hosts, which will provide the source of disease over a long time if unattended (Khoury and Makkouk, 2010; Nicaise, 2014; Nazarov et al., 2020).

ii. Preventing biotic modes of transmission

Biotic agents or animal—especially insects and mites—also spread diseases (**Figure i4-4**). Other animal activity including that of human can mediate the spread of crop diseases, for instance, by using contaminated equipment. Therefore, different phytosanitary measures that eliminate such vectors or protect crops from them are important. These include but not limited to the physical barriers and biological controls using natural enemies. Improving the cultural practices such as by disinfecting the equipment is also part of the measures. Combining these methods reduce the application of chemicals to prevent the transmission of the diseases.

However, even with these methods, it is difficult to completely eliminate the diseases that are rapidly transmitted by the biotic agents. For instance, potyviruses that cause economically important diseases (and are part of the thesis subject) are spread by aphids that acquire the viral particles in a matter of a few seconds to minutes (Moury et

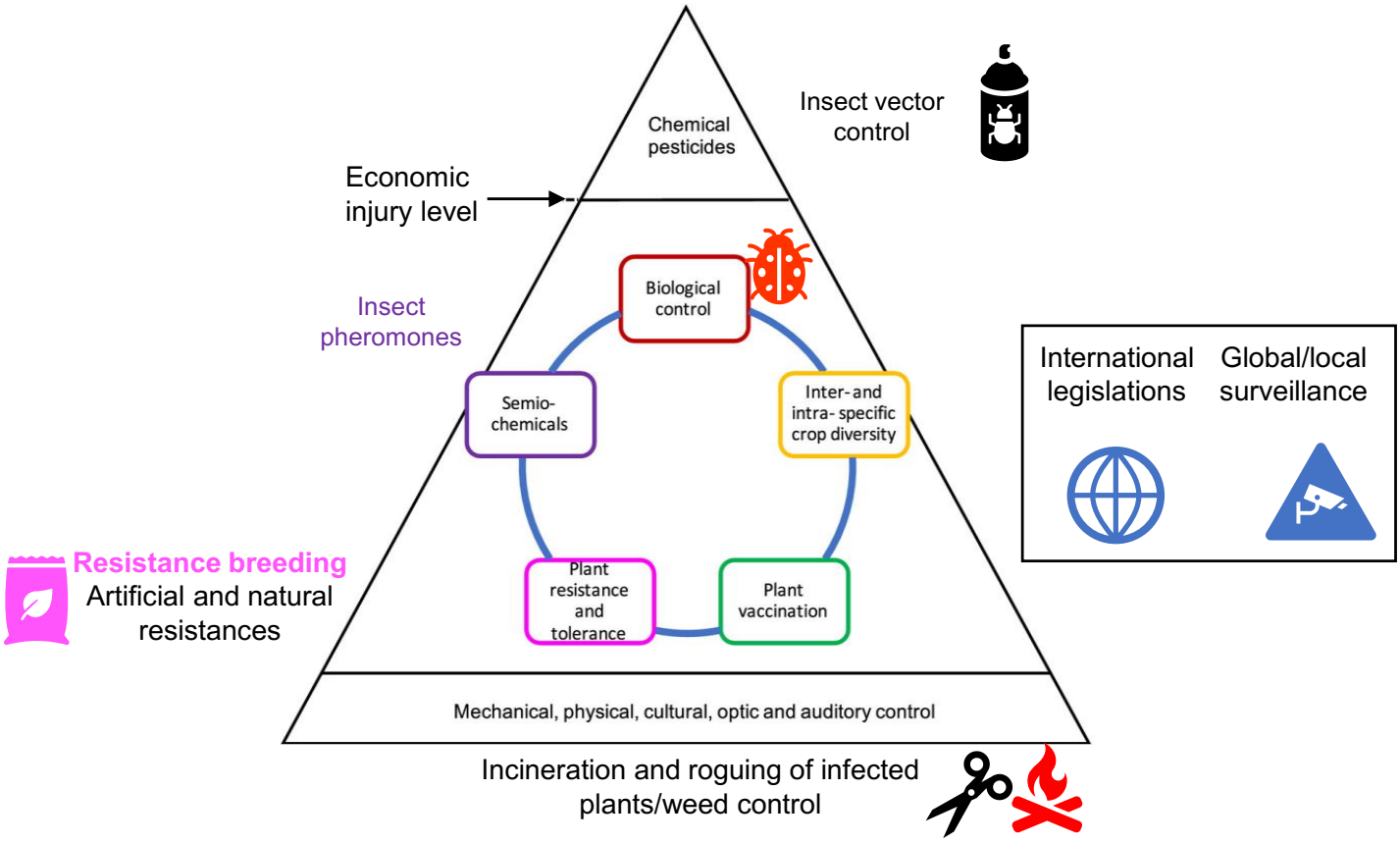


Figure i4-5 The Integrated Pest Management pyramid. Different control methods of IPM are displayed in the order of most to least sustainable from bottom to top. The base of the pyramid includes physical and cultural control. The middle part includes different ecological preventative and curative methods; resistance breeding is part of this tier. The top of pyramid is the chemical pesticide control that could be applied if the Economic Injury Level (EIL), or the smallest amount of injury that will cause yield loss equal to pest management costs (*sensu* Karlsson Green et al., 2020), has been reached. Figures adapted from Karlsson Green et al., 2020.

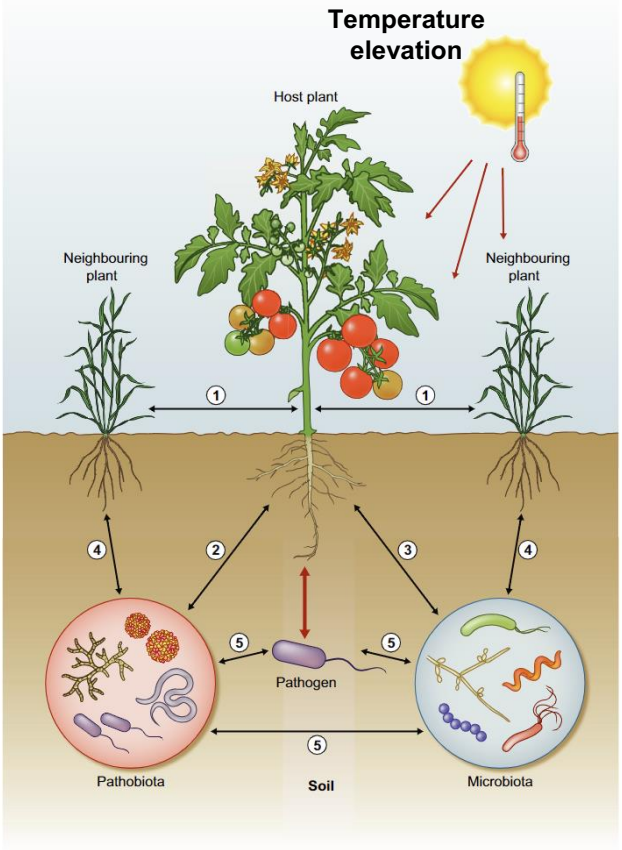


Figure i4-6 The relationships between plants and microorganisms surrounding the crops that are potentially impacted by temperature elevation.

Host plant (crop) of interest, tomato, is in the middle and the pathogen of interst just below. (1) Impact on the neighboring plants. (2) Impact on pathobiota (other pathogens in the environment than and the plant-pathobiota interaction) and (3) on microbiota (beneficial root microbiota) and their interaction with the host plant. (4) Impact on the interaction between the neighboring plant and pathobiota or microbiota. (5) Interactions between the pathogen of interst and the other microorganisms. Figure taken from Desaint et al., 2021.

al., 2007; Yang et al., 2021). The aphids stay on the plants only during the short time of probing[8] and feeding and fly to other plants including non-infected to repeat the same feeding activity, resulting in exchange of the viral particles (Perring et al., 1999; Ng and Falk, 2006; For review, Gadhave et al., 2020). Therefore, it is difficult to eliminate the aphids before virus transmission. For such virus-vector system, application of pesticides have proven unsuccessful in controlling these viruses once the infection occurs in a field (Perring et al., 1999; Gadhave et al., 2020).

iii. Protecting crops by plant breeding

Therefore, in the face of fast-spreading pathogens, the most important part of the IPM is to enhance the plant immunity against the pathogens so as to prevent the occurrence of diseases. Indeed, plant resistance breeding plays a role in the IPM (**Figure i4-5**). Cultivar selection oriented for conferring crops disease resistance has been done incidentally and now more deliberately with improvement in genomics and informatics studies combined with the mutagenesis technologies (Wallace et al., 2018; Gao, 2021). (cf. *Introduction 1.b*). The resulting elite cultivars of genetic resistance against diseases will be advantageous because these provide the basis for the IPM before the culture, allowing to predict and adjust the other IPM methods and reduce the environmental impact. Conversely, the other IPM methods play an important role in preventing overreliance on one or few selected genes to exert disease resistance in the culture, which can induce evolution of resistance breaking pathogens (**Figure i4-5**) (Rojas et al., 2018).

c. Emerging risks

In addition to the already existing pests and pathogens, the climate change and degrading agricultural environment pose a new risk today. The change in atmospheric composition, temperature, and landscapes impact the plant physiology and vector and herbivore lifestyle and biodiversity (Hamann et al., 2021; Zandalinas et al., 2021) (**Figure i4-6**).

Amidst the fluctuating abiotic and biotic parameters, it is progressively important to understand the disease infection mechanisms and how some natural varieties of plants have exerted resistance against pathogens in different environments. Such knowledge

[8] Shallow "sampling" of cell contents by aphid with its needle-like, mouthpart called stylet to see if the plant is a suitable host

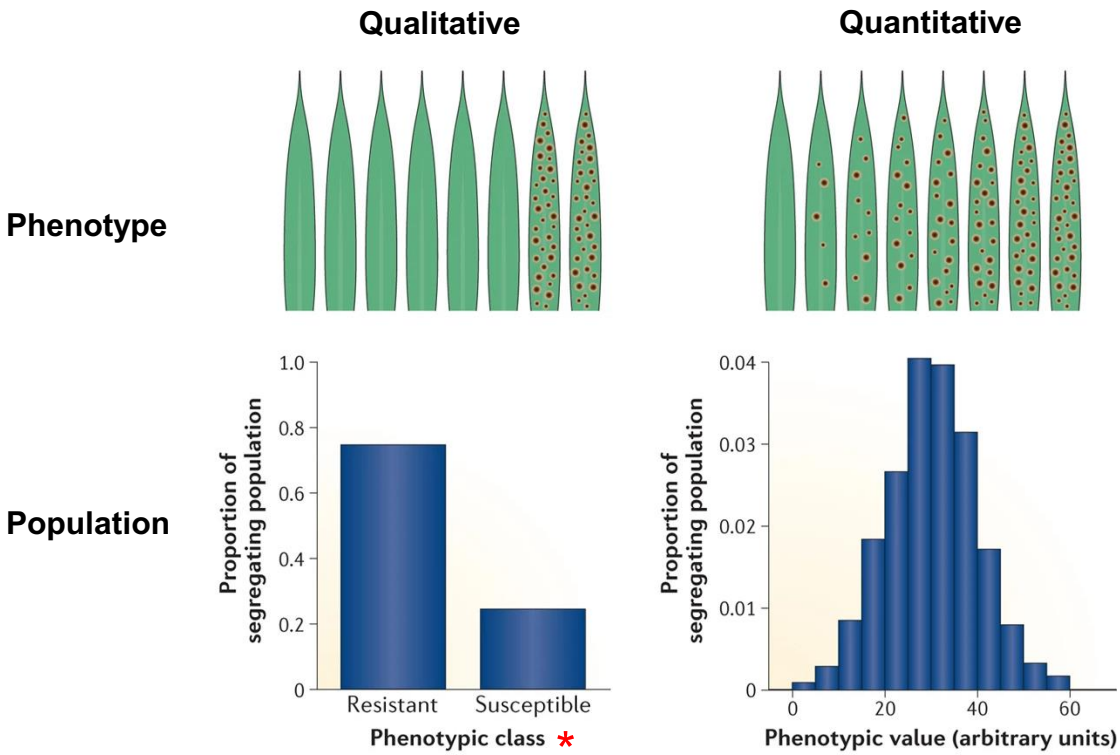


Figure i5-1 Schematics of qualitative and quantitative resistance traits.
Qualitative resistance traits show discrete phenotypes (resistant or susceptible) and segregation pattern in the population, while quantitative resistance traits show continuous phenotypes, i.e. various degrees of resistance and segregation pattern in the population. *In the case of a dominant *R* gene. For a recessive resistance allele, the ratio is reversed (R:S=1:3). Figure taken from Nelson et al., 2018.

will provide hints for generating durable genetic resistance via precise genome-editing and thus could participate to a sustainable IPM-based agriculture.

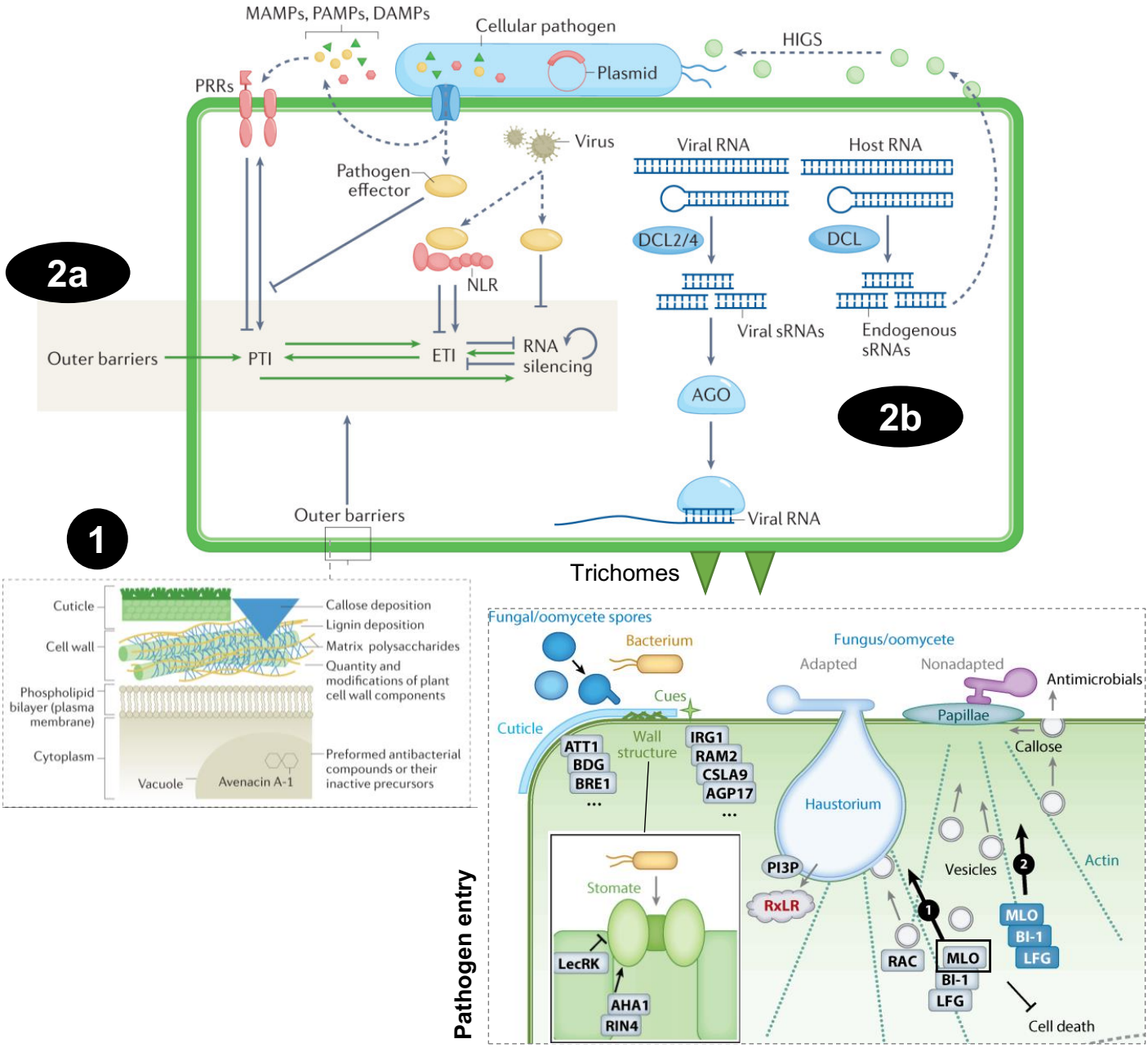


Figure i5-2 First "structural" and second "immune response" layers of plant defense

(1) Plant surface structures compose the first layer of defense against external pathogens. These include trichomes, cuticle, and cell wall. (2a) PTI-ETI immune response, mainly against non-viral pathogens, based on the pathogen recognition. Both PTI and ETI exchange reciprocal enhancement signals between them (2b) RNA silencing system against vial RNA. Here, the overview of the two layers of defense is shown to provide a big picture. Individual systems will be explained in details in Figure i5-3 and i5-4. Pattern recognition receptors (PRRs), pathogen-associated molecular patterns (PAMPs), microorganism-associated molecular patterns (MAMPs), damage-associated molecular patterns (DAMPs), PAMP-triggered immunity (PTI), NOD-like receptors (NLRs), effector-triggered immunity (ETI), small RNAs (sRNAs), host-induced gene silencing (HIGS). Figure adapted from Lopez-Gomollon and Baulcombe, 2022.

Bottom right corner is the close up view of the host defense structures and the structure-related factors associated with plant susceptibility. The suscpetiblity factor MLO supports the formation of haustoria of pathogen that absorb nutrient from the cell. Other details are available in Van Schie and Takken, 2014. Figure adapted from Van Schie and Takken, 2014.

5. Plant pathogen resistance

Plants have numerous layers of structural barriers and innate immune systems to prevent or counter the invasion of pathogens. Resistance breeding is based on the plant defense systems, and its goal is to boost the pathogen resistance of crop cultivars of interest by genetic resistance. This makes an important part of the IPM (cf. *Introduction 4.b*). The level and spectrum of pathogen resistance of plant accessions are largely determined by the genotypes of a given accession. Indeed, the natural variations and artificially induced mutants that confer stronger resistance to pathogens are the great source for resistance breeding.

Resistance trait conferred by a gene can be either qualitative or quantitative, but the combination of the resistance genes of different effect size is known to affect the durability of the major resistance genes (**Figure i5-1**) (Quenouille et al., 2013; Nelson et al., 2018). Also, resistance can be inherited by either dominant or recessive manner; for instance, resistance (*R*) gene-mediated resistance that is involved in pathogen recognition is often dominant, while resistance based on loss of function of susceptibility (*S*) gene, which inherently supports infection, is recessive.

In this section, different layers of plant immunity which can be improved by genetic resistance are overviewed in a chronological order of pathogen infection. Since this thesis is centered around development of genetic resistance against plant viruses, a special focus will be given to the antiviral defense system.

a. First layer of defense: host-pathogen compatibility test for entrance

As a first layer of defense, the plant surface structures like cuticle and trichome as well as the cell wall block the entry of the pathogens which are not capable of entering the plants unless there are injury or natural openings like stomata (Xiang et al. 2021) (**Figure i5-2**). Some pathogens have the ability to adhere to the surface and penetrate inside their compatible host, for example, via formation of appressorium which physically punctures the cell wall by high pressure accumulated inside in the case of fungus and oomycetes (Meng et al. 2009). In other words, genetic mutations that render the surface structure or the cell wall composition unsuitable for pathogens can disrupt the host-pathogen compatibility and prevent infection (Hansjakob et al., 2011; Uppalapati et al., 2012). A prominent example is *mildew resistance locus O* (*MLO*) gene which codes for

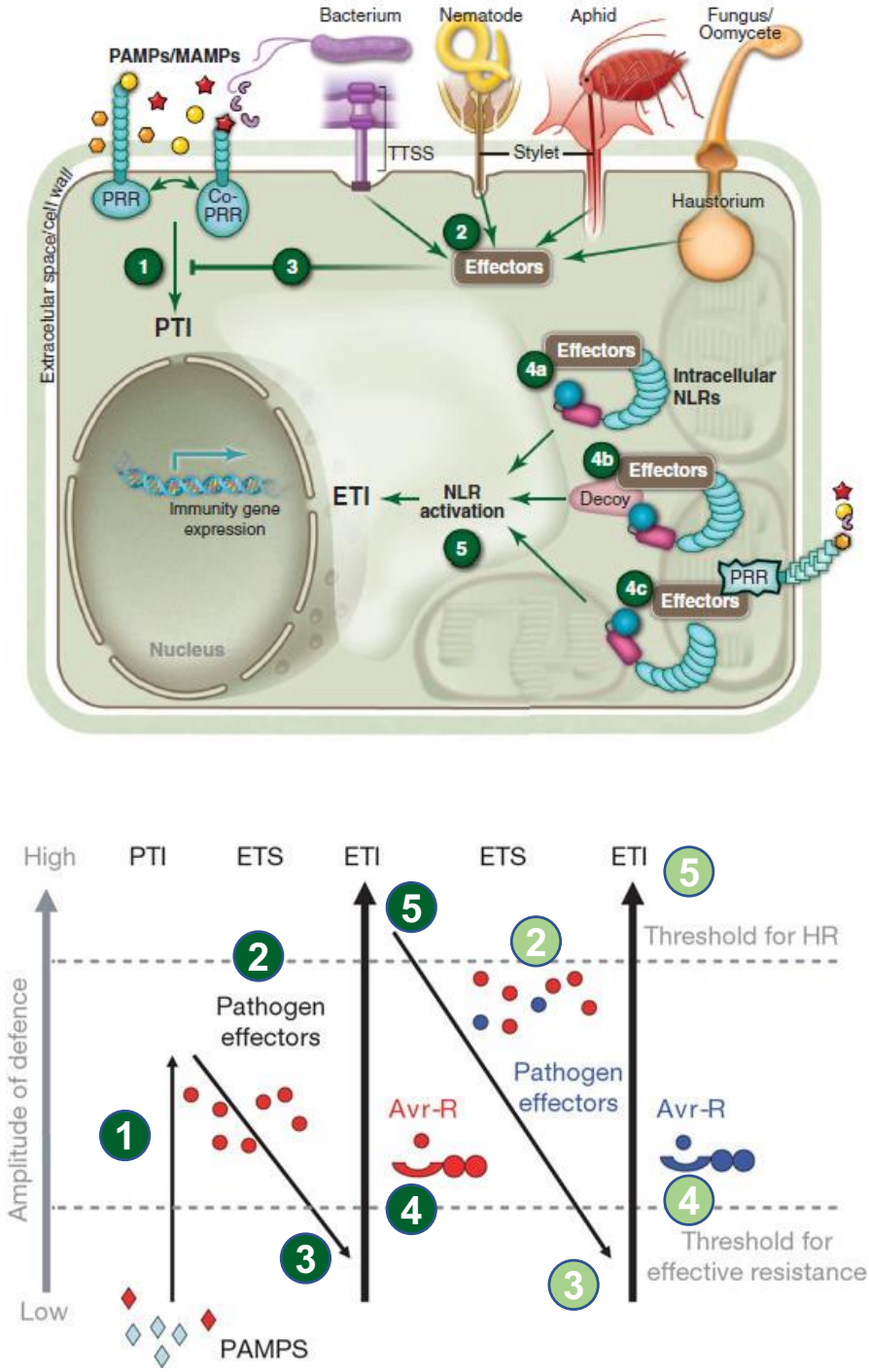


Figure i5-3 Plant immune system based on pathogen recognition: PTI-ETI and the zig-zag-zig model.

Pattern-triggered immunity (PTI) and effector-triggered immunity (ETI) are triggered upon the pathogen invasion and pest attack. Plants perceive the pathogen-/microbe-associated molecular patterns (PAMP/MAMP) by the pathogen recognition receptor (PRR) and this triggers the PTI (1) (DAMP/HAMP are not depicted for simplicity.). In response to host PTI, pathogens secrete effectors (2) and this inhibits the PTI (3). The effectors are directly or indirectly recognized by nucleotide-binding leucine-rich repeat receptors (NLRs) (4a-c). (Detailed explaination on the effector recognition is available in Dangl et al. (2013). The effector recognition triggers the ETI. According to the zig-zag-zig model, the ETI can be suppressed by newly secreted pathogen effectors (2 light green) and this follows the same ETI induction (light green numbers). Figures taken and adapted from Dangl et al., 2013 and Jones and Dangl, 2006.

transmembrane protein and is necessary for penetration of the powdery mildew into the epidermal cells, thereby defined as an *S* gene to this pathogen (**Figure i5-2**). *MLO* is also known to suppress programmed cell death (Wolter et al., 1993; Büschges et al., 1997). Inactivation of *S* gene, *MLO,* provides resistance against powdery mildew in a recessive manner and such *mlo*-based resistance has been used in barley since 1940s (Büschges et al., 1997; Van Schie and Takken, 2014). In the case of virus, unless there is an injury on surface, the entry into the pathogen takes place primarily during the feeding of vectors on the plants; the vector insect penetrates the stylet directly into the phloem for feeding, creating the entry points for the virions (Gadhave et al., 2020).

b. Second layer of defense A: major immune response for non-viral pathogen

The second phase of defense is the immune response to the pathogen that has entered inside of the plant. When the pathogen penetrates into the plants, this entry is recognized by the host which triggers various physiological reactions in an attempt to eliminate the pathogen. As the virus infection does not involve the "penetration" phase, the antiviral defense follows a slightly different path. For this reason, the antiviral immune response will be discussed separately later.

i. Pathogen recognition-based immunity

Non-viral pathogens, i.e. bacteria, fungi, and oomycetes, are often countered by the two or more layers of host defense system based on pathogen recognition: pattern-triggered immunity (PTI) and effector-triggered immunity (ETI) (Many reviews on PTI-ETI are avaialble; e.g., Jones and Dangl, 2006; Ngou et al., 2022) (**Figure i5-3A**).

Briefly, the PTI is triggered upon the recognition of the conserved pathogen-/damage-/microbe-/herbivore-associated molecular patterns (PAMP/DAMP/MAMP/ HAMP) by the host pattern recognition receptor (PRR). This provides a general resistance to many pathogens that are recognized by the host PRR (**Figure i5-3A: 1**). Some pathogens have evolved to evade or suppress the PTI by secreting molecules called effectors (**Figure i5-3A: 2-3**). The effectors trigger the ETI when they are directly or indirectly recognized by the intracellular nucleotide-binding leucine-rich repeat receptors (NLRs) that are coded by *R* genes, with some exceptions of apoplastic effectors (Ngou et al., 2022) (**Figure i5-3A: 4-5**). The PTI and ETI both induce accumulation of reactive oxygen species (ROS) and defense-related hormones like salicylic acid and

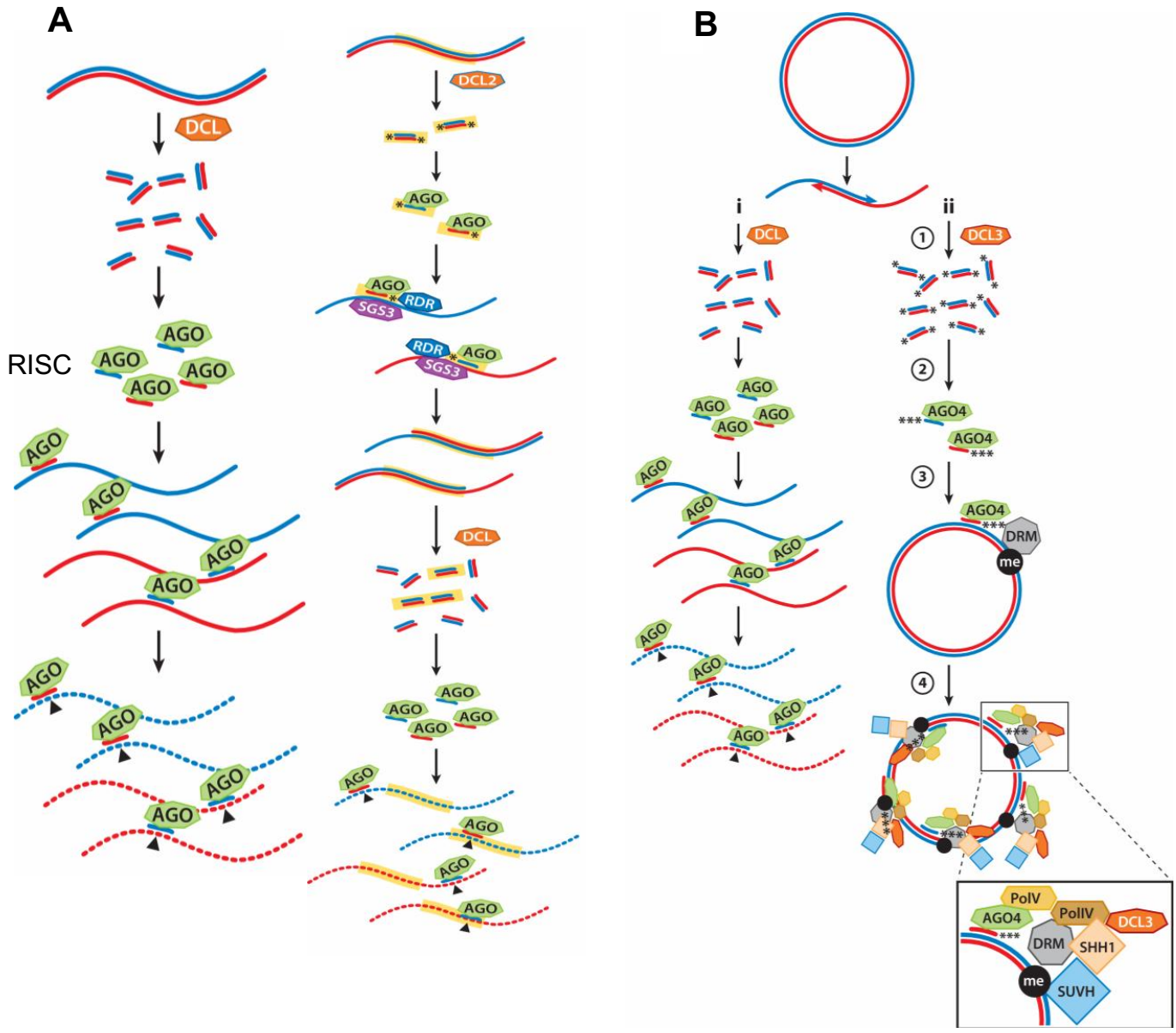


Figure i5-4 Antiviral RNA silencing targets RNA virus and DNA virus.

A (left) RNA silencing process against RNA virus. Virus-derived dsRNA is processed in to 21-nt siRNAs by the dicer-like (DCL) nucleases. One of the two siRNA strands binds to the argonaute (AGO) nucleases to make a RNA-induced silencing complex (RISC) and guides the RISC to their complementary target RNAs (viral genome) for cleavage. The cleaved RNAs are often degraded by non-sequence-specific nucleases in the plant cell. (right) Amplification of siRNA. Virus-derived dsRNA is processed specifically by DCL2 which generates 22-nt siRNAs (indicated by *). When 22-nt siRNA in complex with AGO targets viral RNA, this RISC binds suppressor of gene silencing (SGS)3 which then recruits RNA-dependent RNA polymerase (RDR) 6. RDR6 begins the dsRNA synthesis using the viral RNA as a template. The generated dsRNA is processed by the DCLs like in the initial siRNA synthesis pathway (left). **B** (i) RNA silencing against DNA virus and (ii) RNA-directed DNA methylation (RdDM) of viral genomic DNA. The transcripts of viral genome form a partial dsRNA structure, this is targeted by host DCLs for siRNA production. (i) RNA silencing of DNA virus follows a similar pathway to the A: RNA silencing of RNA virus. In (ii), 24-nt siRNA is generated by DCL3 and this is loaded to AGO4. The mechanism is not fully elucidated, but it is considered to involve the recruitment of domains rearranged methyl transferase (DRM), which initiates methylation (me) and other factors involved in RdDM. Figures taken from Baulcombe, 2022.

jasmonic acid and expression of defense-related genes. The successful ETI leads to the programmed cell death at the infection site, a phenomenon known as hypersensitive reaction (HR).

Each *R*-gene-coded NLR is more specifically evolved to recognize specific effector or the host target effector than the PRR; more than 140 NLRs are characterized compared to more than 60 of PRRs (Ngou et al., 2022). The host resistance by the ETI can be broken down when the pathogens evolve and lose/obtain/diversify effectors to evade or suppress such responses. Plants can in turn acquire a new *R* gene that recognizes the evolved pathogen effectors. This back-and-forth interaction of the plant and pathogen is known as "zig-zag-zig model", reflecting the co-evolution of the two entities through interaction (**Figure i5-3B**) (Jones and Dangl 2006). The arms race between plant and pathogens has given rise to a large number of NLRs, implying the difficulty of establishing durable resistance to evolving pathogens.

ii. PTI-ETI for plant breeding

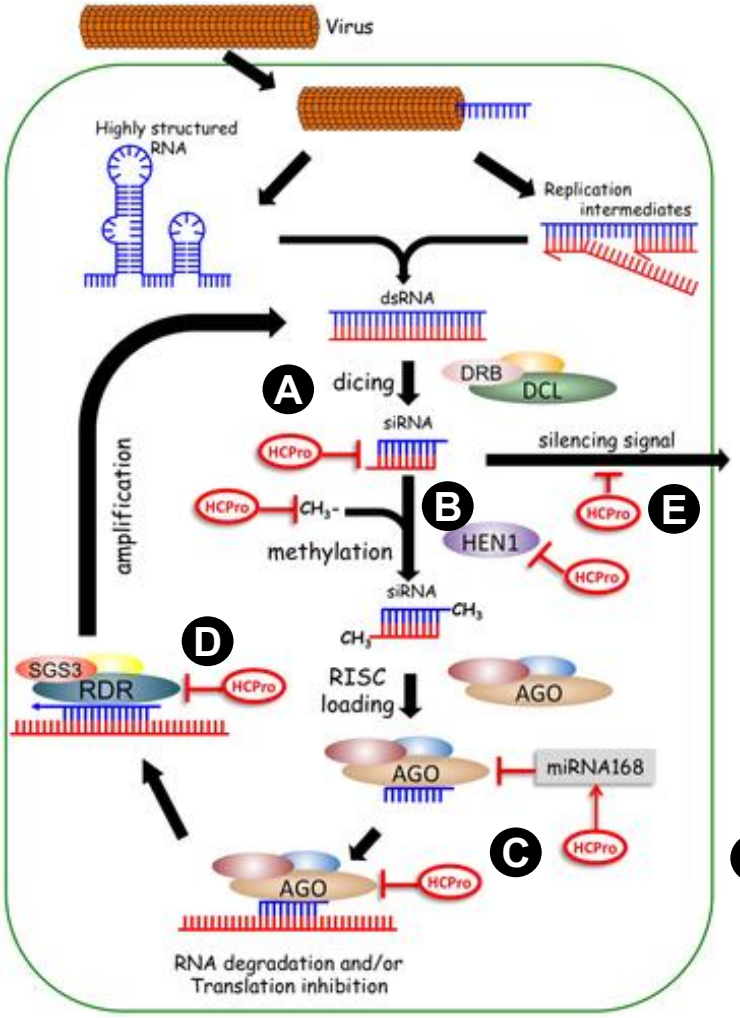
In the aspect of resistance breeding, engineering of both the PTI and the ETI have been attempted using transgenic technology. Heterologous expression of the PRR from one species to has proven successful in many crop species, while that of the NLR has proven relatively difficult due to its specificity and complex mode of function and requires sophisticated modification adapted to the host (Rodriguez-Moreno et al., 2017; For review, Boutrot and Zipfel, 2017; Zdrzałek et al., 2023). One example of the CRISPR/Cas9-based engineering of chimeric NLR by inducing chromosomal rearrangement has been demonstrated in soybean (Nagy et al., 2021).

c. Second layer of defense B: immune response for viruses

i. Classic PTI-ETI and virus

Viruses are intracellular pathogens whose entry is made directly into the cells by vectors or through openings. Therefore, the classic PTI in the context of virus infection is rarely discussed, and indeed, no PRRs are confirmed to be the receptor of viral particles to date (Ngou et al., 2022).

On the other hand, NLR recognition has been reported for multiple viruses. For example, the *N* gene isolated from *N. glutinosa* is a Toll/Interleukin-1 receptor/Resistance (TIR) protein NLR that recognizes the helicase domain of tobacco mosaic virus (TMV)



	Targeted step	Molecular mechanism	Potyvirus
A	vsiRNA uploading	Sequestration of vsiRNAs	TEV, PPV, PRSV, ZYMV, TuMV
B	vsiRNA methylation	Inhibition of production	PVY, PVA
		Binding and inactivation of HEN1	ZYMV
C	Effector	Down-regulation of AGO1	TEV
		Interaction with AGO1	PVA
D	Amplification	Down-regulation of RDR6	SCMV
E	Movement of silencing signal	Sequestration of siRNAs?	PVY, TEV
F	Induction of endogenous silencing suppressors?	Interaction with rgs-CaM and RAV2	TEV, TuMV

Figure i5-5 Potyviral suppressor of RNA silencing HC-Pro interferes at different steps of antiviral RNA silencing.
Steps of simplified RNA silencing pathway potentially targeted by helper component proteinase (HCPro) of potyvirus are displayed and listed below. Simplified schematic representation of the RNA silencing-mediated defences in plants that are deployed against RNA viruses. DRB: double-stranded RNA-binding protein, HEN1: HUA enhancer 1. Figure and table modified from Valli et al., 2018.

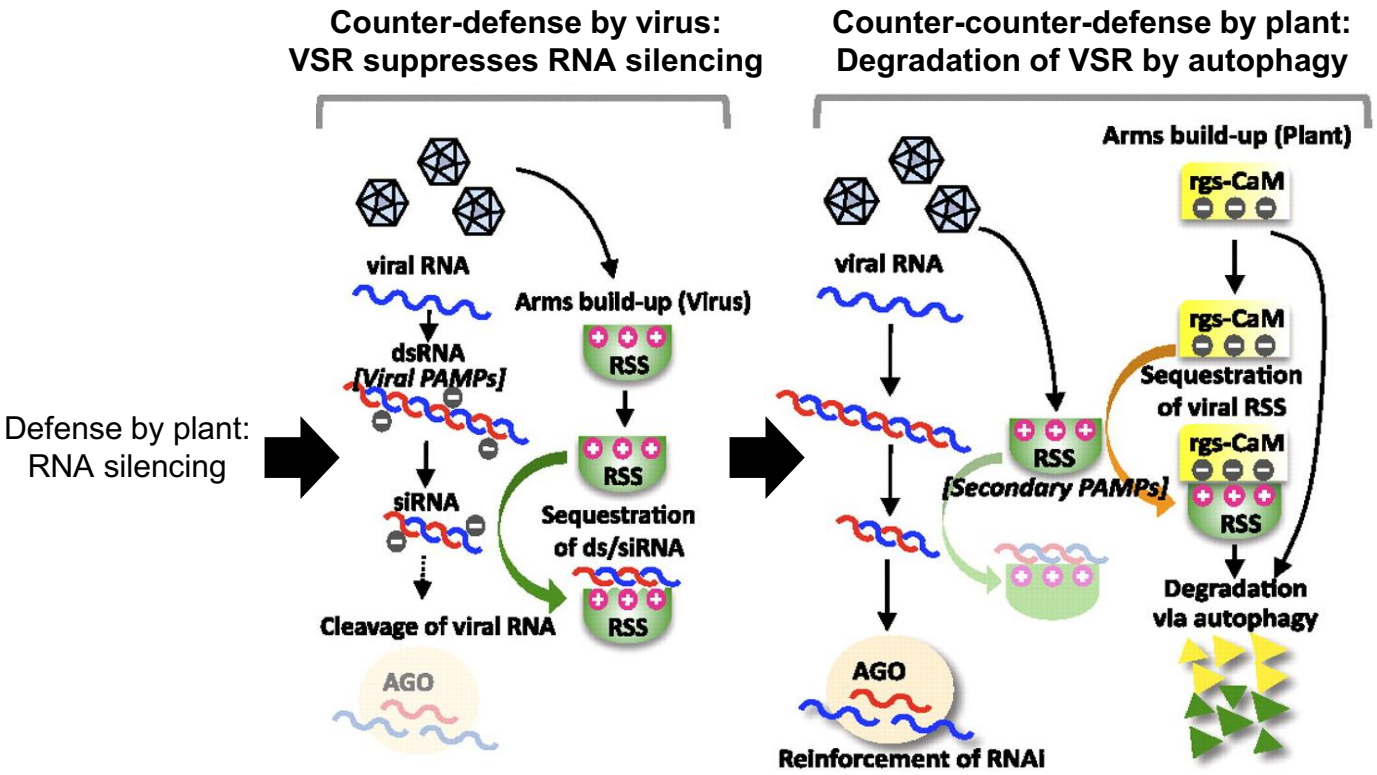
replicase p50 and induces necrotic lesions in leaves (Whitham et al., 1994; Erickson et al., 1999). A potyvirus resistance gene *Pvr4* isolated in *C. annuum* and a tomato spotted wilt virus (TSWV) resistance gene *Tsw* isolated in *C. chinense* are both coiled-coil (CC) NLRs that recognize potyviral RNA-dependent RNA Polymerase (NIb) and TSWV RNA silencing suppressor NS, respectively (de Ronde et al., 2013; Kim et al., 2015). Despite their independent roles in recognition of viral proteins, these two genes were found to be the diverged from a single ancestral gene in *Capsicum* spp. through tandem duplication and neofunctionalization (Kim et al., 2017).

ii. Antiviral RNA silencing

In plants, RNA silencing is an important defense system against viruses though non-viral pathogens also trigger host RNA silencing (Pumplin and Voinnet, 2013). Antiviral RNA silencing is triggered upon the recognition of double-stranded RNA (dsRNA) structure, which is formed as the viral replication intermediate or secondary RNA structure formed from single-stranded RNA (ssRNA) or transcript of DNA virus. Thus, this system generally targets the DNA and RNA viruses (Hohn and Vazquez, 2011; For review, Wang et al., 2012; Pumplin and Voinnet, 2013; Lopez-Gomollon and Baulcombe, 2022).

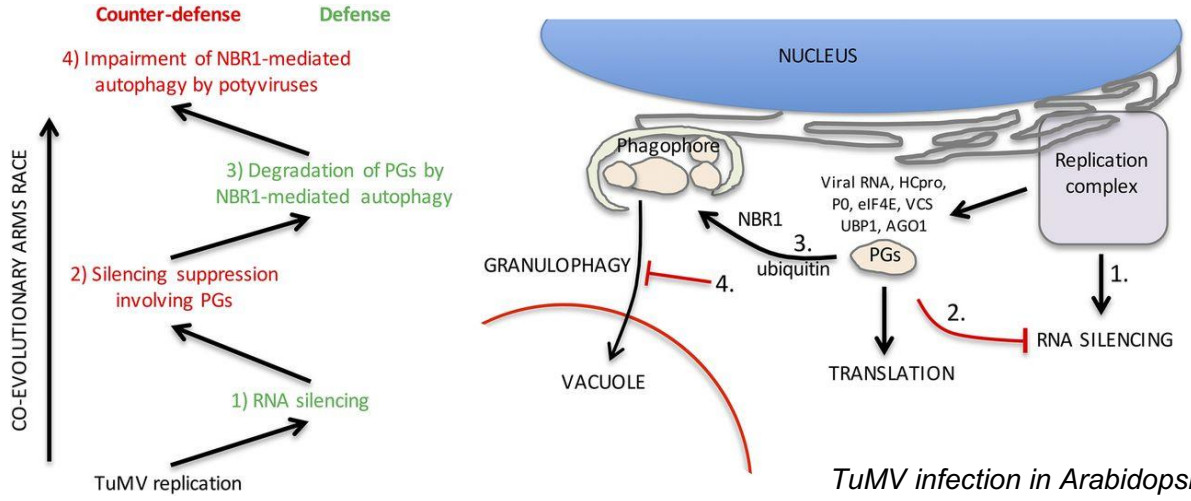
iii. Defense by RNA silencing

Briefly, the virus-derived dsRNA structures are recognized and processed by the host dicer-like proteins (DCLs) into 21-nt to 24-nt small RNAs called small interfering RNA (siRNA) depending on different DCLs (**Figure i5-4**). Then, the 21-nt siRNAs are loaded onto the Argonaute protein (AGO) to form the RNA-induced silencing complex (RISC), which then cleaves viral RNA that is complementary to the loaded siRNA. Meanwhile, the RISC with 22-nt siRNA bound to the target viral genomic RNA binds to the suppressor of gene silencing (SGS) 3; this recruits RNA-dependent RNA polymerase (RDR) 6 and triggers dsRNA, thus amplifying the siRNAs and enhancing the RNA silencing (Reviewed in Pallas and García, 2011; Lopez-Gomollon and Baulcombe, 2022) (**Figure i5-4A**). In the case of DNA viruses, their genome is also targeted for DNA methylation to suppress the virus by the RISC containing 24-nt siRNA with AGO4 (Raja et al., 2008) (**Figure i5-4B.ii**).



CMV-Y infection in tobacco

Counter-counter-counter-defense by virus: Impairement of VSR autophagy



TuMV infection in Arabidopsis

Figure i5-6 Zig-zag-zig model for plant-virus cross talk via RNA silencing machinery.
(Top) From left, Plant's defense antiviral RNA silencing is suppressed by the viral suppressor of silencing (here depicted as RNA-silencing suppressor (RSS)), which can sequester virus-derived dsRNA or siRNA, for example. It was shown in tobacco that during CMV-Y infection, the plant can inhibit such suppressor activity by a calmodulin-like protein rgs-CaM: rgs-CaM binds to the dsRNA/siRNA-binding surface of the VSR of CMV-Y and together targeted for degradation by autophagy. Figure taken from Nakahara et al., 2012. (Bottom) A hypothetical model of autophagy involvement in potyvirus-plant cross talk with TuMV as an example. Plant defense by RNA silencing is countered by potyvirus-induced RNA granules that protect viral RNA from being targeted by RNA silencing. The granules also contain potyviral VSR, HC-Pro. These granules are then targeted for degradation altogether by autophagy. This degradation is suggested be countered by the viral proteins by an unknown mechanism. Figure taken from Hafrén et al., 2018.

Counter-defense by viral suppressor of RNA silencing

To counter the host RNA silencing mechanism, viruses produce viral suppressor of RNA silencing (VSR) and attempt to inhibit different steps of siRNA biogenesis and subsequent RNA silencing. For example, a potyviral VSR, helper component proteinase (HC-Pro) from different potyvirus species is known to interfere at multiple steps by sequestering the virus-derived siRNAs (vsiRNA) from AGO loading, inhibiting HEN1 which stabilizes the vsiRNAs by methylation, downregulating silencing actors like AGO1 and RDR6, sequestering siRNAs from long-distance movement, and promoting endogenous silencing suppressor (**Figure i5-5**) (For review, Pallas and García, 2011; Pumplin and Voinnet, 2013; Valli et al., 2018). The viral genome-liked protein (VPg) is also a potyviral VSR and promotes degradation of SGS3 via ubiquitination and autophagy (Cheng and Wang, 2017).

VSR activity efficiently counteracts host's RNA silencing by inhibiting host machineries that play a dual role in the antiviral mechanism and in the endogenous sRNA-mediated gene regulation. The latter, the endogenous sRNA-mediated gene regulation, is important for normal plant development; therefore, the VSR interference indirectly affects the plant's growth and morphology, resulting in the viral symptom. (Reviewed in Pallas and García, 2011; Wang et al., 2012).

Virus also benefits from the RNA silencing

Ironically, RNA silencing is an antiviral mechanism that can also promote the survival and propagation of viral agents through symptom induction in the nature. It is due to the generated vsiRNA which has fortuitous complementarity to the host gene sequence (Wang et al., 2012; Vaucheret and Voinnet, 2024). For example, co-infection of cucumber mosaic virus (CMV) and Y-satellite RNA[9] on *N. tabacum* induces strong leaf yellowing symptom as a result of RNA silencing on the Y-satellite RNA sequence, which has a complementary sequence to a part of the chlorophyll biosynthesis-related gene *ChlI* (**Figure i5-Box[10]A**) (For review, Kang et al., 2005; Truniger and Aranda, 2009; Wang and Krishnaswamy, 2012). This yellowing increases the attractivity to the vector aphids and furthermore this Y-satellite RNA promotes wing formation of aphids, consequently supporting the proliferation of CMV and satellite RNA in the ecosystem (**Figure i5-BoxB**)

[9] Satellite RNA: subviral agent that depends on the virus (helper virus) for its proliferation through co-infection of a host cell; often attenuates the symptom of helper virus by competing for replication machineries.

[10] Figure i5-Box is located on the left of page 37.

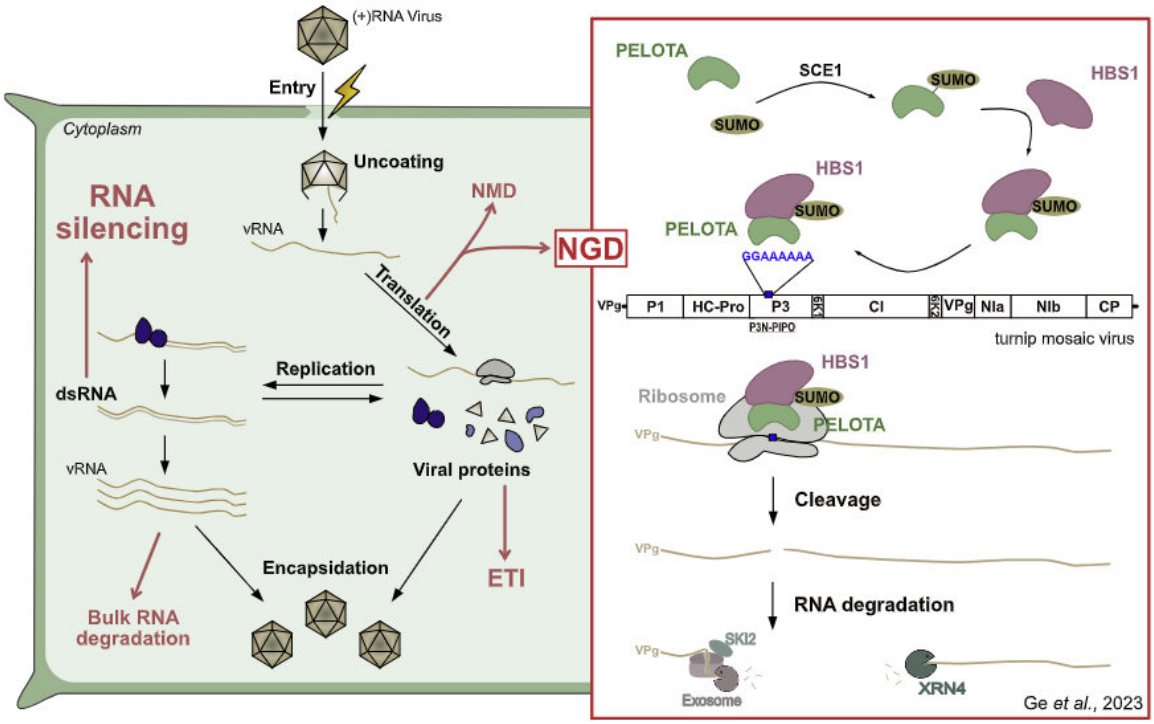


Figure i5-7 mRNA decay as an antiviral mechanism in plants.

Viral RNA (vRNA) is targeted by various host defense mechanisms (in red) including the classic RNA silencing induced by virus-derived dsRNA and the ETI induced by the viral protein as an effector. vRNA under translation can be targeted for non-sense mediated decay (NMD) or no-go decay (NGD) depending on the structure of the viral genomic RNA. For instance, potyviral RNA was shown to be targeted by NGD due to a $G_{1\text{-}2}A_{6\text{-}7}$ motif (in blue). This motif causes ribosome stalling which then recruites RNA quality control machineries such as HBS1 and SUMOylated PELOTA, resulting in RNA degradation HBS1: Hsp70 subfamily B suppressor1. Figure taken from Pouclet et al., 2023.

(Jayasinghe et al., 2021). Therefore, viruses have not only evolved VSRs to counter the antiviral RNA silencing but also utilize the silencing itself for its efficient proliferation.

Counter-counter-…defense

Similarly to the PTI-ETI-based zig-zag-zig model, the plant and virus are in the co-evolutionary arms race to outsmart each other. Indeed, it is observed that the host plant targets the VSRs of different virus families for degradation via autophagy, thereby counter-counter-defending against the virus (**Figure i5-6A**) (Nakahara et al., 2012; Haxim et al., 2017; Hafrén et al., 2018). However, such VSR-targeting degradation is suggested to be inhibited by viral proteins via unknown mechanism (Hafrén et al., 2018) (**Figure i5-6B**).

Despite the viral counter defenses, the RNA silencing is still effective in preventing virus infection and has inspired the RNA-based vaccination for plants in agriculture; a dsRNA containing viral sequence is transgenically expressed or exogenously applied to induce the host RNA silencing (For review, Voloudakis et al., 2022).

iv. RNA decay

In addition to RNA silencing, RNA decay is increasingly attracting attention as another antiviral mechanism. RNA decay is a part of RNA quality control (RQC) that degrades aberrant mRNAs detected during translation (For review, Moon and Wilusz, 2013; Mäkinen et al., 2017; Li and Wang, 2018). Distinct patterns of aberrant mRNAs are degraded via different pathways: mRNA with premature termination codons (PTC) is degraded via nonsense-mediated decay (NMD), mRNA without stop codon via non-stop decay (NSD), and mRNA that causes ribosome stalling[11] via no-go decay (NGD); of which, NMD and NGD are associated with surveillance of plant viral RNA (**Figure i5-7**) (Pouclet et al., 2023).

Recent works suggest that distinct viral genome structures prompt different RNA decay pathways. For instance, potato virus X (PVX) genomic RNA is targeted by NMD due to the several internal termination codons (iTCs) that are recognized in a similar manner to the plant endogenous PTCs (Garcia et al., 2014). On the other hand, a potyvirus, TuMV genomic RNA is not targeted by NMD because it has a long ORF without iTCs. However,

[11] Pausing of ribosomes at a specific position of mRNA due to consecutive adenines or other non-optimal codons for example. Ribosome stalling is perceived as a signal of abnormal translation, hence triggering degradation.

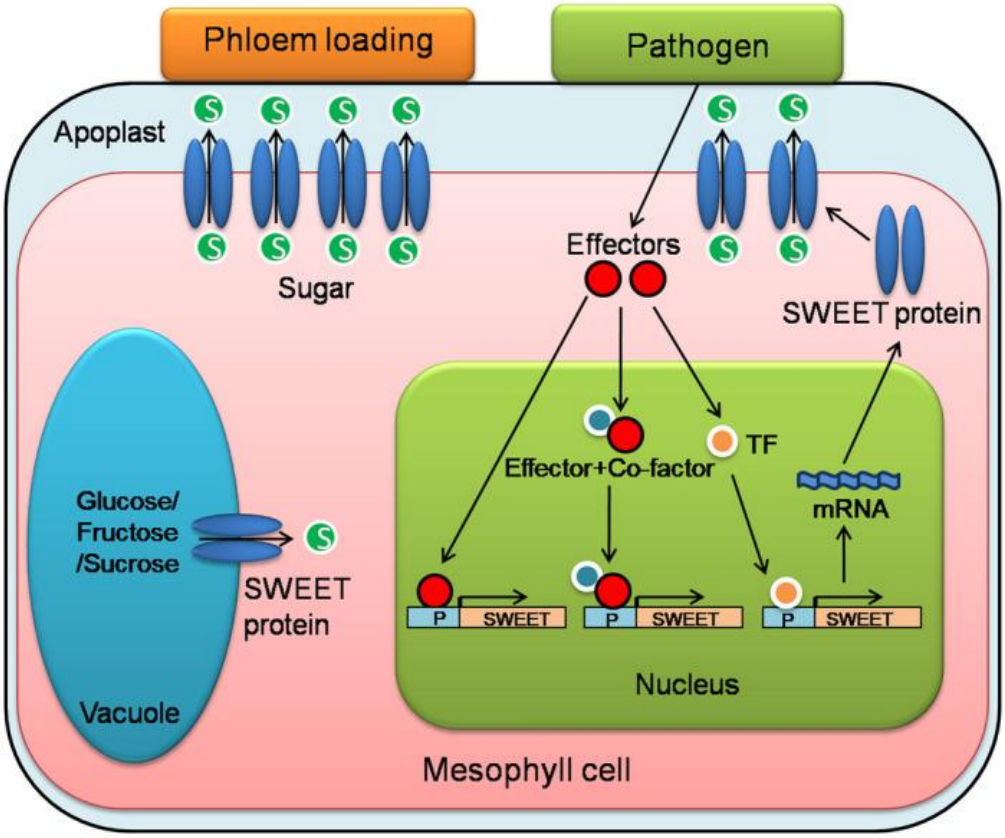


Figure i5-8 An example of susceptibility factor supporting pathogen infection:the SWEET sugar transporters in *Xanthomonas oryzae* pv. *Oryzae* infection.

Bacterial pathogen *Xanthomonas oryzae* pv. *oryzae* inject transcription activator-like (TAL) effectors (red dot) into the cytoplasm of host plant cells. When the TAL effectors enter the nucleus, they induce the expression of host *SWEET* genes either directly or indirectly via the activation of transcription factors/cofactors. The coded SWEET protein then functions as a sugar transporter, releasing sugar ("S" in green circle) into the apoplast as a source of nutrition for *Xoo.* Figure taken from Gupta et al., 2021.

TuMV genomic RNA is targeted by NGD due to a conserved motif in its genome that is important for potyvirus to generate an additional ORF (cf. *Introduction 6*),1 and this motif induces ribosome stalling that is targeted by the RQC machinery (**Figure i5-7** right) (Chung et al., 2008; Ge et al., 2023; Pouclet et al., 2023). Indeed, Arabidopsis plants that are knocked out for genes involved in NMD and NGD show increased susceptibility to PVX and TuMV, respectively (Garcia et al., 2014; Ge et al., 2023).

Interestingly, an opposite observation is reported in the tomato-tomato yellow leaf curl virus (TYLCV) pathosystem where overexpression of Pelota, which plays role in RQC, was associated with increased susceptibility, suggesting that this gene and/or RQC plays a different role during TYLCV infection (Lapidot et al., 2015). *Pelota* knock-out allele, *ty-5*, in tomato was also shown to confer resistance to several other begomovirus species, implying that *pelota*-mediated resistance may be conserved across begomovirus (Ren et al., 2022).

Together, these recent studies on the mRNA-decay-mediated virus resistance suggests a potential path for resistance breeding although the specificity of mechanisms to viral genomic structures may complicate developing resistance traits to different virus families. Further understanding in the mechanisms will be necessary for the implementation.

d. Third layer of defense: host-pathogen compatibility test for sustenance

Finally, the last phase of defense is the compatibility of host cellular machinery which the pathogens require in order to sustain their life cycle and proliferation. After having successfully entered into the plants and evaded or countered host defense response, both non-viral and viral pathogens need to subvert host structures and proteins to obtain nutrient and/or sustain their life cycle. This step is crucial for survival of the obligate parasites, i.e. some fungal species including powdery mildew fungi, oomycetes, or viruses. These host factors targeted by pathogens and support the infection are categorized as susceptibility (S) factors.

The S factors are often important part of host physiology, and multiple copies of *S* genes are identified with redundant functions. For example, bacterial pathogen *Xanthomonas oryzae* pv. *oryzae* is known to modulate the expression of sugar transporter genes *SWEET* by its transcription activator-like (TAL) effectors. *Xoo* secrete specific TAL

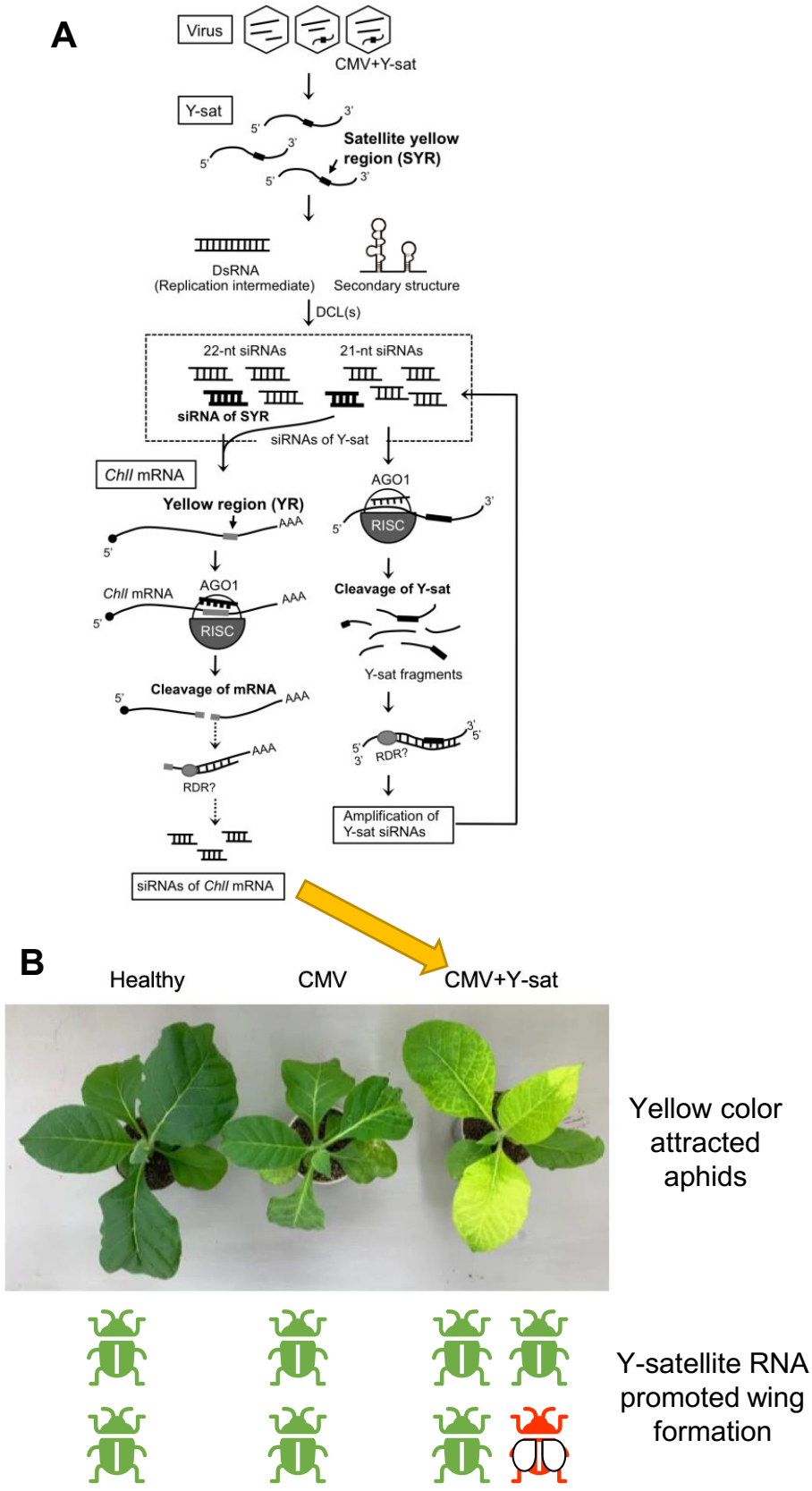


Figure i5-Box Remarkable example of antiviral RNA silencing contributing to the spread of viral pathogen.

A Y-satellite RNA of cucumber mosaic virus (CMV) co-infects plants and triggers host RNA silencing like other virus-derived dsRNA. siRNAs generated from the Y-satellite RNA induces host endogenous mRNA silencing on *Chll* gene due to the high homology. Silencing of *Chll* gene, a key gene in chlorophyll biosynthesis, results in the leaf yellowing symptom characteristic to the CMV+Y-satellite RNA co-infection in *Nicotiana tabacum*. **B** This yellowing was later discovered to play an important role in attracting the vector aphids and in inducing wing formation in the aphid population, consequently promoting long-distance transmission Figures adapted from Shimura et al., 2011 and Jayasinghe et al., 2021.

effectors increase the transcription of SWEET11, 12, 13, 14, 15 in rice and promote sugar efflux into the apoplast to obtain nutrients (**Figure i5-8**) (Reviewed in Gupta et al., 2021).

Viruses are heavily dependent on the host plant's machinery for multiple steps of its proliferation: translation of viral proteins, genome replication, as well as movement between cells. Therefore, the viruses possess compatible factors to target host machinery components. One of the well-studied examples of such S factor in virus infection is eukaryotic translation initiation factor (eIF) 4E, which is part of the translation initiation complex eIF4F and binds to the 5′-cap structure of mRNA to initiate translation. The eIF4E and its homologs are hijacked by potyviral viral genome-linked protein (VPg), which is covalently attached to the 5′-end of viral genome, and this interaction is crucial for successful infection of potyvirus (For review, Robaglia and Caranta, 2006b). (The eIF4E's roles in susceptibility and resistance to potyvirus will be discussed in details in the following section, as this is the main subject of this thesis.)

Because pathogens require compatible S factors for maintaining their life cycle, the disruption of these S factors by artificially or naturally induced inactivation or modification will lead to resistance in the host plants. This is known as resistance by loss-of-susceptibility, and it is mostly conferred and inherited in a recessive manner. Similarly to the zig-zag-zig model for PTI-ETI and RNA silencing, *S*-gene-based resistance is also based on the arms race between the resistance variants of host S factor and the pathogen factors that utilize the S factor, where the resistance has the possibility to be broken down by the evolving pathogens.

e. Plant genetic resistance for breeding

Plants have various defense mechanisms to protect themselves from pathogen invasion and usurpation of their resources. These innate immune systems are results of co-evolution between plants and pathogens, and identifying and elucidating each layer will allow us to implement them in our breeding strategy. Generally speaking, the *R* gene provides dominant and often race-specific resistance to pathogens, while a resistance allele of *S* genes provide recessive resistance to pathogens via conserved mechanisms. Natural variations of these defense-related genes that confer different levels of resistance provide hints for understanding the plant-pathogen crosstalk as well as the resource for the resistance breeding (Zhang et al., 2017; Dempewolf et al., 2017). Maximizing the

defense potential of the plants via breeding will help achieve the sustainable agriculture with less environmental impacts in the coming years.

Among different types of resistance, the *S*-gene-based resistance is interesting for its potential, wide applicability to a larger group of pathogens than a race/strain level due to the conserved mechanism of S factor utilization. Despite the potential, the recessive mode of inheritance and redundancy and significance of *S* genes have made it difficult to implement the resistance by loss-of-susceptibility until now (Van Schie and Takken, 2014). But today, *S*-gene-based resistance can be more easily integrated in the cultivar genome due to the advancement of genome-editing technologies that can directly edit the gene(s) of interest. Together with other resistance genes in the crop cultivars, the edited resistance variant of *S*-gene can be stacked in the future cultivars to reinforce the crop culture. In the last introductory section, the resistance based on loss-of-susceptibility against potyvirus and the difficulties for its implementation will be presented extensively to provide the context for this thesis.

6. Loss-of-susceptibility case file: eIF4E and potyvirus

S-gene-based or loss-of-susceptibility resistance occurs when a host plant S factor that is indispensable for pathogen infection is absent or modified into a form unusable by the pathogen. This implies that such resistance is inherited in recessive manner because the plants heterozygous for a susceptible form and a resistance form of an S factor will end up being susceptible to the pathogen.

This type of resistance is often effective for obligate parasites like viruses that are limited in the genome size and have a smaller number of factors compared to other pathogens. Indeed, natural recessive resistance genes are found more frequently for viruses than other pathogens. Among others, such recessive resistance genes are often found for potyviruses, the largest group of plant RNA viruses, and the majority of these genes were identified to be the key eukaryotic translation initiation factors (*eIF) 4E* and their isoforms (For review, Kang et al., 2005a; Truniger and Aranda, 2009; Wang and Krishnaswamy, 2012). The conservedness of eIF4E-based recessive resistance across plant species has promoted many research groups including our own to study the susceptibility factors to potyviruses and how their variants provide resistance, in an aim to develop novel resistance strategy based on these genes. The following section will

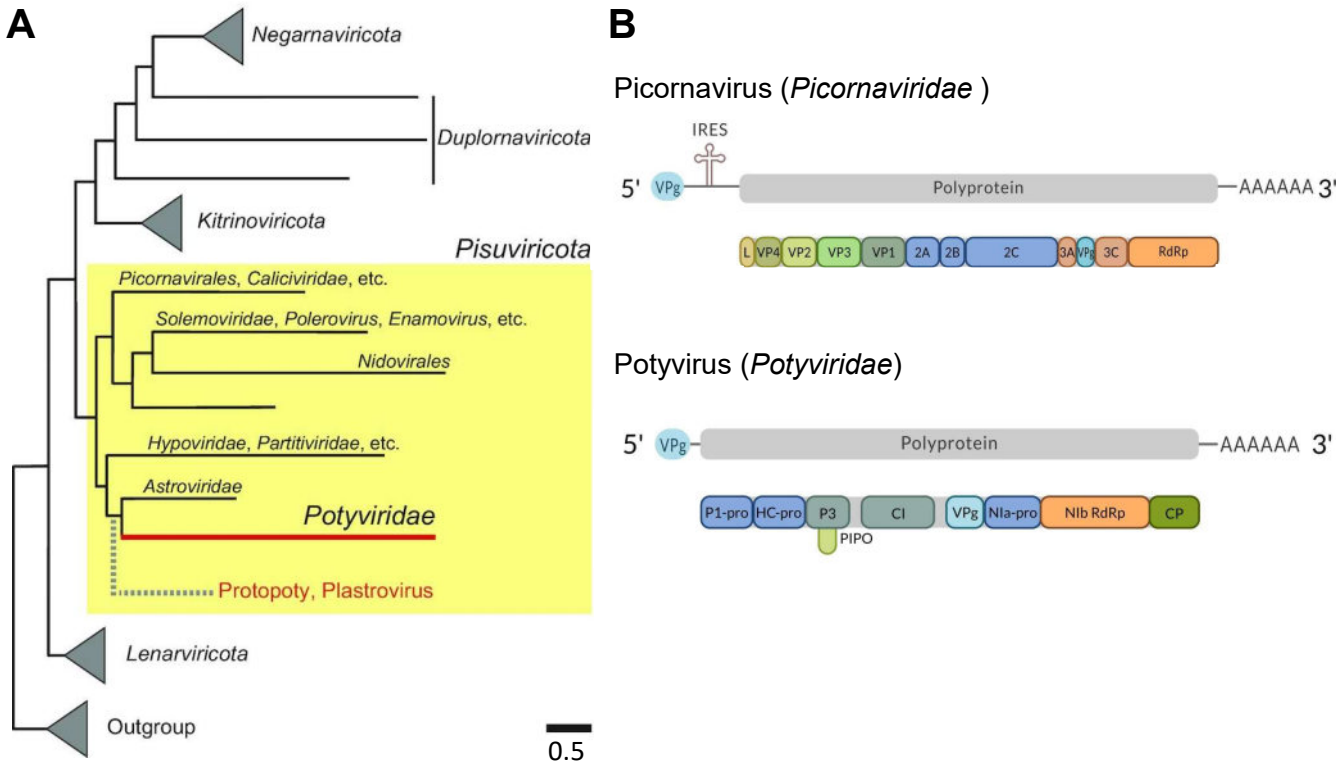


Figure i6-1 Plant-infecting *Potyviridae* and vertebrate-infecting *Picornaviridae* belong to the *Pisuviricota* and share similar genome structures.
A Schematic phylogenetic tree generated based on a viral protein, RNA-dependent RNA polymerases (RdRp), conserved in all RNA viruses. RNA virus phyla are indicated, and the branches of *Pisuviricota* are expanded; reverse transcriptases of group II introns and non-long-terminal-repeat transposons were used as an outgroup as described in Wolf et al., 2018; the amount of genetic change shown by the scale bar. Taken from Pasin, Daròs, and Tzanetakis, 2022. **B** Schematic of typical genomic structures of picornavirus and potyvirus. The colored boxes indicate the proteins that are produced upon cleavage of the polyprotein. Modified from Mesa and Rakotondrafara, 2023.

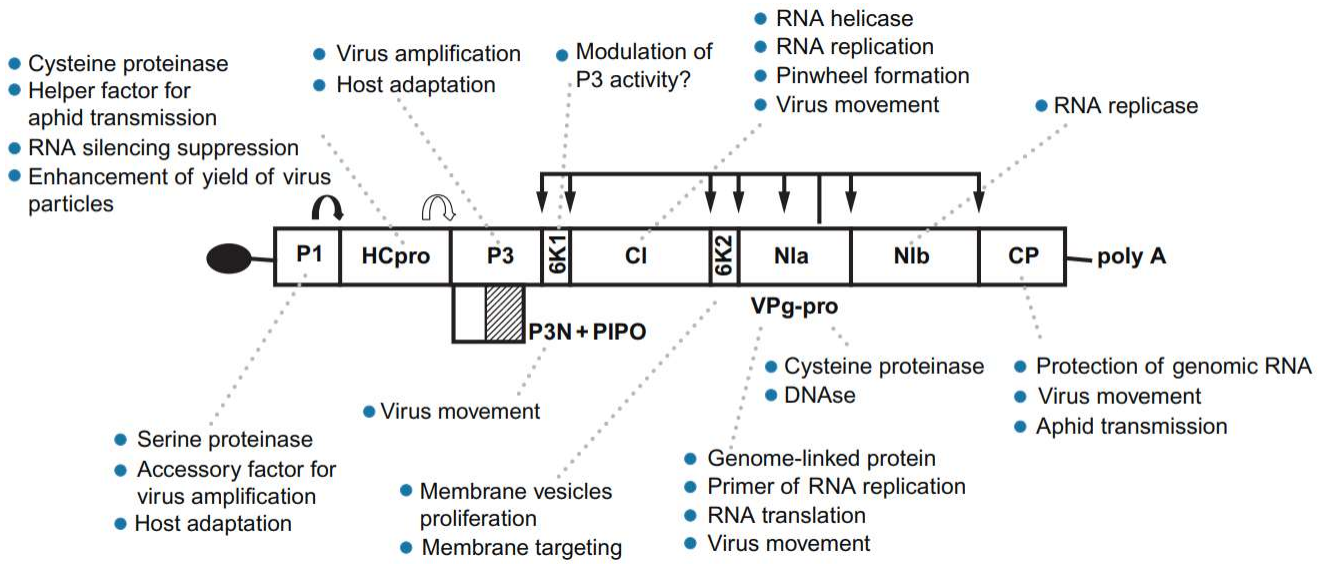


Figure i6-2 Potyviral ORF products (proteins) and their functions. A large rectangle represents the long ORFof potyviral genome encoding polyprotein and the small rectangle below the short ORF. Each box in the rectangles is the functional protein. Black oval represents the VPg linked to the 5'-end of viral genome. P1: protein 1 protease; HC-Pro: helper component protease; P3: protein 3; PIPO: pretty interesting Potyviridae ORF (the shaded rectangle); 6K1: 6 kDa peptide 1; CI: cylindrical inclusion; 6K2: 6 kDa peptide 2; NIa-Pro: nuclear inclusion a-protease; Nib: nuclear inclusion b, RNA-directed RNA polymerase; CP: capsid protein. Taken from Revers and Garcia, 2015.

cover the key discoveries and important concepts in loss of susceptibility to potyvirus that have led to this thesis' core strategy for developing potyvirus resistance alleles by genome-editing.

a. Discovery of eIF4E-based resistance against potyvirus

i. Brief description of potyvirus

Potyvirus, being the largest genus of plant RNA virus, are economically important pathogen. Collectively, they infect more than 50 plant species although the individual potyvirus species have a narrow range of host species (Moury and Desbiez, 2020). Potyvirus is among 12 genera of the family *Potyviridae*, which is most recently placed in the phylum *Pisuviricota*, which comprises of members of the former picorna-like supergroup (Koonin et al., 2020; Inoue-Nagata et al., 2022; Pasin et al., 2022) (**Figure i6-1A**). The genome of potyvirus consists of a large ORF encoding a polyprotein which is cleaved into ten proteins and a short ORF generated by a transcriptional slippage mechanism and encodes one protein (**Figure i6-1B**) (Chung et al., 2008; Rodamilans et al., 2015; Olspert et al., 2015; Hagiwara-Komoda et al., 2016). The potyviral genome is positive-sense[12] single-stranded RNA and has a viral genome-linked protein (VPg) covalently attached to the 5′-end and polyadenylated at the 3′-end (Latest ITCV taxonomy: Inoue-Nagata et al., 2022). The 11 encoded proteins and their functions are summarized in **Figure i6-2** (Revers and García, 2015).

ii. Potyvirus resistance is conferred by translation initiation factor genes

The first complete genome sequences of a few potyviruses were determined in late 1980s, as the sequencing and cloning methods were established in 1970s (For review, Gibbs et al., 2020). Among the first characterized potyvirus species was turnip mosaic virus (TuMV). TuMV was already known to infect a variety of crop species of especially *Brassica* genus as well as the model plant species *Arabidopsis thaliana*, and its genome sequence (of an isolate originated in Quebec, Canada) was determined by Dr. Jean-François Laliberté's team in 1992 (Nicolas and Laliberte, 1992). The characteristics of the genome structure identified as such and the protein expression strategy via polyprotein indicated that the potyviruses belonged to the same supergroup as the vertebrate-

[12] Positive-sense RNA genome can act as a messenger RNA and be directly translated into a protein.

Table i6-1 Resistance to potyviruses based on natural alleleic variants eIF4E and eIFiso4E.

Modified from Sanfaçon, 2015.

eIF4E-based

Transl. Factor	Plant Species	Cause of Resistance	Virus [1]	References
eIF4E	*Capsicum spp*	Natural resistance genes *pvr1*/*pvr1*^{2} (point mutations)	PVY, TEV, PepMoV	[89,106,109, 112,115,130–132,139–141]
eIF4E	*Lactuca sativa*	Natural resistance gene *mo1* (point mutations)	LMV	[111,142]
eIF4E	*Pisum sativum*	Natural resistance genes *sbm1*/*whv*/*cyv2* (point mutations)	PsBMV, BYMV, ClYVV	[88,110,114, 143,144]
eIF4E	*Phaseolus vulgaris*	Natural resistance gene *bc-3* (point mutations)	BCMV, ClYVV	[145,146]
eIF4E	*Solanum habrochaites*	Natural resistance gene *pot-1* (point mutations)	PVY, TEV	[147]
eIF4E	*Citrullus lanatus*	SNP mapping with resistance (point mutations)	ZYMV	[154]
eIF4E	*Prunus armeniaca*	One of several QTLs associated with resistance	PPV	[155]

eIFiso4E-based

Transl. Factor	Plant Species	Cause of Resistance	Virus [1]	References
eIFiso4E	*Capsicum spp*	Natural resistance gene *pvr6* (functions with *pvr1 or pvr1*^{2}) (deletion causing truncation of protein)	PVMV, ChiVMV	[115,128]
eIFiso4E	*Brassica rapa*	Natural resistance gene *retr01* (splicing mutant)	PPV, TuMV	[156–158]

[1]Virus abbreviations not defined in the text: PepMoV (pepper mottle virus), PsBMV (pea seed-borne mosaic virus), BYMV (bean yellow mosaic virus), BCMV (bean common mosaic virus), BaYMV (barley yellow mosaic virus), BaMMV (barley mild mosaic virus), ZYMV (zucchini yellow mosaic virus), MWMV (Maroccan watermelon mosaic virus), CVYV (cucumber vein yellowing virus), PVMV (pepper vein mottling virus), and ChiVMV (Chilli veinal mottle virus).

infecting picornaviruses ("picorna-like supergroup") (Riechmann et al., 1992) (**Figure i6-1A**).

One of the characteristic similarities between potyvirus and picornavirus was the VPg covalently attached to the 5′-end of the viral genome (**Figure i6-1B**). Picornaviral VPg was known to be required for the viral genomic RNA replication, and by analogy, potyviral VPg was suggested to have the similar role. In order to explore this possibility, Dr. Laliberté and his colleagues performed the yeast two-hybrid (Y2H)-based screening of Arabidopsis cDNA library for candidate genes encoding host factors that interact with TuMV VPg (Wittmann et al., 1997). A single candidate host factor was found in this work and identified to be one of the eukaryotic translation initiation family proteins, eIFiso4E, suggesting potyviral VPg's role in RNA translation. This finding was corroborated by (Léonard et al., 2000) that showed that the amino acid substitution in TuMV VPg that abolished the interaction with Arabidopsis eIFiso4E *in vitro* was responsible for preventing infection by TuMV expressing the same mutant VPg *in planta*, in *Brassica perviridis* suggesting that the VPg-eIF4iso4E interaction is integral to the potyvirus infection.

The implication of plant Arabidopsis eIFiso4E in potyvirus infection as a susceptibility factor was confirmed by two independent studies using Arabidopsis (Lellis et al., 2002b; Duprat et al., 2002). Lellis et al. (2002) used the forward genetics approach: the group has generated a population of EMS-mutagenized Arabidopsis plants and screened the progenies by TuMV inoculation; plants resistant to TuMV were selected as susceptibility factor mutants, and the map-based cloning on these plants identified the *eIFiso4E* as the gene encoding for susceptibility factor. On the other hand, Duprat et al. (2002) used the reverse genetics approach: the group has screened *eifiso4e*^{KO} mutants among the defective maize *Spm* transposon insertion Arabidopsis lines and challenged the mutants with TuMV and another potyvirus lettuce mosaic virus (LMV); the *eifiso4e*^{KO} mutants were resistant to TuMV and LMV, confirming that *AteIFiso4E* is required for infection by the tested potyviruses.

In the same year, natural potyvirus resistance locus *pvr2* in pepper (*Capsicum* spp.) that has been long used in pepper breeding program was identified to be an allelic variant of pepper (*C. annuum*) *eIF4E*, using a candidate gene approach based on the presumed role of *eIF4E* isoforms in potyvirus infection (Greenleaf, 1986; Ruffel et al., 2002b). The

A

```
pvr2  (pepper)    MATAEMEKTTTFDEAEKVKLN---ANEADDEVEEGEIVEETDDTTSYLS----KEIATKH
pot1  (tomato)    MAAAEMERTMSFDAAEKLKAADGGGGEVDDELEEGEIVEESNDTASYLG----KEITVKH
mo1   (lettuce)   -MKSEEQKLIDVNKHRGVRSD----GEEDEQLEEGEIVGGDADTLSSSSSSRPGTAIAQH
sbm1  (pea)       --MVVEETPKSIITDDQITTNPNRVIEDDNNLEEGEILDEDDSSATSKP-----VVHQPH

pvr2  (pepper)    PLEHSWTFWFDNPVAKSKQAAWGSSLRNVYTFSTVEDFWGAYNNIHHPSKLVVGADLHCF
pot1  (tomato)    PLEHSWTFWFDNPTTKSRQTAWGSSLRNVYTFSTVEDFWGAYNNIHHPSKLIMGADFHCF
mo1   (lettuce)   PLEHSWTFWFDTPSAKSKQVAWGSSMRPIYTFSSVEEFWSLYNNIHRPSKLAQGADFYCF
sbm1  (pea)       LLENSWTFWFDTPAAKSKQPDWG-SMRPIYTFSTVEEFWSIYNNIHHPGKLAVGADFYCF

pvr2  (pepper)    KHKIEPKWEDPVCANGGTWKMSFSKGKSDTSWLYTLLAMIGHQFDHEDEICGAVVSVRGK
pot1  (tomato)    KHKIEPKWEDPVCANGGTWKMSFSKGKSDTSWLYTLLAMIGHQFDHGDEICGAVVSVRAK
mo1   (lettuce)   KNKIEPKWEDPVCANGGKWTMTFTKAKSDTCWLYTLLAMIGEQFDHGDDICGAVVNVRAR
sbm1  (pea)       KHKIEPKWEDPICANGGKWTANYPKGKSDTSWLYTLLAMIGEQFDHGDEICGAVVNVRGR

pvr2  (pepper)    GEKISLWTKNAANETAQVSIGKQWKQFLDYSDSVGFIFHDDAKRLDRNAKNRYTV
pot1  (tomato)    GEKIALWTKNAANETAQVSIGKQWKQFLDYSDSVGFIFHDDAKRLDRNAKNRYTV
mo1   (lettuce)   QEKIALWTKNAANESAQLSIGKQWKEFIDYNDTIGFIFHEDAKTLDRSAKNKYTV
sbm1  (pea)       AEKISIWTKNASNEAAQVSIGKQWKEFLDYNETMGFIFHDDARKLDRNAKNKYVV
```

B

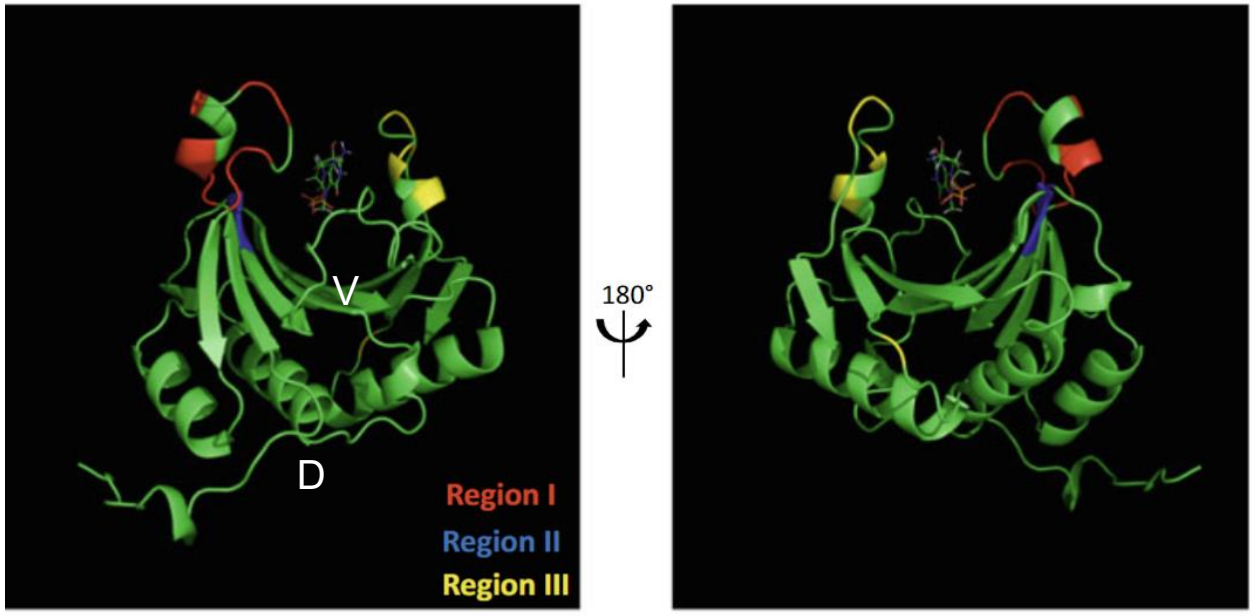


Figure i6-3 Localization of clusters of eIF4E amino acids differing between potyvirus susceptible and resistant genotypes in the predicted amino acid sequences and 3D structure.

A Alignment of predicted eIF4E protein sequences from susceptible genotypes of pepper (*Capsicum* spp.), tomato (*Solanaceae* spp.), lettuce (*Lactuca* spp.) and pea (*Pisum sativum*) . Black boxes indicate amino acids differing between susceptible and resistant genotypes. The two protein regions region I and region II with clustering of amino acid substitutions are highlighted in orange and blue, respectively. Taken from Robaglia and Caranta, 2006. **B** Three dimensional structure model of eIF4E1 protein from the potyvirus-susceptible pepper (*Capsicum annuum*) 'Yolo Wonder' accession carrying eIF4E1-encoding allele *pvr2*+, built on the 2WMC-C02 PDB file from pea (*Pisum sativum*). Amino acid positions whose changes are associated with resistance to PVY, TEV or PTV are colored in red (region I), blue (region II) or yellow (region III). Three regions are located around the cap-binding pocket on the ventral side ("V") of the protein. "D" indicates the dorsal side. The mRNA cap analogue is represented in sticks. The 3D model was build and represented using the softwares yasara (www.yasara.org) and pymol (https://pymol.org/2/), respectively. Taken from Tamisier et al., 2023. Region I and II of **A** and **B** correspond.

pvr2-encoded eIF4E contained two amino acid substitutions compared to the eIF4E from susceptible pepper cultivars.

iii. Natural resistance alleles

After the identification of *pvr2* as pepper *eIF4E* allele, several natural potyvirus resistance alleles coding for eIF4E protein variants were subsequently identified in other crop species and their wild relative (**Table i6-1**). These include *mo1* alleles in lettuce (*Lactuca sativa*), *sbm1* allele in pea (*Pisum sativum*), and *pot1* allele in wild tomato species *S. habrochaites* accession PI247087 (Nicaise et al., 2003; Gao et al., 2004; Ruffel et al., 2005).

All of these alleles generate full-length (except for $mo1^1$ allele which includes in-frame 6-bp deletion) cDNA of *eIF4E* which contains nonsynonymous mutations and is presumed functional. Interestingly, all three works above pointed out that according to the homology-based structural modeling, the majority of the amino acid substitutions were located in two regions close to a cap-binding pocket of eIF4E, which is essential for translation initiation (later referred to as "Region I" and "Region II") (Reviewed in Robaglia and Caranta, 2006b) (**Figure i6-3A,B**). Especially a rich variety of natural resistance alleles have been found in cultivated and wild pepper accessions. At present, there are 34 distinct *eIF4E1* alleles associated with resistance of different spectra to PVY, tobacco etch virus (TEV), and peru tomato mosaic virus (PTV). These alleles possess one or more mutations in Regions I and/or II as well as in another region of pepper *eIF4E1*, presumed Region III, which is also close to the cap-binding pocket (Charron et al., 2008; Poulicard et al., 2016; Region III presented in Tamisier et al., 2023) (**Figure i6-3B**). For some of the alleles, abolished interaction between the resistance eIF4E variants and the VPg have been experimentally validated using *in vitro* binding assay and *in vivo* Y2H assay, proving that the mutations are responsible for the loss of interaction with VPg. However, this was not observed for all pathosystems (Reviewed in Wang and Krishnaswamy, 2012). The eIF4E-VPg interaction will be discussed more in detail later.

In contrast to the great diversity of resistance *eIF4E* alleles with nonsynonymous mutations, only two cases of truncated *eIF4E* isoforms in association with potyvirus resistance have been reported: *pvr6* allele with a deletion in pepper *eIFiso4E* and *retr01* allele that generates mis-spliced *B. rapa eIFiso4Ec* (Ruffel et al., 2006; Nellist et al., 2014; Reviewed in Sanfaçon, 2015) (**Table i6-1**). These are loss-of-function *eIFiso4E* alleles, and therefore the potyvirus are not able to use these factors for infection. It is noteworthy that

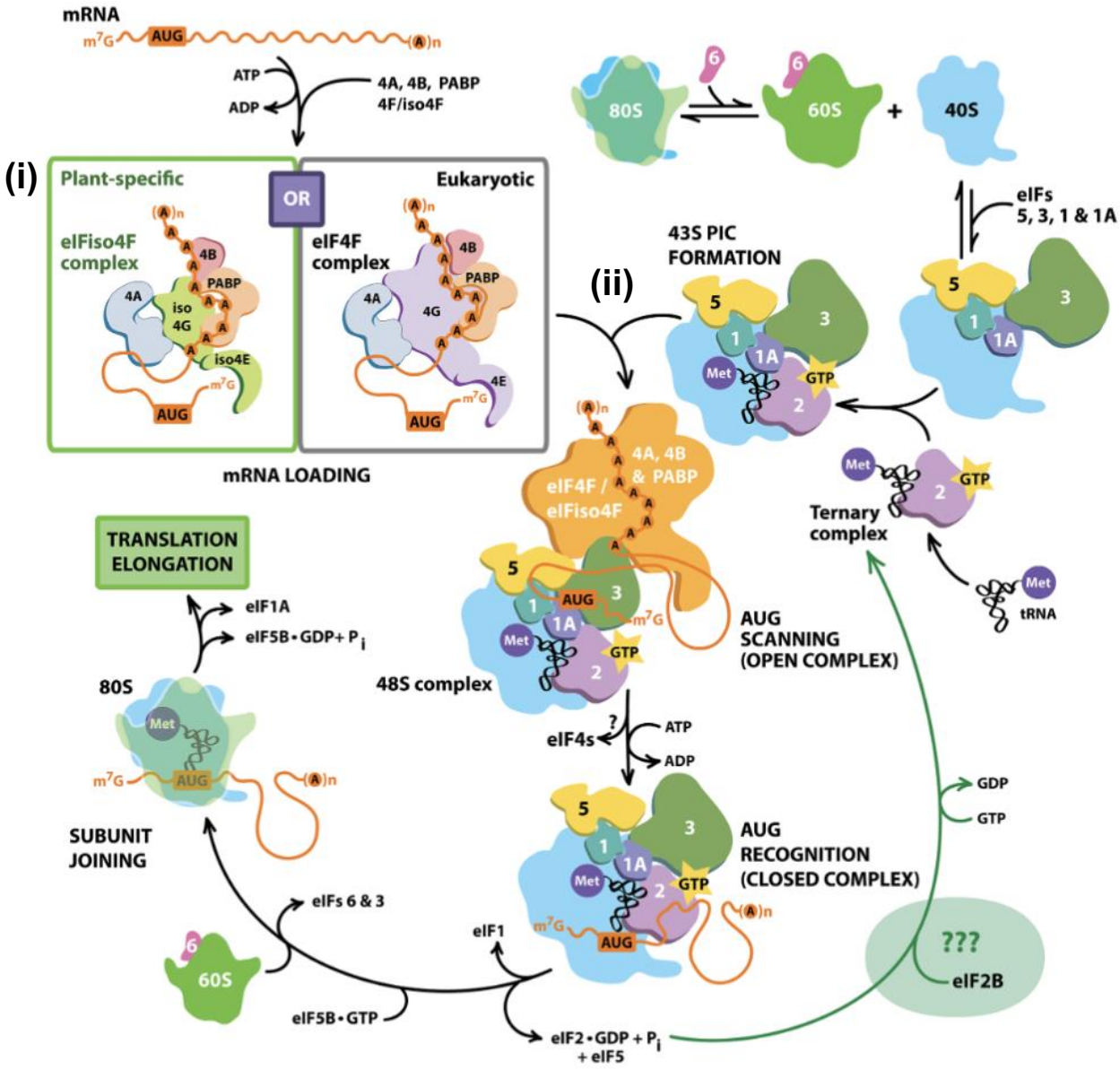


Figure i6-4 Flow of plant mRNA translation initiation.

(i) Formation of the eukaryotic translation initiation complex eIF4F/eIFiso4F. Once the mRNA has been exported from the nucleus into the cytoplasm, it interacts with a cap-binding complex (eIF4F or eIFiso4F) at the 5′ end via mRNA cap and poly(A)-binding protein (PABP) at the 3′ end via poly-A tail. Additional factors, eIF4A and eIF4B are recruited to the mRNA. (ii) Interaction of mRNA and translation initiation factors complex with the 43S PIC (pre-initiation complex). The 43S PIC is formed from the 40S ribosome and its associated factors eIF1, eIF1A, eIF3 and eIF5 plus the ternary complex (TC): eIF2•Met-tRNAi•GTP. This 43S PIC engages the mRNA and its associated factors (eIF4F/eIFiso4F, eIF4A, eIF4B, PABP) to form the 48S scanning complex, which functions to scan the 5' UTR of the mRNA in the 5′ to 3′ direction in order to select the AUG. Taken from Browning and Bailey-Serres, 2015. Refer to the main text of Browning and Bailey-Serres, 2015 for the process after AUG recognition.

the *pvr6-eifiso4e* alone does not show resistance to potyvirus but can be used to expand the potyvirus resistance spectrum to some potyviruses like pepper veinal mottle virus (PVMV) and Chilli veinal mottle virus (ChiVMV) when combined with *pvr2-eIF4E* alleles (Caranta, 1996). However, stacking of *pvr6* with *pvr2*^{3}- or *pvr2*^{4}*-eIF4E1* allele was also associated with increased overcoming of *pvr2*-mediated resistance by PVY isolates (Quenouille et al., 2014; Quenouille et al., 2016).

Both the Arabidopsis studies and discoveries of natural resistance alleles have supported the concept of resistance by loss of susceptibility factor eIF4E against potyvirus either due to the modification in its potential interaction site or the inactivation of the factor. Intriguingly, the profiles of naturally selected alleles indicate the preference of the former, the eIF4E proteins with amino acid substitutions, over the latter, the non-functional eIF4E. This phenomenon can be explained by the roles that eIF4E isoforms play in the plant development and the virus infection. In the following section, these roles will be dissected.

b. Biological roles of plant eIF4E

In this section, the function of eIF4E isoforms in plant development will be covered, primarily based on a comprehensive review (Browning and Bailey-Serres, 2015). In the angiosperm *eIF4E* gene family, there are three major isoforms: eIF4E, eIF(iso)4E, and eIF4E-homologous protein (4EHP ; also known as novel cap-binding protein: nCBP, initially). eIFiso4E is a plant-specific isoform that evolved at the emergence of flowering plants (Patrick and Browning, 2012). All functional isoforms coded by *eIF4E* gene family bind to a 5′ m7G cap structure—an N7 methylated guanine, linked to the first nucleotide of the RNA via triphosphate linkage though at different affinity (Kropiwnicka et al., 2015; mRNA cap reviewed in Ramanathan et al., 2016). This cap-binding function gives eIF4E isoforms the most important role to play in plant mRNA translation, but recent observations in animal suggests additional roles of the plant eIF4E isoforms in nuclear export of mRNA and pre-mRNA splicing (Culjkovic et al., 2007; Wang and Krishnaswamy, 2012; Borden, 2023). The following section will mainly cover eIF4E function in translation and briefly in other cellular process.

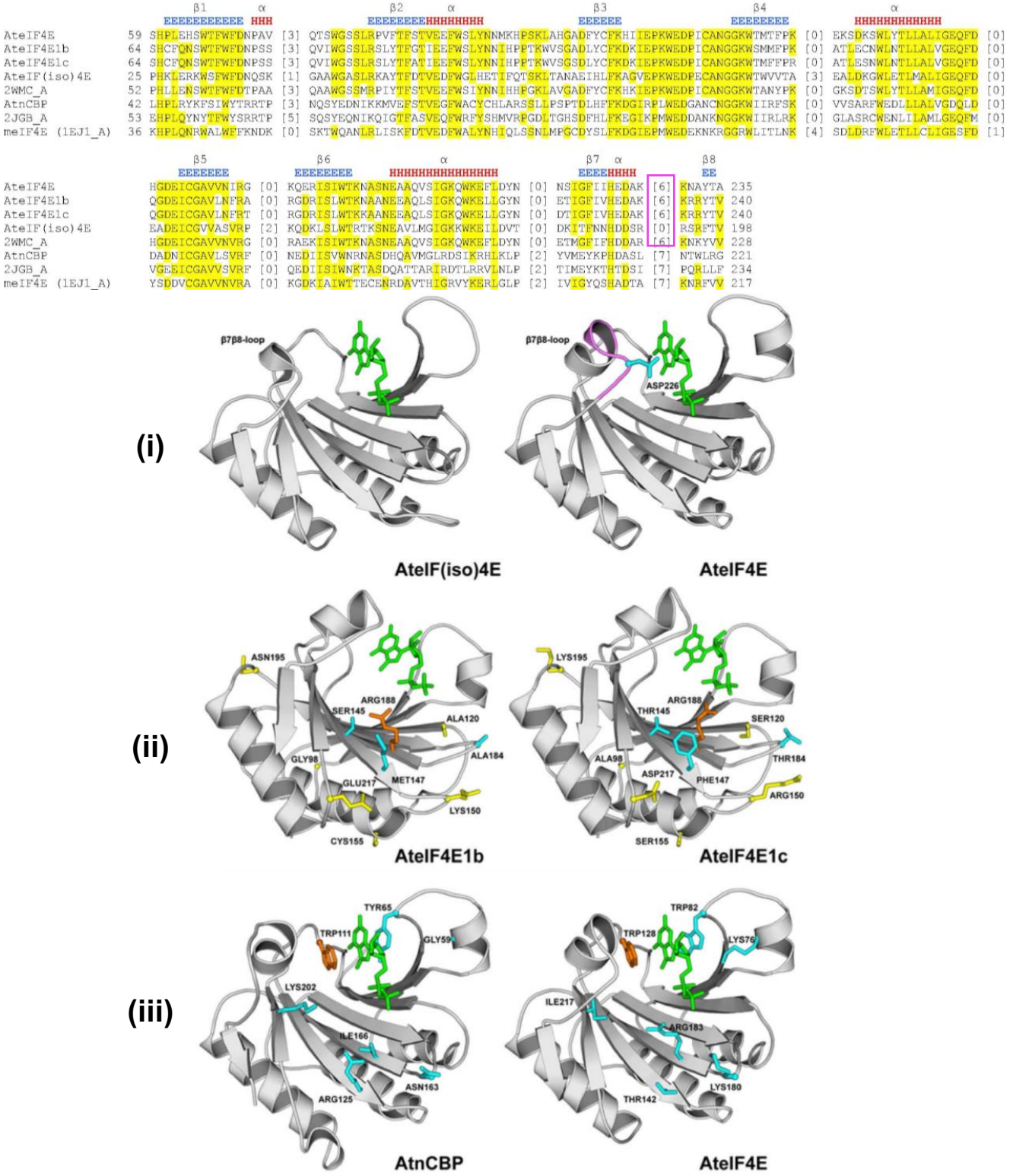


Figure i6-5 Comparison of *A. thaliana* eIF4E partial amino acid sequences and 3-D structures. **(Top)** Partial amino acid sequence alignment for A. thaliana eIF4E isoforms and their closest homologs of known structure (pea eIF4E, PDB: 2WMC; mouse M. *musculus* eIF4E, PDB: 1EJ1; human 4EHP, PDB: 2JGB). The numbers of excluded residues are specified in square brackets. Conserved residues (shared by >50% of the aligned sequences) are highlighted in yellow. Locations of secondary structure elements β-strand and α-helix as identified in PDB: 2WMC are marked above sequences. **(Bottom)** 3-D structures of AteIF4E isoforms. Cap analog (m^7GDP) taken from PDB: 1EJ1 structure is shown in green. Amino acids in cyan are critical for the difference in the binding affinity for cap analogs between the two structures side by side. (i) AteIFiso4E vs. AteIF4E; insertion of six amino acids in the loop between β-strands β7 and β8 of AteIF4E is in magenta and corresponds to the excluded residues in sequence in magenta open box. (ii) AteIF4E1b vs. AteIF4Ec; additional residues that differ between AteIF4E1b and AteIF4E1c are presented in yellow. Conserved Arg188 is in orange (iii) AtnCBP (At4EHP) vs. AteIF4E; conserved tryptophan is in orange. Taken from Kropiwnicka et al., 2015.

i. mRNA translation

Plant eIF4E is a part of the cap-binding complex eIF4F which is integral for initiation of translation of 5′-capped mRNA ("cap-dependent translation"). An eIF4F complex forms as eIF4E binds to the 5′-cap structure of mRNA on its concave, "ventral" side and to the scaffold protein eIF4G on its convex, "dorsal" side (Miras et al., 2017) (**Figure i6-3B**). The concave surface is referred to as the cap-binding pocket. The eIF4G in the complex helps the assembly of eIF4A, eIF4B, and poly(A)-binding protein (PABP) on the mRNA. It also binds to eIF3 which mediates the binding of 43S (pre-initiation complex) PIC to start the scanning of 5′-UTR of mRNA for translation (**Figure i6-4**). eIF4E and eIFiso4E are class I cap-binding proteins and form a complex with eIF4G and eIFiso4G, respectively (Joshi et al., 2005). The resulting eIF4F and plant-specific eIFiso4F complexes are the canonical complex for translation in plants. The third isoform, 4EHP is a class II cap-binding protein and known to form a complex of At4EHP-wheat eIFiso4G *in vitro* which shows a weak translation initiation ability (Ruud et al., 1998; Joshi et al., 2005). Given its lower cap-binding affinity and the lack of observation compared to other eIF4Es, 4EHP has not been considered as a part of canonical complex (Ruffel et al., 2002). Interestingly, the recent studies on mammal 4EHP ortholog suggest its possible vast roles in regulation of gene expression, including participation in the RQC (Hickey et al., 2020; Weber et al., 2020; 4EHP orthologs reviewed in Christie and Igreja, 2021) (RQC has a role in virus infection cf. *Introduction 5.c.iv*).

Number of gene copies of each eIF4E isoform vary between plant species; for example, Arabidopsis has five copies including three *eIF4Es*, one *eIFiso4E*, and one *4EHP*, whereas tomato has two *eIF4Es*, one *eIFiso4E*, and one *4EHP* (Kropiwnicka et al., 2015; Lebaron et al., 2016). The same observation is made for *eIF4G* isoforms; Arabidopsis has one *eIF4G* and two *eIFiso4G* genes, while tomato has two genes of each isoform (Patrick and Browning, 2012). Phylogenetic analysis indicated that the eIF4E isoforms arose from the ancestral gene duplication events (Nishimasu et al., 2018b; Veillet et al., 2020). In terms of structure, all the present isoforms show high homology as seen in the example of five Arabidopsis eIF4E isoforms with few differences in amino acid residues (**Figure i6-5**) (Veillet et al., 2020). On the other hand, the eIF4G has a long N-terminus that lacks in the plant specific eIFiso4G (Patrick and Browning, 2012).

Table i6-2 Loss-of-function *eIF4E* alleles and associated phenotypes.
Taken from Bastet et al., 2017.

Family	Species	Gene	Allele	Origin	Translation initiation	Phenotype associated	Associated resistance[a]	Refs
	Arabidopsis thaliana	*eIF4E1*	*eif4e1*	T-DNA insertion	KO = non functional	Bolting delay	ClYVV	[43], unpublished data
			cum1	EMS	KO = non functional	Bolting delay	ClYVV, CMV	[18], unpublished data
		eIFiso4E	*lsp1*	EMS	KO = non functional	None	TuMV, TEV	[16]
			eifiso4e	Transposon insertion	KO = non functional	None	TuMV, TEV	[17]
			eifiso4e	CRISPR/Cas9	KO = non functional	None	TuMV	[27]
		eIF4E1 × *eIFiso4E*	*eif4e1* × *eifiso4e*	T-DNA × transposon	KO × KO	Male gametophyte lethal	Not available	[43]
			cum1 × *eifiso4e*	EMS × transposon	KO × KO	Lethal	Not available	[42]
Solanaceae	*Lycopersicum* spp.	*eIF4E1*	*eIF4E1-pot1*	Natural variation	Variant = functional	None	Large PVY TEV	[26,44]
			eIF4E1-KO	TILLING/EMS	KO = non functional	None	Narrow PVY	[25]
		eIF4E2	*eIF4E2-KO*	TILLING/EMS	KO = non functional	Slow growth	none	[25]
		eIF4E1 × *eIF4E2*	*eIF4E-RNAi*	Knockdown	Knockdown	Slow growth	Large PVY TEV	[19]
		eIF4E1 × *eIF4E2*	*eIF4E1-KO* × *eIF4E2-KO*	TILLING/EMS	KO × KO	Slow growth, bushy	Large PVY TEV	[44]
		eIFiso4E × *eIF4E2*	*eIFiso4E-KO* × *eIF4E2-KO*	TILLING/EMS	KO × KO	Lethal	Not available	[25], PhD thesis[b]
	Capsicum spp.	*eIF4E1*	*pvr2*^{1} *to pvr2*^{25}	Natural variation	Variant = functional	None	PVY TEV	[34–37]
		eIFiso4E	*pvr6*	Natural variation	KO = non functional	None	None	[38,39]
		eIF4E1 × *eIFiso4E*	*pvr2* × *pvr6*	Natural variation	Functional × KO	None	PVY TEV PVMV	[38,39]
Cucurbits	*Cucumis melo*	*eIF4E*	*nsv*	Natural allele	Variant = functional	None	MNSV	[29]
				RNAi	Knockdown	Affected in growth and fertility	ZYMV, MNSV, CVYV	[20]
	Cucumis sativus	*eIF4E*	*eif4e*	CRISPR/Cas9	KO = non functional	None	ZYMV, PRSV-W	[28]

[a]Abbreviations: ClYVV, *Clover yellow vein virus*; CMV, *Cucumber mosaic virus*; CVYV, *Cucumber vein yellowing virus*; MNSV, *Melon necrotic spot virus*; PRSV-W, *Papaya ring spot mosaic virus-W*; PVMV, *Pepper veinal mottle virus*; PVY, *Potato virus Y*; TEV, *Tobacco etch virus*; TuMV, *Turnip mosaic virus*; ZYMV, *Zucchini yellow mosaic virus*.
[b]C. Gauffier, PhD thesis, Aix-Marseille University, 2015.

ii. Redundancy of eIF4E genes in mRNA translation

Importantly, the canonical eIF4E isoforms, eIF4E and eIFiso4E, both have the cap-binding ability and form eIF4F/eIFiso4F complex; this suggested that these proteins are functionally redundant in plant mRNA translation. Indeed, functional analyses using gene KO and knockdown plants for one or more eIF4E copies in several species have so far demonstrated that the two eIF4E isoforms show functional redundancy in plant growth to some extent; but the inactivation or repression of certain copies are not completely compensated by the other copies (**Table i6-2**) (For review, Bastet et al., 2017). For instance, Arabidopsis *Ateif4e1*^{KO} mutant shows bolting delay, an indication of growth retardation, whereas the *Ateifiso4e*^{KO} mutant does not show any obvious phenotype, implying that the AteIFiso4E can compensate AteIF4E function, but this is not fully reciprocated (Duprat et al., 2002; Bastet et al., 2018) (**Figure i6-6A**). Similar relationship was observed between tomato SleIF4E1 and its homolog SleIF4E2: *Sleif4e1*^{KO} mutant shows no growth defect, while *Sleif4e2*^{KO} mutant shows slower growth, reflecting the incomplete functional redundancy (Piron et al., 2010; Gauffier et al., 2016) (**Figure i6-6B**).

This can be at least partially explained by differential spatiotemporal and stress-induced expression of eIF4E or eIF4G isoforms (i.e. each isoform is expressed preferentially in certain tissues and/or developmental stages and/or in specific stress response) and selectivity of mRNA species by eIF4F or eIFiso4F complex (Rodriguez et al., 1998; Gallie and Browning, 2001; Bush et al., 2009; Salazar-Díaz et al., 2021). The preference of mRNA is determined by both eIF4E and eIF4G isoforms and is based on aspects such as the 5′-UTR secondary structure and the methylation status of the cap of the mRNA. Moreover, Arabidopsis eIF4F, but not eIFiso4F, can translate uncapped mRNA as well as the second cistron of bicistronic mRNA; and this is partially attributed to the mRNA selectivity of eIF4G isoforms, despite the eIF4E isoforms being the direct interactor of 5′-cap structure (Veillet et al., 2020). On the other hand, post-translational modification of eIFiso4G upon hypoxic stress is known to play role in translation of specific stress response gene (Reviewed in Castellano and Merchante, 2021). Altogether, the evidence indicates that the eIF4F/eIFiso4F subunits are partially functionally redundant but that they also have distinct roles in specific mRNA translation.

The functional analyses have also shown that the KO of multiple copies of eIF4E isoforms can be detrimental to the plant growth or even lethal, proving their importance

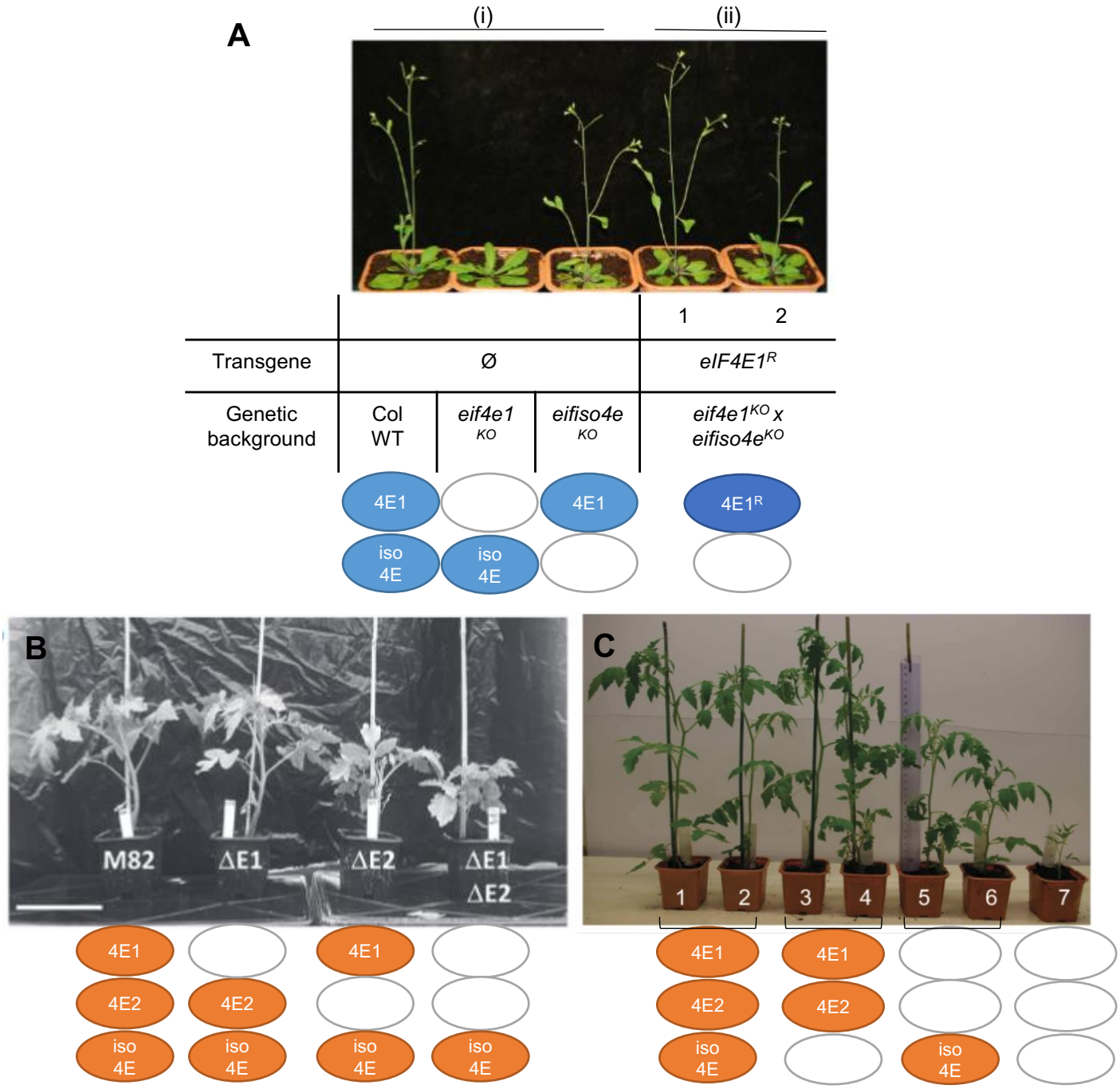


Figure i6-6 Phenotype of single and double mutants of eIF4E isoforms.
A (i) Comparison of *A. thaliana* WT (col-0 ecotype) and single KO mutants *eif4e1*^{KO} and *eifiso4e*^{KO}. The *eif4e1*^{KO} mutant is delayed in bolting (development of flowering stem) compared to the other two genotypes. Double KO *eif4e1*^{KO}x*eifiso4e*^{KO} mutant is not viable and not shown. (ii) Complementation of lethal double KO genotype by transgenic expression of synthetic Arabidopsis eIF4E1^{R} allele. Refer to the text (Bastet et al., 2018) for the *eIF4E1*^{R} allele. Figure adapted from Bastet et al., 2018. **B** Comparison of tomato WT (cv. M82) and TILLING KO mutants *eif4e1*^{KO} *(ΔE1)*, *eif4e2*^{KO} *(ΔE2)*, and double KO *eif4e1*^{KO}x*eif4e2*^{KO} *(ΔE1ΔE2).* WT and *eif4e1*^{KO} mutant show comparative growth, while *eif4e2*^{KO} mutant is smaller. Double KO mutant shows stunted growth. Scale bar = 10 cm. Figure adapted from Gauffier et al., 2016. **C** Comparison of tomato WT (cv. WVa106) and RNAi-KD mutants. 1: WVA106; 2: transgenic control; 3 and 4: *RNAi-eIFiso4E*-KD lines; 5 and 6: RNAi-*eIF4E1xeIF4E2*-KD lines; and 7: F1 progeny between RNAi--*eIF4E1xeIF4E2*-KD *and* RNAi-*eIFiso4E*-KD lines. 5 and 6 show stunted growth. 7 shows severe stunt. Figure adapted from Mazier et al., 2011. In **A-C**, remaining canonical eIF4E copies are shown in ovals below each photo. (For Arabidopsis, “4E1” (eIF4E1) and “iso4E” (eIFiso4E); for tomato, “4E1” (eIF4E1), “4E2” (eIF4E2), and “iso4E” (eIFiso4E).)

for plant physiology. The experimentally validated examples of multiple *eif4e/eifiso4e* KO mutants are listed in **Table i6-2**, including the double KO and RNAi-mediated KD tomato mutants of *eIF4E1* and *eIF4E2* genes which show stunted growth (Mazier et al., 2011; Gauffier et al., 2016) (**Figure i6-6B, C**). In Arabidopsis, double KO of *eIF4E1* and *eIFiso4E* is lethal (Callot and Gallois, 2014; Patrick et al., 2014). Therefore, at least one copy of canonical eIF4E is necessary to promote plant growth.

iii. Proposed roles in nuclear export of mRNA and splicing

In plants, eIF4Es are mainly studied in the context of translation in cytoplasm. However, their localization in the nucleus implies their involvement in other cellular processes due to its cap-binding function such as the nuclear export of specific mRNAs as observed in mammalian cells. In mammals, mRNA export function of eIF4E, combined or not with the translation initiation function, is considered to control the cell cycle progression and cellular proliferation (Veillet et al., 2020a). More recently, the direct and indirect implication of eIF4Es in mRNA splicing was suggested (Borden, 2023) (**Figure i6-7**). Although these functions have not been confirmed in plants so far, given the high conservation of eIF4E structures and localization in eukaryotes, it is possible that plant eIF4E isoforms play similar roles.

iv. Natural resistance alleles are functional

Because the canonical eIF4E isoforms are important for plant translation and healthy growth of plants, the question was posed as to whether the natural potyvirus resistance eIF4E variants were functional as cap-binding protein after their discovery. Natural resistance *eIF4E* alleles code for full-length protein with a few amino acid substitutions; most of which were found in the regions surrounding the cap-binding pocket in the 3-D structure, holding the potential to impact the eIF4E's original function. The biochemical assay to evaluate the cap-binding ability and/or the yeast complementation assay to evaluate the translation initiation of the eIF4E have been carried out for different natural resistance alleles from pepper *pvr* (Kang et al., 2005b; Yeam et al., 2007; Charron et al., 2008), lettuce *mo1* (Ruffel et al., 2002a), and pea *sbm1* (Ashby et al., 2011). These results showed that not all but many natural resistance variants retained the *in vitro* cap-binding ability and the *in vivo* translation initiation capacity despite the amino acid substitutions located around the functionally important regions.

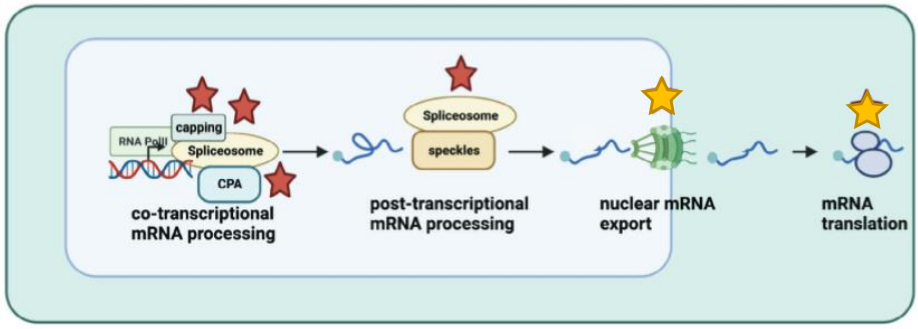


Figure i6-7 Additional validated roles of eIF4E isoforms and suggested roles in animals.

Yellow stars indicate the steps (nuclear mRNA export and mRNA translation) where the involvement of eIF4E is experimentally validated. Red stars indicate the steps (co-transcriptional and post-transcriptional, especially splicing) where the involvement eIF4E is suggested. Modified from Borden, 2023.

More recently, *in planta* complementation was demonstrated by stable transformation of *eif4e1*^{KO} Arabidopsis mutants with a synthetic Arabidopsis *eIF4E1*^{R} allele that contains the equivalent nonsynonymous mutations of pea resistance *sbm1* alleles (Bastet et al., 2018). In this case, the expression of the synthetic *AteIF4E1*^{R} allele restored the wild-type phenotype in the *eif4e1*^{KO} Arabidopsis which showed bolting delay. Furthermore, the *eIF4E1*^{R} allele complemented the lethal phenotype of the double KO *Ateif4e1*^{KO}x*Ateifsio4e*^{KO} (**Figure i6-6A**). Though it is more time-consuming, *in planta* complementation would provide more accurate evaluation the function of the mutant *eIF4E* allele.

c. Role of eIF4E in virus infection – molecular aspect

The cap-binding ability of eIF4E isoforms hint at the utilization of these factors during virus infection. The eIF4E isoforms are perhaps the most studied plant host factor in association with potyviral VPg, as this interaction was the first to be identified in plant-potyvirus system (Wittmann et al., 1997) (*cf. Introduction 6.a.ii*). In the following section, the eIF4E's roles in the viral infection context, especially in relation to the VPg functions, will be covered.

i. VPg's role in potyvirus infection and recruitment of eIF4E

Firstly, VPg's role in potyviral infection will be briefly described. Potyviral has 11 proteins to achieve the infection cycle. Being so few in number, each of these proteins have multiple functions, but the VPg is one of the best characterized as a versatile protein that promote RNA translation, gene silencing suppression, replication, cell-to-cell and long-distance movement, and phloem loading of the virus (For review, Hipper et al., 2013; Ivanov et al., 2014). VPg is an intrinsically disordered protein and known to have a flexible structure which endows the high versatility in function (Rantalainen et al., 2011).

ii. Viral RNA translation

The eIF4E's function in mRNA translation and the resemblance of the VPg-RNA structure to the 5′-cap-mRNA have strongly suggested that the VPg serves as a cap surrogate and recruits eIF4E in order to initiate translation of viral RNA. It is noteworthy that VPg of feline calicivirus (FCV; in the same *Pisuriviricota* phylum) interacts with eIF4E and this interaction is necessary for viral translation although direct analogy must be avoided as the VPg sizes are different and potyviral VPg is larger (20-22 kDa) than the caliciviral one (10-13 kDa) (For review, Goodfellow, 2011).

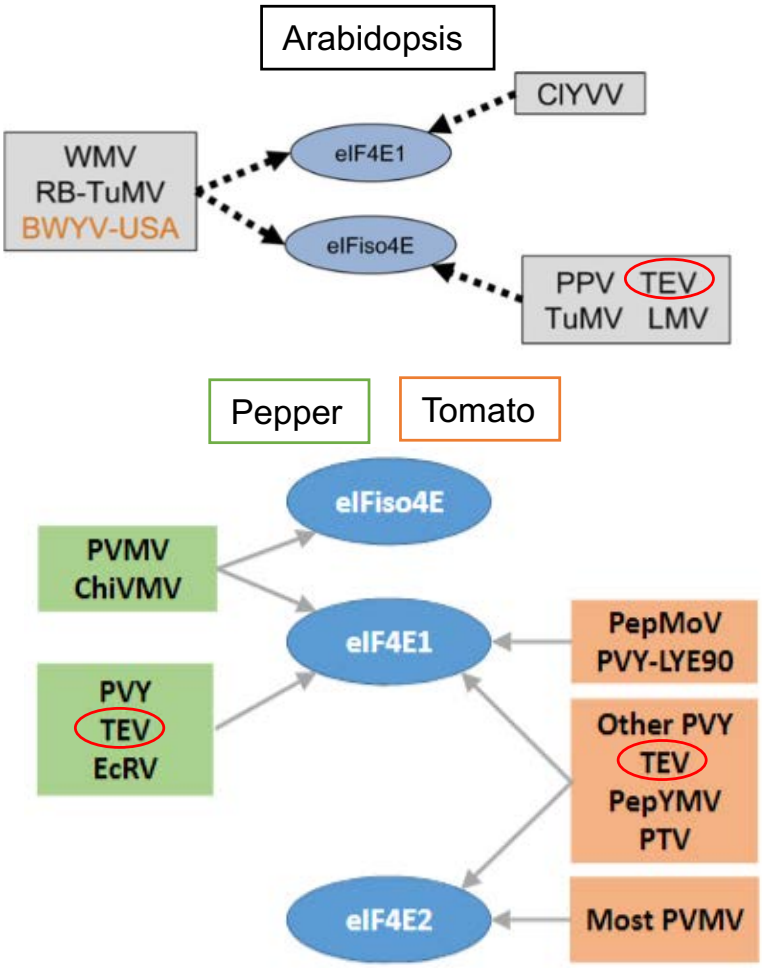


Figure i6-8 Potyvirus isolates have different preferences of eIF4E isoforms and it is dependent on host species.

Arrows point to the eIF4E isoform(s) (blue ovals) used by each virus in each host (indicated in open box). TEV (Tobacco etch virus) is in all pathosystem and is marked by red circle. TEV shows differential preferences of eIF4E isoforms between different host species.

WMV: Water melon mosaic virus; RB-TuMV: resistance-breaking Turnip mosaic virus; ClYVV: Clover yellow vein virus; PPV: Plum pox virus; LMV: Lettuce mosaic virus; BWYV-USA: Beet western yellows virus-USA (a polerovirus exceptionally shown). Modified from Bastet et al., 2018 and Moury et al.,2020.

In plants, the potyviral VPg's role as a cap substitute is supported by several observations. Firstly, it has been shown that the potyviral VPg and a cap analog m7GTP compete in binding the Arabidopsis eIFiso4E and that the VPg had a higher affinity for the eIFiso4E than the cap analog (Léonard et al., 2000; Miyoshi et al., 2006). Inhibition of mRNA translation by VPg was also observed *in vitro* and *in planta* (Miyoshi et al., 2008; Eskelin et al., 2011). Recently, structural resolution of VPg was performed using nuclear magnetic resonance (NMR), and VPg-eIF4E complex was constituted based on the modeling and the biophysical assays supporting the previous observations (Coutinho de Oliveira et al., 2019). Secondly, challenging Arabidopsis mutants with potyvirus shows that potyviral VPg recruits eIF4E isoforms and eIF4G isoforms in the coordinated manner (Wang et al., 2012b; Zhang et al., 2019c). Finally, the amino acid substitutions in natural eIF4E variants that are associated with virus resistance are located close to the cap-binding region; and these are considered responsible for disrupting the physical VPg-eIF4E interaction (For review, Robaglia and Caranta, 2006b; Wang and Krishnaswamy, 2012). The Y2H has been successfully used to experimentally validate the direct interactions in some pathosystems such as PVY-*pvr2* and TEV-*pvr2* in pepper, but not for other combinations of potyvirus and the host susceptible allele (Mazier et al., 2011; for review, Wang and Krishnaswamy, 2012; Estevan et al., 2014).

Interestingly, potyvirus has specific requirement of eIF4E isoforms in order to achieve infection (Nicaise et al., 2007; for review, Wang and Krishnaswamy, 2012). Some potyvirus isolates require one form of eIF4E or the other, while the rest can use both forms (**Figure i6-8**; arrows). Preference of one potyvirus isolate varies between host plant species (**Figure i6-8**; TEV). This indicates that the KO of specific eIF4E isoforms can provide resistance against potyvirus isolates that require these isoforms. Nevertheless, the mutant analyses showed that while inactivation of one copy of eIF4E/eIFiso4E provides resistant to some potyvirus isolates, it can be overcome by other isolates due to their evolution. This phenomenon is called resistance breaking and will be discussed in the next section.

iii. Resistance breaking

eIF4E-mediated resistance breaking (hereafter referred to as RB) is observed for some potyvirus isolates, and this is usually mediated by emergence of virus that contains amino acid substitutions in their VPg sequence (**Figure i6-9A, B**). It occurs when the mutations in VPg either restore the disrupted interaction with the resistance form of

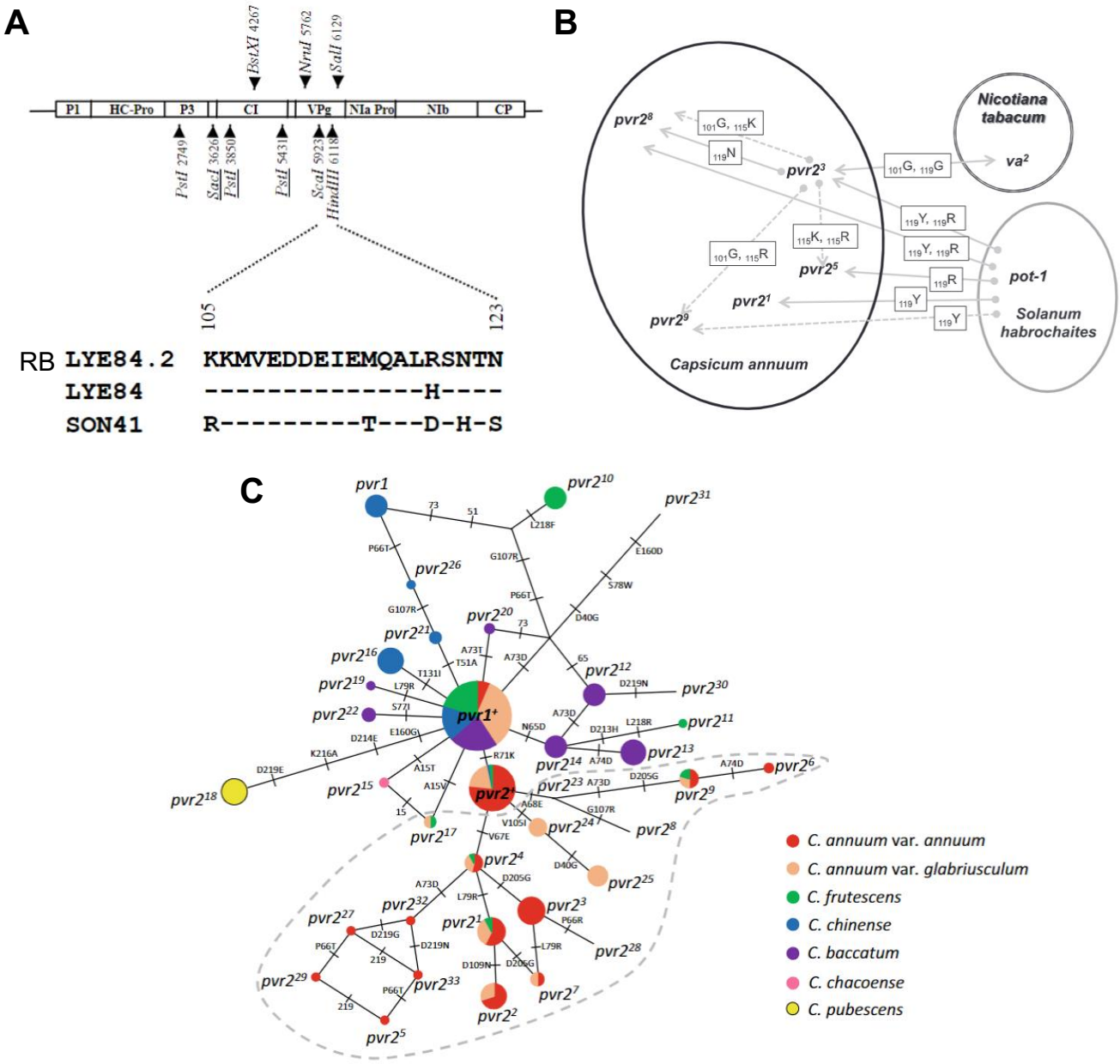


Figure i6-9 VPg mutation overcomes *eIF4E*-based resistance and promotes co-evolution of eIF4E resistance alleles.

A Central region (aa position 105-123) of potyviral VPg is responsible for resistance breaking in a host as well as virulence towards different hosts. LYE84.2 is a resistance breaking isolate of LYE84 and gains stronger virulence toward tomato. SON41 is a pepper-infecting isolate shown as a comparison. Moury et al., 2004. **B** Mutations in VPg gained by resistance breaking PVY isolates as they overcome each resistance alleles. The gained mutations are shown in open box and the number indicates the position of amino acid substitutions in the PVY VPg. Taken from Moury et al., 2014. **C** Minimum spanning network of eIF4E1 of 432 pepper genotypes from six *Capsicum* species including cultivars and wild accessions. Circle indicates each eIF4E1 allele. Colors in the circle indicate the proportion of the number of sequenced individuals from each pepper accession. The amino acid substitutions between each allele are indicated on the branches that connect the alleles (circles). Gray dotted line groups all *C. annuum* genotypes with potyvirus resistance. Refer to Tamisier et al., 2023 for detailed methods and references. Taken from Tamisier et al., 2023.

eIF4E or establish newly the interaction with the eIF4E isoform that is not originally used. The former is observed in a vast collection of pepper *pvr* alleles that show different spectra of resistance which are indeed reflection of co-evolution between PVY and pepper resistance eIF4E (**Figure i6-9B,C**) (Moury et al., 2004; Charron et al., 2008; for review, Moury et al., 2014; Tamisier et al., 2023). The latter has been demonstrated using Arabidopsis: RB TuMV isolate emerged in the *eifiso4e*^{KO} plants as it obtained mutations in VPg which enabled recruitment of the remaining eIF4E1b and eIF4E1c (Gallois et al., 2010; Bastet et al., 2018). Interestingly, many mutations associated with RB cluster in a specific, central part of the VPg protein in different pathosystems (Moury et al., 2004; Charron et al., 2008; Gallois et al., 2010; Zafirov et al., 2020).

There are two exceptional cases in which eIF4E-mediated potyvirus resistance was overcome by mutations in other viral proteins than VPg: one is due to mutations in clover yellow vein virus (ClYVV) P1 in pea and another in LMV CI in lettuce, via unknown mechanisms (Abdul-Razzak et al., 2009; Nakahara et al., 2010). However, most of RB events are accounted for by the mutation in VPg as seen in the clear co-evolutionary relationship between VPg and eIF4E in PVY-*pvr* combination (Charron et al., 2008; Moury et al., 2014).

iv. Other functions as a part of virus-induced granule: RNA silencing

In addition to the viral RNA translation, involvement of eIF4E/eIFiso4E in other processes of viral infection has been proposed due to its interaction with VPg and other viral proteins. For instance, eIF4E/eIFiso4E is a part of a potyvirus-induced granule (PG) which is formed upon TuMV or potato virus A (PVA; potyvirus) infection and composed of viral RNA and viral proteins including HC-Pro and P0 as well as several host factors (Hafrén et al., 2015; Hafrén et al., 2018; Saha and Mäkinen, 2020) (**Figure i6-10**). This PG was shown to be important for suppressing the RNA silencing and thereby protecting viral RNA from degradation. However, the study on eIF4E-based resistance has focused primarily on VPg-eIF4E interaction and the associated mutations. It would be interesting to explore how the eIF4E can be involved with other host or viral proteins which support the infection.

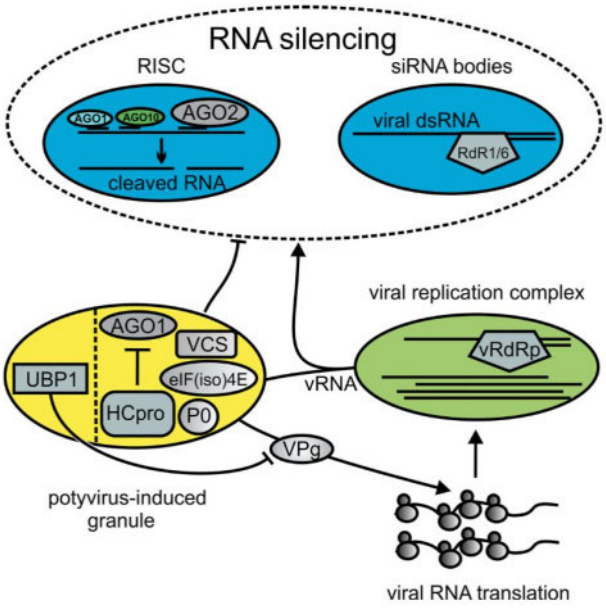


Figure i6-10 eIF4E/eIFiso4E plays a role in RNA silencing suppression as a part of potyvirus-induced granule (PG).

PG in yellow is composed of eIFiso4E and other host factors (VCS, AGO1, and UBP1) and potyviral factors (P0 and HC-Pro) and translationally inactive viral RNA. The PG protects the viral RNA from the host silencing machinery when viral translation is not supported. When the translation is supported, the VPg disrupts the PG and promotes translation together with PG components including eIFiso4E. The promotion of translation and PG formation for protecting viral RNA are interrelated and eIFiso4E seems to play role in both processes. Taken from Hafrén et al., 2015.

d. Implications of eIF4E-based potyvirus resistance for plant breeding

The eIF4E isoforms play a central role in potyvirus infection, and inactivating the protein or impairing its functional domain for viral infection, namely the VPg-interaction site, can bring about the virus resistance. This knowledge has provided us a few strategies to fight off against the potyviruses by cross breeding and mutational breeding of eIF4E variants. Today, such potyvirus resistance breeding can also be generated by using genome-editing which allows full exploitation of the knowledge from the natural eIF4E-based resistance.

i. Natural resistance alleles in cross breeding

Natural eIF4E-based potyvirus resistance has been implemented in the pepper breeding program before the mapping of the resistance locus (Greenleaf, 1986). Owing to the sexual compatibility, a variety of resistance *pvr* alleles found in *Capsicum* spp. could be integrated into the genome of the elite pepper cultivars by cross breeding. Indeed, the resistance *eIF4E* alleles can be used for resistance breeding of crops that are related or sexually compatible to the origin of such alleles. However, as the **Table i6-1** shows, the natural resistance alleles have been identified in a limited number of crop species or their relatives. Distantly related crops or asexually propagated crop species including potato (*S. tuberosum*) cannot be improved easily by this strategy.

ii. KO alleles by random and directed mutagenesis and their caveat

Generation of loss-of-function *eIF4E/eIFiso4E* alleles has proven to provide resistance against some potyvirus species and isolates. Random mutagenesis combined with TILLING is a powerful tool for obtaining plants with mutations in *eIF4E* genes as demonstrated in tomato (Piron et al., 2010). Single *eIF4E/eIFiso4E* KO plants show resistance to some potyvirus species and isolates in different plant species (Reviewed in Bastet et al., 2017) (**Table i6-2**). Although cost-intensive, mutational breeding aiming for inactivation of a copy of *eIF4E/eIFiso4E* gene is interesting for crop species for which the natural alleles are not available, in a transgene-free manner. Today, single KO alleles can also be generated much faster by genome-editing techniques such as CRISPR/Cas9, directly targeting the *eIF4E* gene to knock-out (First potyvirus resistance editing by CRISPR: Chandrasekaran et al., 2016).

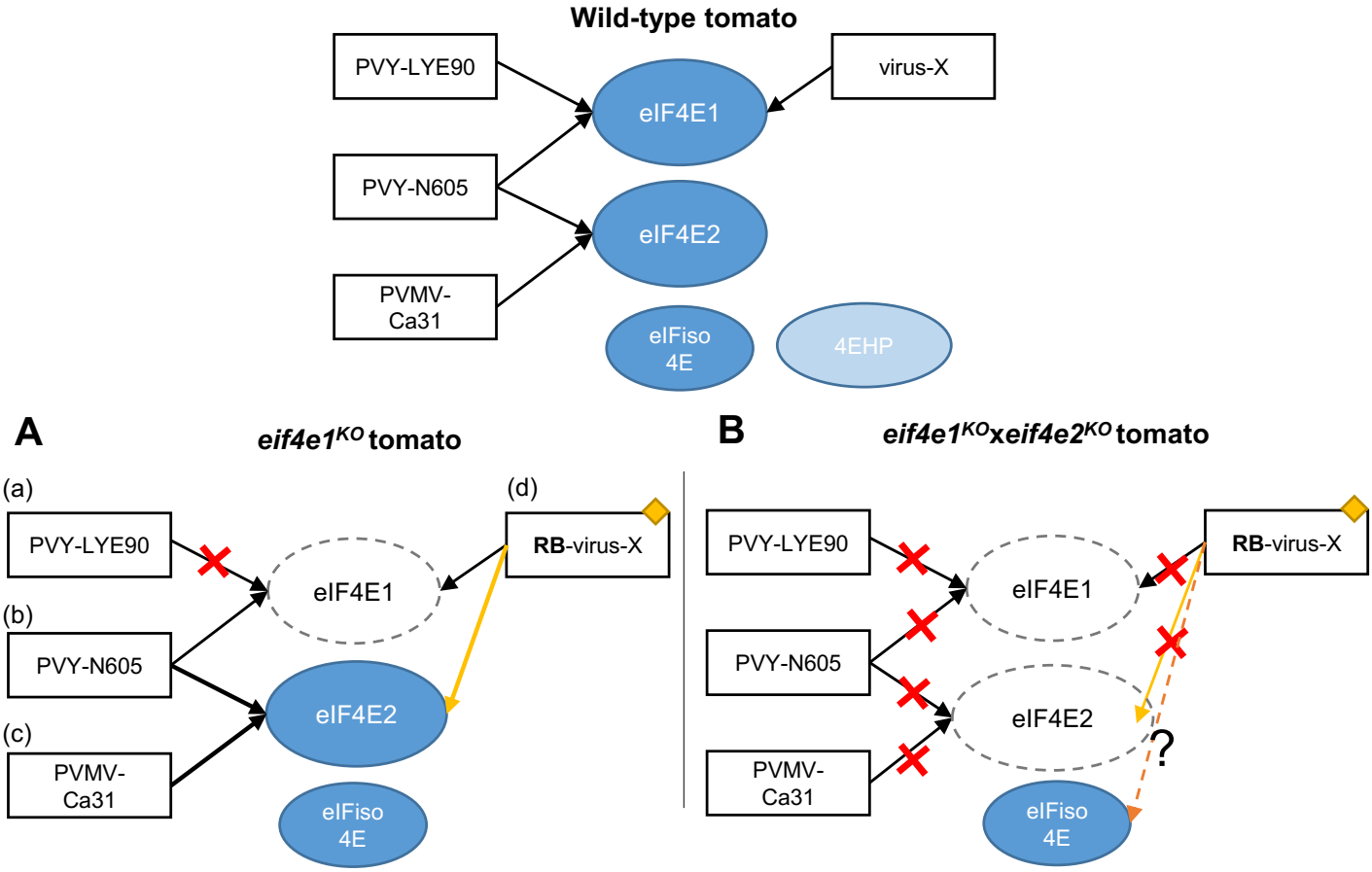


Figure i6-11 Schematics of consequences of single and double KO eIF4E isoforms in tomato in virus infection context.
Ovals represent eIF4E isoforms (canonical forms in blue, non-canonical 4EHP in light blue). In tomato, potyvirus that utilizes eIFiso4E has not been reported. Arrows indicate the preferences of eIF4E isoforms by each potyvirus. **A** Single e*if4e1*^{KO} tomato is resistant to a virus isolate that only uses one isoform (a). It is not resistant to virus isolates that can originally use either of the isoform (b) and that use only the other isoform. It is also not resistant to the isolate that originally uses the isoform targeted for KO but can evolve and gain the ability to use the other isoform (d; yellow rhombus indicates mutations that enable RB). (d) is represented as a hypothetical isolate virus-X. **B** Double e*if4e1*^{KO}x*eif4e2*^{KO} tomato is resistant to all virus isolates that require eIF4E1 or eIF4E2 or both for infection. Theoretically it is possible that a potyvirus isolate can overcome this type of resistance by gaining the ability to use the remaining isoform eIFiso4E, but this has not been observed in tomaot so far. Note that the double KO tomato is severely hampered in its growth.

However, a single KO strategy has a weak point: resistance conferred by single KO of one *eIF4E* isoform is limited to the viruses that can only use the same isoform (**Figure i6-11A**; (a)) and does not apply to virus isolates that only use the remaining isoform (b) and those that can use both isoforms by nature (c) or by evolution into resistance breaking isolate (d). The latter (b and d) is explained by the redundancy of the *eIF4E* genes because in a single KO plant, the remaining eIF4E isoform can compensate the pro-viral function, making the plant susceptible (Gauffier et al., 2016; Moury et al., 2020). Interestingly, the resistance spectrum provided by the single *eif4e1*^{KO} was much narrower than that by introgression of natural resistance *pot1* allele in tomato (Gauffier et al., 2016). In addition, single KO of an eIF4E isoform can induce hypersensibility to potyvirus that hijacks the remaining isoform in Arabidopsis (Zafirov et al., 2020).

On the other hand, inactivation of multiple eIF4E isoforms targeted by potyvirus confers a broad-spectrum resistance, as reported for both double KO and RNAi-mediated KD *eif4e1*x*eif4e2* tomato; however, such plant is hampered for growth, and so this strategy is not agronomically practical (Mazier et al., 2011; Gauffier et al., 2016) (**Figure i6-6, 11B**). Overall, mutational breeding is possible to generate eIF4E-based resistance, but studies have shown that single KO allele of *eIF4E/eIFiso4E* would provide narrow resistance and sometimes hypersusceptibility and that double KO alleles would be detrimental to the plant growth, both proving unable to replace the natural resistance allele. The redundancy and the importance of host eIF4E isoforms impose the trade-off between growth and resistance, making it difficult to deploy the KO strategies efficiently.

iii. New avenue for resistance breeding CRISPR-based base-editing

The caveat of KO strategies for developing potyvirus resistance has prompted switching to transferring the characteristics of natural resistance to the susceptible crops. The previous studies indicated that the natural *eIF4E* variants provide a wide-spectrum potyvirus resistance without apparent impairment in protein function and hence plant growth, unlike the loss-of-function alleles (cf. *Introduction 6.a.iii, 6.b.iv*). Ectopic expression of such resistance alleles from pepper has indeed successfully, and surprisingly due to its recessive nature, conferred TEV resistance in tomato and PVY resistance in potato, in a dominant manner (Kang et al., 2007; Cavatorta et al., 2011) (transgenesis; **Figure i6-12A**). Then, overexpression of designed potato *eIF4E1* containing amino acid substitutions of pepper resistance *pvr1*^{2} or natural resistance potato *eIF4E1* allele in potato (Cavatorta et al., 2011; Duan et al., 2012); and the native-

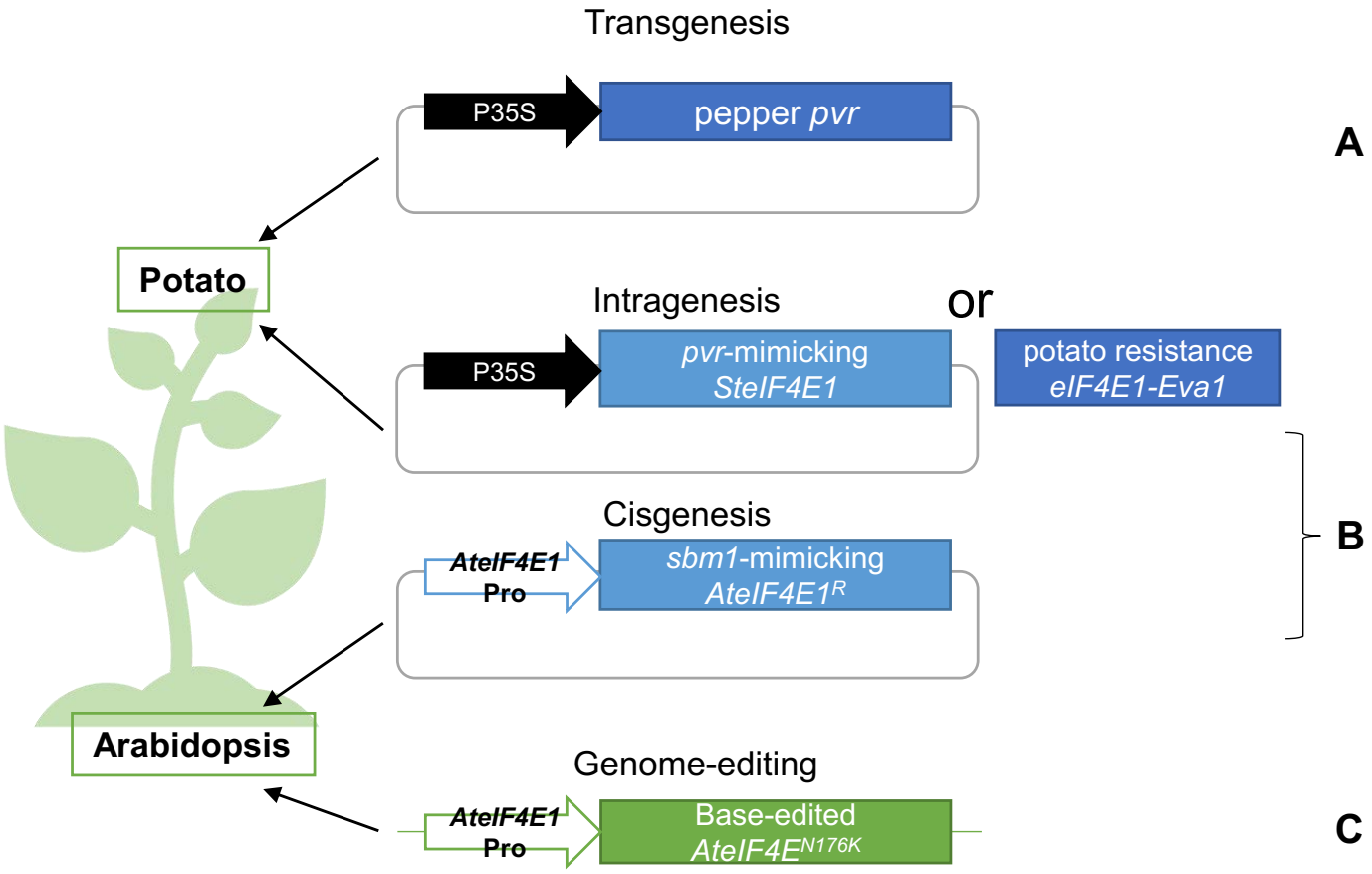


Figure i6-12 Schematics of transgenic and genome-editing strategies for developing *eIF4E-based* resistance.
Examples of experimental validations are presented. **A** Transgenic expression of resistance gene. Ectopic expression of resistance *pvr* alleles from pepper in potato under strong constitutive promoter CaMV P35S. Example from Cavatorta et al., 2011). **B** Intragenic or cisgenic expression of resistance gene. Ectopic expression of designed potato *SteIF4E1* allele mimicking pepper *pvr* allele in potato under strong constitutive promoter CaMV P35S (intragenesis) (Example from Cavatorta et al., 2011). Or ectopic expression of natural potato resistance *eIF4E1-Eva1* allele from wild potato relative species. (Example from Duan et al., 2012). Ectopic expression of designed Arabidopsis *AteIF4E1* allele mimicking pea *sbm1* allele in potato under native Arabidopsis *eIF4E1* promoter (cisgenesis) (Example from Bastet et al., 2018.) **C** Base-editing of susceptible gene into resistance gene. Direct, targeted modifications of susceptible *eIF4E1* gene in the Arabidopsis genome to a resistance allele (Example from Bastet et al., 2019). **A,B** schematics of plasmid constructs for ectopic expression. **C** schematic of the genomic structure of the edited gene in comparison.

level expression of the designed Arabidopsis *eIF4E1* allele based on pea resistance *sbm1* allele have also been shown to confer potyvirus resistance (Bastet et al., 2018) (intragenesis and cisgenesis; **Figure i6-12B**). These transgenic studies further reinforced the notion that the characteristics of natural resistance eIF4E alleles are transferrable and applicable to various plants across species.

The ensemble of natural resistance alleles found in a few crop species and their wild relatives point to the fact that majorly two regions clustered with these amino acid substitutions in the eIF4E/eIFiso4E are determinant of the VPg-eIF4E interaction and the virus infection (Robaglia and Caranta, 2006b) . Furthermore, several studies dissecting the roles of signature amino acids in the resistance eIF4E variant have shown that only one or few amino acid substitutions are enough to confer the plants virus resistance, while retaining the original function (German-Retana et al., 2008; Ashby et al., 2011; Bastet et al., 2019).

In Bastet et al. (2019), our team identified one amino acid substitution in eIF4E1 derived from *sbm1*-eIF4E that is responsible for functionality and resistance to three potyvirus species (and two isolates for one of them), N176K (Asparagine to Lysine at position 176 in *At*eIF4E1). Then, the CRISPR-based CBE was successfully used to directly modify the codon in Arabidopsis *eIF4E1* and induce the same amino acid substitution N176K. The base-edited *eIF4E1*^{N176K} mutant Arabidopsis showed resistance to ClYVV which only uses eIF4E1 as a susceptibility factor in Arabidopsis. This work was the first of proof-of-concept of generating potyvirus resistance by using the CRISPR-based base editor to modify only one or a few bases in susceptibility gene *eIF4E* without impacting the protein function (**Figure i6-12C**). The PoC also underscored the importance of curation of genetic information from the natural varieties which may serve as a rich source of various natural resistance, yet to be discovered.

The biggest advantages of engineering eIF4E-based resistance by base-editing are that it is possible to target a small region of the gene for editing, hence minimizing potential negative impact on the protein function, and that it is theoretically applicable to any crop cultivars and species which can be transformed. Potyvirus species altogether affect a large number of crop species, but they require the host eIF4E/eIFiso4E as susceptibility factor. Therefore, designing one or more host *eIF4E* isoform alleles into resistance variant by base-editing should be commonly effective in any host species. Because the base-editing does not insert or delete the genome, it is possible to add the

resistance *eIF4E* allele in the elite cultivar in which other pathogen resistance or other beneficial traits are cumulated (gene stacking). The PoC in Arabidopsis in Bastet et al. (2019) would serve as the basis of strategy for such wide application of genetic resistance against potyvirus.

Resistance by loss of susceptibility to potyviruses holds a great potential for today's crop breeding. Resistance allele of *S* gene can be integrated in the elite cultivars by cross breeding or transgenic breeding or generated by mutational breeding and genome-editing. Especially using the genome-editing, such resistance is engineerable by inactivating or modifying the susceptibility factors, in this case eIF4E/eIFiso4E, which are highly conserved in plants. To achieve this, functional analyses of the *S* gene and its natural or induced resistance variant are necessary to understand the mode of action of these factors during the infection and to pinpoint the characteristics of these factors that contribute to the differences in susceptibility/resistance to the virus. Such analyses found the basis for developing the resistance by random and targeted mutagenesis as seen in the extensive examples of of the eIF4E/eIFiso4E as S factor to the potyvirus. Based on this, in this thesis, two methods of engineering resistance *eIF4E* alleles using the genome-editing techniques will be demonstrated and evaluated in a crop model, cherry tomato, in an aim to provide the eIF4E-based resistance editing scheme that can be applied to other crop species.

7. Thesis objectives

This thesis comprises of four parts that altogether aim to evaluate the effectiveness of *eIF4E* genome-editing for developing potyvirus resistance without impairing plant growth in a crop species. Two major questions will be answered in main chapters: a) How can we generate eIF4E-based virus resistance in tomato using genome-editing techniques?; and b) Are the base-edited tomatoes free from any phenotypic consequences?

a. How can we generate eIF4E-based resistance using genome-editing?

To answer the first question, two strategies of developing eIF4E-based resistance were explored: generation of a single *eif4e*^{KO} allele and a base-edited *eIF4E* allele, in cherry tomato (*S. lycopersicum* var. *cerasiforme*) cv. West Virginia 106 (WVa106) as a fruit crop

model. Both strategies were aimed to confer resistance against potyvirus by loss-of-susceptibility, either via inactivation of the *S* gene or via modification of the interaction site with virus in the *S* gene, directly in the tomato genome.

i. Inactivation of the S gene SleIF4E2

In the first strategy, *SleIF4E2* gene was targeted for gene KO using CRISPR/Cas9 in cherry tomato (*Chapter 1*). This was in an attempt to evaluate the efficiency of the CRISPR/Cas9 in recreating the potyvirus resistance that was previously generated by the EMS KO of the *SleIF4E2* in another cultivated tomato variety, *S. lycopersicum* var. M82 (Moury et al., 2020). The results from this part underscore the efficiency of the CRISPR/Cas9 in transferring the resistance by gene KO established in one variety to another in tomato. In the meantime, the results also recapitulate the limitation of the single *eIF4E*^{KO}-based resistance, regardless of the methods, that has been previously suggested by different studies (as described in *Introduction 8.d.ii*) and calls for an alternative that bypasses the gene KO that is explored in *Chapter 2*.

ii. Modification of the S gene SleIF4E1 to mimic natural resistance allele

In the second strategy, *SleIF4E1* gene was sequentially targeted for introducing nonsynonymous mutations that share the similar characteristics with the natural resistance *eIF4E1* variant, using TALEN technologies and the CBE (*Chapter 2*). Two regions of *SleIF4E1* associated with the protein function in both mRNA translation and the potyvirus susceptibility/resistance were specifically targeted. One region (region I) was targeted by TALEN and the second (region II) by the CBE. In the region II, co-base-editing was performed, targeting the *SleIF4E1* and the reporter gene *SlALS1*, which can be selected based on the herbicide resistance (cf. *Introduction 2.f.ii*). The diversity of *de novo eIF4E1* alleles generated through the two editing experiments were then evaluated for their virus resistance spectra and the functionality of the coded protein.

The results from this part demonstrate that it is possible to generate a genetic diversity in the *SleIF4E1* allele with complex of several amino acid substitutions through iterative editing, and the different edited *SleIF4E1* alleles are associated with varying potyvirus spectra. The comparative studies of the edited alleles show that mutations in the two regions are necessary for resistance, which is in consistent with the characteristics of natural *eIF4E1* alleles.

Together, the two schemes (KO and base-editing) prove the efficacy of the genome-editing, especially CRISPR/Cas-based technologies, in generating loss-of-function or gain-of-function *eIF4E* mutants, thereby conferring potyvirus resistance. The comparison also proves that the gain-of-function (resistance) *eIF4E1* allele generated by base substitutions in the regions associated with virus susceptibility confers an extended resistance spectrum compared to the previously demonstrated loss-of-function *eIF4E1* allele. However, the base-edited resistance *eIF4E1* allele did not exhibit the same wide spectrum of potyvirus resistance as the natural resistance *pot1* allele from tomato wild relative, implying the room of improvement in the future editing.

b. Are the base-edited tomatoes free from any phenotypic consequences?

After we successfully generated functional resistance *SleIF4E1* allele by genome-editing (*Chapter 2*), we proceeded to answer important questions for practical application of the allele in the agronomic setting as a follow-up. First question to be addressed is whether the *SleIF4E1* base-editing would maintain *in planta* function in growth and development, and the second is whether the resistance conferred by this allele would be durable.

In an initial evaluation of the *SleIF4E1*-edited plant phenotypes, we noticed that some of the genotypes exhibited slower growth in greenhouse. Given the functionality of mutant *SleIF4E* alleles, we suspected that the *SlALS1* allele that was co-edited as a reporter of edition was responsible for the growth phenotype. To carry out the proper assessment of impact of each allele on agronomic traits, we generated progeny populations from the *SleIF4E1*x*SlALS1* co-base-edited populations: one with *SleIF4E1* mutation and free of *SlALS1* mutation, and another with *SlALS1* mutation and free of *SleIF4E1* mutation. In this thesis manuscript, preliminary, whole-plant phenotyping of these first segregating population will be presented (*SleIF4E1* mutant population in *Chapter 3*, *SlALS1* mutant population in *Chapter 4*). The *SleIF4E1* mutant population is also evaluated for the durability of resistance conferred by the *SleIF4E1* edition, an important aspect for implementation of new cultivars in agriculture (*Chapter 3*). The *SlALS1* mutant population is also analyzed at metabolomic level in order to identify the potential alterations in plant physiology leading to the growth phenotypes (*Chapter 4*).

At present (June 2024), the repetitions of the phenotypic evaluations by an engineer school student Anae Laurens (INRAE GAFL) and of the metabolomic profiling by

Anae Laurens and an assistant engineer Marie-Noëlle Corre (INRAE GAFL), using a larger number of plants and/or genotypes are in progress in order to confirm these preliminary results from the greenhouse assay in the spring of 2023 presented in the manuscript.

Chapter 1: Manuscript 1

Brief introduction

The present manuscript includes the CRISPR KO mutant generation by the team and the experiments I started as a part of my M2 internship project in spring 2021 and concluded during the first months into my PhD. This manuscript was submitted and published from the journal *Plant Science* on January 29th, 2022 (DOI: 10.1016/j.plantsci.2021.111160). Complete data including the supplementary data is provided in this chapter. A published version of the manuscript (main text without supplementary data) is available in the annex.

CRISPR-based knock-out of eIF4E2 in a cherry tomato background successfully recapitulates resistance to pepper veinal mottle virus

Kyoka Kuroiwa[1]*, Christina Thenault[1]*, Fabien Nogué[2], Laura Perrot[3], Marianne Mazier[1], Jean-Luc Gallois[1].**

* These authors contributed equally

** corresponding author

[1]INRAE, GAFL, 84143 Montfavet, France

[2]Institut Jean-Pierre Bourgin, INRAE, AgroParisTech, Université Paris-Saclay, 78000 Versailles, France

[3]Toulouse Biotechnology Institute, Université de Toulouse, 135 avenue de Rangueil, 31077 Toulouse CEDEX 04, France

Abstract

The host susceptibility factors are important targets to develop genetic resistances in crops. Genome editing tools offer exciting prospects to develop resistances based on these susceptibility factors, directly in the cultivar of choice.

Translation initiation factors 4E have long been known to be a susceptibility factor to the main genus of *Potyviridae*, potyviruses, but the inactivation of the *eIF4E2* gene has only recently been shown to provide resistance to some isolates of pepper veinal mottle virus (PVMV) in big-fruit tomato plants. Here, using CRISPR-Cas9-NG, we show how *eIF4E2* can be targeted and inactivated in cherry tomato plants. Three independent knockout alleles caused by indel in the first exon of *eIF4E2*, resulted in the complete absence of the eIF4E2 protein. All three lines displayed a narrow resistance spectrum to potyvirus, similar to the one described earlier for an *eIF4E2* EMS mutant of M82, a big-fruit tomato cultivar; the plants were fully resistant to PVMV-Ca31, partially to PVMV-IC and were fully susceptible to two isolates of PVY assayed: N605 and LYE84.

These results show how easily a resistance based on *eIF4E2* can be transferred across tomato cultivar, but also confirm that gene redundancy can narrow the resistances based on eIF4E knockout.

Keywords

eIF4E, CRISPR-Cas9, genome editing, virus resistance, Tomato, Cas9-NG, potyvirus

Introduction

The recently developed techniques of genome editing have already proved themselves to be essential tools for crop breeding. Among the various traits that can be modified through these techniques, resistance to pathogens appears as being particularly promising. One of the main ways to do so and generate resistance is by the ability to suppress or modify in the plant host genome the genes encoding the susceptibility factors that the pathogens do require to infect the plant. The modification of these susceptibility factors makes the plant resistant to the pathogen thereby helping reduce the use of pesticide during culture (Veillet et al., 2020b).

Among those susceptibility factors, the translation initiation factors eIF4E, which are encoded by a small multigene family, appear as very meaningful targets for crop breeding (Bastet et al., 2017). Initially, a translation initiation factor eIF4E was characterized among the natural variability of pepper (*Capsicum* species) as being responsible for resistance to potyviruses, the largest group of viruses affecting plants, including several isolates from potato virus Y (PVY) and tobacco etch virus (TEV) (Ruffel et al., 2002). Indeed, many alleles of the *eIF4E1* gene have been selected during pepper domestication: they are characterized by non-synonymous mutations that lead to resistance to potyviruses. It was interesting to note that those mutations do not affect the role of eIF4E1 in translation initiation (Charron et al., 2008). Translation initiation is essential for the plant physiology as it is an essential step to produce new proteins, so the selection of functional eIF4E factors could be a way of maintaining a trade-off between resistance and plant development during plant domestication. Following the characterization of *eIF4E* as a resistance gene to potyvirus in pepper, translation initiation factors have been shown to play a pivotal role in many crops by providing genetic resistance to potyviruses and related viruses (Robaglia and Caranta, 2006a; Bastet et al., 2017).

The ubiquitous role of eIF4E in resistance to potyvirus makes it a very exciting model for translational biology across species, or how a function that has been revealed in one species can be transferred to another one (Jacob et al., 2018; Schmitt-Keichinger, 2019). While pepper was instrumental in studying the co-evolution between natural eIF4E-based resistances and potyviruses, it is not well suited for biotechnology studies. On the other hand, the closely related Tomato (*Solanum lycopersicum*) was ideally suited and presents the same redundancy in the small multigene family encoding eIF4E as pepper does, with two genes encoding eIF4E proteins—eIF4E1 (Solyc03 g005870) and

eIF4E2 (Solyc02 g021550)—and one gene encoding a smaller protein named eIFiso4E (Solyc09 g090580)(Piron et al., 2010; Gauffier et al., 2016; Lebaron et al., 2016). Besides being a model plant, it is also a crop of major economical importance at the forefront of biotechnological innovations (Rothan et al., 2019). Thanks to the numerous genetic tools available in this species, it was possible to generate eIF4E-based resistance in cultivated tomato accessions and compare them with the natural resistance alleles that had been characterized in wild accessions (Ruffel et al., 2005; Lebaron et al., 2016). In addition to providing a better understanding of the virus/eIF4E interaction, the knowledge obtained can then be transferred to other cultures for which natural resistance to potyviruses is not available. Also, given the specifics of market requirements by the consumer, tomato posits essential challenges for the breeders, such as combining new resistances with other traits. Of those are fruit shape and size that make immediate appeal to the consumer (Rothan et al., 2019).

Initial works to generate resistances based on eIF4E in tomato were carried out using gene inactivation, through gene silencing or by chemical mutagenesis through TILLING (Piron et al., 2010; Mazier et al., 2011). Interestingly, those studies revealed important limitation of the gene knock-out strategies. Natural resistances to the two important potyviruses affecting tomato, PVY and TEV, consist in non-synonymous mutations in *eIF4E1* that can be associated with a large resistance spectrum to several virus isolates, whereas the *eIF4E1* knock-out (KO) tomato plants did not mimic the same resistance spectrum. These KO plants displayed a very narrow resistance spectrum to one isolate of PVY (LYE90), and to pepper mottle virus (PepMoV) while being fully susceptible to other PVY and TEV isolates tested (Piron et al., 2010; Gauffier et al., 2016). In addition, it was shown that, when *eIF4E1* is knocked-down or knocked-out, most PVY and TEV isolates can recruit eIF4E2, a very close homolog of eIF4E1. The emerging picture shows that in tomato, some potyvirus isolates can selectively hijack either eIF4E1 or eIF4E2 for their infectious cycle, whereas most can recruit both. Unfortunately, knocking out both *eIF4E1* and *eIF4E2* genes cannot be an option for tomato improvement because of detrimental effects on plant growth and development (Mazier et al., 2011; Gauffier et al., 2016). Interestingly, it was shown recently, that some isolates of pepper veinal mottle virus (PVMV), which is an important constraint on tomato production, did rely on eIF4E2 alone and big-fruit tomato plants knocked out for *eIF4E2* were resistant to several isolates of PVMV, making this gene an important target to develop resistance trait in tomato (Moury et al., 2020).

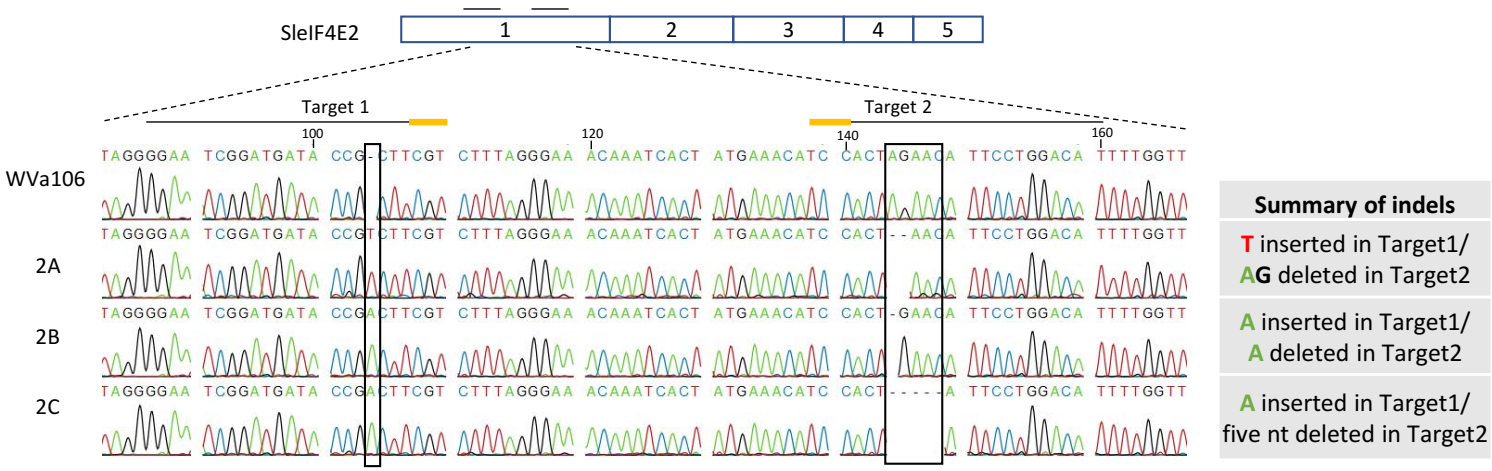


Figure 1 Editing at two targets in *eIF4E2* exon1 caused nucleotide insertion or deletion in the genomic *eIF4E2*. Genomic sequence of alleles of *eIF4E2* CRISPR mutants from lines 2A, 2B, and 2C. Target 1 and Target 2 represent the sequences of the two guide RNAs used in [24] and yellow bar indicates the PAM sequence for each target. The blue numbered boxes indicate the five exons of eIF4E2. Nucleotide position in respect to the WVa106 sequence is indicated at the top of the alignment. Summary of the modifications in each line is indicated on the right. nt:nucleotide.

The rapid development of genome editing technologies, based on the use of the CRISPR-Cas9 system soon appeared as an ideal tool to knock out genes in a given genotype. The expression *in planta* of a specific guide RNA and the Cas9 nuclease can lead to insertion or deletion of nucleotides at the targeted site in the genome. This system has been widely used to target gene coding sequence: the eventual indel leading to frameshift or inaccurate splicing results in incorrect protein translation, creating loss-of-function alleles. Interestingly those edited lines can be transgene-free by segregating away or not integrating the transgene harboring the Cas9 and sgRNA machinery (Veillet et al., 2019; Danilo et al., 2019). Such approaches have already been used to target translation initiation factors 4E. CRISPR-Cas9 inactivation of *eIFiso4E* in *Arabidopsis thaliana*, allowed to generate a resistance to turnip mosaic virus similar to the one that had been previously characterized using chemical or insertional mutation of the same gene (Lellis et al., 2002a; Duprat et al., 2002; Pyott et al., 2016). In cucumber (*Cucumis sativa*), inactivation of eIF4E resulted in resistance to zucchini yellow mosaic virus and papaya ring spot mosaic virus-W (Chandrasekaran et al., 2016). Likewise, inactivation of the two genes encoding the atypical translation initiation factors nCBP1 and nCBP2 resulted in reduced susceptibility and reduced symptoms severity to cassava brown streak disease (CBSV) in cassava (Gomez et al., 2019).

Recently, two studies in tomato showed that inactivation through a CRISPR-Cas9 strategy of *eIF4E1* leads to a narrow resistance spectrum to potyvirus (Yoon et al., 2020; Atarashi et al., 2020). In order to complete these studies, we present here the inactivation through CRISPR-Cas9 of *eIF4E2*. We could demonstrate that tomato plants knocked out for the *eIF4E2* gene are fully resistant to one isolate of PVMV but only partially to another one and fully susceptible to the two tested PVY isolates. These results confirm and complete what has been observed in big-fruit tomato plants in which the *eIF4E2* gene was inactivated through EMS (Gauffier et al., 2016; Moury et al., 2020). Finally, by developing the *eIF4E2* KO mutant in a cherry fruit accession, we show how resistance to PVMV can be easily translated into a genome background of choice using a CRISPR-Cas9 strategy.

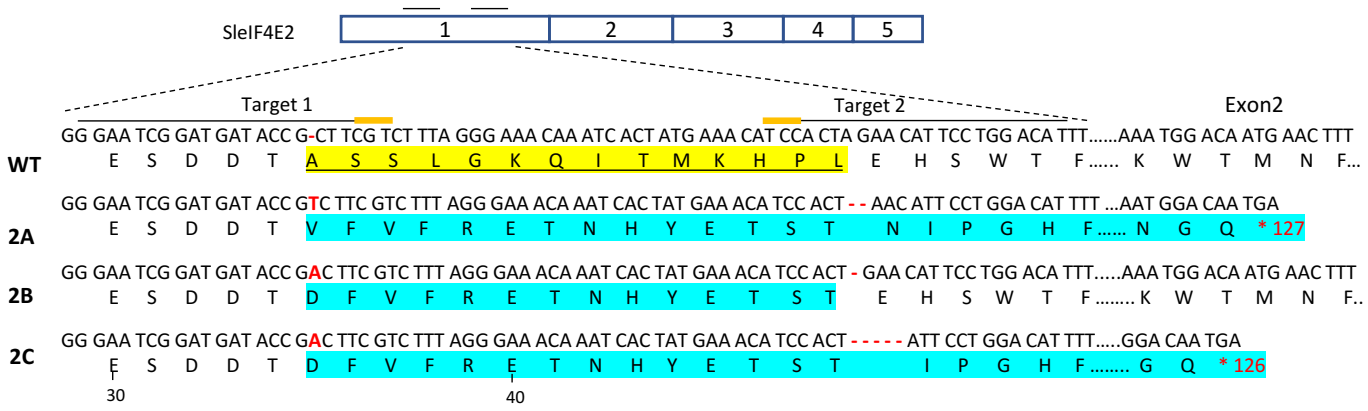


Figure 2 Indels in *eIF4E2* exon1 induced by CRISPR/Cas9 cause extensive amino acid changes and premature stop codons.

Genomic sequence of alleles of *eIF4E2* CRISPR mutants from lines 2A, 2B, and 2C. Target 1 and Target 2 represent the sequences of the two guide RNAs used in [24] and yellow bar indicates the PAM sequence for each target. The blue numbered boxes indicate the five exons of eIF4E2. Red letters or hyphens are the insertions or deletions (edited sites) in respect to WT sequence in the 2A, 2B and 2C genomic sequences. Red asterisks represent the premature stop codons in plant 2A and 2C. Highlighted in blue are amino acid substitutions in respect to WT amino acid sequence in 2A, 2B and 2C. Underlined (highlighted in yellow) is the amino acid sequence of WT that corresponds to the 14 AA modified in 2B. Amino acid position number is indicated at the bottom.

Results

Confirmation of efficient editing at two targets in eIF4E2 in the cherry tomato WVa106 background

We previously reported the editing at two closely located sequences of the tomato *SleIF4E2* gene, using a rationally engineered SpCas9 variant, the SpCas9-NG, that recognizes NGN protospacer adjacent motifs (Nishimasu et al., 2018; Veillet et al., 2020a). Based on the sequencing of both targeted regions in *eIF4E2* on primary tomato plant transformants (Veillet et al., 2020a), three independent T0 regenerated individuals were selected with potential mutations (indel) at both sites. The plants were allowed to self in greenhouse for seed production. Two successive screenings on T1 and T2 individualized progenies, monitored by HRM and sequencing, resulted in selecting T2 genotypes homozygous for the modifications in *eIF4E2* gene (Supplementary Fig. 1). A third selfing produced T3 seeds in which the T-DNA–harboring the SpCas9-NG construct—has been eliminated by segregation, as verified by PCR (Supplementary Fig. 2). The resulting plants are therefore edited but do not contain the transgene.

The three independent edited lines are hereafter named as 2A, 2B and 2C, in which genome modifications (insertion or deletion of bases) had occurred at both target regions. The respective alleles will be named $eIF4E2^{2A}$, $eIF4E2^{2B}$ and $eIF4E2^{2C}$. Sequence data from the screening was re-analyzed closely to determine the indels caused by genome editing (Fig. 1). Finally, cDNAs were amplified from all three lines to check 1/ that the expression of the *eIF4E2* was not significantly affected, and 2/ that the mutations were indeed present on the expressed mRNA (Supplementary Fig. 3). The *eIF4E2* cDNA from the $eif4e2^{KO}$ EMS mutant, that had been isolated in the M82 big-fruit tomato by TILLING, was sequenced and showed accordingly a G254A mutation resulting in a tryptophan to stop mutation at position 85 (Piron et al., 2010; Gauffier et al., 2016).

eIF4E2 encoded-protein may suggest different outcomes for the three edited alleles.

Based on the cDNA sequences obtained, the resulting translation product was predicted for each allele $eIF4E^{2A}$, $eIF4E^{2B}$ and $eIF4E^{2C}$, respectively (Fig. 2). For the three edited alleles, a single-base insertion at the first targeted site leads to a frameshift in translation resulting in AA changes starting at the 35th residue. However, based on modification occurring at the second target, two cases could be considered.

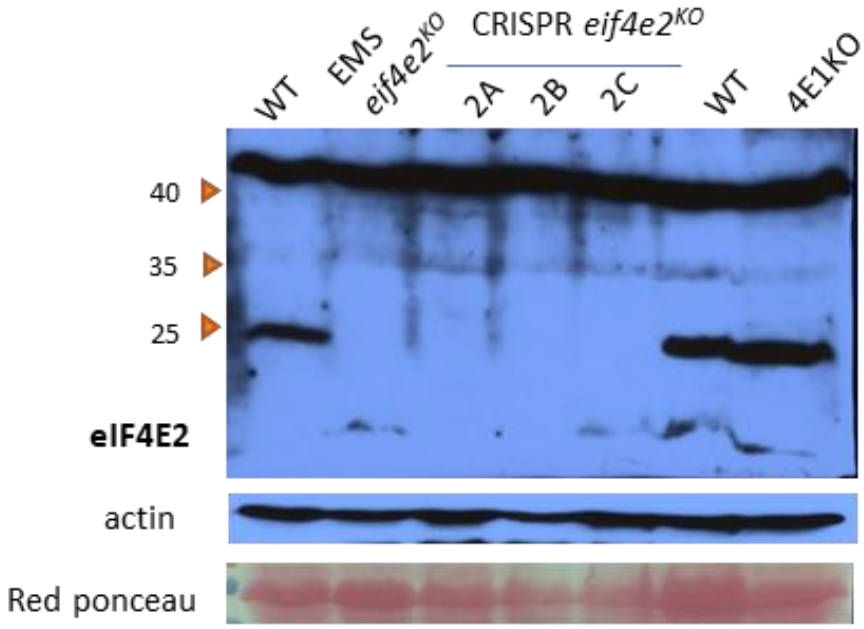


Figure 3 All three *eIF4E2* CRISPR-edited alleles are loss-of function alleles that do not accumulate eIF4E2.
Analysis of eIF4E2 protein accumulation in tomato leaves by western blot. WT, the wildtype tomato and the EMS *eif4e2*^{KO} mutant, are used as positive and negative controls, respectively, for eIF4E2 protein detection. The band relative to eIF4E2 stands at 25 kDa. Actin detection and red ponceau are shown as loading controls.

For *eIF4E*^{2A} and *eIF4E*^{2C}, the modifications occurring at the second targeted region did not restore the frame and the frameshift resulted in a premature stop codon at AA position 127 and 126 in the exon 2 out of total 221 amino acids, respectively. As a result, only the 34 first AA of the wild-type eIF4E2 are conserved in plants 2A and 2C, and more than 40 % of the C-terminus of the protein is missing. Based on that, *eIF4E2* alleles of 2A and 2C can be expected to be loss-of-function.

On the contrary, in plant 2B, the frameshift caused by the insertion of an A in target 1 was shifted back to the correct reading frame due to the deletion of another A in target 2. Thus, *eIF4E2*^{2B} encodes a predicted protein with modification of 14 amino acids within exon1-encoded region of the eIF4E2 protein, in the N-term region, while the rest of the protein–spanning AA 49 to 221—is similar to the wild-type eIF4E2 protein. The wild-type *eIF4E2* cDNA encodes a 221 Amino Acid (AA) protein, whose structure can be modelized based on its strong homologies with crystallized eIF4E proteins (Supplementary Fig. 4, (Moury et al., 2020)). As the structure of the 54 first amino acids of eIF4E protein have not been resolved by protein 3D crystallization, it is not possible to foresee whether this limited peptide change (highlighted in yellow on Supplementary Fig. 4) will have any effect on the protein expression or function.

Overall, protein predictions suggest that out of the three selected alleles, at least two are full loss-of-function alleles, while the question remains open on the third one.

Protein expressed from the three edited eIF4E2 alleles do not accumulate in planta.

To confirm the loss-of-function of alleles *eIF4E2*^{2A}and *eIF4E2*^{2C,} as well as to investigate the potential effect of the stretch of 14 AA changes in the N-term part of eIF4E2^{2B}, the accumulation of the eIF4E2 protein was investigated in tomato leaf tissues from the three genotypes. To this aim, we used a specific polyclonal antibody that had been raised against the tomato eIF4E2 full protein. The serum has been shown to be highly specific for the tomato eIF4E2 protein, as attested by analysis of protein accumulation in wild type plants and in the EMS *eif4e2*^{KO} EMS mutant ((Gauffier et al., 2016) and Fig. 3, first two lanes). Analysis of eIF4E2 protein accumulation in leaf tissues showed that none of the three edited alleles mutants accumulate any detectable eIF4E2 protein (Fig. 3). While this could be expected from plants carrying the *eIF4E2*^{2A}or *eIF4E2*^{2C} alleles, the same holds true for *eIF4E2*$^{2B.}$ This suggests that the 14 AA changes in the N-term part of the protein is sufficient to suppress the eIF4E2 protein accumulation, probably by triggering

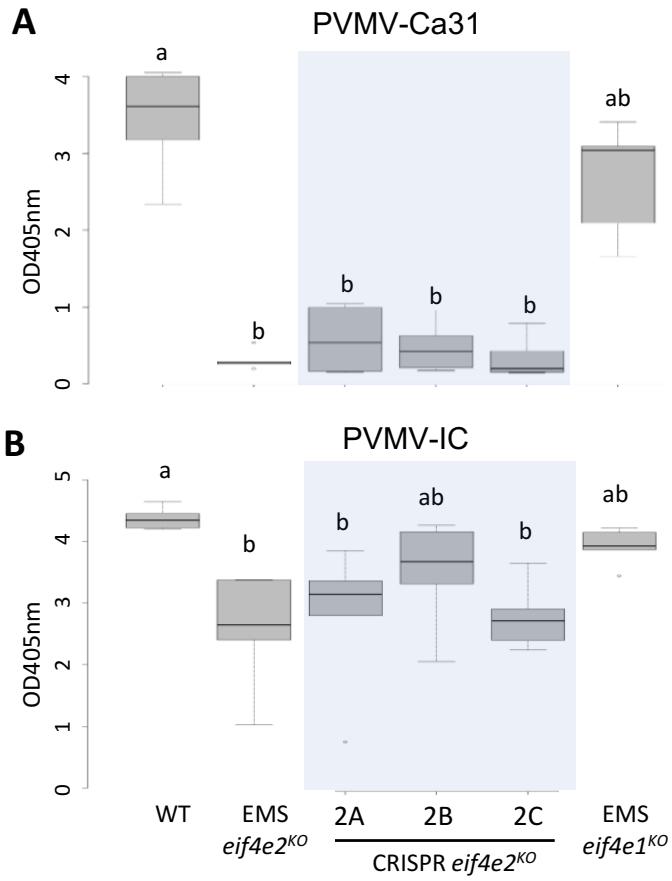


Figure 4 eIF4E2 CRISPR-edited loss-of function alleles are associated with resistance to PVMV-Ca31 but not to PVMV-CI.

(A, B) Resistance assay to PVMV isolates in tomato plants. The accumulation of viral coat proteins was detected by DAS-ELISA in tomato plant tissues from plants challenged with PVMV-Ca31 (A) or PVMV-IC (B). Samples were collected at 21 dpi. Different alphabets indicate the statistically significant difference ($p < 0.05$) after Kruskal-Wallis analysis. Six plants were tested par genotype.

its degradation. In conclusion, all three edited alleles can be considered as independent loss-of-function alleles for *eIF4E2* in the cherry tomato WVa106.

All three edited eIF4E2 alleles are associated with resistance to some PVMV isolates.

We next assessed whether the inactivation of *eIF4E2* in the cherry tomato WVa106 was associated with resistance to PVMV as it has been shown in big-fruit tomato. Previously, we showed that the *eif4e2^{KO}* EMS plants showed very contrasted response to different isolates of PVMV, ranging from full resistance, intermediate resistance, to full susceptibility (Moury et al., 2020). Out of those isolates we selected two, Cameroon 31 (Ca31) and Ivory Coast (IC), towards which the *eif4e2^{KO}* EMS plants responded with full and intermediate resistance, respectively. The edited plants were challenged with these isolates and the virus accumulation was assayed by DAS-ELISA after 21 days (Fig. 4). Results showed that the three edited lines, carrying either *eif4e2^{2A}*, *eif4e2^{2B}*, or *eif4e2^{2C}*, behaved similarly to the *eif4e2^{KO}* EMS plants, by being fully resistant to PVMV-Ca31 while accumulating significantly less PVMV-IC virus. By directly editing *eIF4E2* using CRISPR tools, we show therefore that we can reproduce the specific resistance to PVMV isolates in another cultivar of choice.

In contrast to PVMV, Moury et al. (Moury et al., 2020) showed that the plants carrying the EMS knock out in *eIF4E2* are fully susceptible to all other isolates of potyvirus tested, including PVY. We therefore challenged the three *eIF4E2*-edited lines with two isolates of PVY, N605 and LYE84. We found all three lines to be fully susceptible to the two PVY isolates (Fig.5). As for the *eif4e2^{KO}* EMS plants, it suggests that knocking out *eIF4E2* alone by CRISPR-Cas9 in the WVa106 background results in a very narrow resistance spectrum.

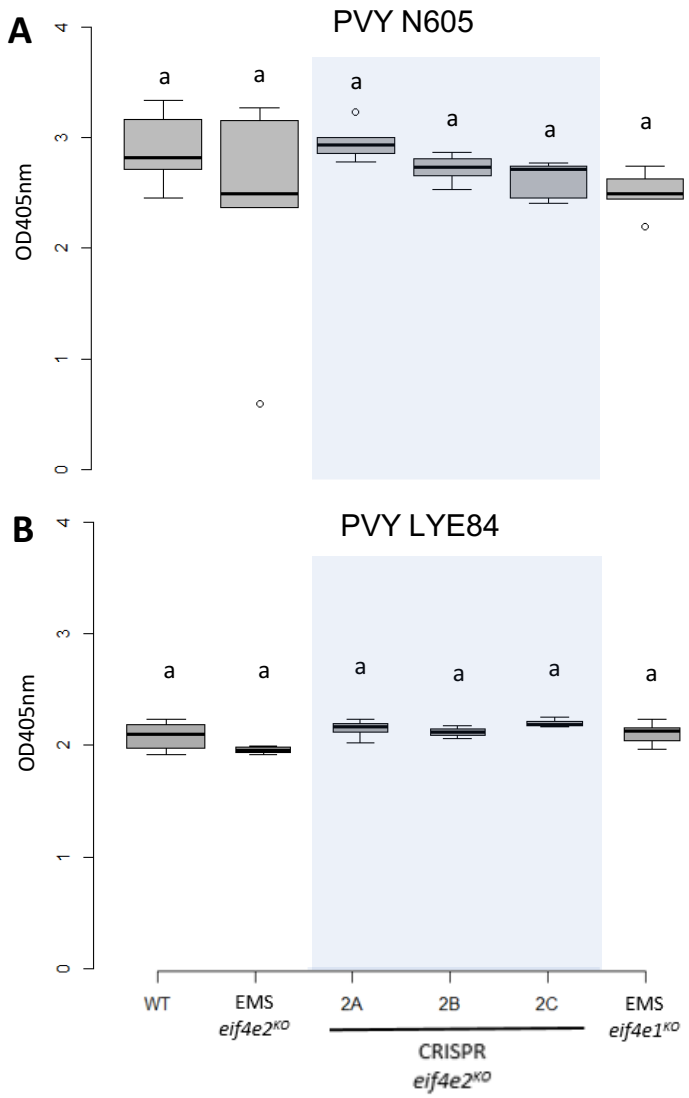


Figure 5 eIF4E2 CRISPR-edited loss-of function alleles are fully susceptible to PVY LYE84 and PVY N605.

(A, B) Resistance assay to PVY isolates in tomato plants. The accumulation of viral coat proteins was detected by DAS-ELISA in tomato plant tissues from plants challenged with PVY-LYE84 (A) and PVY-N605 (B). Samples were collected at 21 dpi. Different alphabets indicate the statistically significant difference ($p < 0.05$) after Kruskal-Wallis analysis. Six plants were tested par genotype.

Discussion

In this work, using genome editing of the susceptibility allele *eIF4E2* in tomato, we show quick and efficient development of genetic resistance to PVMV based on CRISPR-Cas gene knockout.

Here, we used a very straightforward approach by targeting two regions of eIF4E2, using the recently developed SpCas9-NG. The use of a Cas9 with a broader PAM selection (NGN instead of NGG), makes it easier to directly target the first exon of *eIF4E2*, in order to avoid the production of truncated protein with remnant activities. Indeed, we could select three independent tomato lines harboring loss-of-function in *eIF4E2*, as attested by the absence of the encoded protein accumulation in leaf tissues. At the same time, by selecting efficient editing at both targeted regions, we could isolate two types of mutants. In the first type, editing resulted in frameshifts and subsequent premature stop codons, leading to loss-of function alleles. For the other type, we isolated one allele (*eIF4E2*^{2B}) where the frameshift caused at the first target is canceled by a nucleotide insertion at the second target, leading to a short stretch of AA changes (spanning AA 35 to 48). We show that this modification is sufficient to obliterate the accumulation of the edited *eIF4E2* in tomato leaves although its mRNA is correctly produced. Little is known about the N-term part of eIF4E proteins. The first N-term part of eIF4E1 has not been crystallized yet, as its presence is detrimental to the crystallization process (Monzingo et al., 2007; Ashby et al., 2011; Miras et al., 2017). In addition, linking other proteins to the N-term or C-term parts of eIF4E is detrimental to its function (Levins et al., 2016). This could be caused either by destabilizing the protein or impairing protein-protein association. Here, our results show that modifying a short region of the N-term part of the eIF4E2 protein can result in its destabilization. Incidentally, this work here shows how combining several guides using low complexity PAM sequences could give opportunities to selectively mutate specific regions coding for domains, which could help characterizing their function.

One of the advantages of genome editing is the possibility to introduce mutations directly in the genetic background of choice. For the breeder, it implies the possibility to modify directly pre-breeding or breeding material and avoid reshuffling of traits that have already been carefully selected. Previous reports of edition of *eIF4E1*, the homolog of *eIF4E2*, have been carried out in the model *S. lycopersicum* cultivar Micro-Tom (Yoon et

al., 2020) and in a big-fruit tomato inbred line (Atarashi et al., 2020). Here, we implement a similar strategy in a cherry-tomato fruit, WVa106, belonging to the *Solanum lycopersicum* var. *cerasiforme,* and very closely related to the *Solanum pimpinellifolium* species (Ranc et al., 2008). Instead of eIF4E1, we decided to target eIF4E2, that has been previously characterized as a susceptibility factor to PVMV in a big-fruit tomato mutant issued from an EMS mutagenesis. We show here that we can easily translate such resistance from a big-fruit tomato to this new background. It has been documented that the genetic background of an accession can be instrumental for the efficiency of the genetic resistance trait or its durability (for review, (Gallois et al., 2018)). For recessive resistances such as the one relying on eIF4E, such effects are generally caused by modification of gene redundancy among the *eIF4E* gene family which can provide or remove alternative host factors that the virus hijacks. We show here that there is no genetic background effect between M82 and WVa106 affecting resistance to PVMV and thus, that there is not major reshuffling of redundancy between *eIF4E1* and *eIF4E2* between M82 and WVa106. This is in line with the very close relationship between *S. lycopersicum* and cherry tomatoes, such as WVa106 which results from the admixture of the *S lycopersicum* and *S pimpinellifolium* genomes (Ranc et al., 2008). However, it will be interesting to see whether these two genes, *eIF4E1* and *eIF4E2*, are fully redundant or not during the plant development: it has been shown that an EMS loss-of-function of *eIF4E2* is associated with a slight developmental delay (Gauffier et al., 2016), and it will be interesting to see whether this effect is conserved in WVa106 background.

Potyviruses have long been known to selectively target specific translation initiation factors eIF4E or eIFiso4E (Robaglia and Caranta, 2006a; Bastet et al., 2017). In tomato, such knowledge has been acquired through different biotechnological approaches, ranging from RNA interference, TILLING to now genome editing. These studies showed that most potyvirus isolates are able to recruit either eIF4E1 or eIF4E2 in tomato. A smaller subset of isolates does rely on eIF4E1 alone (PVY-LYE90 and PepMoV) while it was shown recently that some isolates of PVMV did rely on eIF4E2 alone (Piron et al., 2010; Mazier et al., 2011; Gauffier et al., 2016; Moury et al., 2020). These virus/host factor relationships are confirmed by the recent analysis of *eIF4E1* and *eIF4E2* knock-out alleles generated by genome editing ((Yoon et al., 2020; Atarashi et al., 2020) ; this study). Concerning the *eIF4E1* CRISPR-Cas9 knock-out Micro-Tom lines produced by Yoon et al. (Yoon et al., 2020), although the eIF4E1 protein accumulation was not confirmed in these

lines, the deletions caused by the edition presumably induced premature stop codon in the edited alleles. The resulting plants were found to be resistant to PepMoV, but not to TEV. This is in accordance with results obtained previously on tomato lines generated through TILLING and knocked-out for the same *eIF4E1* gene (Piron et al., 2010). In those plants, a splicing defect caused by an EMS mutation resulted in the absence of eIF4E1 protein accumulation. Those plants were fully resistant to PepMoV as well. From the more extensive resistance spectrum assayed with the TILLING lines, it can be expected that the plants edited for eIF4E1 will be also resistant to one isolate of PVY-LYE90, but susceptible to most PVY isolates and all TEV isolates tested so far (Piron et al., 2010; Gauffier et al., 2016). Concerning the *eIF4E2* CRISPR-Cas9 knock-out, we could confirm in this study the very limited resistance spectrum observed in a *eIF4E2* knock-out obtained by EMS (Moury et al., 2020). The three independent lines we generated showed full resistance to one PVMV isolate (PVMV-Ca31), intermediate resistance to another one (PVMV-IC) and full susceptibility to two PVY isolates -N605 and LYE84 that are known to rely on eIF4E1 and eIF4E2. As previously shown for other plant/potyviruses pathosystems (Gauffier et al., 2016; Bastet et al., 2018; Moury et al., 2020), it is very likely that differences between the PVMV genome sequences, and variability in the proteins they encode, are the basis for these differences in resistance by modulating the virus selectivity for one or several eIF4E factors.

Our results, together with those from Yoon et al. (Yoon et al., 2020) on tomato, highlight the fact that CRISPR-based genome editing is not going to be the silver bullet to achieve genetic resistances if we limit ourselves to gene knock-out. Beyond the very exciting opportunities provided by the CRISPR technology, what will matter is the mechanism by which the virus or other pathogens do subvert the host machinery. In the case of eIF4E-based resistance, the redundancy among 4E factors, as well as the virus's ability to break down resistance by recruiting other 4E homologs should be taken into considerations. As we previously stated, by taking advantage of the coevolution that happened during crop domestication, *eIF4E* natural resistance alleles could help provide a blueprint for generating resistance alleles while avoiding the drawback associated with loss-of-functions such as limited resistance spectrum, higher resistance breaking, or over-susceptibility mechanism (Gauffier et al., 2016; Bastet et al., 2017; Zafirov et al., 2021). It is all the more because pyramiding KO in 4E genes, that could extend resistance, is likely to be associated with plant developmental default due to the importance of translation

impairing the eIF4E translation initiation. A proof of concept for such an approach has been done in *Arabidopsis*, where a single non-synonymous mutation in the *eIF4E1* gene was found to be sufficient to trigger resistance to clover yellow vein virus while not impairing the plant yield (Bastet et al., 2019). A recent study provides exciting prospect along this line: by targeting the genomic sequence of *eIF4E1* in a tomato breeding line, Atarashi et al., (Atarashi et al., 2020) achieved targeted deletion of one or two codon in a specific region of the encoded eIF4E1 protein. Strikingly, the targeted mutation managed to reduce significantly the plant susceptibility to cucumber mosaic virus, a cucumovirus. Likewise, it is tempting to expect that a combination of multiplexed base editing, prime editing, or codon deletion are likely to provide new alleles of interest by generating a high level of diversity in the genetic backgrounds of choice. There again, by constituting a master plan for designing genes, it is to be expected that such strategy can be extended to other susceptibility factors to pathogens. This is of major importance as other host susceptibility factors targeted by pathogens may have very important functions for the plant development as well (Van Schie and Takken, 2014).

Overall, the feasibility of transferring resistances by loss-of-susceptibility based on CRISPR-Cas9 technologies is well acted, but idiosyncratic properties of these susceptibility factors can jeopardize their efficiency in the field. A major challenge for the years to come will be to assess on a case-by-case the efficiency to deliver efficient resistance traits for plant breeding. To do so, it will be worth providing more extensive resistance spectrum studies to characterize resistances, and at term one should start tackling the durability of those resistances.

Materials and methods

Plant material

West Virginia 106 (WVa106), a cherry fruit tomato (*Solanum lycopersicum* var. *cerasiforme*) which is highly susceptible to potyvirus, was used as background for the editing of *eIF4E2* (Solyc02g021550). The construct and methods to edit *eIF4E2* has been described previously (Veillet et al., 2020a). Briefly, two closely located loci of *eIF4E2* first exon were targeted through *Agrobacterium*-mediated transformation, using a construct expressing two guide RNAs and a *Streptococcus pyogenes* Cas9 NG (SpCas9-NG). Primary transformants edited in *eIF4E2* were selected through High Resolution Melting (HRM) analysis. The TILLING tomato lines *eIF4E1*^{KO} and *eIF4E2*^{KO} are knocked out for *eIF4E1* (Solyc03g005870) and *eIF4E2*, respectively. They have been obtained through Ethyl Methyl Sulfonate (EMS) mutagenesis in the *Solanum lycopersicum* M82 background and their characterization has been described earlier (Piron et al., 2010; Gauffier et al., 2016). All plants were grown in a growth chamber at 22°C during the day, 20°C during the night, 70% relative humidity and 16h day length.

Genotyping and selection of edited tomato lines

Genomic DNA was extracted from fresh leaf samples according to (Fulton et al., 1995).

Detection of the stable integration of the T-DNA was detected by PCR analysis from genomic DNA using the following 6 primers: AHAF, AHAR, Cas9F, Cas9R, kanaF and KanaR in the same 20 µl reaction containing the GoTaq® G2 Flexi DNA Polymerase (Promega, Madison, USA). All primer sequences are available in Supplementary data (Appendix A)
The High Resolution Melting-curve analysis was mostly performed as described in (Veillet et al., 2016). Briefly, PCR amplification of small amplifications products (131 bp) was carried out from genomic DNA using 4e2HRMF and 4e2HRMR primers and the Precision Melt Supermix (BioRad, Hercules, USA) with the CFX96™ Real-Time PCR Detection System (BioRad, Hercules, USA), according to the manufacturer's instructions. Melting-curves analysis was performed using the Precision Melt Analysis software (BioRad, Hercules, USA). Sequence information was obtained by amplifying the target locus with 4e2F and 4e2R primers and GoTaq® G2 Flexi DNA Polymerase (Promega, Madison, USA). PCR products were purified and Sanger sequenced (Genoscreen, Lille, France).

Total RNA extraction and cDNA synthesis

Total RNA was extracted from young tomato leaf tissue (< 100 mg) of each plant in 500 µL of TRI-Reagent (Sigma-Aldrich, St. Louis, USA) and quality-checked by concentration measurement on Nanodrop ND-1000 (ThermoFisher, Waltham, USA). Reverse transcription (RT) of the extracted RNA sample (1 µg) was performed using AMV (Avian Myeloblastosis Virus) Reverse Transcription Kit (Invitrogen, Waltham, USA) and oligo-dT18 (1 µM) to obtain coding DNA (cDNA).

The cDNA for *eIF4E2* and control *GAPDH* (Glyceraldehyde-3-Phosphate Dehydrogenase) were amplified by PCR using primer pairs Z1379F/Z1380F and Q3F/Q3R, respectively, and Sanger sequenced (Genoscreen, Lille, France).

Protein extraction and western blot analysis

Total protein was extracted from 100 mg of young tomato leaves of each plant, and 150 µL of Laemmli 2x DTT buffer was added to the crushed leaves. The homogenized solution was boiled for 5 min. The supernatant after centrifugation at 4 °C was collected and conserved at -20 °C.

The extracted total protein samples were run on the 12 % polyacrylamide gel in the SDS-PAGE system at 80 V for 2.5 hr to separate proteins by their sizes. Then, the separated proteins were transferred from the gel on to the nitrocellulose membrane (GE Healthcare, Chicago, USA) using semi-dry transfer method in 30-45 min. The membranes were locked with TBS 1x + 5 % milk overnight and then incubated with primary antibody for 2 hr and after with secondary antibody for 1 hr. The anti-Sl-eIF4E2 (Gauffier et al., 2016) serum was diluted 1/500, and combined with secondary goat horseradish peroxidase-labelled anti-rabbit serum (Sigma-Aldrich, St. Louis, USA) diluted 1/2000. As loading controls, monoclonal anti-plant actin antibodies (1/10000 dilution) (Sigma-Aldrich, St. Louis, USA) were used with horseradish peroxidase-labelled rabbit anti-mouse serum (1/10000 dilution) (Sigma-Aldrich, St. Louis, USA). The results of immunoblotting were revealed on autoradiography films Blu-Lite (MTC Bio, Sayreville, USA) using LumiGLO Reserve® Chemiluminescent Substrate Kit (LGC Seracare, Milford, USA).

3D modelization of eIF4E2 structure

Homology modelling of the eIF4E2 protein were carried out as previously (Moury et al., 2020) using the YASARA software (http://www.yasara.org/), using structural data from pea (*Pisum sativum*) eIF4E (GenBank ID: AY423375, PDB ID: 2WMC-C) as the template. Protein structure was visualized using PyMol software (https://pymol.org/).

Virus isolates and infectious assays

Potyvirus infection assays were performed using PVY-N605 (Parrella et al., 2002), PVY-LYE84 (Moury et al., 2004), PVMV isolates Cameroon 31 (PVMV-Ca31) and Ivory Coast (PVMV-IC) (Moury et al., 2020). Isolates were maintained in *Nicotiana benthamiana* plants 12 days before inoculation.

Inocula were prepared from 1 g (fresh weight) of the infected tobacco leaves ground with 4 mL of grinding buffer: potassium phosphate buffer (0.03 M, pH = 7) with 0.2% of diethyldithiocarbamate, 80 mg active charcoal and 80 mg carborundum. Six plants per genotype were mechanically inoculated by manually rubbing cotyledons of 14 day-old tomato plants with inoculum extract and rinsed with water 5 min later. Infected tomato plants were grown in a growth chamber in the same condition of photoperiod and temperature as before inoculation. At 21 days post-inoculation (dpi), 1 g of leaves from each plant was collected and ground in 4 mL of grinding buffer. For the PVMV tests only, the extracts were subjected to a serial dilution (1:25) with the same grinding buffer. To evaluate the accumulation for each virus, DAS-ELISA were performed using anti-PVY and anti-PVMV antisera (Sediag, www.sediag.com) according to the manufacturer's instructions. A pool of non-inoculated plants was used as healthy control. The absorbance was measured by a spectrophotometer at wavelength 405 nm.

Statistical analysis

Kruskal–Wallis non-parametric tests were performed to identify significant differences between genotypes using the pgirmess package on the free software R (http://www.r-project.org/).

Acknowledgments

We thank Benoît Moury for advising and supplying virus isolates as well as for helpful advices for ELISA, Mathilde Causse for discussions, Delyan Zafirov for advices on western blots, Emmanuel Botton for taking care of tomato plants and Nathalie Truglio for supplying tobacco plants.

Funding

This work was supported by the French National Research Agency (ANR11-BTBR-0001-GENIUS). The IJPB benefits from the support of the LabEx Saclay Plant Sciences-SPS (ANR-

10-LABX-0040-SPS). KK was supported by the Ito foundation (Japan, N0257) and Bourses du Gouvernement Français 2020-2021 (B-20001 JCA/NH).

Author contributions

FN, MM and JLG conceived and designed the experiments. KK, CT, LP, MM performed and interpreted the experiments. KK and CT designed the figures. JLG wrote the manuscript with inputs from all authors. All authors have approved the final manuscript.

Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Ashby, J. A., Stevenson, C. E. M., Jarvis, G. E., Lawson, D. M., and Maule, A. J. (2011). Structure-based mutational analysis of eIF4E in relation to sbm1 resistance to pea seed-borne mosaic virus in pea. PLoS One 6:e15873.

Atarashi, H., Jayasinghe, W. H., Kwon, J., Kim, H., Taninaka, Y., Igarashi, M., Ito, K., Yamada, T., Masuta, C., and Nakahara, K. S. (2020). Artificially edited alleles of the eukaryotic translation initiation factor 4E1 gene differentially reduce susceptibility to cucumber mosaic virus and potato virus Y in tomato. Front. Microbiol. 11:564310.

Bastet, A., Robaglia, C., and Gallois, J. L. (2017). eIF4E Resistance: Natural Variation Should Guide Gene Editing. Trends Plant Sci. 22:411–419.

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

Bastet, A., Zafirov, D., Giovinazzo, N., Guyon-Debast, A., Nogué, F., Robaglia, C., and Gallois, J. L. (2019). Mimicking natural polymorphism in eIF4E by CRISPR-Cas9 base editing is associated with resistance to potyviruses. Plant Biotechnol. J. 17:1736–1750.

Callot, C., and Gallois, J. L. (2014). Pyramiding resistances based on translation initiation factors in Arabidopsis is impaired by male gametophyte lethality. Plant Signal. Behav. 9:e27940.

Chandrasekaran, J., Brumin, M., Wolf, D., Leibman, D., Klap, C., Pearlsman, M., Sherman, A., Arazi, T., and Gal-On, A. (2016). Development of broad virus resistance in non-transgenic cucumber using CRISPR/Cas9 technology. Mol. Plant Pathol. 17:1140–1153.

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

Danilo, B., Perrot, L., Mara, K., Botton, E., Nogué, F., and Mazier, M. (2019). Efficient and transgene-free gene targeting using Agrobacterium-mediated delivery of the CRISPR/Cas9 system in tomato. Plant Cell Rep. 38:459–462.

Duprat, A., Caranta, C., Revers, F., Menand, B., Browning, K. S., and Robaglia, C. (2002). The Arabidopsis eukaryotic initiation factor (iso)4E is dispensable for plant growth but required for susceptibility to potyviruses. Plant J. 32:927–934.

Fulton, T. M., Chunwongse, J., and Tanksley, S. D. (1995). Microprep protocol for extraction of DNA from tomato and other herbaceous plants. Plant Mol. Biol. Rep. 13:207–209.

Gallois, J.-L., Moury, B., and German-Retana, S. (2018). Role of the Genetic Background in Resistance to Plant Viruses. Int. J. Mol. Sci. 19.

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

Gomez, M. A., Lin, Z. D., Moll, T., Chauhan, R. D., Hayden, L., Renninger, K., Beyene, G., Taylor, N. J., Carrington, J. C., Staskawicz, B. J., et al. (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 Biotechnol. J. 17:421–434.

Jacob, P., Avni, A., and Bendahmane, A. (2018). Translational Research: Exploring and Creating Genetic Diversity. Trends Plant Sci. 23:42–52.

Lebaron, C., Rosado, A., Sauvage, C., Gauffier, C., German-Retana, S., Moury, B., and Gallois, J. L. (2016). A new eIF4E1 allele characterized by RNAseq data mining is associated with resistance to potato virus Y in tomato albeit with a low durability. J. Gen. Virol. 97:3063–3072.

Lellis, A. D., Kasschau, K. D., Whitham, S. A., and Carrington, J. C. (2002). Loss-of-susceptibility mutants of Arabidopsis thaliana reveal an essential role for eIF(iso)4E during potyvirus infection. Current Biology 12.

Levins, E., Tseng, C.-Y., Patrick, R. M., Mayberry, L. K., Cole, N., and Browning, K. S. (2016). Fusion proteins of Arabidopsis cap-binding proteins: Cautionary “tails” of woe. Translation 4.

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

Miras, M., Truniger, V., Silva, C., Verdaguer, N., Aranda, M. A., and Querol-Audí, J. (2017). Structure of eIF4E in Complex with an eIF4G Peptide Supports a Universal Bipartite Binding Mode for Protein Translation. Plant Physiol. 174:1476–1491.

Monzingo, A. F., Dhaliwal, S., Dutt-Chaudhuri, A., Lyon, A., Sadow, J. H., Hoffman, D. W., Robertus, J. D., and Browning, K. S. (2007). The structure of eukaryotic translation initiation factor-4E from wheat reveals a novel disulfide bond. Plant Physiol. 143:1504–1518.

Moury, B., Morel, C., Johansen, E., Guilbaud, L., Souche, S., Ayme, V., Caranta, C., Palloix, A., and Jacquemond, M. (2004). Mutations in potato virus Y genome-linked protein determine

virulence toward recessive resistances in Capsicum annuum and Lycopersicon hirsutum. Mol. Plant. Microbe. Interact. 17:322–329.

Moury, B., Lebaron, C., Szadkowski, M., Ben Khalifa, M., Girardot, G., Bolou Bi, B. A., Koné, D., Nitiema, L. W., Fakhfakh, H., and Gallois, J.-L. (2020). Knock-out mutation of eukaryotic initiation factor 4E2 (eIF4E2) confers resistance to pepper veinal mottle virus in tomato. Virology 539:11–17.

Nishimasu, H., Shi, X., Ishiguro, S., Gao, L., Hirano, S., Okazaki, S., Noda, T., Abudayyeh, O. O., Gootenberg, J. S., Mori, H., et al. (2018). Engineered CRISPR-Cas9 nuclease with expanded targeting space. Science 361:1259–1262.

Parrella, G., Ruffel, S., Moretti, A., Morel, C., Palloix, A., and Caranta, C. (2002). Recessive resistance genes against potyviruses are localized in colinear genomic regions of the tomato (Lycopersicon spp.) and pepper (Capsicum spp.) genomes. Theor. Appl. Genet. 105:855–861.

Patrick, R. M., Mayberry, L. K., Choy, G., Woodard, L. E., Liu, J. S., White, A., Mullen, R. A., Tanavin, T. M., Latz, C. A., and Browning, K. S. (2014). Two arabidopsis loci encode novel eukaryotic initiation factor 4E isoforms that are functionally distinct from the conserved plant eukaryotic initiation factor 4E. Plant Physiology 164.

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

Pyott, D. E., Sheehan, E., and Molnar, A. (2016). Engineering of CRISPR/Cas9-mediated potyvirus resistance in transgene-free Arabidopsis plants. Mol. Plant Pathol. 17:1276–1288.

Ranc, N., Mũos, S., Santoni, S., and Causse, M. (2008). A clarified position for solanum lycopersicum var. cerasiforme in the evolutionary history of tomatoes (solanaceae). BMC Plant Biology 8.

Robaglia, C., and Caranta, C. (2006). Translation initiation factors: a weak link in plant RNA virus infection. Trends in plant science 11:40–45.

Rothan, C., Diouf, I., and Causse, M. (2019). Trait discovery and editing in tomato. Plant J. 97:73–90.

Ruffel, S., Dussault, M. H., Palloix, A., Moury, B., Bendahmane, A., Robaglia, C., and Caranta, C. (2002). A natural recessive resistance gene against potato virus Y in pepper corresponds to the eukaryotic initiation factor 4E (elF4E). Plant Journal 32.

Ruffel, S., Gallois, J. L., Lesage, M. L., and Caranta, C. (2005). The recessive potyvirus resistance gene pot-1 is the tomato orthologue of the pepper pvr2-eIF4E gene. Mol. Genet. Genomics 274:346–353.

Schmitt-Keichinger, C. (2019). Manipulating Cellular Factors to Combat Viruses: A Case Study From the Plant Eukaryotic Translation Initiation Factors eIF4. Front. Microbiol. 10:17.

Van Schie, C. C. N., and Takken, F. L. W. (2014). Susceptibility genes 101: How to be a good host. Annu. Rev. Phytopathol. 52:551–581.

Veillet, F., Gaillard, C., Coutos-Thévenot, P., and La Camera, S. (2016). Targeting the AtCWIN1 Gene to Explore the Role of Invertases in Sucrose Transport in Roots and during Botrytis cinerea Infection. Front. Plant Sci. 7:1899.

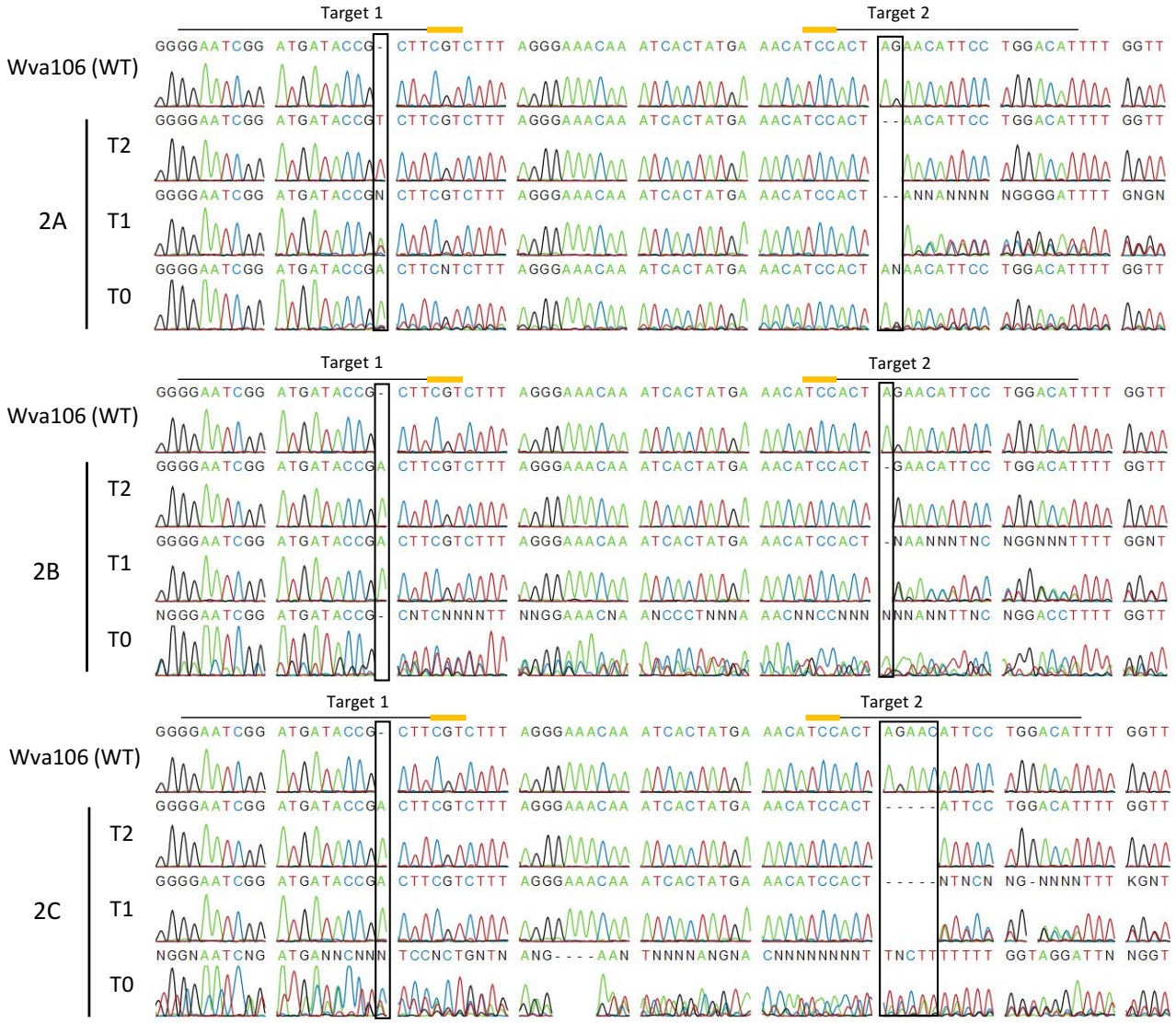
Veillet, F., Perrot, L., Chauvin, L., Kermarrec, M. P., Guyon-Debast, A., Chauvin, J. E., Nogué, F., and Mazier, M. (2019). Transgene-free genome editing in tomato and potato plants using Agrobacterium-mediated delivery of a CRISPR/Cas9 cytidine base editor. Int. J. Mol. Sci. 20.

Veillet, F., Perrot, L., Guyon-Debast, A., Kermarrec, M.-P., Chauvin, L., Chauvin, J.-E., Gallois, J.-L., Mazier, M., and Nogué, F. (2020a). Expanding the CRISPR Toolbox in P. patens Using SpCas9-NG Variant and Application for Gene and Base Editing in Solanaceae Crops. Int. J. Mol. Sci. 21.

Veillet, F., Durand, M., Kroj, T., Cesari, S., and Gallois, J. L. (2020b). Precision Breeding Made Real with CRISPR: Illustration through Genetic Resistance to Pathogens. Plant Communications 1.

Yoon, Y.-J., Venkatesh, J., Lee, J.-H., Kim, J., Lee, H.-E., Kim, D.-S., and Kang, B.-C. (2020). Genome editing of eIF4E1 in tomato confers resistance to pepper mottle virus. Front. Plant Sci. 11:1098.

Zafirov, D., Giovinazzo, N., Bastet, A., and Gallois, J.-L. (2021). When a knockout is an Achilles' heel: Resistance to one potyvirus species triggers hypersusceptibility to another one in Arabidopsis thaliana. Molecular Plant Pathology 22:334–347.



Supplementary Figure 1 Genomic sequence of parental selected individuals that generated the *eIF4E2* CRISPR T3 mutant lines.

Chromatograms covering the *eIF4E2* targeted region. Target 1 and Target 2 represent the sequences of the two guide RNAs used in the SpCas9-NG-SleIF4E2 construct.

The genealogies are presented for the 2A, 2B, and 2C genotypes (respectively **A**, **B**, and **C**).

The black open boxes show the position in the sequences of the final insertions/deletions obtained in the selected T3 lines.

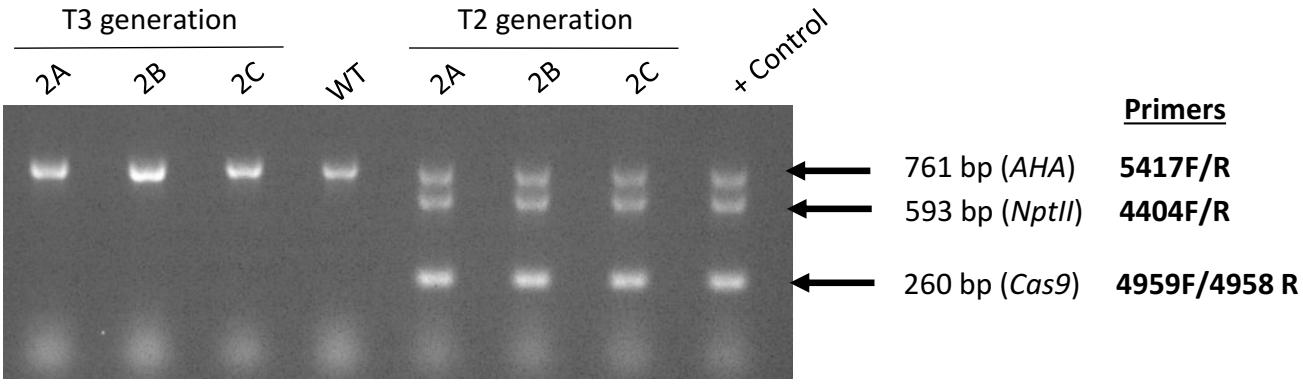
WT: wild type sequence of the Wva106 genotype used for genetic transformation.

T0: primary transformed plant directly regenerated after genetic transformation experiment using the SpCas9-NG-SleIF4E2 construct.

T1: progeny obtained by selfing of the T0 plant.

T2: progeny obtained by selfing of the T1 plant (2 successive generations from the T0). By selfing, this individual gave the T3 progenies homozygous for the modifications used in the present study.

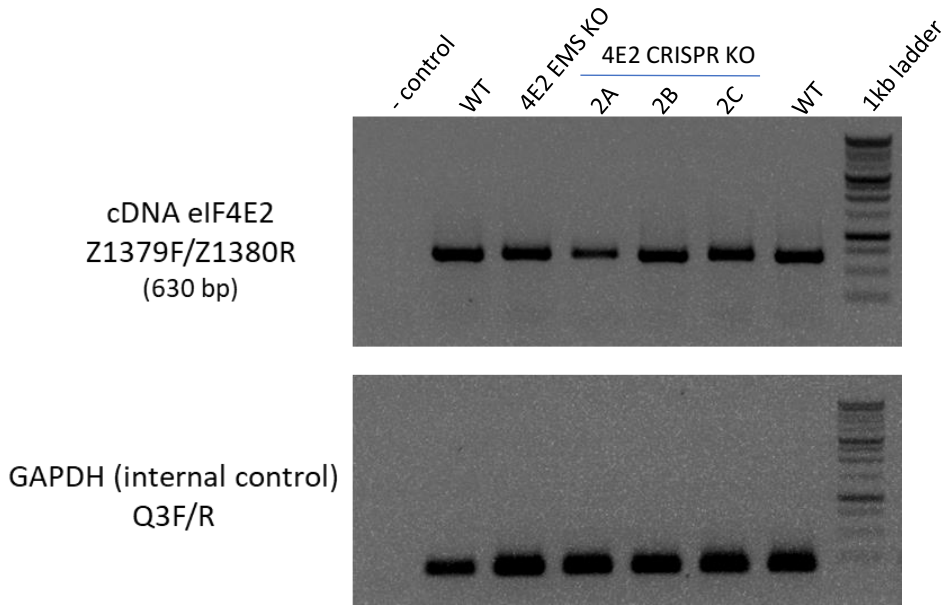
Due to overlapping sequences caused by indel, some T0 sequences (T0 2C) present double peaks in the mid section.



Supplementary Figure 2 Triplex-PCR genotyping of genome-edited 2A, 2B, and 2C shows the absence of T-DNA in generation T3.

PCR amplification on genomic DNA extracted from plants from T2 and T3 generations for the three selected edited lines: 2A, 2B, and 2C. DNA wild-type WVa106 (WT) and a from a plant harboring the Cas9 construct (+ Control) are shown. *AHA* (Acetolactate synthase) is amplified as an endogenous gene control while *NptII* and *Cas9* are both located on the T-DNA that was transformed. Plants for T3 do not harbor anymore the transgene.

A



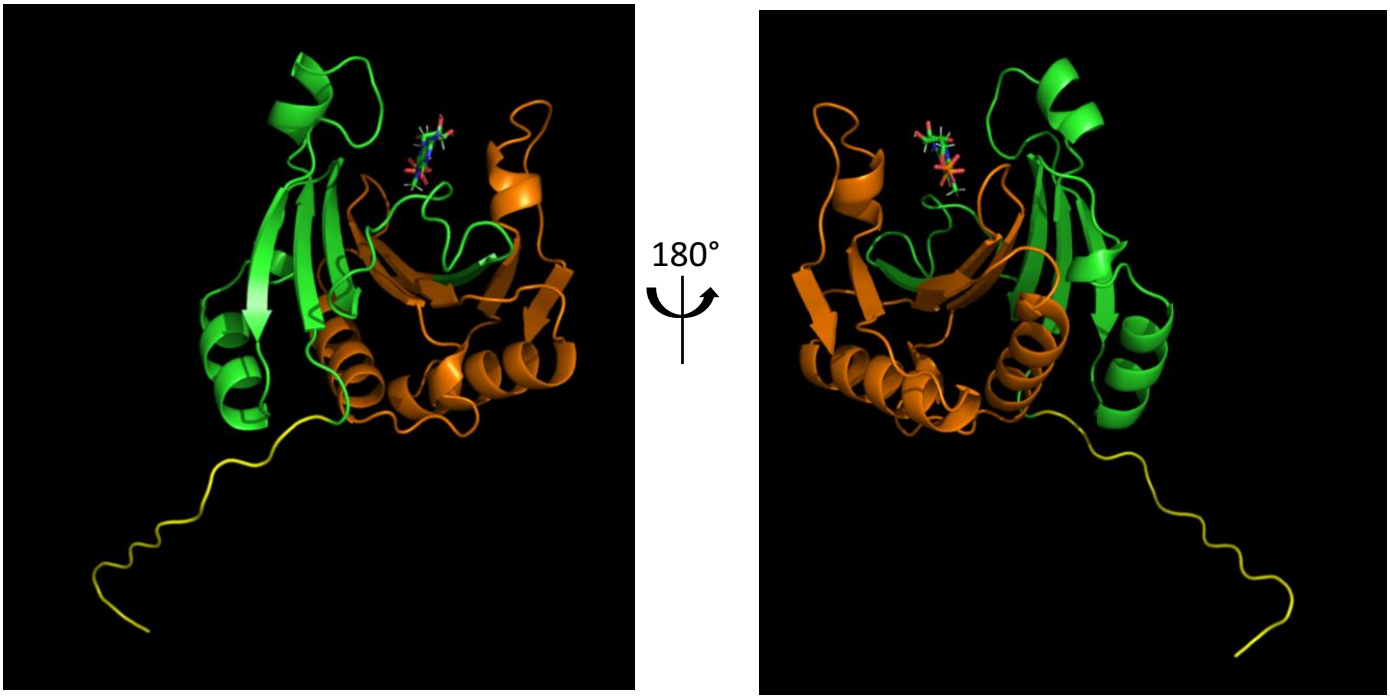
B



Supplementary Figure 3 Analysis of the edited eIF4E2 cDNAs by RT-PCR and Sanger sequencing.

(A) RT-PCR amplification of *eIF4E2* cDNA in T3 plants for the three lines edited for *eIF4E2*: 2A, 2B, and 2C. GAPDH is amplified as an internal control. However, the RT-PCR cannot be considered as quantitative. Negative control is water. WT are WVa106 and 4E2 EMS KO is the TILLING mutant.

(B) Verification of the edited sites on the full length eIF4E2 cDNA sequence, from two independent plants per genotype, only one is presented here. cDNA fragments amplified in (A) were Sanger-sequenced. Only the part covering both target zones are represented here.



Supplementary Figure 4 Tomato eIF4E2 3D structure model as built using the Pea eIF4E1 structure (PDB 2WMC).

Two views of the model are presented. The Yellow string includes the region of 35th-48th amino acid in WT sequence underlined in Figure 2. As this peptide is has not been crystallized in the Pea 2WMC 3D structure, its sequence has not been resolved. The eIF4E2 part that is deleted due to premature stop codon in plants 2A and 2C are represented in orange. The structure in sticks represents the cap analog.

Supplementary Table 1 List of primers used in this study

name	sequence	code	Target
4e2HRMF	GACAGAGGTGAGGAAGGAGAGATC	Z5770F	amplification of eIF4E2 sequences (131bp) for HRM analysis
4e2HRMR	GATTTCCCTGAAGGGTTATCGA	Z5770R	
AHAF	gtggttttgccgttgccaat	Z5417F	amplification in Acetolactate synthase gene (AHA, 761 bp)
AHAR	cagctcctcacttgattgcg	Z5417R	
Cas9F	TCCCTTACTACGTGGGACCTC	Z4959F	amplificaion in SpCas9-NG gene (260 bp)
Cas9R	cattccctcggtcacgtact	Z4958R	
KanaF (NptII)	AGACAATCGGCTGCTCTGAT	Z4404F	amplification in NptII gene (593 bp)
KanaR (NptII)	AGCCAACGCTATGTCCTGAT	Z4404R	
4e2F	CAAATCGGAGTTCTCCCTCA	Z5768F	amplification in eIF4E2 gene (450 bp)
4e2R	CAACAAATATCCGACCACCATT	Z5768R	
	AgagaattcatgGCTGATGAACTAAACAAAGCT	Z1379F	*eIF4E2* full-length cDNA amplification and sequencing
	AgactcgagCTACACGGAATAACGATTTTTGGC	Z1380F	
	GGTTACAGTTCCCGTGTGATTGA	Q3F	*GAPDH* (Solyc05g014470) cDNA amplification
	TCAAAGATAACAACAGGGTTCCGTT	Q3R	

Chapter 2: Manuscript 2

Brief introduction

The generation of materials in this manuscript was initiated about more than ten years ago by my team using the then state-of-art techniques. Thanks to a decade of their effort, an exciting set of base-edited tomato *eIF4E1* mutants were generated, ready to be tested just about the time I arrived in the lab as a master 2 student.

This manuscript includes the experiments on these mutants that I started as a main part of my M2 internship project in spring 2021 and concluded at the end my first year of PhD. This period included exactly 12 months of optimization of an experiment—from the very fast assay on April 6th, 2021 until the quality result from an assay on April 19th, 2022—which finally bore a fruit as the smallest figure in the manuscript (Figure 4c).

The manuscript was submitted and published from the journal *Plant Biotechnology Journal* on January 30th, 2023 (DOI: 10.1111/pbi.14003). Complete data including the supplementary data is provided in this chapter. A published version of the manuscript (main text without supplementary data) is available in the annex.

The mutants described in this manuscript have also prompted the work of two chapters that follow.

An iterative gene editing strategy broadens eIF4E1 genetic diversity in Solanum lycopersicum and generates resistance to multiple potyvirus isolates

Kyoka Kuroiwa[1], Benoit Danilo[1], Laura Perrot[2], Christina Thenault[1], Florian Veillet[3], Fabien Delacote[4], Philippe Duchateau[4], Fabien Nogué[5], Marianne Mazier[1] and Jean-Luc Gallois[1]

[1]INRAE, GAFL, F-84143 Montfavet, France

[2]Toulouse Biotechnology Institute, Université de Toulouse, 135 avenue de Rangueil, 31077 Toulouse CEDEX 04, France

[3]INRAE, Agrocampus Ouest, Université de Rennes, IGEPP, F-29260 Ploudaniel, France

[4]CELLECTIS S.A., 8 Rue de la Croix Jarry, 75013 Paris, France

[5]Université Paris-Saclay, INRAE, AgroParisTech, Institut Jean-Pierre Bourgin (IJPB), 78000, Versailles, France.

Abstract

Resistance to potyviruses in plants have been largely provided by the selection of natural variant alleles of *eukaryotic initiation factors* (*eIF*) *4E* in many crops. However, the sources of such variability for breeding can be limited for certain crop species, while new virus isolates continue to emerge. Different methods of mutagenesis have been applied to inactivate the *eIF4E* genes to generate virus resistance, but with limited success due to the physiological importance of translation factors and their redundancy. Here, we employed genome editing approaches at the base level to induce non-synonymous mutations in the e*IF4E1* gene and create genetic diversity in cherry tomato (*Solanum lycopersicum* var. *cerasiforme*). We sequentially edited the genomic sequences coding for two regions of eIF4E1 protein, located around the cap-binding pocket and known to be important for susceptibility to potyviruses. We show that the editing of only one of the two regions, by gene knock-in and base-editing, respectively, is not sufficient to provide resistance. However, combining amino acid mutations in both regions resulted in resistance to multiple potyviruses without affecting its functionality in translation initiation. Meanwhile, we report that extensive base editing in exonic region can alter RNA splicing pattern, resulting in gene knockout. Altogether our work demonstrates that precision editing allows to design plant factors based on the knowledge on evolutionarily selected alleles and enlarge the gene pool to potentially provide advantageous phenotypes such as pathogen resistance.

Introduction

Recent genome editing techniques allow to directly inactivate or modify genomes at specific loci, thereby opening the possibility to control specific agronomic traits (Eisenstein, 2022). These techniques, including transcription activator-like effector nucleases (TALEN) and CRISPR (clustered, regularly interspaced, short palindromic repeat)-associated (Cas) endonucleases, allow the targeting of genomic regions of interest and cut the nucleic acid strand(s) to introduce random mutations or to precisely customize target loci with elements like donor sequences and deaminases (Veillet et al., 2020b; Gao, 2021). Such precise targeting is essential to minimize side effects especially for genes that code for proteins involved in several physiological activities through different domains.

Pathogen resistance in certain accessions is a result of evolution of the interaction between plants and pathogens. This is often represented by the natural diversity in resistance alleles of susceptibility genes. These genes can be hijacked by viruses in susceptible plant accessions and are often essential for plant physiology: plants have to develop efficient resistance without impairing their growth. *Eukaryotic translation initiation factor* (*eIF) 4E* is one of such genes co-evolving with the viral-genome linked protein (VPg)-coding gene of potyvirus, the largest plant RNA virus group. *eIF4E*s make a small multi-gene family of mRNA 5'-cap binding proteins that are essential for translation initiation in eukaryotes. The interaction of the potyviruses' VPg with host eIF4E(s) is essential to establish the infection although the precise mechanism is still unknown (Miras et al., 2017; Tavert-Roudet et al., 2017; Coutinho de Oliveira et al., 2019; Saha and Mäkinen, 2020). Hence, the mutant forms of *eIF4E* have been found as resistant alleles in several crop species including lettuce (*Lactuca sativum*), pea (*Pisum sativum*), and tomato (Ruffel et al., 2002b; Nicaise et al., 2003; Gao et al., 2004; Ruffel et al., 2005) though they are not available in all species (Bastet et al., 2017). In pepper (*Capsicum* spp.), 25 allelic forms of *eIF4E1* have been reported and associated with various spectra and durabilities of potyvirus resistance (Charron et al., 2008; Moury et al., 2014; Poulicard et al., 2016). Interestingly, those natural mutants harbor only substitutions in their eIF4E1 amino acid sequences compared to their susceptible counterparts. These substitutions are highly clustered in two domains—named Region I and II—of the eIF4E1 protein surrounding the cap-binding pocket (Robaglia and Caranta, 2006b; Charron et al., 2008). Despite the importance of this domain for the protein's original function, many resistance genes have been proven functional reflecting the natural selection of the resulting amino acids

(Charron et al., 2008; Moury et al., 2014). This can provide a blueprint to be implemented in resistance breeding.

Similar resistance is found among the ancestral species of tomato. The recessive resistance gene *pot1,* isolated from the wild tomato relative *S. habrochaites* PI247087 accession, encodes an eIF4E1 protein which possesses eight amino acid mutations (L48F, N68K, P69S, A77D, V85L, M109I, K123Q, and N224S) compared to the susceptible cultivated tomato, including five within and in the proximity of Regions I and II. The *pot1*-encoded eIF4E1 protein is associated with a broad resistance spectrum to multiple potyvirus species and strains, while it retains its function to initiate translation (Ruffel et al., 2005; Gauffier et al., 2016).

Various eIF4E knock-out mutants targeting eIF4E1 and/or eIF4E2 have been generated by methods such as TILLING, RNAi, and CRISPR/Cas9 to generate potyvirus resistance in tomato as a model crop (Piron et al., 2010; Mazier et al., 2011; Gauffier et al., 2016; Schmitt-Keichinger, 2019; Yoon et al., 2020; Kuroiwa et al., 2022). Although the mutant genotypes exhibited resistance to potyviruses, the resistance spectrum associated with single-gene knock-out was limited. Only the severely stunted, double *eif4e1; eif4e2* knock-out (KO) mutants had a resistance spectrum as large as the one provided by the natural allele *pot1.* This is because most potyviruses can use interchangeably both tomato *Sl*eIF4E1 and *Sl*eIF4E2 (Gauffier et al., 2016; Lebaron et al., 2016). Consequently, KO strategies were achievable, but unlike the natural resistance eIF4Es, neither the single nor double KO of tomato eIF4Es fully met the agronomic demand due to the functional redundancy and the importance of this protein. These results have prompted us to study how more precise genome editing techniques can help designing resistant phenotypes based on the positively selected non-synonymous mutations rather than producing KO alleles in crops like tomatoes (Bastet et al., 2017).

Previously, we provided a proof-of-concept by showing how six mutations selected in the pea (*Pisum sativum*) *eIF4E1*: *sbm1* alleles could be transferred through transgenesis in the *Arabidopsis eIF4E1* susceptible allele and provide resistance to a potyvirus, clover yellow vein virus (Bastet et al., 2018). Moreover, this resistance could be pinpointed to discrete mutations. And when one of those—N176K—was introduced in the susceptible wild-type *Arabidopsis* through base editing, the resulting plants showed the same resistance. This was to our knowledge the first example of using genome precise edition

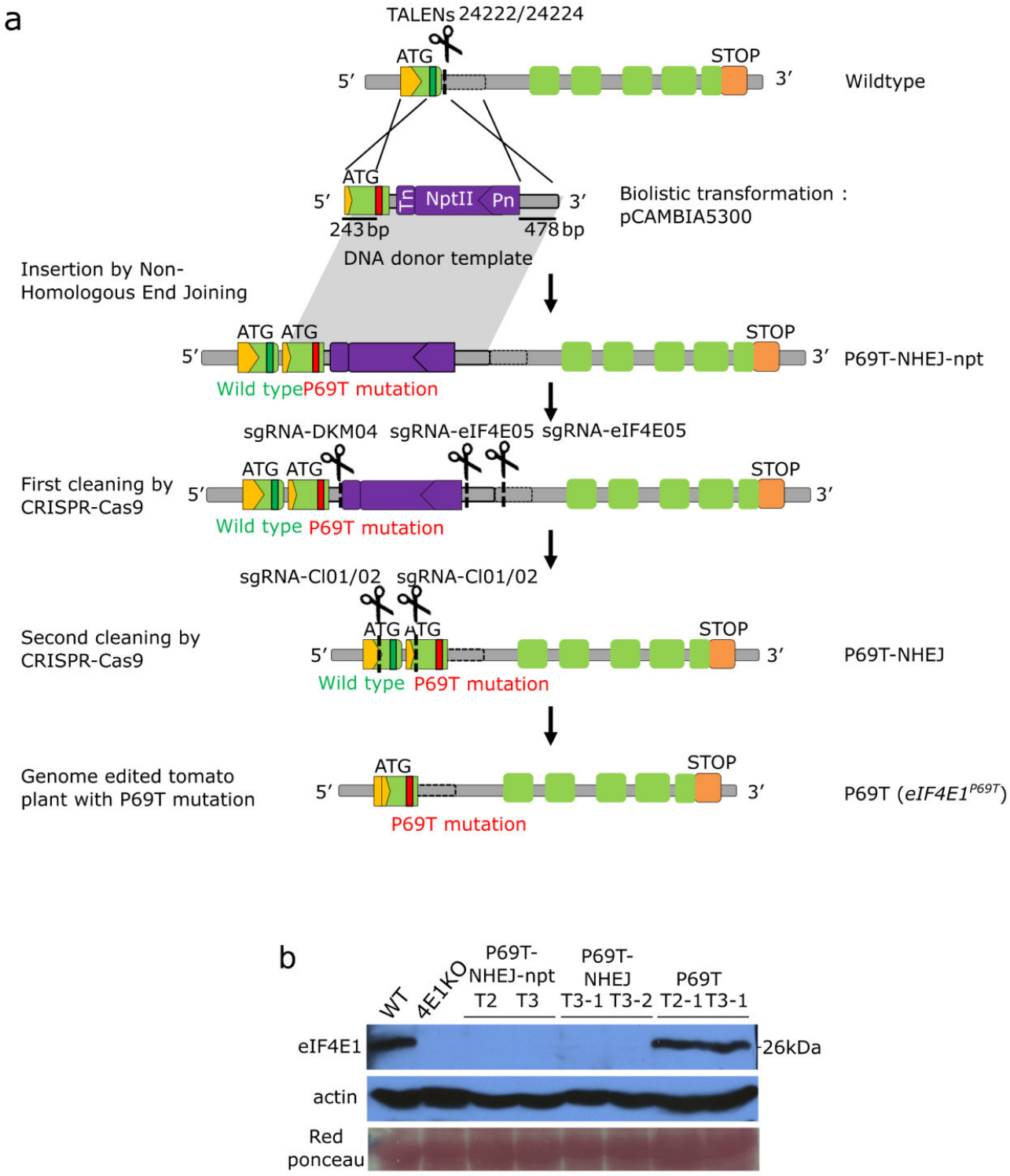


Figure 1 A gene knock-in strategy using TALEN® followed by the targeted elimination of undesirable inserted parts using CRISPR/Cas9 generates a single AA change in *Sl*eIF4E1 exon 1.

a Schematic diagram of the targeted gene Knock-In and the successive targeted deletion of genomic region. Dotted line under the scissors indicate each targeted site of TALEN or CRISPR sgRNA indicated above. The codon position for P69T mutation is indicated by the strips in green or red (wildtype or mutant). Green boxes represent the exons and yellow pentagon the 5′-UTR. The crossed lines between the genome and the donor DNA template indicate recombined region. Pn: Nos promoter, Tn: nos terminator.

b Western blot analysis to detect the *Sl*eIF4E1 accumulation in the homozygous mutant plants from successive genetic modification steps. T2 and T3 indicate the transformant generation. Actin accumulation and red ponceau are shown as the loading control. WT: wildtype, 4E1KO: TILLING *eif4e1*^{KO}.

to copy a non-synonymous mutation associated with resistance to viruses (Bastet et al., 2019).

In the present work, we decided to introduce in cultivated tomato a potential amino acid mutation by gene knock-in in the *eIF4E1* Region I, which is known to be important for resistance. Facing the difficulty of such approach, we next decided to randomly tinker another genomic domain in *eIF4E1*—Region II, also known to be involved in resistance, using cytosine base editor (CBE). We show that by iteratively generating diversity in *eIF4E1* genomic regions known to be involved in resistance across plant species, we succeeded in developing a functional potyvirus resistance allele. Incidentally, we observed that heavily mutating a coding region with the CBE can impair the proper gene splicing leading to knock out. Altogether, this work presents the promising potential and a caveat in using genome editing for improvement of crops such as for virus resistance by generating targeted variability in alleles.

Results

A codon change in *SleIF4E1* associated with a P69T mutation is not sufficient to confer resistance to potyviruses

The broad-spectrum resistance allele *eIF4E1-pot1* differs from the wild-type susceptible tomato allele by eight non-synonymous codon substitutions, and the role of individual substitutions in the resistance process has not been fully elucidated (Gauffier et al., 2016). Because a related *Solanaceae*, pepper, displays a wide range of resistant alleles, we looked at this *eIF4E1* gene pool to identify the causative mutations for the resistance. In many pepper resistant varieties, one or more mutations at amino acid position 66 to 68, such as P66T, V67E, and A68E within the eIF4E1 Region I, are present (Kang et al., 2005b; Charron et al., 2008). As P69 (pepper P66 equivalent) is also mutated in the *pot1*-encoded protein, we decided to introduce a P69T mutation in the susceptible tomato *eIF4E1* allele to generate a new resistant tomato (Figure S1).

In order to mimic the pepper P66T by mutating codon #69 from CCT to ACA, we attempted a gene knock-in strategy, using a pair of TALENs® to generate a double-strand break in the intron 1 of the *SleIF4E1* gene and to recombine with a donor DNA template. The template contained the *eIF4E1* exon 1 with substituted codon ACA as well as a kanamycin resistance *NptII* gene for selection, flanked by left and right homology arms to

Table 1 Resistance phenotypes against four potyvirus isolates of plants at different stages during the knock-in strategy.

Viruses tested	Plant genotypes						
	WVa106 (wildtype)	TILLING *eif4e1*^{KO}	P69T-NHEJ-npt T3-1	P69T-NHEJ T3		P69T	
				1	2	1	2
PepMoV	S	R	R	R	R	S	S
PVY-LYE90	S	PR	PR/S	PR	PR	S	S
PVY-N605	S	S	S	S	S	S	S
PVY-SON41	S	S	S	S	S	S	S

The accumulation of viral coat protein (CP) 21 days post inoculation (dpi) was measured by ELISA. Six plants are used for each mutant line represented by column (two lines for the P69T-NHEJ and P69T genotypes) and 12 for wildtype. S: susceptible, R: resistant (colored in yellow), PR: partially resistant (CP level lower than S but higher than R; colored in orange).

promote homologous recombination (HR) (Fig. 1a). Both the TALEN® T-DNA vectors and the donor DNA template were introduced into cells of cherry tomato cultivar (cv.) WVa106 by the biolistic method. Thirty-two regenerating buds were assessed by PCR and sequencing, but none of them harbored an integration of the DNA template via HR (Figure S2). Nevertheless, one plant had integrated the donor template through non-homologous end joining (NHEJ) and was named P69T-NHEJ-npt. (Fig. 1a P69T-NHEJ-npt, Figure S2, Table S1, Data S1). This NHEJ-based insertion in the native intron 1 has resulted in the expected insertion of *NptII* gene as well as an undesirable, partial duplication within the *eIF4E1* gene, namely the entire exon 1 deriving from the native and recombined sequences.

The P69T-NHEJ-npt was selfed to produce TALEN-free plants, homozygous for the modified *eIF4E1* locus. As expected, eIF4E1 protein accumulation was not detected in the progenies using specific polyclonal antibodies (Fig. 1b, P69T-NHEJ-npt): this was consistent with the presence of an approximately 1.4 kb *NptII* expression cassette in the native exon 1. Thus, in order to restore a functional *eIF4E1* gene, two successive CRISPR/Cas9-mediated editing strategies were used to eliminate undesirable sequences of the recombined *eIF4E1* gene at the 3' and the 5' of the insert, respectively (Fig. 1a, alleles P69T-NHEJ and P69T; Figure S3,). Several independent tomato lines were recovered in which both the *NptII* cassette and the endogenous exon 1 of *eIF4E1* were removed. In these plants, the resulting edited *eIF4E1* gene harbored the T69 codon, and the eIF4E1 protein accumulation was restored to wild-type level (Fig. 1b, P69T lanes). These plant lines are hereafter referred to as $eIF4E1^{P69T}$ lines. This suggests that the targeted modification of an exon in *eIF4E1* through knock-in with a DNA donor template is achievable but might be labor-intensive due to low frequency of homologous recombination and possible rearrangements of the targeted locus.

To evaluate the effect of the eIF4E1 Region I modification on the virus resistance, we used four potyviruses—pepper mottle virus (PepMoV) and three different isolates of potato virus Y (PVY)—to challenge the plants homozygous at the three editing stages (Fig. 1a) and quantified the viral accumulation by ELISA 21 days after infection (Table 1). In accordance with the lack of eIF4E1 expression in the first steps of this editing scheme, intermediate lines (P69T-NHEJ-npt and P69T-NHEJ in Fig. 1a and b) were found to be fully resistant to PepMoV and PVY LYE90 isolate but not to other PVY isolates. They therefore present, as expected, the same resistance spectrum as a TILLING $eif4e1^{KO}$ line (Piron et al.,

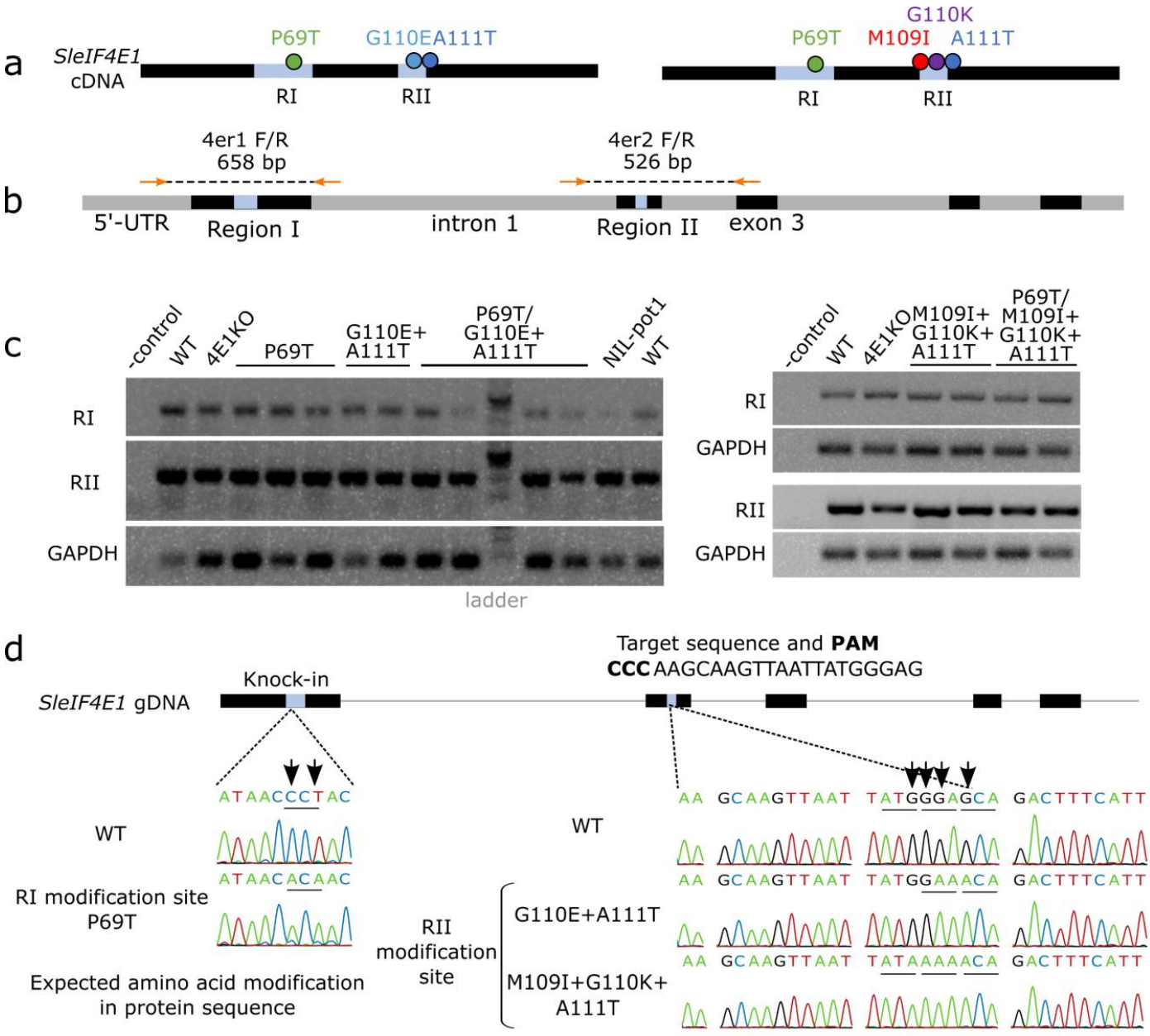


Figure 2 Base-editing in *eIF4E1* Region II generates two combinations of amino acid substitutions in different Region I backgrounds.

a Schematic diagram of the relative positions of amino acid substitutions induced in this work. Light blue boxes represent Region I and II.

b Primer positions used in the genomic sequence analyses of *SleIF4E1* Region I and II. Orange arrows are primers. Gray and black box represent introns and exons, respectively. Light blue boxes represent Region I and II.

c Amplification of genomic sequence of *SleIF4E1* Region I and II. Bands of correct lengths were verified. Two to four independent lines, including those not used in the further analyses, were used per allele. Mutant lines are indicated by their expected amino acid modifications in the *Sl*eIF4E1 protein. -control is water. GAPDH is used as internal control. WT: wildtype, 4E1KO: TILLING *eif4e1*^{KO}.

d Sanger-sequencing results of the amplified genomic regions confirm the expected modifications in the edited plants. Black boxes indicate the exons in *SleIF4E1* genomic DNA. Representative chromatogram from each modified region (light blue boxes) is shown. Analyses were performed for at least two plants of independent lines of each genotype. Arrows indicate the positions of nucleotide substitutions. The sgRNA in Reverse Complementary sequence used in the base-editing of the Region is shown.

2010; Gauffier et al., 2016). Interestingly, the *eIF4E1*^{P69T} lines in which the eIF4E1 expression was restored were fully susceptible to all four potyvirus isolates tested (Table 1). This implied that introducing a single P69T mutation in the tomato eIF4E1 is not sufficient to induce resistance to the potyvirus infection. However, based on the natural *eIF4E* resistance alleles, which often possess multiple non-synonymous mutations, we reasoned that additional mutations could confer resistance. We thus proceeded to the editing of another genomic region of *eIF4E1*, Region II, where non-synonymous mutations are documented in resistance alleles from multiple species in combination with the Region I.

A base editing strategy efficiently generates diverse alleles with multiple substitutions in *eIF4E1* Region II

Since many natural potyvirus resistance alleles possess amino acid substitutions in both Region I and II of eIF4E1, we decided to target the Region II genomic sequence in *eIF4E1*^{P69T} plants. The same editing strategy was carried out in parallel on wild-type WVa106 plants. In order to bypass the labor-intensive knock-in strategy, we chose a base-editing strategy using nCas9-cytidine deaminase enzyme (or cytosine base editor; CBE). Because a CBE can introduce mutations at one or more cytosines, provided they are located at appropriate position related to a PAM sequence (Nishida et al., 2016), it can trigger varied amino acid changes in this region instrumental for potyvirus resistance. An *SleIF4E1* guide RNA was designed to modify a maximum of four cytosines situated in the editing window at 16-20 nt relative to the PAM on the reverse strand, with potential to induce amino acid changes for M109, G110 and A111 (Figure S4a). Moreover, to facilitate the selection, we implemented a co-editing strategy which can simultaneously mutate cytosines in the *acetolactate synthase SlALS1* targeted genomic regions (Figure S4c). As shown previously, plants edited for *SlALS1* can be selected on ALS inhibitors such as chlorsulfuron, allowing a selection of potentially edited plantlets by CRISPR activity (Veillet et al., 2019).

Explants that grew on the chlorsulfuron selection media were recovered from both WVa106 and *eIF4E1*^{P69T} backgrounds. Based on the sequencing of the genomic DNA covering *eIF4E1* exon 2 that includes the targeted Region II, two edited profiles were selected 1) a double amino acid change of G110E + A111T (hereafter referred to as ET) and 2) a triple change of M109I + G110K + A111T (hereafter referred to as IKT) (Fig. 2a). We therefore obtained four original combinations of mutations in eIF4E1 that have never

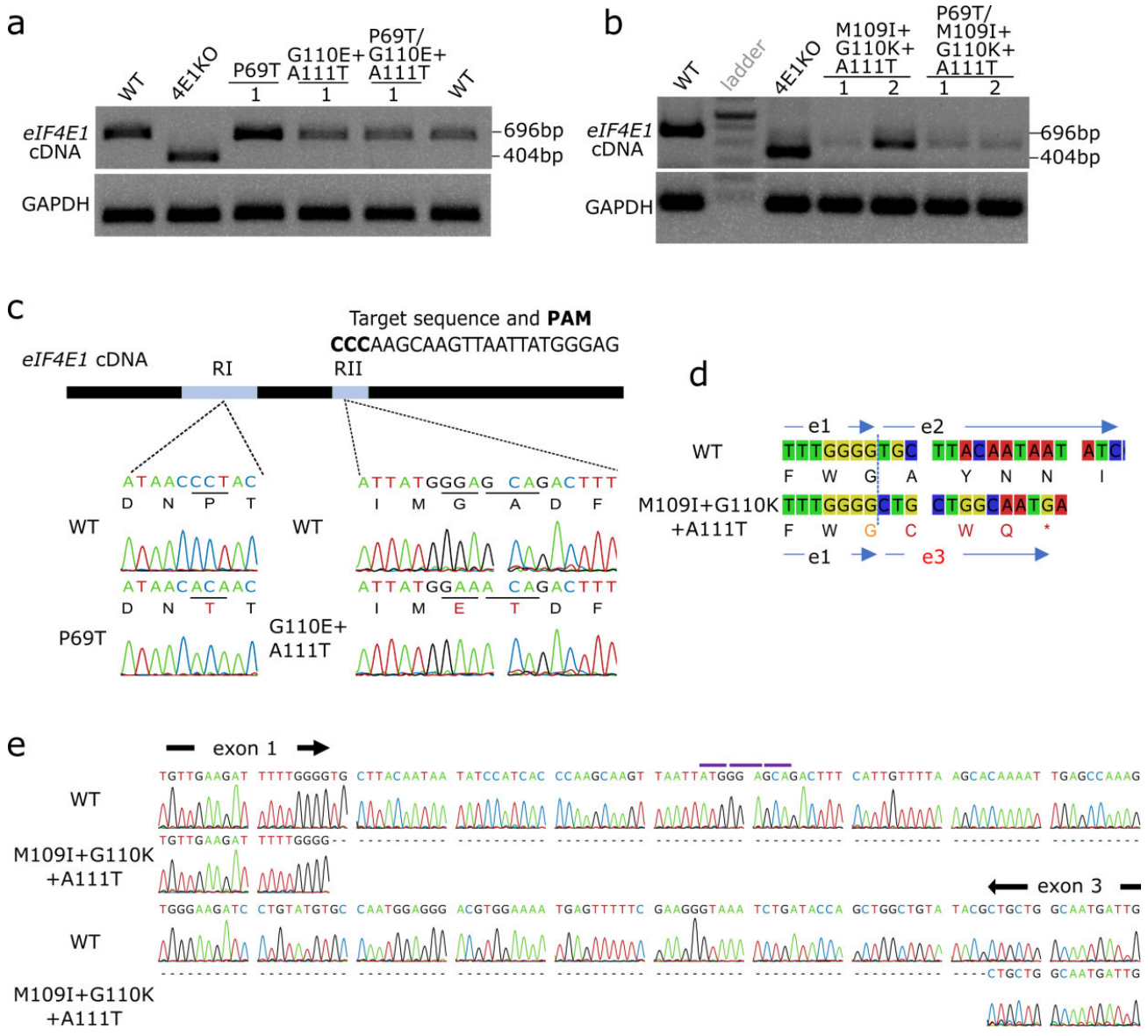


Figure 3 *eIF4E1* cDNA analysis confirms correct mRNA expression for moderately edited plants but reveals abnormal splicing for the triple-codon mutants in Region II.

a, b Amplification of *eIF4E1* cDNA following reverse transcription; the expected size of the PCR fragment is 696 bp for WT and 404 bp for the TILLING *eif4e1*^{KO}. The GAPDH cDNA is amplified as internal control. Representative sample from two independent mutant lines are shown. WT: wildtype, 4E1KO: TILLING *eif4e1*^{KO}.

c Sanger-sequencing of the amplified *SleIF4E1* cDNA confirms the expected modifications in the P69T- and G110E+A111T-coding mutant lines. Representative chromatogram from each modified region is shown. Analyses were performed for at least two plants of independent lines of each genotype. Mutant lines are indicated by their expected amino acid modifications in the *Sl*eIF4E1 protein. Substitution targets are underlined. Predicted translation of each sequence is shown under the nucleotide sequence. Amino acid substitution relative to WT sequence is indicated in red.

d Predicted protein translation of cDNAs from the WT plant and from an IKT-containing edited plant. The premature stop codon in the eIF4E1 cDNA from IKT plants is represented by a star. e1, e2, and e3 are exons 1, 2, and 3.

e Chromatogram of the representative cDNA sequence of IKT group shows an unexpected deletion corresponding to the wild-type exon 2. Shown from the 3′ end of the exon 1 to the 5′ end of the exon 3. Purple bars indicate the positions of the codons expected for mutations based on the genomic DNA sequencing.

been described among the natural diversity of eIF4Es. Moreover, this setup allows to potentially monitor the effect of mutations in the two regions, independently or together. Two independent lines for each of the four new alleles (*eIF4E1*^{ET}, *eIF4E1*$^{P69T/ET}$, *eIF4E1*^{IKT}, and *eIF4E1*$^{P69T/IKT}$) were selected and after three selfings, transgene-free homozygous lines for the *eIF4E1* modification were obtained (Fig. 2b-d).

Extensive base editing in exon 2 is associated with mis-splicing of *eIF4E1* mRNA

To characterize the effects of modifications on *eIF4E1* gene function in each mutant line, we performed RT-PCR on the *eIF4E1* mRNA using specific primers. In the *eIF4E1*^{P69T}, *eIF4E1*^{ET}, and *eIF4E1*$^{P69T/ET}$-edited lines, cDNA of the same size as wildtype (696 bp), was amplified (Fig. 3a). A shorter cDNA of 404 bp was amplified for *eif4e1*^{KO} mutant which is a product of an aberrant splicing due to a mutation in a 5′-splice site flanking the exon 3 (Piron et al., 2010). In contrast, in the *eIF4E1*^{IKT} and *eIF4E1*$^{P69T/KT}$ lines, cDNA with a size smaller than wildtype but larger than that of *eif4e1*^{KO} mutant was amplified (Fig. 3b). Based on the Sanger-sequencing and alignment with the wild-type sequence, all but the *eIF4E1* cDNA derived from *eIF4E1*^{IKT} and *eIF4E1*$^{P69T/KT}$ contained the predicted sequences including the correct nucleotide substitutions already verified in genomic DNA (Fig. 2d, 3c). In contrast, the shorter cDNAs amplified from all of the IKT modification mutants were missing a 166-bp-long segment which entails Region II of *eIF4E1*, rendering the cDNA shorter (530 bp-long) (Fig. 3e). In fact, this missing segment perfectly matched the entire exon 2.

To gain insight into this event, we analyzed closely the amplified gDNA sequences surrounding the Region I and II and the aberrant cDNA sequence obtained from the edited tomato lines in comparison with previous studies on regulatory elements. In the gDNA fragment that spans the 526-bp genomic region from the last 100 bp of the intron 1 to the first 30 bp of the exon 3 (amplified by 4e2F/R in Fig. 2b), we did not observe any other mutations than our predicted target codon modifications that code for the IKT substitutions. The sequencing also revealed that the 3′-splice sites of exons 1 and 2 and the 5′-splice sites of exons 2 and 3 were conserved in the edited plant compared to the wild-type, with the conserved dinucleotides known as the donor and acceptor sites for the pre-mRNA splicing (Brown and Simpson, 1998; Li et al., 2019) (Figure S5a). Thus, we concluded that the modification of a wild-type sequence ATGGGAGCA into an IKT-coding sequence ATAAAACA in the exonic region has altered the splicing pattern and induced an exon skipping in *eIF4E1* (Figure S54).

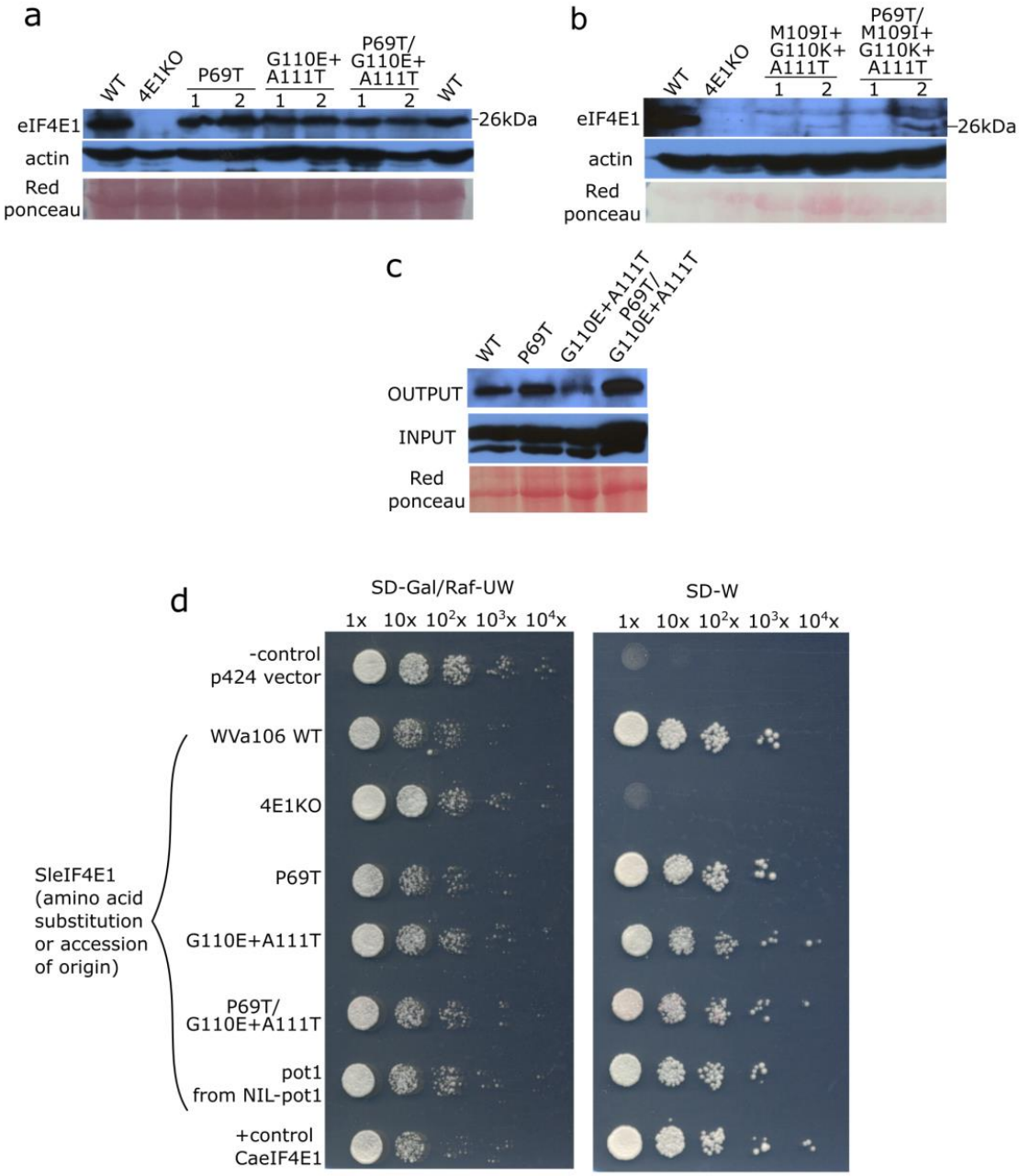


Figure 4 Moderate editing in eIF4E1 region II does not affect the protein accumulation level *in planta* and the protein functionality *in vitro* and *in vivo*.

a, b Western blot analysis to detect the *Sl*eIF4E1 accumulation in the edited lines. Actin accumulation and red ponceau are shown as the loading control. Mutant lines are indicated by their expected amino acid modifications in the *Sl*eIF4E1 protein. Two independent edited lines were used per genotype. The expected molecular weight of *Sl*eIF4E1 is 26 kDa. WT: wildtype, 4E1KO: TILLING *eif4e1*^{KO}.

c Cap-affinity purification and following western blot shows the cap-binding ability of the edited eIF4E1 proteins at wild-type level. OUTPUT is immunoblot using anti-SleIF4E1 specific antibody on the eluted samples after incubation with m^{7}GTP-cap analog and repeated washing. INPUT is immunoblot using anti-actin antibody on the total soluble protein extracts as the loading control. Red ponceau staining of the INPUT loading is also shown as the loading control.

d Yeast complementation assay demonstrates that the edited *SleIF4E1* proteins correctly expressed *in planta* retain the functionality in translation initiation. Gal/Raf-UW is the control media and Glu-W is the selective media for the eIF4E function. J055 transformed with p424 empty vector is shown as a negative control. *Cae*IF4E1 is the positive control. All *Sl*eIF4E1 are the transformants with the cDNA obtained from the edited plants. X indicates the dilution level of the yeast culture. Ca: *Capsicum annuum.* WT: wildtype, 4E1KO: TILLING *eif4e1*^{KO}.

At translational level, this exon deletion would cause a frameshift at position 97 in amino acid sequence, resulting in a premature termination codon corresponding to position 100 in the wild-type sequence (Fig. 3d). This possibly leads to truncated eIF4E1 production and/or the degradation of such aberrant form in *eIF4E1*^{IKT} and *eIF4E1*$^{P69T/IKT}$ lines.

Mis-splicing resulting from heavy editing at *eIF4E1* Region II affects the eIF4E1 protein accumulation while a moderate editing maintains accumulation and functionality of eIF4E1

Not only the mis-splicing, but also the introduction of amino acid modifications itself can destabilize the protein structure affecting the accumulation and/or functionality of the protein. To examine the protein accumulation, we performed western blot analysis on the proteins extracted from leaves for each edited line. All independent lines of plants with the *eIF4E1*^{P69T}, *eIF4E1*^{ET}, and *eIF4E1*$^{P69T/ET}$ alleles expressed eIF4E1 protein (26 kDa) at a level comparable to that of the wildtype (Fig. 4a). In contrast, lines harboring three mutations in eIF4E1 Region II (IKT) showed no to very low accumulation of eIF4E1 at the expected size (Fig. 4b). This suggests that the amino acid substitutions in the eIF4E1^{P69T}, eIF4E1^{ET}, and eIF4E1$^{P69T/ET}$ do not affect the protein accumulation. The results also implied that the mis-splicing due to the IKT modification in Region II has led to lack of accumulation or expression of the eIF4E1 protein. As a consequence, the *eIF4E1*^{IKT} and *eIF4E1*$^{P69T/IKT}$ alleles could be considered *eIF4E1* knock-out alleles.

Given the importance of eIF4E1 protein in plant translation, we further checked, in addition to the protein accumulation, whether the functionality of the edited eIF4E1 proteins was not affected. To verify the ability of the edited proteins to bind the 5′-cap structure and initiate translation, we performed both *in vitro* and *in vivo* methods. We first assessed the ability of the eIF4E1 proteins to bind the m^{7}GTP-cap analog. After the incubation of total protein extract with cap analog, eIF4E1 protein was detected for the wildtype as well as for all three edited lines, namely *eIF4E1*^{P69T}, *eIF4E1*^{ET}, and, *eIF4E1*$^{P69T/ET}$, showing that the edited alleles code for eIF4E1 with cap-binding ability (Fig. 4c). Additionally, we used a yeast complementation assay, which relies on the conditional complementation of a yeast strain lacking its native *eIF4E* gene. In accordance with the cap affinity purification results, the heterologous expression of the three edited tomato alleles complemented the lack of a native yeast *eIF4E* (Fig. 4d). The serial dilution suggests that the complementation efficiency with the edited version of eIF4E1 is comparable to

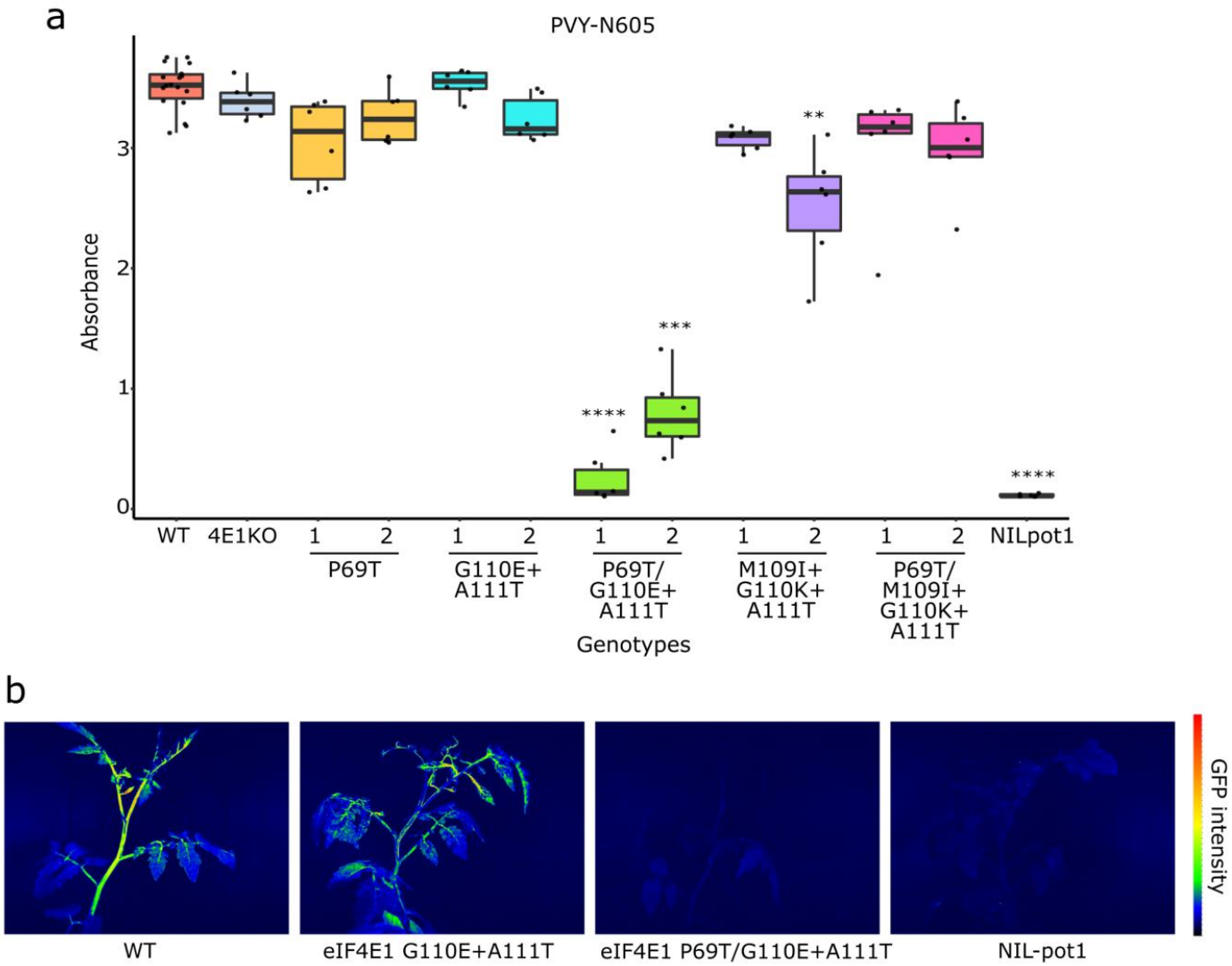


Figure 5 A combination of mutations at eIF4E1 region I and II triggers resistance to PVY N605.

a DAS-ELISA quantification of the accumulation of PVY-N605 in tomato plants at 21 dpi. Points overlaid on boxplots represent the individual plants. Wild-type WVa106 and KO4E1 (TILLING *eif4e1*^{KO}) lines were used as positive controls, while NIL-pot1 plants are used as controls for resistance. Two independent lines were tested for the five edited genotypes. Colored by alleles. Six plants were tested per mutant and 12 for wildtype. Significance of difference compared to WVa106 based on Kruskal-Wallis and Dunn-test as a *post-hoc* test is indicated by asterisks.< 10^{-4} ****; 10^{-4} to 10^{-3} ***; 10^{-3} to 10^{-2} **. WT: wildtype, 4E1KO: TILLING *eif4e1*^{KO}.

b Monitoring of the virus infection using a GFP-tagged PVY-N605 clone corroborates the resistance of the edited *eIF4E1*$^{P69T/ET}$ line demonstrated by ELISA results. Wild-type WVa106 was used as the positive control, while NIL-pot1 was used as the resistant control. GFP signal was observed at 15 dpi under blue light (420 nm) by using GFP camera imaging. The bar next to the photos indicate the intensity of GFP perceived by the device, from blue (no GFP) to red (high intensity). Each photo is a representative of 10-15 plants observed for each genotype. Two independent lines were tested for the edited alleles. WT: wildtype.

that achieved with the wild-type eIF4E1 as well as the *pot1*-encoded eIF4E1 (Fig. 4d). Therefore, the *in vitro* and *in vivo* assays strongly suggest that the editing carried out in *eIF4E1*^{P69T}, *eIF4E1*^{ET}, and *eIF4E1*$^{P69T/ET}$ alleles did not affect their functionality in translation initiation processes, in a similar fashion to the natural *eIF4E* varieties.

Taken together, these results show that the P69T editing in Region I does not affect eIF4E1 expression or function. In contrast, while the editing of two codons in Region II (G110E+A111T) does not result in any apparent change, a more extensive, three-codon mutation (M109I+G110K+A111T) results in a lack of wild-type level eIF4E1 protein expression due to the altered splicing in the genomic sequence.

A combination of mutations in Region I and II of *eIF4E1* associates with resistance to several isolates of potyviruses

Finally, we evaluated the resistance spectrum to potyviruses of all the 10 edited tomato lines—two for each allele, along with plants of wild-type WVa106, TILLING *eif4e1*^{KO} mutant, and a *pot1* near-isogenic line (NIL-pot1) (Lebaron et al., 2016). Plants were challenged with three isolates of PVY and two other potyviruses, PepMoV and tobacco etch virus (TEV) CAA10 isolate, previously tested (Gauffier et al., 2016). Resistance of each genotype was based on viral accumulation assessed by ELISA after 21 days.

All eight lines harboring Region II mutations ET or IKT, regardless of the presence of P69T in Region I, showed complete resistance to PepMoV and complete or partial resistance to PVY-LYE90 (Table 2). This is significant but not any better than the limited resistance spectrum associated with knocking out *eIF4E1*. Indeed, the limited resistance spectra of the *eIF4E1*^{IKT} and *eIF4E1*$^{P69T/IKT}$ mutants were consistent with their loss of function of the *eIF4E1* gene just like in the TILLING *eif4e1*^{KO}. In contrast, while *eif4e1*^{KO} plants are fully susceptible to PVY-N605 and PVY-SON41, we show that the combination of mutations P69T and ET in *eIF4E1* (eIF4E1$^{P69T/ET}$) triggered resistance to PVY-N605, as well as partial resistance to the PVY-SON41 isolate (Table 2, Fig. 5a). The infection of PVY-N605 was further assessed by inoculating a GFP-tagged infectious clone of this isolate and monitoring the GFP fluorescence. GFP fluorescence could be observed on the inoculated and systemic leaves of the wild-type WVa106 and of the *eIF4E1*^{ET} line, mutated only in Region II. On the contrary, the fluorescence was observed neither on the inoculated nor the systemic leaves of the *eIF4E1*$^{P69T/ET}$ lines and NIL-pot1, thereby confirming the

Table 2 Schematic diagram of the selected edited alleles and resistance phenotypes associated.

	Region I	Region II		Virus isolates tested				
Genotypes	P69T	G110E+ A111T	M109I+ G110K+ A111T	PepMoV	PVY-LYE90	PVY-N605	PVY-SON41	TEV-CAA10
WVa106 (wildtype)				S	S	S	S	S
TILLING *eif4e1*^{KO}				R	PR	S	S	S
NIL-pot1				R	R	R	R	R
P69T	X			S	S	S	S	S
ET		X		R	PR	S	S	S
P69T/ET	X	X		R	R	R	PR	S
IKT			X	R	PR	S	S	S
P69T/IKT	X		X	R	PR	S	S	S

Summary of virus infection testing results of each mutant against five potyvirus isolates. The accumulation of viral coat protein (CP) after 21 dpi was measured by ELISA. Six to ten plants were used for each mutant line and 12-15 plants for wildtype. Two independent lines were tested for the five edited genotypes. S: susceptible, R: resistant (colored in yellow), PR: partially resistant (CP level lower than S but higher than R; colored in orange).

resistance of the specific Region I + II edited tomato (Fig. 5b). Nevertheless, the *eIF4E1*$^{P69T/ET}$ allele was not associated with resistance to TEV-CAA10 unlike the natural resistance *pot1* allele in which a larger set of amino acid substitutions has been selected (Table 2).

Discussion

In this work, we have used TALEN-based gene knock-in and CRISPR-mediated base editing techniques on the tomato *eIF4E1* allele in an attempt to generate genetic resistance against a broad range of potyviruses. Our targeted approach to the specific regions in *eIF4E1* for multiple base substitutions has brought about a *de novo* functional resistance tomato *eIF4E1* allele. Edited plants with this allele showed the broadest potyvirus resistance spectrum achieved by genetic engineering in tomato so far. Meanwhile, we have observed a previously unreported effect on RNA splicing by certain base modifications in a target exon resulting in abnormal expression.

Since the identification of pepper PVY-resistance alleles *pvr1/pvr2* as alleles of the *eIF4E1* gene in early 2000s, there has been a number of potyvirus resistance *eIF4E* alleles found among genetic resources, including *pot1* in tomato (Ruffel et al., 2002b; Kang et al., 2005b; Ruffel et al., 2006). Apart from using transgenic expression, direct implementation of these genes to other species in resistance development was restricted because they are found in a limited number of species and the conventional breeding is often bound by the interspecies barriers. Moreover, the functional significance and redundancy of the *eIF4E* genes in plants make the gene KO or knock-down strategies insufficient (Piron et al., 2010; Mazier et al., 2011; Gauffier et al., 2016; Chandrasekaran et al., 2016; Kuroiwa et al., 2022; Pechar et al., 2022). One intriguing feature of natural *eIF4E1*-based resistance is how the presence of an eIF4E1 protein variant is sufficient to bar the potyvirus from hijacking eIF4E2 although the latter can be expressed. Several hypothesis have been put forward, such as the fact that the functional eIF4E1 protein encoded by the *pot1* resistance allele might titrate other host factors which are also required by the virus (Kang et al., 2007; Gauffier et al., 2016; Bastet et al., 2017). These results have suggested that an alternate strategy, based on the generation of resistant functional alleles, similar to the ones selected among the natural variability, should be more efficient than knocking out *eIF4E* genes. Thus, more delicate engineering at the base level is clearly needed to achieve resistance and keep eIF4E function at the same time (Bastet et al., 2017).

Though this work confirms the high efficiency of base editing to generate diversity in a specific region of a gene, we also highlight a potential drawback of base editing following the mutation of four bases in the *eIF4E1* Region II. While the genomic modification in the IKT-coding mutants was done correctly, the expression of the resulting full-length mRNA and protein was impaired because the genomic modification

induced an exon skipping. So far, there are no reported cases of such aberrant splicing due to CRISPR-mediated editing in exons of a gene although there are some examples after editing in introns or inducing indels in exons (Li et al., 2019; Tang et al., 2021). It is possible that this set of modifications has converted the exon 2 into an intron-like sequence. Introns are often defined by UA-richness of the sequence (Simpson et al., 1996). The modified sequence does not completely correspond to known consensus of branchpoint or intronic splicing enhancer sequences which are important intronic *cis*-elements in splicing events (Wang et al., 2012; Zhang et al., 2019). Nevertheless, it is tempting to associate the increased A-richness in an otherwise exonic region, with the false recognition of the intron. On a practical point, our results stress that the sequencing of the genomic target is not sufficient for inferring the effect of a base editing strategy on a gene. This holds an important message that i) the results of the genome-editing should be confirmed by assessing cDNA and protein expressions and that ii) codon choices should not be neglected.

Here, by inducing multiple amino acid substitutions in eIF4E1, we have designed cherry tomato mutant lines which present resistance against a wider range of potyviruses than the single *eif4e1*^{KO} mutant. To our knowledge, this is the first example of virus resistance conferred to a crop species via base-editing technology by generating a *de novo* functional resistance *eIF4E1* allele. The amino acid changes induced in the resistance eIF4E1 include side-chain-lacking Glycine to negatively charged Glutamic Acid at position 110 (G110E). Since potyviral VPg is negatively charged, such substitution may repel the VPg protein from the two regions and reduce the interaction. Nevertheless, this substitution does not abolish the functionality of the protein as shown in the cap-affinity purification and yeast complementation assay. Remarkably, tomatoes possessing mutant eIF4E1$^{P69T/ET}$ have accumulated a very low level of PVY-N605 which is known to use either *Sl*eIF4E1 or *Sl*eIF4E2 in the absence of one of them in susceptible wild-type cultivars. The resistance against this isolate had been achieved only when the two eIF4Es were knocked-out or when the *SheIF4E1-pot1* allele from the wild relative *S. habrochaites*, was expressed (Gauffier et al., 2016; Lebaron et al., 2016). Our results show that with the exception of TEV-CAA10, the eIF4E1$^{P69T/ET}$ protein behaves similarly to the natural resistant *pot1*-encoded protein that appears to make the *Sl*eIF4E2 unavailable to the potyvirus infection by a yet unknown mechanism. Indeed, we confirm that editing eIF4E1 does not affect the accumulation of eIF4E2 mRNA or protein in leaves (Figure S6), suggesting a same mechanism that is yet to be characterized, rendering the eIF4E2 unavailable to the virus.

In all, our results prove that it is possible to pyramid modifications in two separate regions in one gene to establish resistance while maintaining the functionality and that indeed a combination of mutations in the two regions is necessary to establish resistance. To induce targeted genetic variation in the tomato *eIF4E1*, we have focused on two separate domains in the genome that are in the proximity in the predicted 3D protein structure (Figure S7) and achieved virus resistance.

Our work summarizes the progress of base editing in the recent years: from gene knock-in using TALENs to the CRISPR-based base editors. In the coming years, we expect that a similar strategy could be designed and carried on in a faster and more accurate manner, by using prime editor and/or CRISPR/Cas proteins with relaxed PAM requirement (Veillet et al., 2020b; Ren et al., 2021). These techniques should allow us to tinker the protein function as we design and to minimize the unfavorable effects brought about by the insertions/deletions or lack of precision. By taking the susceptibility factor eIF4E as an example, we exemplify how genome editing can be used to design complex alleles and generate useful traits that could boost plant breeding.

Material and methods

Plant material

Cherry tomato (*Solanum lycopersicum* var. *cerasiforme*) cv. West Virginia 106 (WVa106), was used for the co-editing of *eIF4E1* (Solyc03g005870) and *ALS1* (Solyc03g044330) and as the wildtype. The TILLING line *eif4e1*^{KO} was previously obtained through ethyl methanesulfonate (EMS) mutagenesis in the *S. lycopersicum* cv. M82 background (Piron et al., 2010). NIL-pot1 is a near-isogenic line of the *eIF4E1* allele *pot1* from *S. habrochaites* PI247087 introgressed in *S. lycopersicum* cv. Mospomorist (Lebaron et al., 2016). All plants were grown in a growth chamber at 24°C/18°C in the 16 hr day/8 hr night cycle.

Constructs for genome editing using TALEN®, CRISPR/Cas9, and Cytosine Base Editors (CBE)

TALENs® designed to target 5′-TTTCCTCTTTCAAATTggtgatagtgtagtgTAAGGGAAAACAGGGA (nucleotides recognized by the TAL effectors are capitalized) in the *SleIF4E1* genomic sequence were cloned into vectors, pCLS23222 and pCLS23224, and provided by Cellectis (Paris, France) (Data S2, S3). The DNA donor template for gene knock-in was synthesized by GenScript (Piscataway, NJ, USA) and cloned into pBGWD0 vector (Gateway vector VIB) through LR Gateway recombination. The linear DNA template designed for HR was generated by digestion with PacI and PmeI (Fig. 1b, Data S4). For the two targeted deletion steps using CRISPR/Cas9 after the NHEJ event, different pairs of single guide RNAs: pair 1 [sgRNA-DKM04 and sgRNA-eIF4E05] and pair 2 [sgRNA-Cl-01 and sgRNA- Cl-02] were cloned into pDe-Cas9-Hpt binary vector and pDe-Cas9-NptII binary vector, respectively (Danilo et al. 2018). The sgRNA cassettes were synthesized by Integrated DNA Technologies (Coralville, IA, USA). For co-base-editing of the *SleIF4E1* and *SlALS1*, an *SleIF4E1* sgRNA cassette with *StU6* promoter was cloned into pDicAID_nCas9-PmCDA1_NptII binary vector inserted with *AtU6p-SlALS1* guide (Veillet et al., 2019) (Figure S4a and b, Data S5) and named CBE-4E1-RII. The *SleIF4E1* sgRNA cassette was synthesized by TwistBioscience (San Francisco, CA, USA). TALEN targets and CRISPR guide RNA sequences are listed in Table S2.

Genetic transformation of tomato

TALEN® plasmids and the linear DNA template were delivered into leaves from four-week-old tomato plants by biolistic method. Other transformations with CRISPR/Cas9 or CBE vectors were performed on cotyledon pieces from 8-12-day-old tomato seedlings by

Agrobacterium-mediated transformation using *Agrobacterium tumefaciens* C58pGV2260 strain. Plant regeneration and transformant selection following both transformations were performed as previously described (Mazier et al., 2011; Danilo et al., 2018; Veillet et al., 2019). *In vitro* culture was carried out in a controlled environment at the temperature of 22°C/18°C with 16 hr/8 hr (day/night) photoperiod.

Genotyping by PCR and High-Resolution Melting analysis

Genomic DNA was extracted from fresh tomato leaf samples according to (Fulton et al., 1995). T0 plantlets after editing with TALEN and CRISPR/Cas9 were screened by PCR using specific primers for each transformation. T0 plantlets after co-base-editing using CBE were screened by HRM analysis using specific primers and Precision Melt Supermix (BioRad, Hercules, USA) with the CFX96™ Real-Time PCR Detection System (BioRad), as previously described (Veillet et al., 2019). Melting-curve analysis was performed using the Precision Melt Analysis software (BioRad, Hercules, USA). PCR was also performed on the *SleIF4E1* genomic sequence to further verify. *GAPDH* (*Glyceraldehyde-3-Phosphate Dehydrogenase*) gene was amplified as the technical control with gapdhF/R. PCR products were Sanger-sequenced by Genoscreen (Lille, France). T-DNA detection was done by PCR. All the primer sequences and pairs used for detection are available in Table S3.

Total RNA extraction and cDNA analysis

Total RNA was extracted from young tomato leaf tissue (< 100 mg) of each plant in 500 µL of TRI-Reagent (Sigma-Aldrich, St. Louis, USA) and quality-checked by concentration measurement on Nanodrop ND-1000 (ThermoFisher, Waltham, USA). Reverse transcription (RT) of the extracted RNA samples (1 µg) was performed using AMV (Avian Myeloblastosis Virus) Reverse Transcription Kit (Invitrogen, Waltham, USA) and oligo-dT18 (1 µM) to obtain coding DNA (cDNA). The cDNA for *eIF4E1* was amplified with a primer pair c4e1F/R and *GAPDH* (*Glyceraldehyde-3-Phosphate Dehydrogenase*) gene as the internal control with gapdhF/R.

Protein extraction and western blot analysis

Total protein was extracted from 100 mg of young tomato leaves, and 150 µL of Laemmli 2x DTT buffer was added to the crushed leaves. The homogenized solution was boiled for 5 min. The supernatant after centrifugation at 4°C was collected and conserved at -20°C. Western blot was performed as previously described using specific antibodies (Kuroiwa et al., 2022). The anti-*Sl*eIF4E1 polyclonal serum (Gauffier et al., 2016) was diluted 1/2000, and combined with secondary goat horseradish peroxidase-labelled anti-

rabbit serum (Sigma-Aldrich, St. Louis, USA) diluted 1/5000. As loading controls, monoclonal anti-plant actin antibodies (1/10000 dilution) (Sigma-Aldrich) were used with horseradish peroxidase-labelled anti-mouse serum (1/10000 dilution) (Sigma-Aldrich). Dilution was done in TBS 1x + 5% milk (Tween 0.05% + Triton X-100 0.2% for cap-affinity purification).

Virus isolates and infection assays using DAS-ELISA quantification

Potyvirus infection assays were performed using PepMoV (Mazier et al., 2011), PVY-N605 (Parrella et al., 2002), PVY-LYE84, PVY-SON41 (Moury et al., 2004), and TEV-CAA10 (Charron et al., 2008). Isolates were maintained in *Nicotiana benthamiana* plants 12 days before inoculation. For the PVY infection monitoring, the infectious clone of PVY-N605 tagged with GFP (hereafter, PVY-N605-GFP) incorporated in pCambia vector was multiplied on *N. tabacum* cv. Xanthi plants before inoculation of 21-day-old tomato plants. Inocula were prepared from 1 g (fresh weight) of the infected tobacco leaves ground with 4 mL of grinding buffer: potassium phosphate buffer (0.03 M, pH = 7) with 0.2% of diethyldithiocarbamate, 80 mg active charcoal and 80 mg carborundum. Six to twelve plants per genotype were mechanically inoculated on the cotyledons of 14-day-old tomato plants. At 21 days post-inoculation (dpi), 1 g of leaves from each plant was collected and ground in 4 mL of grinding buffer. To measure the virus accumulation, DAS-ELISA was performed using anti-PVY or anti-TEV antiserum (Sediag, Bretenière, France) or anti-polypoty antiserum (Agdia, Grigny, France) for PepMoV according to the manufacturer's instructions. The absorbance was measured by a spectrophotometer at wavelength 405 nm.

Virus detection by fluorometric camera analysis

Monitoring of PVY-N605-GFP infection was carried out with a closed fluorometric camera FluorCam FC 800-C/1010- GFP (Photon System Instruments, Drasov, Czech Republic) equipped with a GFP filter. For each genotype, 10-15 plants were used. GFP fluorescence was captured by the camera upon excitation at 420 nm and the intensity of the GFP signal was indicated by the color scale.

m^{7}GTP cap-affinity purification

Total soluble proteins were extracted from 300 mg leaf tissue in 1200 µL binding buffer with protease inhibitor (+PI) (performed in three sets 100 mg:400 µL each). After centrifugation at 13,000 rpm for 10 min at 4 °C, 50 µL of the supernatant (INPUT) was recovered, and the rest was incubated with 150 µL of γ-aminophenyl-m^{7}GTP (C10-

spacer)-agarose beads (Jena Bioscience, Jena, Germany) pre-equilibrated with binding buffer without protease inhibitor at 4°C for 1.5 h. The beads were pelleted for 1 min at 13,000 rpm and washed three times with binding buffer +PI at 4°C to remove unbound proteins. The INPUT was diluted two-fold with Laemmli DTT 2x buffer. The proteins linked to the cap analog were eluted using 100 µL Laemmli DTT 2x buffer (OUTPUT) and analyzed by western blot assay as described above.

Yeast complementation assay

Amplified cDNA from the different *eIF4E1* alleles was subcloned using the gateway technology (ThermoFisher) in a converted p424-GDP-DEST plasmids. Yeast complementation method was adapted from (Gauffier et al., 2016). Briefly, attB sequence was attached to cDNA for BP recombination reaction to generate the entry clones. The entry clone was then used for LR recombination reaction with the destination vector p424-GPD to generate the tryptophan-selectable expression vector with a constitutive promoter pGAD. Then, yeast (*Saccharomyces cerevisiae*) strain J055 (cdc33-Δ: LEU2 leu2 ura3 his3 trp1 ade2 [YCp33supex-h4E URA3]) that lacks its native SceIF4E and expresses human eIF4E (H eIF4E) under galactose-inducible, glucose-repressible promoter pGAL was transformed with each expression vector. The transformants were selected on galactose-containing, uracil/tryptophan deficient media (-UW); then, the transformed colonies were spotted on glucose-containing, tryptophan deficient media (-W) to see if the eIF4E expressed from the expression vector complements the lack of its native eIF4E or repressed H eIF4E. Incubation for the complementation assay was done at 20°C/16°C with 16 hr/8 hr cycle. The empty vector is p424-RfA.

3D modelization of eIF4E structures

Homology modelling of the eIF4E1 protein were carried out as previously (Moury et al., 2020) using the YASARA software (http://www.yasara.org/), using structural data from pea (*Pisum sativum*) eIF4E (GenBank ID: AY423375, PDB ID: 2WMC-C) as the template. Protein structure was visualized using PyMol software (https://pymol.org/).

Statistical analysis

All the statistical tests were performed on the free software R (https://www.R-project.org/). Kruskal-Wallis non-parametric tests were performed to evaluate the significance of differences between genotypes using the "pgirmess" package. As the *post-hoc* test, Dunn test with p-value adjustment by Benjamini-Yekutieli method or Benjamini-

Krieger-Yekutieli method was performed using “dunn.test” package to determine the significance.

Acknowledgments

The authors thank Emmanuel Botton (INRAE GAFL) for taking care of tomato plants, Nathalie Truglio (INRAE Pathologie Vegetale) for supplying tobacco plants, Pierre Barret, Caroline Tassy, and Anne Partier (INRAE GDEC) for helpful advices to set up biolistic experiments, and Luc Mathis (Cellectis) for TALEN®s design and production.

Funding

This work was supported by the French national Research Agency (ANR11-BTBR-0001-GENIUS). The IJPB benefits from the support of Saclay Plant Sciences-SPS (ANR-17-EUR-0007).

Author contributions

KK, FV, FN, MM, JLG designed the experiments. KK, BD, LP, CT, FV, FD, MM carried out the experiments. FD, PD contributed new reagents. KK and JLG wrote the manuscript with contributions from all authors. All authors read and approved the final manuscript.

Competing interests

TALEN® is a registered trademark owned by Cellectis and a patented technology.

References

Bastet, A., Robaglia, C., and Gallois, J. L. (2017). eIF4E Resistance: Natural Variation Should Guide Gene Editing. Trends Plant Sci. 22:411–419.

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

Bastet, A., Zafirov, D., Giovinazzo, N., Guyon-Debast, A., Nogué, F., Robaglia, C., and Gallois, J. L. (2019). Mimicking natural polymorphism in eIF4E by CRISPR-Cas9 base editing is associated with resistance to potyviruses. Plant Biotechnol. J. 17:1736–1750.

Brown, J. W. S., and Simpson, C. G. (1998). SPLICE SITE SELECTION IN PLANT PRE-mRNA SPLICING. Annu. Rev. Plant Physiol. Plant Mol. Biol. 49:77–95.

Cermak, T., Doyle, E. L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J. A., Somia, N. V., Bogdanove, A. J., and Voytas, D. F. (2011). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res. 39:7879–7879.

Chandrasekaran, J., Brumin, M., Wolf, D., Leibman, D., Klap, C., Pearlsman, M., Sherman, A., Arazi, T., and Gal-On, A. (2016). Development of broad virus resistance in non-transgenic cucumber using CRISPR/Cas9 technology. Mol. Plant Pathol. 17:1140–1153.

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

Coutinho de Oliveira, L., Volpon, L., Rahardjo, A. K., Osborne, M. J., Culjkovic-Kraljacic, B., Trahan, C., Oeffinger, M., Kwok, B. H., and Borden, K. L. B. (2019). Structural studies of the eIF4E-VPg complex reveal a direct competition for capped RNA: Implications for translation. Proc. Natl. Acad. Sci. U. S. A. 116:24056–24065.

Danilo, B., Perrot, L., Botton, E., Nogué, F., and Mazier, M. (2018). The DFR locus: A smart landing pad for targeted transgene insertion in tomato. PLoS One 13:e0208395.

Eisenstein, M. (2022). Base edit your way to better crops. Nature 604:790–792.

Fulton, T. M., Chunwongse, J., and Tanksley, S. D. (1995). Microprep protocol for extraction of DNA from tomato and other herbaceous plants. Plant Mol. Biol. Rep. 13:207–209.

Gao, C. (2021). Genome engineering for crop improvement and future agriculture. Cell 184:1621–1635.

Gao, Z., Johansen, E., Eyers, S., Thomas, C. L., Noel Ellis, T. H., and Maule, A. J. (2004). The potyvirus recessive resistance gene, sbm1, identifies a novel role for translation initiation factor eIF4E in cell-to-cell trafficking. Plant J. 40:376–385.

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

Kang, B.-C., Yeam, I., Frantz, J. D., Murphy, J. F., and Jahn, M. M. (2005). The pvr1 locus in Capsicum encodes a translation initiation factor eIF4E that interacts with Tobacco etch virus VPg. Plant J. 42:392–405.

Kang, B.-C., Yeam, I., Li, H., Perez, K. W., and Jahn, M. M. (2007). Ectopic expression of a recessive resistance gene generates dominant potyvirus resistance in plants. Plant Biotechnol. J. 5:526–536.

Kuroiwa, K., Thenault, C., Nogué, F., Perrot, L., Mazier, M., and Gallois, J.-L. (2022). CRISPR-based knock-out of eIF4E2 in a cherry tomato background successfully recapitulates resistance to pepper veinal mottle virus. Plant Sci. 316:111160.

Lebaron, C., Rosado, A., Sauvage, C., Gauffier, C., German-Retana, S., Moury, B., and Gallois, J. L. (2016). A new eIF4E1 allele characterized by RNAseq data mining is associated with resistance to potato virus Y in tomato albeit with a low durability. J. Gen. Virol. 97:3063–3072.

Li, Z., Xiong, X., Wang, F., Liang, J., and Li, J.-F. (2019). Gene disruption through base editing-induced messenger RNA missplicing in plants. New Phytol. 222:1139–1148.

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

Miras, M., Truniger, V., Silva, C., Verdaguer, N., Aranda, M. A., and Querol-Audí, J. (2017). Structure of eIF4E in Complex with an eIF4G Peptide Supports a Universal Bipartite Binding Mode for Protein Translation. Plant Physiol. 174:1476–1491.

Moscou, M. J., and Bogdanove, A. J. (2009). A simple cipher governs DNA recognition by TAL effectors. Science 326:1501.

Moury, B., Morel, C., Johansen, E., Guilbaud, L., Souche, S., Ayme, V., Caranta, C., Palloix, A., and Jacquemond, M. (2004). Mutations in potato virus Y genome-linked protein determine virulence toward recessive resistances in Capsicum annuum and Lycopersicon hirsutum. Mol. Plant. Microbe. Interact. 17:322–329.

Moury, B., Charron, C., Janzac, B., Simon, V., Gallois, J. L., Palloix, A., and Caranta, C. (2014). Evolution of plant eukaryotic initiation factor 4E (eIF4E) and potyvirus genome-linked protein (VPg): a game of mirrors impacting resistance spectrum and durability. Infect. Genet. Evol. 27:472–480.

Moury, B., Lebaron, C., Szadkowski, M., Ben Khalifa, M., Girardot, G., Bolou Bi, B. A., Koné, D., Nitiema, L. W., Fakhfakh, H., and Gallois, J.-L. (2020). Knock-out mutation of eukaryotic initiation factor 4E2 (eIF4E2) confers resistance to pepper veinal mottle virus in tomato. Virology 539:11–17.

Nicaise, V., German-Retana, S., Sanjuán, R., Dubrana, M.-P., Mazier, M., Maisonneuve, B., Candresse, T., Caranta, C., and LeGall, O. (2003). The eukaryotic translation initiation factor 4E controls lettuce susceptibility to the Potyvirus Lettuce mosaic virus. Plant Physiol. 132:1272–1282.

Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K. Y., et al. (2016). Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science 353.

Parrella, G., Ruffel, S., Moretti, A., Morel, C., Palloix, A., and Caranta, C. (2002). Recessive resistance genes against potyviruses are localized in colinear genomic regions of the tomato (Lycopersicon spp.) and pepper (Capsicum spp.) genomes. Theor. Appl. Genet. 105:855–861.

Pechar, G. S., Donaire, L., Gosalvez, B., García-Almodovar, C., Sánchez-Pina, M. A., Truniger, V., and Aranda, M. A. (2022). Editing melon eIF4E associates with virus resistance and male sterility. Plant Biotechnol. J. Advance Access published July 1, 2022, doi:10.1111/pbi.13885.

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

Poulicard, N., Pacios, L. F., Gallois, J.-L., Piñero, D., and García-Arenal, F. (2016). Human Management of a Wild Plant Modulates the Evolutionary Dynamics of a Gene Determining Recessive Resistance to Virus Infection. PLoS Genet. 12:e1006214.

Ren, Q., Sretenovic, S., Liu, S., Tang, X., Huang, L., He, Y., Liu, L., Guo, Y., Zhong, Z., Liu, G., et al. (2021). PAM-less plant genome editing using a CRISPR-SpRY toolbox. Nat Plants 7:25–33.

Robaglia, C., and Caranta, C. (2006). Translation initiation factors: a weak link in plant RNA virus infection. Trends Plant Sci. 11:40–45.

Ruffel, S., Dussault, M.-H., Palloix, A., Moury, B., Bendahmane, A., Robaglia, C., and Caranta, C. (2002). A natural recessive resistance gene against potato virus Y in pepper corresponds to the eukaryotic initiation factor 4E (eIF4E). Plant J. 32:1067–1075.

Ruffel, S., Gallois, J. L., Lesage, M. L., and Caranta, C. (2005). The recessive potyvirus resistance gene pot-1 is the tomato orthologue of the pepper pvr2-eIF4E gene. Mol. Genet. Genomics 274:346–353.

Ruffel, S., Gallois, J.-L., Moury, B., Robaglia, C., Palloix, A., and Caranta, C. (2006). Simultaneous mutations in translation initiation factors eIF4E and eIF(iso)4E are required to prevent pepper veinal mottle virus infection of pepper. J. Gen. Virol. 87:2089–2098.

Saha, S., and Mäkinen, K. (2020). Insights into the Functions of eIF4E-Binding Motif of VPg in Potato Virus A Infection. Viruses 12.

Schmitt-Keichinger, C. (2019). Manipulating Cellular Factors to Combat Viruses: A Case Study From the Plant Eukaryotic Translation Initiation Factors eIF4. Front. Microbiol. 10:17.

Simpson, C. G., Clark, G., Davidson, D., Smith, P., and Brown, J. W. (1996). Mutation of putative branchpoint consensus sequences in plant introns reduces splicing efficiency. Plant J. 9:369–380.

Tang, Y., Abdelrahman, M., Li, J., Wang, F., Ji, Z., Qi, H., Wang, C., and Zhao, K. (2021). CRISPR/Cas9 induces exon skipping that facilitates development of fragrant rice. Plant Biotechnol. J. 19:642–644.

Tavert-Roudet, G., Anne, A., Barra, A., Chovin, A., Demaille, C., and Michon, T. (2017). The Potyvirus Particle Recruits the Plant Translation Initiation Factor eIF4E by Means of the VPg covalently Linked to the Viral RNA. Mol. Plant. Microbe. Interact. 30:754–762.

Veillet, F., Perrot, L., Chauvin, L., Kermarrec, M. P., Guyon-Debast, A., Chauvin, J. E., Nogué, F., and Mazier, M. (2019). Transgene-free genome editing in tomato and potato plants using Agrobacterium-mediated delivery of a CRISPR/Cas9 cytidine base editor. Int. J. Mol. Sci. 20.

Veillet, F., Durand, M., Kroj, T., Cesari, S., and Gallois, J. L. (2020). Precision Breeding Made Real with CRISPR: Illustration through Genetic Resistance to Pathogens. Plant Communications 1.

Wang, Y., Ma, M., Xiao, X., and Wang, Z. (2012). Intronic splicing enhancers, cognate splicing factors and context-dependent regulation rules. Nat. Struct. Mol. Biol. 19:1044–1052.

Yoon, Y.-J., Venkatesh, J., Lee, J.-H., Kim, J., Lee, H.-E., Kim, D.-S., and Kang, B.-C. (2020). Genome editing of eIF4E1 in tomato confers resistance to pepper mottle virus. Front. Plant Sci. 11:1098.

Zhang, X., Zhang, Y., Wang, T., Li, Z., Cheng, J., Ge, H., Tang, Q., Chen, K., Liu, L., Lu, C., et al. (2019). A Comprehensive Map of Intron Branchpoints and Lariat RNAs in Plants. Plant Cell 31:956–973.

Manuscript 2 Supporting Figures



Figure S1 Alignment of the amino acid sequences from potyvirus susceptible and resistant tomato and pepper accessions.
"SleIF4E1" and "CaeIF4E" are the representative *eIF4E1* sequence from the susceptible cultivars. "pot1", "pvr1", and "pvr2 1" (pvr2[1]) are resistance allele sequences. Green and blue bars indicate the position of the Region I and II, respectively. Tomato P69T and the equivalent pepper P66T are shaded in gray.

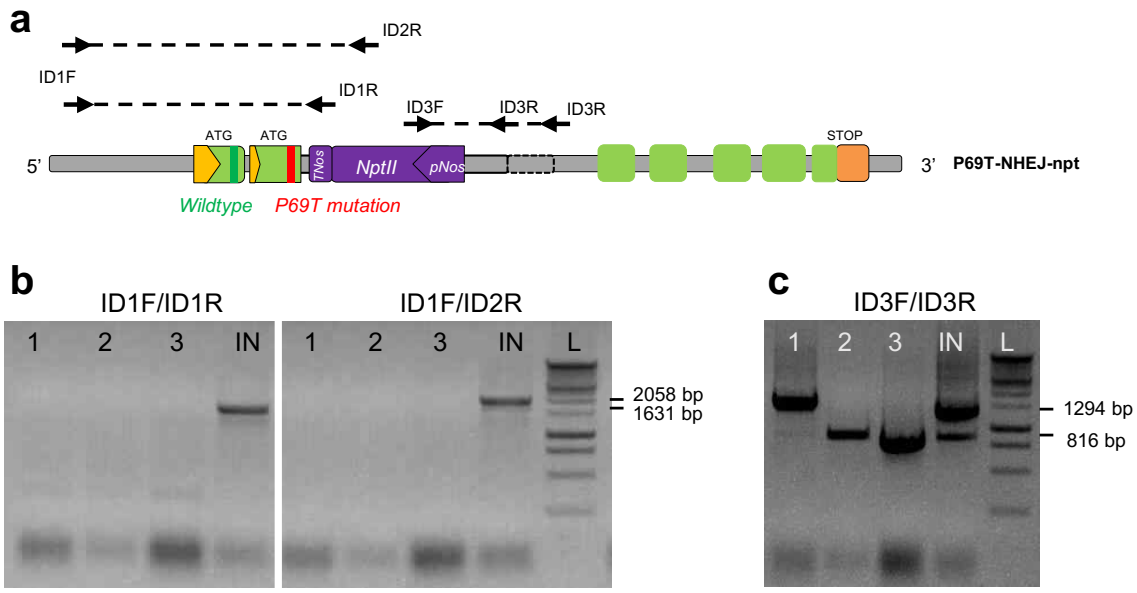


Figure S2 Detection of insertion of the DNA donor template after the biolistic experiment with TALEN.
a Primers designed to detect the DNA donor template insertion.. **b** Detection of the 5′-end of the insert. **c** Detection of the 3′-end of the insert. The indicated base pairs correspond to the expected band size in an NHEJ event. 1-3 are plant lines with incomplete insertion on the 5′-end. (In C, 1 shows an insertion by NHEJ and 2 and 3 by HR.) IN: NHEJ inserted line (P69T-NHEJ-npt line), L: ladder.

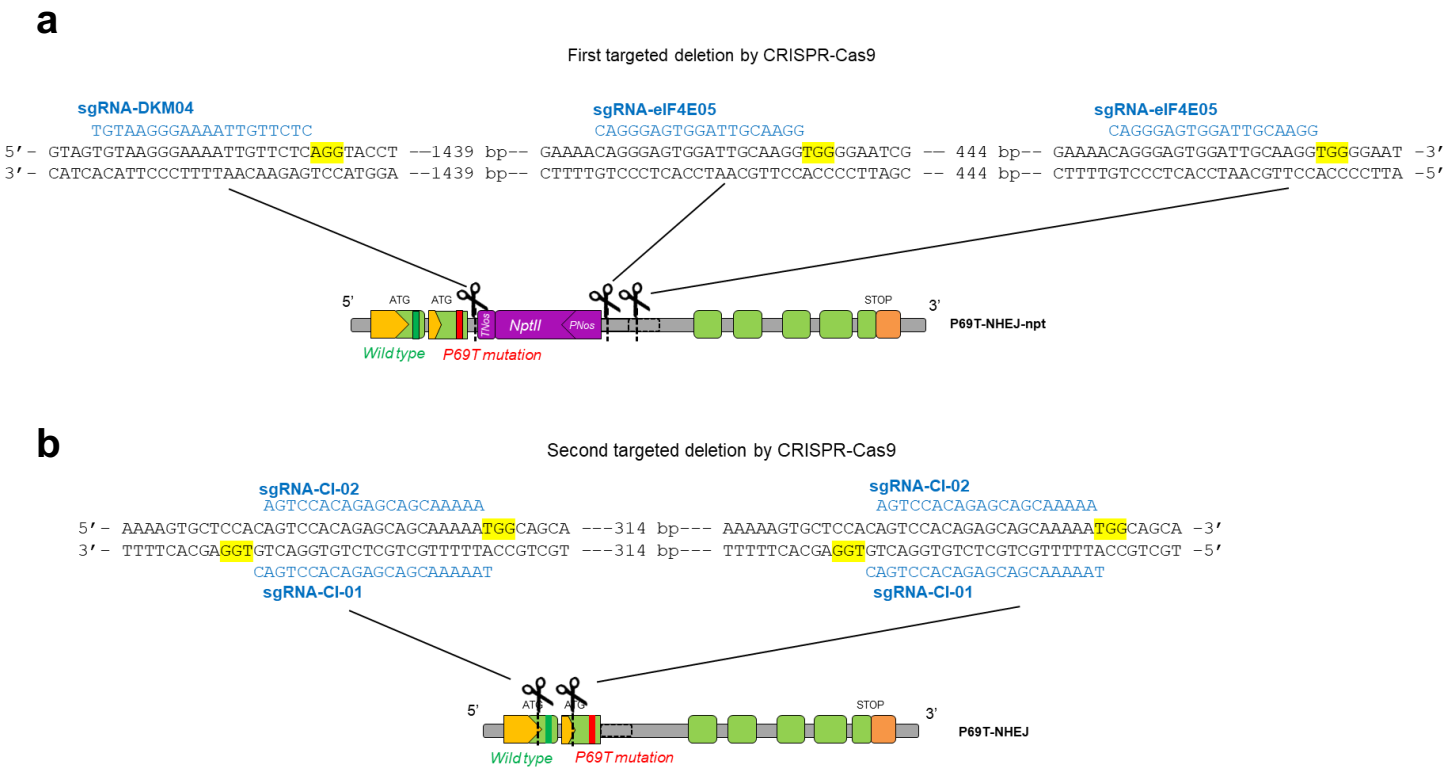


Figure S3 Details on the target deletion using CRISPR-Cas9 to produce eIF4E1^{P69T}-expressing plants after the TALEN method.

Scissors indicate the approximate cleavage sites in the targeted genome. The number of nucleotides between each guide RNA is indicated in the figure. **a** First targeted deletion via CRISPR-Cas9 on the 3′-end of the inserted sequence. The guide RNA sequences of Pair 1: sgRNA-DKM04 and sgRNA-eIF4E05 are in blue and their PAM sequences are highlighted in yellow. sgRNA-eIF4E05 targeted two regions of the same sequence in the eIF4E1 genome because of the partial redundancy of intron I caused by the NHEJ insertion. This resulted in a 1949-bp deletion in the targeted genome, eliminating the 3′-end of the inserted donor including the *NptII* gene and a part of native intron I. **b** Second targeted deletion via CRISPR-Cas9 on the 5′-end of the inserted sequence. The guide RNA sequences of Pair 2: sgRNA-CI-01 and sgRNA-CI-02 are in blue and the PAM sequences are highlighted in yellow. Both guide RNAs targeted two regions in the eIF4E1 genome due to the partial redundancy of exon I caused by the NHEJ insertion. This resulted in a 356-bp deletion, eliminating the redundant part including P69 codon from the native exon I.

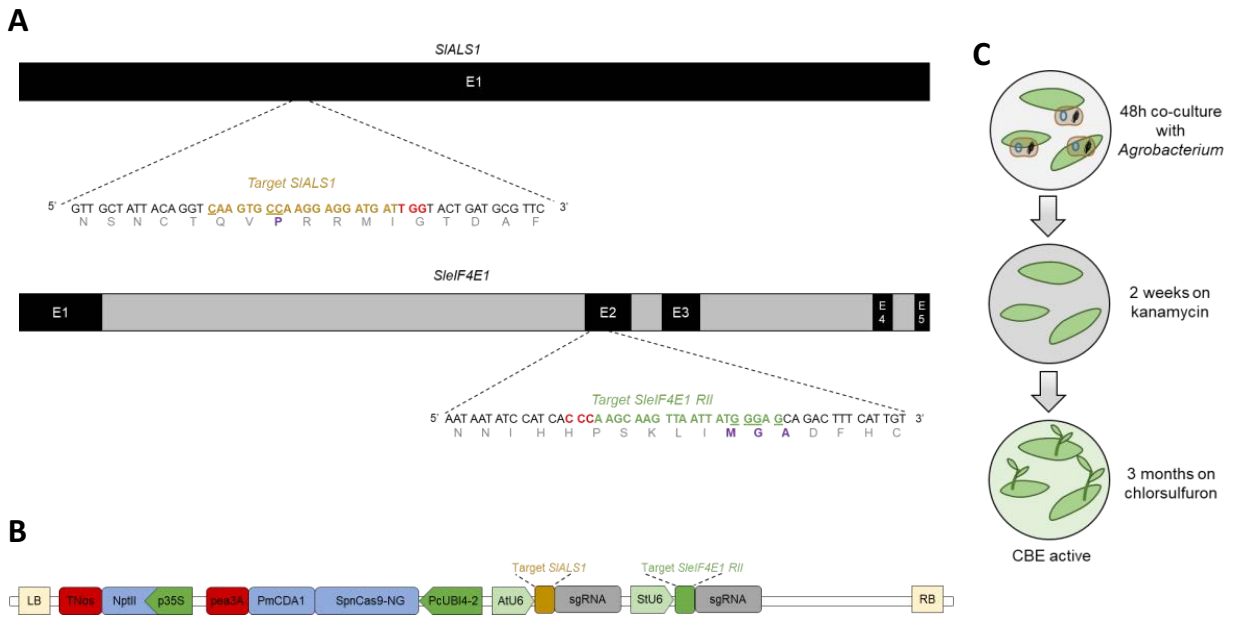


Figure S4 The targets, constructs, and scheme for the co-base-editing of *SlALS1* and *SleIF4E1* genes.

a Target sequences located in the *SlALS1* and *SleIF4E1* genes. Dark and gray boxes represent exons and introns, respectively. The PAM sequence is indicated in red. Targeted nucleotides are underlined and the resulting amino acids are indicated in purple. **b** T-DNA of the CBE-4E1-RII construct used to edit *SlALS1* and *SleIF4E1* genes. LB: left border; RB: right border. **c** Schematic of the production of ALS-co-edited tomato plants through *Agrobacterium*-mediated transformation.

A

···ATACGGCATCGTATTTAGGGAAAGAAATCACAGTGAAGCATCCATTGGAGCATTCATGGACTTT
TTGGTTTGATAACCCTACCACTAAATCTCGACAAACTGCTTGGGGAAGCTCACTTCGAAATGTCTA
CACTTTCTCCACTGTTGAAGATTTTTGGGGGTAAGTTTTTTT·····AATCTTGAACCATGTTGCTTA
TTCTACAAACAGTGCTTACAATAATATCCATCACCCAAGCAAGTTAATT**ATAAAAACA**GACTTTCA
TTGTTTTAAGCACAAAATTGAGCCAAAGTGGGAAGATCCTGTATGTGCCAATGGAGGGACGTGGA
AAATGAGTTTTTCGAAGGGTAAATCTGATACCAGCTGGCTGTATACGGTATTCCGAAGTTATTTCC
ATCCAGCCCTTAATGATAGGTCATTCTAGTAATGTTATTTTCCCCTTTGATATAATTTCCACTCTTG
TTTTCTTATATGGAATTATTGTAGCTGCTGGCAATGATTGGACATCAATTCGATCATGGAGATGAA
ATTTGTGGAGCAG···

B

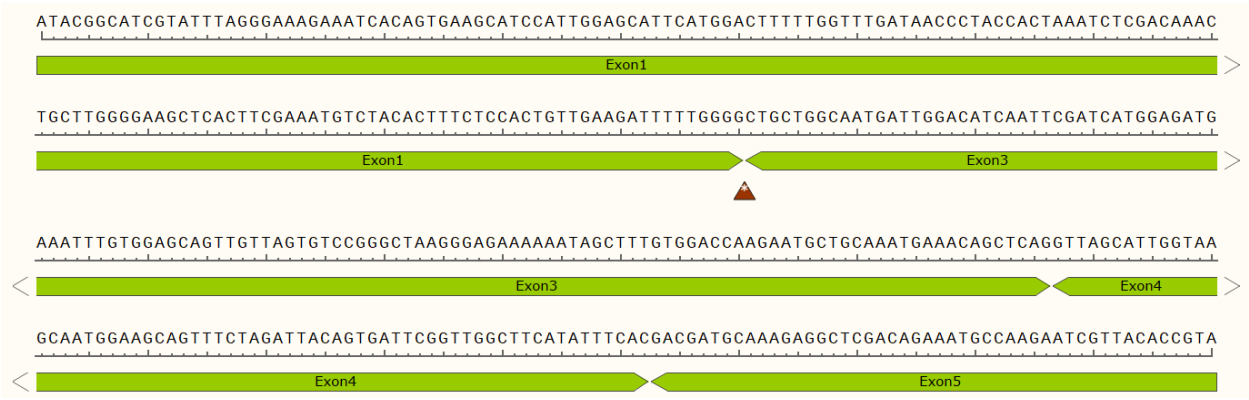


Figure S5 Genomic and cDNA sequencing of P69T/M109I+G110K+A111T indicate the abnormal splicing of the *eIF4E1* gene.

a Genomic sequencing of the Region I and II periphery from an IKT-modified plant line (*eIF4E1*^{IKT} line). Shaded in gray are the known splicing donor and acceptor sites. Colored sequences correspond to the exons: exon 1 and 3 in green and exon 2 in red. Underlined sequence is identified in the cDNA sequencing. Bolded is the IKT-coding sequence. **b** A part of cDNA sequenced from the same IKT-modified plant line. Red triangle indicates the site where the wild-type exon 2 is normally found.

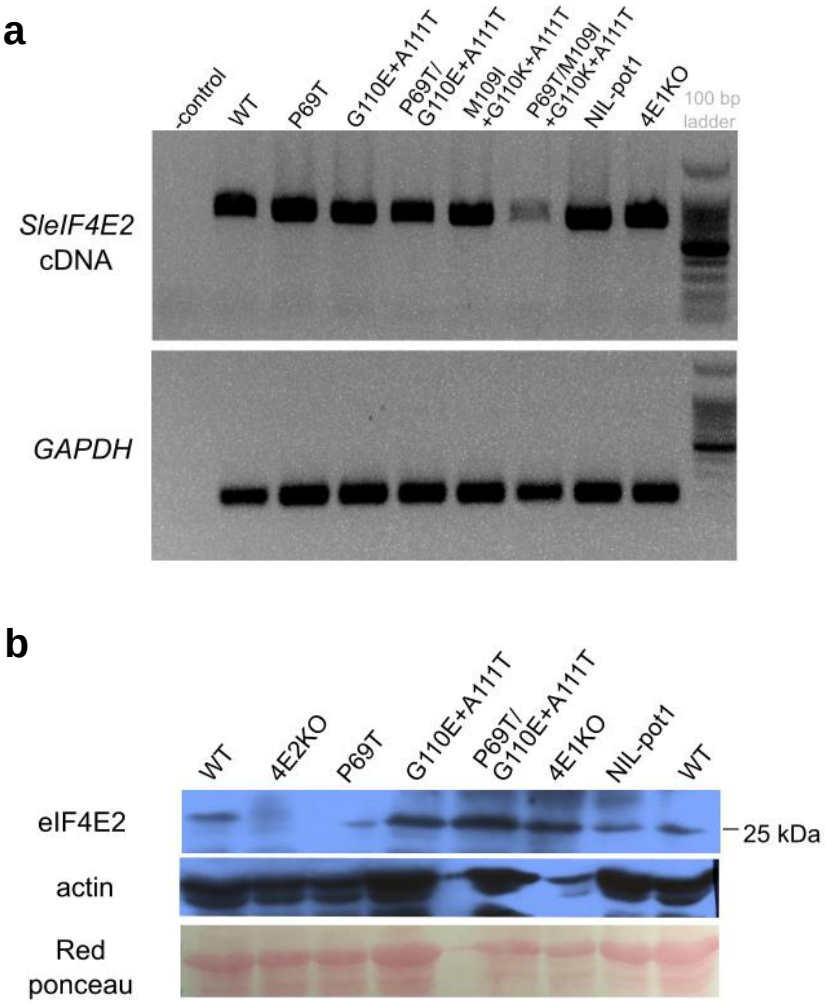


Figure S6 eIF4E editing does not affect the *eIF4E2* mRNA expression and eIF4E2 protein accumulation level *in planta*.

a Amplification of *eIF4E2* cDNA following reverse transcription; the expected size of the PCR fragment is 666 bp for WT. The GAPDH cDNA is amplified as the internal control. Representative sample from two independent mutant lines are shown. Primers used for *SleIF4E2* are forward: AgagaattcatgGCTGATGAACTAAACAAAGCT and reverse: AgactcgagCTACACGGAATAACGATTTTTGGC and gapdhf/r pair. **b** Western blot analysis to detect the *Sl*eIF4E2 accumulation in the cotyledons and true leaf of the edited lines. Actin accumulation and red ponceau of each blot are shown as the loading control. Mutant lines are indicated by their expected amino acid modifications in the SleIF4E1 protein. Two independent edited lines were assayed per genotype. The expected molecular weight of SleIF4E2 is 25 kDa. M109I+G110K+A111T and P69T/M109I+G110K+A111T samples were omitted from the western blot due to loading constraint.

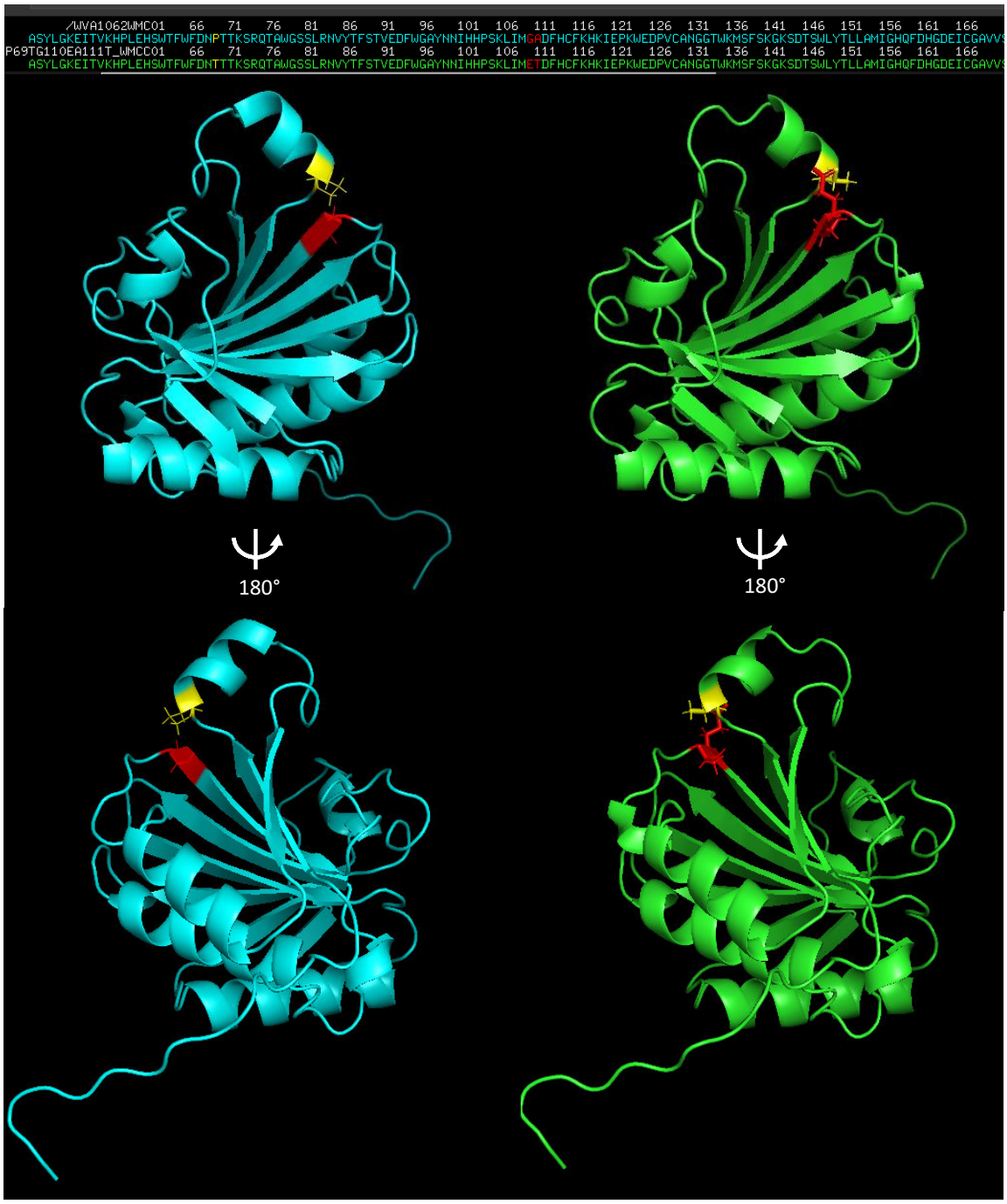


Figure S7 Three-dimensional homology modelling of the tomato eIF4E1 protein, based on crystallography data from the *Pisum sativum* eIF4E 3D structure (PDB ID: 2WMC-C).
WVa106 is shown on the left in blue and the edited P69T/G110E+A111T-eIF4E1 protein on the right in green. P69 and T69 are colored in yellow and G110 A111 and E110 T111 in red, with apparent side chains showing. 3D modelization was performed using Yasara and shown with PyMol as in (Moury et al., 2020). 180° rotation along the y axis is represented below.

Supporting Data

Data S1 Sequence analysis of plant #266 with insertion of the donor template

GCACTTCAAACAGGAGCAACTTAAGTTTGATAGAAGCTTTTCTTGTATTTTTTTTTTTGGTTAATTACTATCCTAAAGTTTTGTAATATTGAAA
TGTTTGCATTACTCATCTCTCATCTATCCTAGCCTCACTGAACAAATTTACATGATACATGTTGCTCCCGCGGTGATAGATATCCTTTAGAATT
AGTTGAGATTTGCGAAAACTGAACTTGTGCGAAATGTTCGCATTATTCTCTTCTCATCTCTCATCAGTCCTAGCCTCGCTGAATAAATTTACAT
GACATATGTTGCTAGTGCTGGAGGTGATAGATATACTTTAAAATTAGTTGAAATTTGAGAAAACTGAACTTGTTCGAAATGTTTGCATTGTTCT
GTTCTCATCTCTCATCTATTATAGCTTCGCAGAGTTTACCTAATAGTTATTACACCACTCTATGTGAAATGTTTGCTTTATTTTCTTCTCATCT
CTCATTTGTCATAGTCTCGGTAAACAGAGTTATCTGATACCTGTTATTAATGAGAGGTGATAGAAATCCTTTAAAATTAATTGAGATGTGTGAA
AACTGACCGAAACACCAACATTATAAATATAAAATATTTATAACACATGAAATCTCATTTAAATATTTATTTGAATTTCTTTCTTATTATAAAA
ATCACTTTTTTACTTATAAATAAAAGAAGTTCAAACTATATGATTAGCCTTACATAATTAACCAAGAAACTAAAAATTTGTAATTACATTATGA
CAAACAAACAAATTAATATAACTGTTCATGTGACATTATTGGTATCAAACTTCTTAATTGCTAATTTTGAAAAATAAAGCTTCCATATTCTGAA
CTAGATTAAAATTAAAATCCACGTAAATTAATTATTTGATTCTCACGATACCTAATTTAGGTATTTACAACATTAGTACTTCTACAATTTAAGT
ATTACACTATGGTCCAAACAGTTCTTATAAATACCATTTTTCTTAATTAAGTAAAAAAAAAGTGCTCCA**CAGTCCACAGAGCAGCAAAAATGG**C
AGCAGCTGAAATGGAGAGAACGATGTCGTTTGATGCAGCTGAGAAGTTGAAGGCCGCCGATGGAGGAGGAGGAGAGGTAGACGATGAACTTGAA
GAAGGTGAAATTGTTGAAGAATCAAATGATACGGCATCGTATTTAGGGAAAGAAATCACAGTGAAGCATCCATTGGAGCATTCATGGACTTTTT
GGTTTGATAACCCTACCACTAAATCTCGACAAACTGCTTGGGGAAGCTCACTTCGAAATGTCTACACTTTCTCCACTGTTGAAGATTTTTGGGG
GTAAGTTTTTTTTTTTCCTCTTTCAAATTG[GTAAAAAAAAAAGTGCT**CCACAGTCCACAGAGCAGCAAAAATGG**CAGCAGCTGAAATGGAGAGA
ACGATGTCGTTTGATGCAGCTGAGAAGTTGAAGGCCGCCGATGGAGGAGGAGGAGAGGTAGACGATGAACTTGAAGAAGGTGAAATTGTTGAAG
AATCAAATGATACGGCATCGTATTTAGGGAAAGAAATCACAGTGAAGCATCCATTGGAGCATTCATGGACTTTTTGGTTTGATAACACAACCAC
TAAATCTCGACAAACTGCTTGGGGAAGCTCACTTCGAAATGTCTACACTTTCTCCACTGTTGAAGATTTTTGGGGGTAAGTTTTTTTTTTTCCT
CTTTCAAATTGGTGATAGTGTAG**TGTAAGGGAAAA**ttgttctc**agg**tacctcagccagacagcttgcatgccggtcgatctagtaacatagatg
acaccgcgcgcgataatttatcctagtttgcgcgctatattttgttttctatcgcgtattaaatgtataattgcgggactctaatcataaaaac
ccatctcataaataacgtcatgcattacatgttaattattacatgcttaacgtaattcaacagaaattatatgataatcatcgcaagaccggca
acaggattcaatcttaagaaacttattgccaaatgtttgaacgatctgcttgactctagctagagtccgaaccccagagtcccgctcagaaga
actcgtcaagaaggcgatagaaggcgatgcgctgcgaatcgggagcggcgataccgtaaagcacgaggaagcggtcagcccattcgccgccaag
ctcttcagcaatatcacgggtagccaacgctatgtcctgatagcggtccgccacacccagccggccacagtcgatgaatccagaaaagcggcca
ttttccaccatgatattcggcaagcaggcatcgccgtgggtcacgacgagatcctcgccgtcgggcatccgcgccttgagcctggcgaacagtt
cggctggcgcgagcccctgatgctcttcgtccagatcatcctgatcgacaagaccggcttccatccgagtacgtgctcgctcgatgcgatgttt
cgcttggtggtcgaatgggcaggtagccggatcaagcgtatgcagccgccgcattgcatcagccatgatggatactttctcggcaggagcaagg
tgagatgacaggagatcctgccccggcacttcgcccaatagcagccagtcccttcccgcttcagtgacaacgtcgagcacagctgcgcaaggaa
cgcccgtcgtggccagccacgatagccgcgctgcctcgtcttggagttcattcagggcaccggacaggtcggtcttgacaaaaagaaccgggcg
cccctgcgctgacagccggaacacggcggcatcagagcagccgattgtctgttgtgcccagtcatagccgaatagcctctccacccaagcggcc
ggagaacctgcgtgcaatccatcttgttcaatcatgcctcgatcgagttgagagtgaatatgagactctaattggataccgaggggaatttatg
gaacgtcagtggagcattttgacaagaaatatttgctagctgatagtgaccttaggcgacttttgaacgcgcaataatggtttctgacgtatg
tgcttagctcattaaactccagaaacccgcggctgagtggctccttcaacgttgcggttctgtcagttccaaacgtaaaacggcttgtcccgcg
tcatcggcgggggtcataacgtgactcccttaattctcatgtatgattgttctcaggtacctcagccagaTAGTGTAAGGGAAAA**CAGGGAGTG**
GATTGCAAGGTGGGGAATCGAATCCTCGAGAAAGTTTAGATAGTCAATTAATTGAGCTACTGAGATTCCTCGGATTTTTTAAAAAAAATAAATT
GGGGATAGAGGAAGAGGAAATGGGGGAGACTTTTATAAGGTTATACCGTTGTTAACAAGGTGAAATTTTAGGTAGTTAATCAATTGAGCTACTA
AGATTCTTTAGATACGATTTTCGTAAGCTCAAATTGGAAAATAAAAAATTAGATTATTACTAAGGAGCTGAAGAAATTCAGAAATGAGTTACCT
TTTGAGCCAGCAGAACCAGTTAAGTTGAGATTTCATTGATTTTAGCTTGTTTTTAATGTTGGTTGTGAATTATGGGGTTTTATTATAATGTTCC
AACAATTTTGTTGGTGTTGTAGGACTTTTTTGATTTAGCGTTTTGAATGTGTTCTGATTATCTGCCTTCACATAGGTG[TAGTGTAAGGGAAAA
CAGGGAGTGGATTGCAAGG**TGG**GGAATCGAATCCTCGAGAAAGTTTAGATAGTCAATTAATTGAGCTACTGAGATTCCTCGGATTTTTTAAAAA
AAATAAATTGGGGATAGAGGAAGAGGAAATGGGGGAGACTTTTATAAGGTTATACCGTTGTTAACAAGGTGAAATTTTAGGTAGTTAATCAATT
GAGCTACTAAGATTCTTTAGATACGATTTTCGTAAGCTCAAATTGGAAAATAAAAAATTAGATTATTACTAAGGAGCTGAAGAAATTCAGAAAT
GAGTTACCTTTTGAGCCAGCAGAACCAGTTAAGTTGAGATTTCATTGATTTTAGCTTGTTTTTAATGTTGGTTGTGAATTATGGGGTTTTATTA
TAATGTTCCAACAATTTTGTTGGTGTTGTAGGACTTTTTTGATTTAGCGTTTTGAATGTGTTCTGATTATCTGCCTTCACATAGGTGATATTGA
GGTAAAAATGGCATCTTGATTGATTTTGCTCTCATAATTGTTAGATTTTCTTCTTGTACATTTTGCAGTTTTTAGTTAACTTGCCAAGGTGTTT
AGTGGATGCACAATATGCGGTATTACTACGGAATGTCTGAATTGGTTTGAGCTTTAGTTTACCCAAGCTGTATTAGTAATTGCACATTGCTTTG
TCCTTTTTTAGTTAGTACTTTCCTTAAAACTGGAATTTTGATCTCTTGGTTACTTGCCAATTTAAATTTTAGTTCATGGATCGGATTCAATGAA
TAGATATATGCTCTTTGGTGGTGGGAGGACAAGGATAGATAAATTAAACTCGGTAAGGAGAGAAAGTGCAGAATGAATGGAAGTTGAGACATTT
TGGAGTATGGGGTGAAGAATGACCTGTGAAGTTTTAGGAAAGAAAAAGTTAACCTTGTTGAATAGTCATATTTCTCTTCCTTGCACAATCTTGT
GTAGGAATGCCTTTAATTGGGTTTATATCGTCGGATAGTTTATATTGCTTCAAATCTAAGAAACTCTTTCATATTTGTCAAATGCTCCTATTTT
CAACAAAGGATATTTTGAGCTGAAAAGGTAGTTGAATAATGCATTGGTCGTAGTTTTCTCTCCTAAGCAAGATAATATAGTTACTATTTTCAAG
TACAACTACTTCTTTTTCTTTAGTTTACGTATAAGAGCAGGTC

SleIF4E1 gene is in uppercase. In the pink bracket is the inserted donor sequence including 5′-UTR shaded in light yellow, exon1 in green, intron 1 in light gray, and kanamycin resistance marker colored in purple and lowercase. The P69T codon is highlighted in red. Outside of the bracket is the endogenous *SleIF4E1* sequence (native exon1 shaded in blue, 5′-UTR and intron1 not shaded).

The TALEN target region with left (red) and right (blue) binding domains are underlined. The CRISPR-gRNA targets for the deletion following the NHEJ are bolded and PAM sites are highlighted in yellow.

Data S2 *SleIF4E*1 genomic sequence and TALEN® target sequences pCLS24222 and 24224

AGTGCTCCACAGTCCACAGAGCAGCAAAAATGGCAGCAGCTGAAATGGAGAGAACGATGTCGTTTGATGCAGCTGAGAAGTTGAAGGCCGCCGA
TGGAGGAGGAGGAGAGGTAGACGATGAACTTGAAGAAGGTGAAATTGTTGAAGAATCAAATGATACGGCATCGTATTTAGGGAAAGAAATCACA
GTGAAGCATCCATTGGAGCATTCATGGACTTTTTGGTTTGATAACCCTACCACTAAATCTCGACAAACTGCTTGGGGAAGCTCACTTCGAAATG
TCTACACTTTCTCCACTGTTGAAGATTTTTGGGGGTAAGTTTTTTTTTTTCCTCTTTCAAATTGGTGATAGTGTAGTGTAAGGGAAAACAGGGA
GTGGATTGCAAGGTGGGGAATCGAATCCTCGAGAAAGTTTAGATAGTCAATTAATTGAGCTACTGAGATTCCTCGGATTTTTTAAAAAAAATAA
ATTGGGGATAGAGGAAGAGGAAATGGGGGAGACTTTTATAAGGTTATACCGTTGTTAACAAGGTGAAATTTTAGGTAGTTAATCAATTGAGCTA
CTAAGATTCTTTAGATACGATTTTCGTAAGCTCAAATTGGAAAATAAAAAATTAGATTATTACTAAGGAGCTGAAGAAATTCAGAAATGAGTTA
CCTTTTGAGCCAGCAGAACCAGTTAAGTTGAGATTTCATTGATTTTAGCTTGTTTTTAATGTTGGTTGTGAATTATGGGGTTTTATTATAATGT
TCCAACAATTTTGTTGGTGTTGTAGGACTTTTTTGATTTAGCGTTTTGAATGTGTTCTGATTATCTGCCTTCACATAGGTGATATTGAGGTAAA
AATGGCATCTTGATTGATTTTGCTCTCATAATTGTTAGATTTTCTTCTTGTACATTTTGCAGTTTTTAGTTAACTTGCCAAGGTGTTTAGTGGA
TGCACAATATGCGGTATTACTACGGAATGTCTGAATTGGTTTGAGCTTTAGTTTACCCAAGCTGTATTAGTAATTGCACATTGCTTTGTCCTTT
TTTAGTTAGTACTTTCCTTAAAACTGGAATTTTGATCTCTTGGTTACTTGCCAATTTAAATTTTAGTTCATGGATCGGATTCAATGAATAGATA
TATGCTCTTTGGTGGTGGGAGGACAAGGATAGATAAATTAAACTCGGTAAGGAGAGAAAGTGCAGAATGAATGGAAGTTGAGACATTTTGGAGT
ATGGGGTGAAGAATGACCTGTGAAGTTTTAGGAAAGAAAAAGTTAACCTTGTTGAATAGTCATATTTCTCTTCCTTGCACAATCTTGTGTAGGA
ATGCCTTTAATTGGGTTTATATCGTCGGATAGTTTATATTGCTTCAAATCTAAGAAACTCTTTCATATTTGTCAAATGCTCCTATTTTCAACAA
AGGATATTTTGAGCTGAAAAGGTAGTTGAATAATGCATTGGTCGTAGTTTTCTCTCCTAAGCAAGATAATATAGTTACTATTTTCAAGTACAAC
TACTTCTTTTTCTTTAGTTTACGTATAAGAGCAGGTCCTGTATCTACATAGCTGTGATATCTCTATGTGCATTTGTCTTAATATACTATTCACT
TGTCTTAATATATTATTCACTTTCAAAAATAAAGACAGATGTGTTTCTATTTCTTAATTAGGCTAAAGTATGCTCATCTCCTCTGTAGATTGAT
GAGACTTAGCTGCAGGCCAGTTTTTTCAATCTCACGCATAGACTTTTTTATTAAGTCCTATGATTTGAGTAATGTAGGTGGTAATGGTGTGAAT
TTGATGTCTGCCATGGATGCATGCACCTTGTTTGGTGAGTTCTTTATTGGTCTAATTACTCAAGGCACTTGAGTTATTGTACAACTTGAACGAT
GACATACCTGTTTGATATTCCACATCATGGATTAGGTCTTTGAAATGCTATTATCCTTTTGGCTCATGATGAAATCTTGAACCATGTTGCTTAT
TCTACAAACAGTGCTTACAATAATATCCATCACCCAAGCAAGTTAATTATGGGAGCAGACTTTCATTGTTTTAAGCACAAAATTGAGCCAAAGT
GGGAAGATCCTGTATGTGCCAATGGAGGGACGTGGAAAATGAGTTTTTCGAAGGGTAAATCTGATACCAGCTGGCTGTATACGGTATTCCGAAG
TTATTTCCATCCAGCCCTTAATGATAGGTCATTCTAGTAATGTTATTTTCCCCTTTGATATAATTTCCACTCTTGTTTTCTTATATGGAATTAT
TGTAGCTGCTGGCAATGATTGGACATCAATTCGATCATGGAGATGAAATTTGTGGAGCAGTTGTTAGTGTCCGGGCTAAGGGAGAAAAAATAGC
TTTGTGGACCAAGAATGCTGCAAATGAAACAGCTCAGGTAATTTGCTTTATATTTTTGGTGTCAACGGCCACGTTTGTCATTTTCTTGGATCAA
GCGGACAGACATTTTTTGTTGTGTACATACTGTAGTGCTGATGTTTATTCAAAATATACTATGAAATACATTGTACCGTTGAGTATATTGGAAG
TTAAGAGGGGAAGTCATCAGATATATAGGAAGTGCTCCTTGTGAGAAATTTCTCTAGTAGGGCTCAAACCCGAGAACCATGGTTAAGGACGGAG
GGATCCCTTTGGTGGTATCAAGGGTTAATTATATTATTTCATGTTTCATATGATTGAAACATTTGGATATCCTTGCAAAGTGTGATATTGTTTT
GTCTCGTCATTTGTTTTATGCTTCATAAGAATTATCTTTTCTTTCGTCCCCTTCTCTTTGTTTCTTTTTCTCCAGTGTACTTCATTCTTCTAGA
ATTATATAGTGGTCGGCGTTCCCTTTTGTGCTTCATTGATAATATTGAATCCTTTCAGCTGTTTACCGGTTGATTATTTTTATGTTGGTGCTTA
TTGCTAGTTTTTTGTTTTTATATCTTTGCAACTATACCTACCATCTTCCTGAAACTTCTCTTTACAGGTTAGCATTGGTAAGCAATGGAAGCAG
TTTCTAGATTACAGTGATTCGGTTGGCTTCATATTTCACGTATGAAATCTTGGTTATCATACGCCTTTAATTCAGTTTCTCTTCAATTAGCAAG
ACTCATAAAGAATCATCTTCTTTTGCAGGACGATGCAAAGAGGCTCGACAGAAATGCCAAGAATCGTTACACCGTATAGTTCTTGATGCAGTGT
GGGATTGCAAGAAACACAATTCGTACTGGAAAGGTTGGTAGGTACTAGTTTAGTTTCTCATTTGATAAGCTTCTGGTTTGAGTAACTCGTGTGT
TGGTGTTTACACTTTCTAATCGTGGAAAATTGTTTGATTTGAATCCATGCCTCTATGTTTCGTGTGTTCTGTGTTACTTTCCCAGCTTCCCTAT
TATTTCTCTGCAGTAAAAATCTCTGAAAATATTGTTGATGGAGGTGAAAGTGATACCAAGCACATAACAAAACACAAAT

The TAL effector target region is underlined. And the left and the right TAL effector DNA-binding domains are in red and blue, respectively. The exons are shaded in blue.

Data S3 The left and the right TALEN® cassette sequences in pCLS24222 and pCLS24224, respectively

(The left TALEN® cassette)

aatcccacaaaaatctgagcttaacagcacagttgctcctctcagagcagaatcgggtattcaacaccctcatatcaactactacgttgtgtat
aacggtccacatgccggtatatacgatgactggggttgtacaaaggcggcaacaaacggcgttcccggagttgcacacaagaaatttgccacta
ttacagaggcaagagcagcagctgacgcgtacacaacaagtcagcaaacagacaggttgaacttcatcccccaaaggagaagctcaactcaagcc
caagagctttgctaaggccctaacaagcccaccaaagcaaaaagcccactggctcacgctaggaaccaaaaggcccagcagtgatccagccccca
aaagagatctcctttgccccggagattacaatggacgatttcctctatctttacgatctaggaaggaagttcgaaggtgaaggtgacgacacta
tgttcaccactgataatgagaaggttagcctcttcaatttcagaaagaatgctgacccacagatggttagagaggcctacgcagcaggtctcat
caagacgatctacccgagtaacaatctccaggagatcaaataccttcccaagaaggttaaagatgcagtcaaaagattcaggactaattgcatc
aagaacacagagaaagacatatttctcaagatcagaagtactattccagtatggacgattcaaggcttgcttcataaaccaaggcaagtaatag
agattggagtctctaaaaaggtagttcctactgaatctaaggccatgcatggagtctaagattcaaatcgaggatctaacagaactcgccgtga
agactggcgaacagttcatacagagtcttttacgactcaatgacaagaagaaaatcttcgtcaacatggtggagcacgacactctggtctactc
caaaaatgtcaaagatacagtctcagaagaccaaagggctattgagacttttcaacaaaggataatttcgggaaacctcctcggattccattgc
ccagctatctgtcacttcatcgaaaggacagtagaaaaggaaggtggctcctacaaatgccatcattgcgataaaggaaaggctatcattcaag
atctctctgccgacagtggtcccaaagatggacccccacccacgaggagcatcgtggaaaaagaagacgttccaaccacgtcttcaaagcaagt
ggattgatgtgacatctccactgacgtaagggatgacgcacaatcccactatccttcgcaagacccttcctctatataaggaagttcatttcat
ttggagaggacacgcacgagatctgcaggtcgactctagaggcgcgccacacgcaaacacaaatacacagcggccttgccacc**atgggc**gatcc
taaaaagaaacgtaaggtcatcgattacccatacgatgttccagattacgctatcgatatcgccgatctacgcacgctcggctacagccagcag
caacaggagaagatcaaaccgaaggttcgttcgacagtggcgcagcaccacgaggcactggtcggccacgggtttacacacgcgcacatcgttg
cgttaagccaacacccggcagcgttagggaccgtcgctgtcaagtatcaggacatgatcgcagcgttgccagaggcgacacacgaagcgatcgt
tggcgtcggcaaacagtggtccggcgcacgcgctctggaggccttgctcacggtggcgggagagttgagaggtccaccgttacagttggacaca
ggccaacttctcaagattgcaaaacgtggcggcgtgaccgcagtggaggcagtgcatgcatggcgcaatgcactgacgggtgccccgctcaact
tgaccccccagcaggtggtggccatcgccagcaatggcggtggcaagcaggcgctggagacggtccagcggctgttgccggtgctgtgccaggc
ccacggcttgacccccagcaggtggtggccatcgccagcaatggcggtggcaagcaggcgctggagacggtccagcggctgttgccggtgctg
tgccaggcccacggcttgaccccggagcaggtggtggccatcgccagccacgatggcggcaagcaggcgctggagacggtccagcggctgttgc
cggtgctgtgccaggcccacggcttgaccccggagcaggtggtggccatcgccagccacgatggcggcaagcaggcgctggagacggtccagcg
gctgttgccggtgctgtgccaggcccacggcttgacccccagcaggtggtggccatcgccagcaatggcggtggcaagcaggcgctggagacg
gtccagcggctgttgccggtgctgtgccaggcccacggcttgaccccggagcaggtggtggccatcgccagccacgatggcggcaagcaggcgc
tggagacggtccagcggctgttgccggtgctgtgccaggcccacggcttgacccccagcaggtggtggccatcgccagcaatggcggtggcaa
gcaggcgctggagacggtccagcggctgttgccggtgctgtgccaggcccacggcttgacccccagcaggtggtggccatcgccagcaatggc
ggtggcaagcaggcgctggagacggtccagcggctgttgccggtgctgtgccaggcccacggcttgacccccagcaggtggtggccatcgcca
gcaatggcggtggcaagcaggcgctggagacggtccagcggctgttgccggtgctgtgccaggcccacggcttgaccccggagcaggtggtggc
catcgccagccacgatggcggcaagcaggcgctggagacggtccagcggctgttgccggtgctgtgccaggcccacggcttgaccccggagcag
gtggtggccatcgccagcaatattggtggcaagcaggcgctggagacggtgcaggcgctgttgccggtgctgtgccaggcccacggcttgaccc
cggagcaggtggtggccatcgccagcaatattggtggcaagcaggcgctggagacggtgcaggcgctgttgccggtgctgtgccaggcccacgg
cttgaccccggagcaggtggtggccatcgccagcaatattggtggcaagcaggcgctggagacggtgcaggcgctgttgccggtgctgtgccag
gcccacggcttgaccccccagcaggtggtggccatcgccagcaatggcggtggcaagcaggcgctggagacggtccagcggctgttgccggtgc
tgtgccaggcccacggcttgaccccccagcaggtggtggccatcgccagcaatggcggtggcaagcaggcgctggagacggtccagcggctgtt
gccggtgctgtgccaggcccacggcttgacccctcagcaggtggtggccatcgccagcaatggcggcggcaggccggcgctggagagcattgtt
gcccagttatctcgccctgatccggcgttggccgcgttgaccaacgaccacctcgtcgccttggcctgcctcggcgggcgtcctgcgctggatg
cagtgaaaaagggattgggggatcctatcagccgttcc**cagctggtgaagtccgagctggaggagaagaaatccgagttgaggcacaagctgaa**
gtacgtgccccacgagtacatcgagctgatcgagatcgcccggaacagcacccaggaccgtatcctggagatgaaggtgatggagttcttcatg
aaggtgtacggctacaggggcaagcacctgggcggctccaggaagcccgacggcgccatctacaccgtgggctcccccatcgactacggcgtga
tcgtggacaccaaggcctactccggcggctacaacctgcccatcggccaggccgacgaaatgcagaggtacgtggaggagaaccagaccaggaa
caagcacatcaaccccaacgagtggtggaaggtgtacccctccagcgtgaccgagttcaagttcctgttcgtgtccggccacttcaagggcaac
tacaaggcccagctgaccaggctgaaccacatcaccaactgcaacggcgccgtgctgtccgtggaggagctcctgatcggcggcgagatgatca
aggccggcaccctgaccctggaggaggtgaggaggaagttcaacaacggcgagatcaacttcgcggccgactgataagtcgaccaccaccacca
ccaccactgataagagctcctcgagcgccggcgagcctcgaatttccc**cgatcgttcaaacatttggcaataaagtttcttaagattgaatcct**
gttgccggtcttgcgatgattatcatataatttctgttgaattacgttaagcatgtaataattaacatgtaatgcatgacgttatttatgagat
gggtttttatgattagagtcccgcaattatacatttaatacgcgatagaaaacaaaatatagcgcgcaaactaggataaattatcgcgcgcggt
gtcatctatgttactagatcgggaattcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacg
agccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgc

(The right TALEN® cassette)

aatcccacaaaaatctgagcttaacagcacagttgctcctctcagagcagaatcgggtattcaacaccctcatatcaactactacgttgtgtat
aacggtccacatgccggtatatacgatgactggggttgtacaaaggcggcaacaaacggcgttcccggagttgcacacaagaaatttgccacta
ttacagaggcaagagcagcagctgacgcgtacacaacaagtcagcaaacagacaggttgaacttcatccccaaaggagaagctcaactcaagcc
caagagctttgctaaggccctaacaagcccaccaaagcaaaaagcccactggctcacgctaggaaccaaaaggcccagcagtgatccagcccca
aaagagatctcctttgccccggagattacaatggacgatttcctctatctttacgatctaggaaggaagttcgaaggtgaaggtgacgacacta
tgttcaccactgataatgagaaggttagcctcttcaatttcagaaagaatgctgacccacagatggttagagaggcctacgcagcaggtctcat
caagacgatctacccgagtaacaatctccaggagatcaaataccttcccaagaaggttaaagatgcagtcaaaagattcaggactaattgcatc
aagaacacagagaaagacatatttctcaagatcagaagtactattccagtatggacgattcaaggcttgcttcataaaccaaggcaagtaatag
agattggagtctctaaaaaggtagttcctactgaatctaaggccatgcatggagtctaagattcaaatcgaggatctaacagaactcgccgtga
agactggcgaacagttcatacagagtcttttacgactcaatgacaagaagaaaatcttcgtcaacatggtggagcacgacactctggtctactc
caaaaatgtcaaagatacagtctcagaagaccaaagggctattgagacttttcaacaaaggataatttcgggaaacctcctcggattccattgc
ccagctatctgtcacttcatcgaaaggacagtagaaaaggaaggtggctcctacaaatgccatcattgcgataaaggaaaggctatcattcaag
atctctctgccgacagtggtcccaaagatggacccccacccacgaggagcatcgtggaaaaagaagacgttccaaccacgtcttcaaagcaagt
ggattgatgtgacatct<u>ccactgacgtaagggatgacgcacaatcccactatccttcgcaagacccttcctctatataaggaagttcatttcat</u>
<u>ttggagaggacacgcacgagatctgcaggtcgactctagaggcgcgccacacgcaaacacaaatacacagcggccttgccacc</u>atgggcgatcc
taaaaagaaacgtaaggtcatcgataaggagaccgccgctgccaagttcgagagacagcacatggacagcatcgatatcgccgatctacgcacg
ctcggctacagccagcagcaacaggagaagatcaaaccgaaggttcgttcgacagtggcgcagcaccacgaggcactggtcggccacgggttta
cacacgcgcacatcgttgcgttaagccaacacccggcagcgttagggaccgtcgctgtcaagtatcaggacatgatcgcagcgttgccagaggc
gacacacgaagcgatcgttggcgtcggcaaacagtggtccggcgcacgcgctctggaggccttgctcacggtggcgggagagttgagaggtcca
ccgttacagttggacacaggccaacttctcaagattgcaaaacgtggcggcgtgaccgcagtggaggcagtgcatgcatggcgcaatgcactga
cgggtgccccgctcaacttgaccccggagcaggtggtggccatcgccagccacgatggcggcaagcaggcgctggagacggtccagcggctgtt
gccggtgctgtgccaggcccacggcttgaccccggagcaggtggtggccatcgccagccacgatggcggcaagcaggcgctggagacggtccag
cggctgttgccggtgctgtgccaggcccacggcttgaccccggagcaggtggtggccatcgccagccacgatggcggcaagcaggcgctggaga
cggtccagcggctgttgccggtgctgtgccaggcccacggcttgaccccccagcaggtggtggccatcgccagcaatggcggtggcaagcaggc
gctggagacggtccagcggctgttgccggtgctgtgccaggcccacggcttgaccccccagcaggtggtggccatcgccagcaataatggtggc
aagcaggcgctggagacggtccagcggctgttgccggtgctgtgccaggcccacggcttgaccccccagcaggtggtggccatcgccagcaatg
gcggtggcaagcaggcgctggagacggtccagcggctgttgccggtgctgtgccaggcccacggcttgaccccccagcaggtggtggccatcgc
cagcaatggcggtggcaagcaggcgctggagacggtccagcggctgttgccggtgctgtgccaggcccacggcttgaccccccagcaggtggtg
gccatcgccagcaatggcggtggcaagcaggcgctggagacggtccagcggctgttgccggtgctgtgccaggcccacggcttgaccccccagc
aggtggtggccatcgccagcaatggcggtggcaagcaggcgctggagacggtccagcggctgttgccggtgctgtgccaggcccacggcttgac
cccggagcaggtggtggccatcgccagccacgatggcggcaagcaggcgctggagacggtccagcggctgttgccggtgctgtgccaggcccac
ggcttgaccccggagcaggtggtggccatcgccagccacgatggcggcaagcaggcgctggagacggtccagcggctgttgccggtgctgtgcc
aggcccacggcttgaccccggagcaggtggtggccatcgccagccacgatggcggcaagcaggcgctggagacggtccagcggctgttgccggt
gctgtgccaggcccacggcttgaccccccagcaggtggtggccatcgccagcaatggcggtggcaagcaggcgctggagacggtccagcggctg
ttgccggtgctgtgccaggcccacggcttgaccccccagcaggtggtggccatcgccagcaatggcggtggcaagcaggcgctggagacggtcc
agcggctgttgccggtgctgtgccaggcccacggcttgaccccggagcaggtggtggccatcgccagcaatattggtggcaagcaggcgctgga
gacggtgcaggcgctgttgccggtgctgtgccaggcccacggcttgacccctcagcaggtggtggccatcgccagcaatggcggcggcaggccg
gcgctggagagcattgttgcccagttatctcgccctgatccggcgttggccgcgttgaccaacgaccacctcgtcgccttggcctgcctcggcg
ggcgtcctgcgctggatgcagtgaaaaagggattgggggatcctatcagccgttccagctggtgaagtccgagctggaggagaagaaatccga
gttgaggcacaagctgaagtacgtgccccacgagtacatcgagctgatcgagatcgcccggaacagcacccaggaccgtatcctggagatgaag
gtgatggagttcttcatgaaggtgtacggctacaggggcaagcacctgggcggctccaggaagcccgacggcgccatctacaccgtgggctccc
ccatcgactacggcgtgatcgtggacaccaaggcctactccggcggctacaacctgcccatcggccaggccgacgaaatgcagaggtacgtgga
ggagaaccagaccaggaacaagcacatcaaccccaacgagtggtggaaggtgtacccctccagcgtgaccgagttcaagttcctgttcgtgtcc
ggccacttcaagggcaactacaaggcccagctgaccaggctgaaccacatcaccaactgcaacggcgccgtgctgtccgtggaggagctcctga
tcggcggcgagatgatcaaggccggcaccctgaccctggaggaggtgaggaggaagttcaacaacggcgagatcaacttcgcggccgactgata
agtcgaccaccaccaccaccaccactgataagagctcctcgagcgccggcgagcctcgaatttcccgatcgttcaaacatttggcaataaagt
ttcttaagattgaatcctgttgccggtcttgcgatgattatcatataatttctgttgaattacgttaagcatgtaataattaacatgtaatgca
tgacgttatttatgagatgggtttttatgattagagtcccgcaattatacatttaatacgcgatagaaaacaaaatatagcgcgcaaactagga
taaattatcgcgcgcggtgtcatctatgttactagatcgggaattcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcac
aattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactg
cccgctttccagtcgggaaacctgtcgtgccagctgcattaat

Codons corresponding to the repeat Variable Diresidue (RVD), specific to A (NI) are highlighted in blue, specific to T (NG) in red, specific to G (NN) in yellow, and specific to C (HD) in green. Nucleotide sequence recognized by pCLS24222 is TTTCCTCTTTCAAATT**T** and by pCLS24224 is TCCCTGTTTTCCCTTA**T** (its reverse complement: **A**TAAGGGAAAACAGGGA. The target sequence is preceded by a **T** required for TALEN activity (Moscou and Bogdanove, 2009; Cermak et al., 2011). The pCaMV35S is in light blue, the TAL effector repeat array in black, the nonspecific DNA cleavage domain of the FokI endonuclease in orange. Nos terminator in green. The SV40 NLS is underlined.

Data S4 DNA donor template with P69T mutation

aaataatgattttattttgactgatagtgacctgttcgttgcaacacattgatgagcaatgctttttataatgccaactttgtacaaaaaagc
agTTTTGAAAAATAAAGCTTCCATATTCTGAACTAGATTAAAATTAAAATCCACGTAAATTAATTATTTGATTCTCACGATACCTAATTTAGGT
ATTTACAACATTAGTACTTCTACAATTTAAGTATTACACTATGGTCCAAACAGTTCTTATAAATACCATTTTTCTTAATTAAGTAAAAAAAAAG
TGCTCCACAGTCCACAGAGCAGCAAAAATGGCAGCAGCTGAAATGGAGAGAACGATGTCGTTTGATGCAGCTGAGAAGTTGAAGGCCGCCGATG
GAGGAGGAGGAGAGGTAGACGATGAACTTGAAGAAGGTGAAATTGTTGAAGAATCAAATGATACGGCATCGTATTTAGGGAAAGAAATCACAGT
GAAGCATCCATTGGAGCATTCATGGACTTTTTGGTTTGATAACACAACCACTAAATCTCGACAAACTGCTTGGGGAAGCTCACTTCGAAATGTC
TACACTTTCTCCACTGTTGAAGATTTTTGGGGGTAAGTTTTTTTTTTTCCTCTTTCAAATTGGTGATAGTGTAGTGTAAGGGAAAAttgttctc
aggtacctcagccagacagcttgcatgccggtcgatctagtaacatagatgacaccgcgcgcgataatttatcctagtttgcgcgctatatttt
gttttctatcgcgtattaaatgtataattgcgggactctaatcataaaaacccatctcataaataacgtcatgcattacatgttaattattaca
tgcttaacgtaattcaacagaaattatatgataatcatcgcaagaccggcaacaggattcaatcttaagaaactttattgccaaatgtttgaac
gatctgcttgactctagctagagtccgaaccccagagtcccgctcagaagaactcgtcaagaaggcgatagaaggcgatgcgctgcgaatcggg
agcggcgataccgtaaagcacgaggaagcggtcagcccattcgccgccaagctcttcagcaatatcacgggtagccaacgctatgtcctgatag
cggtccgccacacccagccggccacagtcgatgaatccagaaaagcggccattttccaccatgatattcggcaagcaggcatcgccgtgggtca
cgacgagatcctcgccgtcgggcatccgcgccttgagcctggcgaacagttcggctggcgcgagcccctgatgctcttcgtccagatcatcctg
atcgacaagaccggcttccatccgagtacgtgctcgctcgatgcgatgtttcgcttggtggtcgaatgggcaggtagccggatcaagcgtatgc
agccgccgcattgcatcagccatgatggatactttctcggcaggagcaaggtgagatgacaggagatcctgccccggcacttcgcccaatagca
gccagtcccttcccgcttcagtgacaacgtcgagcacagctgcgcaaggaacgcccgtcgtggccagccacgatagccgcgctgcctcgtcttg
gagttcattcagggcaccggacaggtcggtcttgacaaaaagaaccgggcgcccctgcgctgacagccggaacacggcggcatcagagcagccg
attgtctgttgtgcccagtcatagccgaatagcctctccacccaagcggccggagaacctgcgtgcaatccatcttgttcaatcatgcctcgat
cgagttgagagtgaatatgagactctaattggataccgaggggaatttatggaacgtcagtggagcatttttgacaagaaatatttgctagctg
atagtgaccttaggcgacttttgaacgcgcaataatggtttctgacgtatgtgcttagctcattaaactccagaaacccgcggctgagtggctc
cttcaacgttgcggttctgtcagttccaaacgtaaaacggcttgtcccgcgtcatcggcgggggtcataacgtgactcccttaattctcatgta
tgattgttctcaggtacctcagccagaTAGTGTAAGGGAAAACAGGGAGTGGATTGCAAGGTGGGGAATCGAATCCTCGAGAAAGTTTAGATAG
TCAATTAATTGAGCTACTGAGATTCCTCGGATTTTTTAAAAAAAATAAATTGGGGATAGAGGAAGAGGAAATGGGGGAGACTTTTATAAGGTTA
TACCGTTGTTAACAAGGTGAAATTTTAGGTAGTTAATCAATTGAGCTACTAAGATTCTTTAGATACGATTTTCGTAAGCTCAAATTGGAAAATA
AAAAATTAGATTATTACTAAGGAGCTGAAGAAATTCAGAAATGAGTTACCTTTTGAGCCAGCAGAACCAGTTAAGTTGAGATTTCATTGATTTT
AGCTTGTTTTTAATGTTGGTTGTGAATTATGGGGTTTTATTATAATGTTCCAACAATTTTGTTGGTGTTGTAGGACTTTTTTGATTTAGCGTTT
TGAATGTGTTCTGATTATCTGCCTTCACATAGGTGgtttaaacagctttcttgtacaaagttggcattataagaaagcattgcttatcaattt
gttgcaacgaacaggtcactatcagtcaaaataaaatcattattt

eIF4E1 gene is in uppercase. 5′-UTR is shaded in light yellow, exon1 in green, intron 1 in light gray. PacI and PmeI restriction sites are highlighted in pink and dark green, respectively. The P69T codon is highlighted in red.
The kanamycin resistance marker that consists of the *Nos* promoter, the *NptII* gene, and the *Nos* terminator is in purple. Two RAG1 target sequences are underlined (not used in this work).

Data S5 The single guide RNA cassette sequence for targeting *SleIF4E1*

GGGGACAAGTTTGTACAAAAAAGCAGGCTTCACTAGTGTTCAGTTGCATTATGTCTTTATACACCCCTACCTAGATGATTAAGTTTTACTTTAG
TTGGTGTGAAATGGATAAATTCTAAAATATGAGGTGTGGAATGAAGGATTGTCATCAATTAGTTGGCCCCAACCAAGTAAAATAAGAAGGCCGG
CCCATTACAATTAAGTCGTCACACAAGTGGGCTTCATTGAAACAAGCGCAAAAAGGAGTCCAGGCCCGTGTTAGCGTGAAGACTCAACCAGCGA
TTTCTCCCTCATCGGTTGCACAGAAAAGCTGTGTGTTGTTTATATGGCGAAACCTAACAGTCTGACTTGCTCCCATAATTAACTTGCTTGTTTA
AGAGCTATGCTGGAAACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTGTTTTT
TATGTCTAGCGGAGGTAGTGAGCTCGGCTCTAGCGGGCCATTTATATGGGAGAGTGCCACCACGCGTGAAACTCCTGGCCCTGAGGACTAGTAC
CCAGCTTTCTTGTACAAAGTGGTCCCC

Sequence of the guide cassette cloned into the pDicAID_nCas9-PmCDA1_NptII-SlALS1 through *SpeI* restriction/ligation, resulting into the CBE-4E1-RII. Gateway AttB1/B2 sequences are colored in blue, *SpeI* restriction sites in red, *StU6* promoter sequence in yellow, 20-bp target sequence in green, sgRNA scaffold sequence in purple.

Supporting Tables

Table S1: Number of the individuals used and obtained for the Region I codon insertion.

Initial leaves used in biolistic	Leaf pieces cut out	Buds that regenerated on the selective media	NHEJ (both 5′ and 3′ ends) event	HR event
40	233	32	1	0

Table S2: Synthesis of all nucleases and target sites used.

Guide name	Target	Detection
sgRNA-DKM04	TGTAAGGGAAAATTGTTCTC**AGG**	Pair 1; NHEJ 3′-end deletion
sgRNA-eIF4E05	CAGGGAGTGGATTGCAAGG**TGG**	
sgRNA-Cl-01	**CCA**CAGTCCACAGAGCAGCAAAAAT	Pair 2; NHEJ 5′-end deletion
sgRNA-Cl-02	AGTCCACAGAGCAGCAAAAA**TGG**	

Table S3: List of primers used in this study.

Primer	Sequence 5′-3′	Amplicon size (bp)	Detection
ID1F	TGCGAAAACTGAACTTGTGC	1275 if HR, 1631 if NHEJ	DNA donor template insertion in *eIF4E* gene 5′ end
ID1R	CAAAATATAGCGCGCAAACTAGG		
ID1F	TGCGAAAACTGAACTTGTGC	1702 if HR, 2058 if NHEJ	DNA donor template insertion in *eIF4E* gene 3′ end
ID2R	ATGGCCGCTTTTCTGGATTC		
ID3F	GAACCTGCGTGCAATCCATC	816 if HR, 816 and 1294 if NHEJ	DNA donor template insertion in *eIF4E* gene 3′ end
ID3R	AGGCAGATAATCAGAACACATTCA		
Fok1F	AATCCGAGTTGAGGCACAAG	445	TALENs Fok1 domain
Fok1R	TTGCAGTTGGTGATGTGGTT		
Cas9F	TCCCTTACTACGTGGGACCTC	1438	*Cas9* gene
Cas9R	ATCTGCCTGGTTTCCACAAG		
DD1F	AACGATGTCGTTTGATGCAG	2267; 788 if deleted	*NptII* gene deletion (NHEJ 3′-end deletion)
DD2R	TCAAGATGCCATTTTTACCTCA		
DD3F	AACTGTTCATGTGACATTATTGGT	1340; 940 if deleted	NHEJ 5′-end deletion
DD4R	TCAAGATGCCATTTTTACCTCA		
Hyg1F	ATAGGTCAGGCTCTCGCTGA	561	*Hpt* gene (T-DNA)
Hyg1R	ATCATACATGAGAATTAAGGG		
KanaF	AGACAATCGGCTGCTCTGAT	593	*NptII* gene (T-DNA)
KanaR	AGCCAACGCTATGTCCTGAT		
Hrm2F	TGCTTACAATAATATCCATCACCCA	94	HRM on *eIF4E1* region II
Hrm2R	AGGATCTTCCCACTTTGGCTC		
4er1F	GGTCCAAACAGTTCTTATAAATACCA	658	genomic *eIF4E1* Region I
4er1R	TTTTCCAATTTGAGCTTACGAA		
4er2F	CATGCACCTTGTTTGGTGAG	526	genomic *eIF4E1* Region II
4er2R	CCCGGACACTAACAACTGCT		
c4e1F	ATGGCAGCAGCTGAAATGGAGAGAACGATGTCG	696	*SleIF4E1* cDNA
c4e1R	CTATACGGTGTAACGATTCTTGGCATTTCTGTCG		
gapdhF	GGTTACAGTTCCCGTGTGATTGA	190 /195	*GAPDH* genomic/cDNA
gapdhR	TCAAAGATAACAACAGGGTTCCGTT		

Preface to *Chapter 3* and *4*

In *Chapter 2*, we demonstrated successfully in cherry tomato that the precise genome-editing in the *SleIF4E1* gene confers potyvirus resistance of a larger spectrum than the single *SleIF4E*^{KO} genotypes. Part of the genome-editing scheme was achieved through co-base-editing in the *SleIF4E1* and *SlALS* genes in an attempt to facilitate the selection of the edited transformants using the latter as a reporter gene. Then, we were curious to see if this *eIF4E1*-co-base-editing would be a useful strategy in the real agricultural setting where not only the pathogen resistance but also the crop productivity is essential.

When we performed a test round of phenotyping of the co-edited mutants in 2022, we unexpectedly observed inhibited growth in some genotypes and not in the others. This has made us wonder and determine whether this was due to the *SleIF4E1* mutation or *SlALS* mutation.

The following two chapters will provide therefore the preliminary results from the greenhouse assay in 2023 on such investigation, simultaneously aiming to evaluate of the co-base-editing scheme. In *Chapter 3*, we evaluate the practicality of the edited resistance *SleIF4E1* allele, the gene of interest, in agriculture; and in *Chapter 4*, we evaluate the potential impacts of a reporter *SlALS* gene-editing on the plant growth. In order to confirm these preliminary results from 2023, the repetitions of the assays are in progress, performed by an engineer school student Anaë Laurens at present (June 2024). The repetition of the metabolomics assay in *Chapter 4* is also in progress with a larger panel of mutants by Anaë L. and an assistant engineer Marie-Noëlle Corre (INRAE GAFL) at present.

Chapter 3: *eIF4E1* precision editing provides a solution to the trade-off between the plant development and potyvirus resistance albeit with a low durability.

Potential list of authors for future publication:
Kyoka Kuroiwa[1], Anaë Laurens[1], Marianne Mazier[1], and Jean-Luc Gallois[1]

[1]INRAE, GAFL, F-84143 Montfavet, France

Introduction

Resistance by loss-of-susceptibility against potyviruses has been studied and developed in search of application for crop breeding. The *eukaryotic translation initiation factor* (*eIF*) *4E* gene family has been the prominent example of the susceptibility gene whose loss or modification leads to recessive resistance: the natural potyvirus resistance genes are often the variants of this gene family. These natural variants are mostly functional copy of *eIF4E1* with nonsynonymous mutations, with a few exceptions of *eIFiso4E* loss-of-function gene (For review Sanfaçon, 2015). Many of the mutations in those functional resistance *eIF4E* alleles code for amino acid substitutions near two regions—region I and II—surrounding the cap-binding pocket, which is important for the original protein function (Reviewed in Robaglia and Caranta, 2006; Charron et al., 2008). Some crop species like pepper has a rich gene pool of natural resistance *eIF4E* alleles, e.g. *pvr2* genes, within the species or in close relatives, and these alleles have already been integrated in breeding program (Charron et al., 2008). However, many crop species or their wild relatives lack such genetic materials that can be used for conventional crossing, leaving them vulnerable to the potyviruses. Also, because this *eIF4E*-mediated resistance is recessive, it is quite laborious to implement in the elite cultivar of crops like potatoes (*Solanum tuberosum*) that are vegetatively propagated.

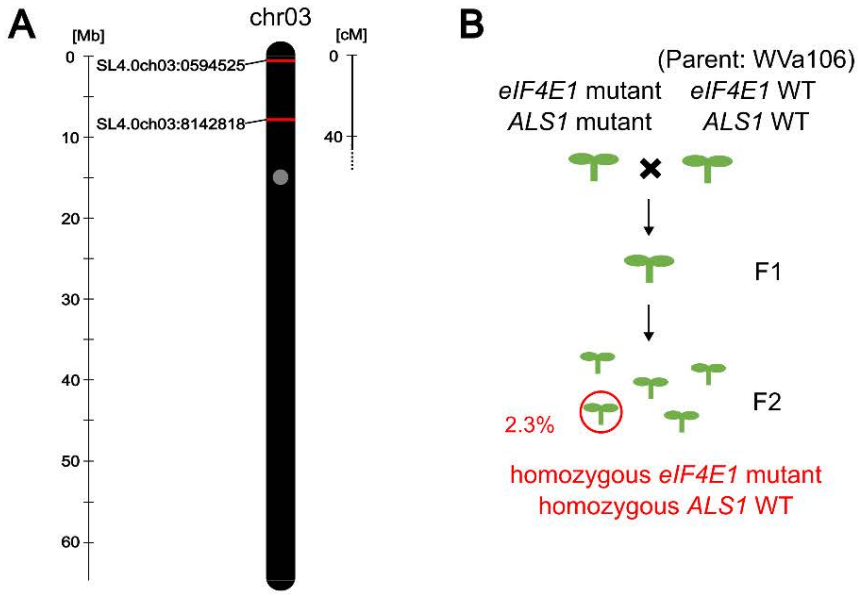
Gene knock-out (KO) of *eIF4E* has been demonstrated in different model and crop species as an alternative strategy to provide resistance. Nonetheless, a single KO genotype proved to be limited in terms of the resistance spectrum as well as the durability, due to the redundant nature of the *eIF4E* gene family members for its function, while a double KO genotype was either not viable or severely affected in its growth (Reviewed in Bastet et al., 2017). In order to bypass these barriers to the implementation of loss-of-susceptibility by KO in plants, mimicking of natural functional resistance *eIF4E* alleles by genome-editing has been proposed (Bastet et al., 2017). As most natural resistance alleles

coded for amino acid substitutions, similar mutations could be introduced by gene knock-in or CRISPR-based base-editing, in an attempt to recreate the resistance. Indeed, a signature amino acid substitution for potyvirus resistance found in pea resistance *eIF4E1* variant could be successfully copied to the endogenous *eIF4E1* in a model species *Arabidopsis thaliana* to confer resistance by cytosine base-editor (CBE) (Bastet et al., 2019).

Based on this Arabidopsis study, previously, we generated mutant *eIF4E1* alleles by gene knock-in using TALEN® technology in region I and cytosine base-editor (CBE) in the region II, in cherry tomato (*Solanum lycopersicum* var. *cerasiforme*) cultivar (cv.) WVa106 as a model crop (*Chapter 2* (Kuroiwa et al., 2023)). The edition of region II by CBE was performed using the co-editing strategy of a selectable marker that facilitates the identification of editing events. We showed that the introduction of several amino acid substitutions, namely P69T in region I and G110E and A111T in the region II, in the susceptibility factor eIF4E1 can confer resistance to multiple potyvirus isolates in cherry tomato. On top of conferring virus resistance, the edited tomato resistance *eIF4E1* allele coding for amino acid substitutions P69T/G110E+A111T were proven functional by the *in vivo* and *in vitro* assays like the natural functional resistance allele.

After successfully generating a *de novo* resistance tomato *eIF4E1* gene by genome-editing, we were interested in studying the possibility to use these edited plants in an agricultural setting. Thus, in this work, we launched a whole-plant phenotyping on a set of tomato genotypes generated partially, if not entirely, by the CBE, possessing the *eIF4E1* alleles coding for substitutions in the region II: G110E+A111T and in the region I and II: P69T/G110E+A111T, along with the wild-type WVa106 plants. We show that the *eIF4E1*-editing in tomatoes do not affect the plant growth and development in a greenhouse setting.

To further assess the practicality of the edited resistance *eIF4E1* genotype P69T/G110E+A111T in agricultural setting, we proceeded with the evaluation of the virus resistance durability. We monitored virus accumulation in plants over a course of nine weeks to observe whether the edited-*eIF4E1* based resistance would be overcome. We show that the P69T/G110E+A111T-coding *SleIF4E1* suppresses the viral accumulation to a significantly lower level than the wild-type *SleIF4E1* but the resistance can be broken down by PVY isolate N605 during the first nine weeks after inoculation when the virus obtains signature mutations of resistance breaking. Altogether, this work



C

Mutant lines	*Sl*eIF4E1	*Sl*ALS1	*Sl*ALS2
WVa106 (WT)	WT	WT	WT
ET-1	G110**E**+A111**T**	WT	WT
ET-2	G110**E**+A111**T**	WT	WT
ET-3	G110**E**+A111**T**	WT	WT
ET-4	G110**E**+A111**T**	WT	Del184-186/WT
ET-5	G110**E**+A111**T**	WT	Del184-186/WT
P69T/ET-1	**P69T**/G110**E**+A111**T**	WT	WT
P69T/ET-2	**P69T**/G110**E**+A111**T**	WT	WT
P69T/ET-3	**P69T**/G110**E**+A111**T**	WT	WT
P69T/ET-4	**P69T**/G110**E**+A111**T**	WT	WT
P69T/ET-5	**P69T**/G110**E**+A111**T**	WT	Q182E/WT

Figure 1 The generation of *eIF4E1*-edited population from the co-edited population.

A Locations of the two genes *eIF4E1* (SL4.0ch3:0594525) and *ALS1* (SL4.0ch3:8142818) on the physical map of the chromosome 3 of tomato (left scale in megabase pair (Mb)) and the genetic distance to represent the linkage of the two genes (right scale in centimorgan (cM)) which were co-edited in Kuroiwa et al. (2023). The figure was created based on the tomato reference genome SL4.0 and the ITAG4.0 annotation. **B** The schematic of how the *eIF4E1*-edited, *ALS1*-edition-free population was obtained. The co-edited population "*eIF4E1* mutant *ALS1* mutant" was backcrossed with the wild-type parent WVa106 to obtain the F1 seeds. F1 population were selfed to obtain the F2 population in which ratio of plants homozygous for both *eIF4E1* mutant allele and *ALS1* wild-type allele (edition-free) was estimated to be 2.3% according to the estimated recombination frequency 30%. **C** The table of *eIF4E1*-edited, *ALS1*-edition-free mutant genotypes obtained from the strategy depicted in B. The expected amino acid mutations due to the genetic mutations verified in each plant are listed. Del 184-186: deletion of amino acids from position 184 to 186. Note that some mutants were heterozygous for deletions or nonsynonymous mutations in *ALS2* (the *ALS1* homolog at chromosome 7), described as "mutation/WT".

assesses the efficacy of resistance breeding strategy by the base-editing in tomato *SleIF4E1* in an applicational point of view.

Results and discussion

Plants with the edited *eIF4E1* allele and the wild-type allele of the co-editing target *ALS1* were obtained from the F2 populations of co-edited parents.

Upon base-editing in the cherry tomato *SleIF4E1* gene using CBE, the *acetolactate synthase* (*ALS)*-co-editing strategy was adapted to facilitate the selection of edited mutants (Kuroiwa et al., 2023). In this strategy, *SlALS1* gene is targeted for base substitution at its Proline-186 codon since such modification confers resistance to herbicides, which then can be used in the culture media to select the edited plants (Veillet et al., 2019; Danilo et al., 2019; Huang et al., 2023). Hence, the *SlALS1* gene serves as a reporter gene for the transformation and the CRISPR/Cas9 activity during the generation of T0 plants without the need of introducing a foreign gene, i.e. an antibiotic resistance gene, for selection.

Because the *SleIF4E1* and *SlALS1* genes are both located on the short arm of the chromosome 3 and genetically linked, we reasoned that the edited alleles of the two genes would co-segregate in the progenies of the edited T0 plants and this could obscure the true phenotype of *SleIF4E1*-edited plants (**Figure 1A**). To confirm this, the T3 generation of plants edited in *SleIF4E1* region II used in Kuroiwa et al. (2023) was subjected to the genomic sequencing of the target region in *SlALS1*. The results showed that all the CBE-edited *SleIF4E1* mutant lines contained nonsynonymous mutations in *SlALS1* at homozygous state (**Supplementary data 1**). We also analyzed the genomic sequence of a homolog, *SlALS2* gene, because simultaneous base-editing in the *SlALS2* with the *SlALS1* have been reported due to the high homology of the target region between the two homologs (Veillet et al., 2019). The sequencing confirmed that all the mutant lines also contained a missense mutation or a deletion of a few nucleotides in *SlALS2* at homozygous state (**Supplementary data 1**).

In order to properly assess the potential impact of the *SleIF4E1* mutations on the plant phenotypes, we sought to remove the edited *SlALS* alleles from the *SleIF4E1*-edited population through genetic recombination. Though linked, the *SleIF4E1* and *SlALS1* genes were located in the highly recombinant region of the tomato chromosome 3, and the recombination frequency of the two genes was estimated to be 30% (Tomato Genome Consortium, 2012; Sim et al., 2012) (**Figure 1A**). Based on this, the theoretic proportion

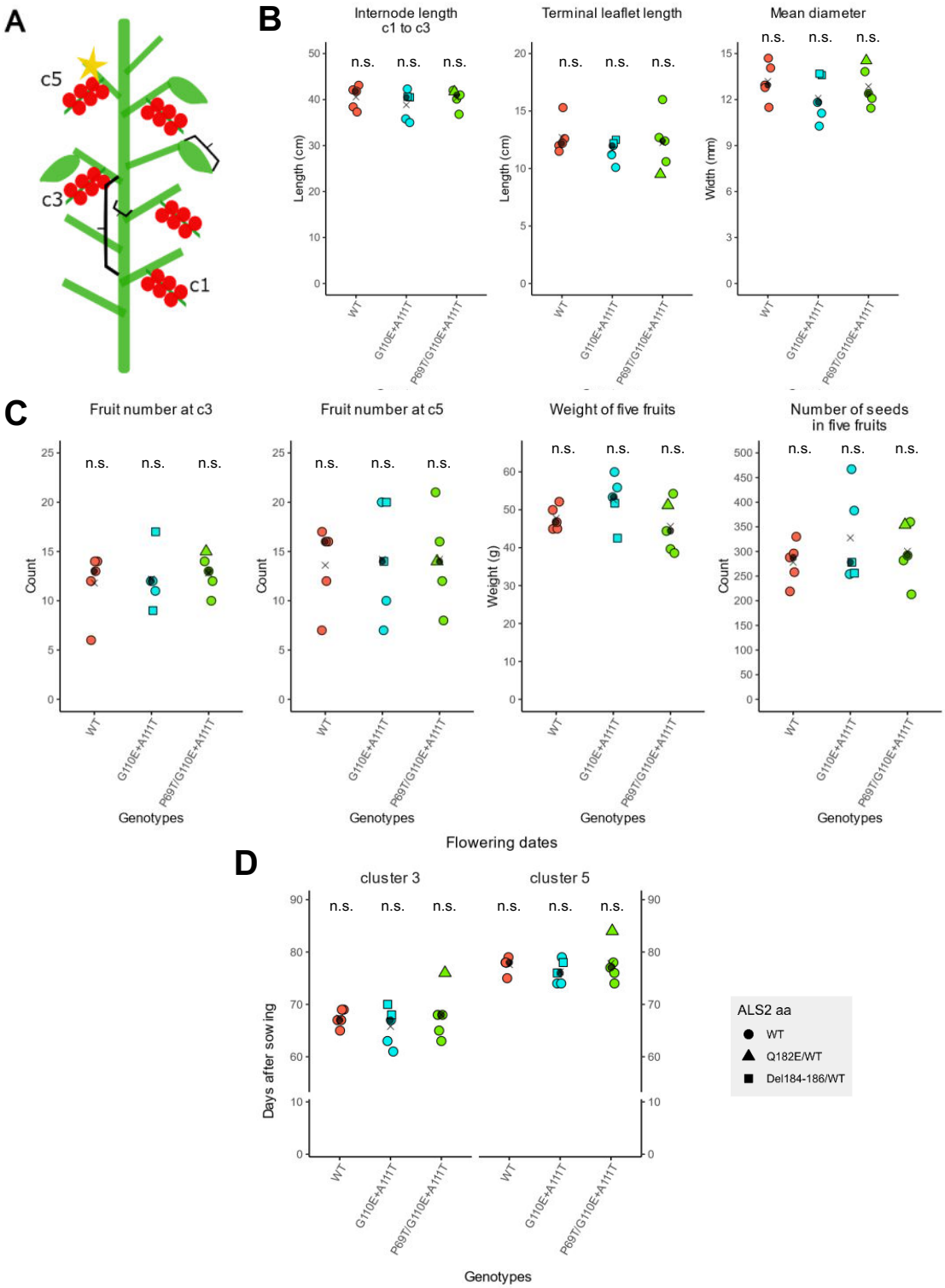


Figure 2 Vegetative and reproductive phenotypes are comparable between the wild-type cv.WVa106 and the e*IF4E1* edited mutants.

A The schematic to show the positions of measurement of each parameter. c1, c3, and c5 indicate the cluster 1, 3 and 5 of flowers/fruits. **B** Measurement of vegetative parameters. Internode length between the cluster 1 and 3; length of terminal leaflet, a leaf at the extremeties of a compound leaf, at one stem above cluster 3; Mean diameter of the two measurements of diagonal axis of main stem above the cluster 2. **C** Measurement of reproductive parameters. Fruit number at cluster 3 and at cluster 5; total weight of five representative fruits from each plant; and the number of seeds in the same five representative fruits on the plant. **D**: Flowering date recorded for the cluster 3 and 5 of each plant. **B-D** No statistically significant differences were observed in any of the parameters (denoted as n.s.) according to the one-way ANOVA test and the Kruskal-Wallis test (applied to the internode length). Shapes indicate the ALS2 amino acid substituttions. Black circle is median, black cross is the mean. Five WT plants and all the plants listed in Figure 1C for the *eIF4E1* mutant genotypes were analyzed.

of the homozygous *SleIF4E1* mutant/*SlALS1* wild-type (WT) individuals in an F2 population after backcrossing of mutants was estimated to be 2.3%, without considering the unlinked *SlALS2* gene on chromosome 7 (**Supplementary Data 2A**). Therefore, we performed backcrossing of the *SleIF4E1*-*SlALS*-co-edited T3 plant lines with wild-type WVa106 plants, the cultivar used for the co-editing, and a subsequent selfing and generated the F2 population for each mutant line (**Figure 1B**). Here, we did not consider the *SlALS2* gene state as it could easily be segregated afterwards. The generated F2 population from each parental *SleIF4E1* mutant line was screened, and the progenies that were homozygous for *SleIF4E1* mutation and homozygous wild-type *SlALS1* allele were selected (**Figure 1B, C**).

As a result, five plants with the *SleIF4E1* region II mutant allele (coding for G110E+A111T substitutions) and the wild-type *SlALS1* allele were obtained, and of which three were also homozygous for the wild-type *SlALS2* allele and two heterozygous for deletion of 9-bp (resulting in ALS2 with deletion of position 184 to 186) (**Figure 1C** blue). Five plants with *SleIF4E1* region I + II mutant allele (coding for P69T/G110E+A111T substitutions) and the wild-type *SlALS1* were obtained, and of which four were also homozygous for the wild-type *SlALS2* allele and one heterozygous for substitution of one base (resulting in ALS with substitution Q182E) (**Figure 1C** green). Although some plants segregated with the edited *SlALS2* alleles, these alleles were at heterozygous state; therefore, we presumed that the potential impact of these mutations on the plant phenotypes may be alleviated. Thus, all the selected individuals and the wild-type WVa106 were subsequently used for the whole-plant phenotyping.

The eIF4E1 precision editing does not affect the whole-plant growth and development at a greenhouse scale.

In order to assess the plant growth of the obtained *SleIF4E1* mutant/*SlALS1* WT plants, we carried out a whole-plant phenotyping in greenhouse. All the selected plants with mutant *SleIF4E1* alleles (listed in **Figure 1C**) were analyzed along with five wild-type plants[13]. Different parameters of both vegetative and reproductive growth were chosen to obtain an overview of the plant growth (Vicente et al., 2015; Bouzroud et al., 2020; Pons

[13]As listed in the Figure 1C, the selected mutants are heterogenous for the *SlALS2* background. Thus, progenies of these mutants with WT *SlALS1* and *SlALS2* alleles are currently phenotyped at a larger number for further validation of the presented assay (June 2024). The situation is explained at the end of the section.

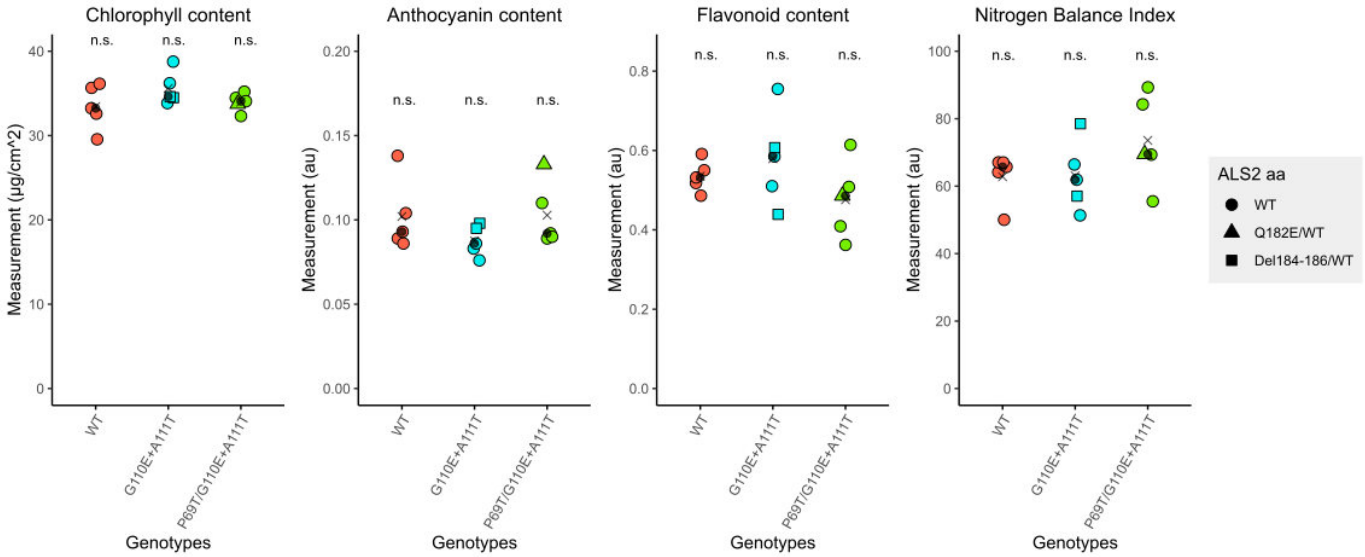


Figure 3 Leaf contents are comparable between the wild-type cv.WVa106 and the e*IF4E1* edited mutants.

Chloropyll content, anthocyanin content, flavonoid content and Nitrogen Balance Index (NBI) were measured by an optical sensor Dualex®. au: absorbance unit for anthocyanin and flavonoid; arbitrary unit for NBI. No statistically significant differences were observed in any of the parameters (denoted as n.s.) according to the one-way ANOVA test and the Kruskal-Wallis test (applied to the anthocyanin content). Shapes indicate the ALS2 amino acid substituttions. Black circle is median, black cross is the mean. The result is a representative of three measurements on the same leaf at a similar developmental stage (a terminal leaflet above the second to the top flower cluster). Five WT plants and all the plants listed in Figure 1C for the *eIF4E1* mutant genotypes were analyzed.

et al., 2022). The measured growth parameters are described in details in materials and methods section and the measured location on the plant is shown in **Figure 2A**.

The phenotyping results showed that there were no statistically significant differences between the WT WVa106 and two edited *eIF4E1* genotypes: G110E+A111T and P69T/G110E+A111T, for all the vegetative and reproductive growth parameters including the flowering time at a greenhouse scale (**Figure 2B-D**). Measurement of leaf contents showed that the WT and the G110E+A111T and P69T/G110E+A111T genotypes accumulated comparable levels of chlorophyll, anthocyanin, and flavonoid, and the NBI, with no statistically significant differences (**Figure 3**). The results demonstrate that the precise modifications in the region II or in both region I and II of eIF4E1 protein do not influence the plant growth. Importantly, this phenotyping complements the results from Kuroiwa et al. (2023); and provides evidence that the precise genome-editing which introduces nonsynonymous mutations in the susceptibility gene *eIF4E1* confers an enlarged potyvirus resistance spectrum compared to the single *eif4e1*^{KO} without compromising the plant growth and productivity.

Due to time constraint, I am presenting the data from the preliminary phenotypic analyses carried out in spring 2023. Because the analyses were done directly on the plants after selection of different genotypes, the small number of plants analyzed makes it difficult to conclude on the phenotypic trends. At present (June 2024), new measurements are being carried out by an engineer school student, Anaë Laurens. Data analyses are in progress, and I am hoping to present the complete data at my defense. The list of assayed genotypes and the experimental design in the ongoing assay is provided in Supplementary Data 8.

Long-term infection test using a stringent isolate PVY-N605 indicates the weak durability of the edited eIF4E1-based resistance

Another important aspect when implementing the genetic resistance in crop breeding is the durability of the resistance against the pathogen of interest. Genetic resistance can be overcome when the resistance breaking mutants of pathogen emerges due to its rapid evolution, adapting to the host resistance allele. The durability of a resistance allele to resistance breakdown is ideally estimated by the amount of time during which the allele could protect the crops when extensively used in the field. However, in the case of *eIF4E1* gene, the durability of resistance alleles have been also successfully estimated by the virus infection tests using a panel of potyvirus species or isolates at a laboratory scale; and the positive correlation between the resistance

durability and spectrum have been demonstrated for pepper *eIF4E1* variants (Moury et al., 2014).

We previously showed that a combination of nonsynonymous mutations in the edited *SleIF4E1* alleles were associated with a larger potyvirus resistance spectrum than each mutation alone or inactivation of the gene. Precisely, the P69T/G110E+A111T-coding (region I + II mutant) *SleIF4E1* displayed a wider resistance spectrum than the single-region mutant allele (P69T-coding allele or G110E+A111T-coding allele) or the single KO *Sleif4e1*^{KO} allele generated through TILLING (Piron et al., 2010; Kuroiwa et al., 2023); and the spectrum included a PVY isolate N605 (PVY-N605), which can use the tomato eIF4E1 and eIF4E2 proteins interchangeably (**Supplementary Data 1**). Nevertheless, the resistance spectrum of P69T/G110E+A111T allele was neither as wide as that of the natural resistance allele *SheIF4E1*^{PI24}*-pot1* derived from the wild tomato relative *S. habrochaites* accession PI247087 (hereafter referred to as PI24) nor the two KO alleles *Sleif4e1*^{KO} and *Sleif4e2*^{KO} combined (Gauffier et al., 2016; Kuroiwa et al., 2023). Notably, the P69T/G110E+A111T lines were resistant to PVY-N605 which are virulent to either the single KO *Sleif4e1*^{KO} or *Sleif4e2*^{KO} mutant lines (Gauffier et al., 2016). The P69T/G110E+A111T plants accumulated significantly lower levels of PVY-N605 CP than the WT WVa106; however, the CP levels were slightly elevated compared to the near-isogenic line of natural resistance *pot1* allele (NIL-pot1) in the *S. lycopersicum* cv. Mospomorist background (Kuroiwa et al., 2023). These differences in resistance spectra and in the suppression of a stringent PVY-N605 isolate between the edited P69T/G110E+A111T-coding *SleIF4E1* allele and the natural resistance *SheIF4E1*^{PI24}*-pot1* allele have incited us to determine whether the resistance conferred by the edited allele was as durable as that by the natural allele.

In order to estimate the durability of the resistance *SleIF4E1* alleles, we carried out an assay to monitor the potyvirus infection for a longer period of time than the routine virus infection assay in which the susceptibility/resistance is determined at three weeks post-inoculation (wpi). We hypothesized that if the edited P69T/G110E+A111T-coding *SleIF4E1* allele is not durably resistant, we may be able to detect the viral accumulation in the plants due to resistance breaking, by assessing at later time points than 3 wpi (**Figure 4A**). PVY-N605 isolate was used in this long-term infection assay, as it differentiated the resistance profile of the plants carrying the edited allele and those carrying the robust natural resistance allele. Moreover, PVY-N605 has previously been used to determine the

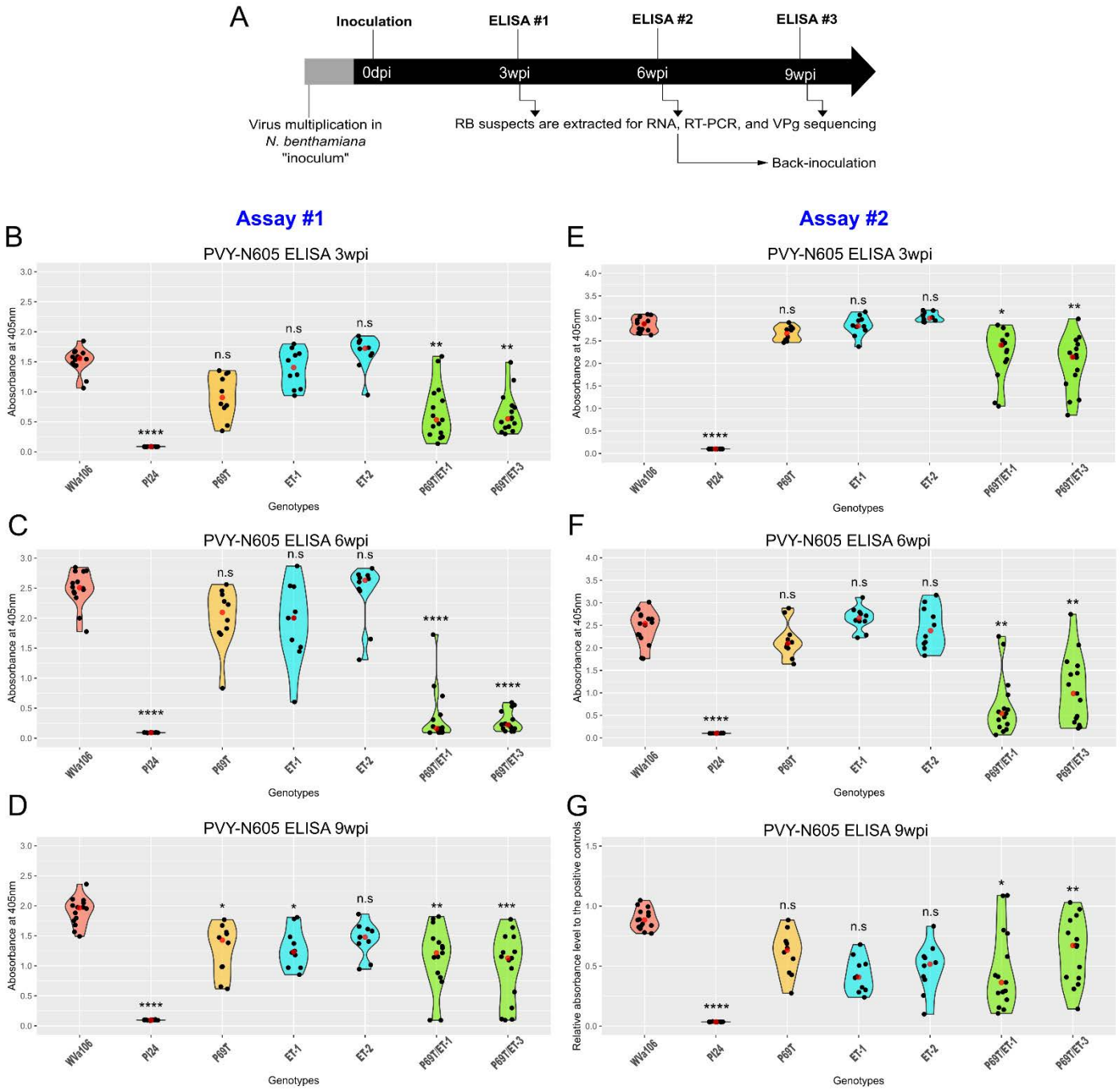


Figure 4 Edited resistance *eIF4E1* mutants are overcome in long-term virus infection in two assays.

A the workflow of the long-term infection test. Counting from the day the virus from the inoculum (*N. benthamiana*) is inoculated to the tomato population « 0dpi », the viral accumulation is quantified at three-, six-, and nine-week-post-inoculation (wpi) from each plant. When the resistance breaking (RB) event is suspected due to elevated viral accumulation, the infected plants are subjected to RNA extraction, RT-PCR on VPg sequence, and Sanger sequencing of the amplicon. If the mutation is found in VPg sequence from a sample, the same plant is used for backinoculation (ref. Figure 6). **B-D** the viral coat protein (CP) accumulations measured by ELISA at 3, 6, and 9 wpi, respectively, in Assay #1. **E-G** Assay #2. The colors are according to the eIF4E1 allele. Blue: G110E+A111T and green: P69T/ G110E+A111T. WVa106 is the susceptible wild-type, PI24 the resistant accession PI247087 of *S. habrochaites.* Red point is median. Significance of difference compared with WVa106 based on Kruskal–Wallis and Dunn-test as a post-hoc test is indicated by asterisks. P-value <10^{-4} ****;10^{-4} to 10^{-3} ***;10^{-3} to 10^{-2} **; 0.05 *. n.s.: no significant differences.

durability of the weak *eIF4E1* resistance allele *pot1*^{2} harboring D112G mutation in the eIF4E1 region II in the *S. pimpinellifolium* accession LA0411 (Lebaron et al., 2016).

Briefly, PVY-N605 was first propagated on *Nicotiana benthamiana* plants and then mechanically inoculated on the WT WVa106 plants, the edited P69T, G110E+A111T, and P69T/G110E+A111T genotypes (WVa106 background), and the *S. habrochaites* accession PI24 possessing *SheIF4E1*^{PI24}*-pot1*. Then, the systemic leaves[14] at different levels of the inoculated plants were sampled at 3-, 6-, and 9 wpi for viral accumulation quantification by double antibody sandwich (DAS)-ELISA, which specifically detects the PVY coat protein (CP) (**Figure 4A**). Independent tests were carried out twice and referred to as assay #1 and #2. For the G110E+A111T, and P69T/G110E+A111T genotypes, the progenies of plants (ET-1 and ET-2; and P69T/ET-1 and P69T/ET-3, respectively) used in the phenotyping and only edited in the *SleIF4E1* gene were used (**Figure 1C**).

The DAS-ELISA results showed that at all three time points, susceptible WT WVa106 plants accumulated higher level of the viral CP than most genotypes and the resistant accession PI24 the undetectable level of viral CP, thereby respectively serving as the susceptible and resistant controls at each time point (**Figure 4B-D, E-G**). This also validated the durable resistance conferred by the *SheIF4E1*^{PI24}*-pot1* allele as previously reported (Lebaron et al., 2016). We defined the scale of susceptibility/resistance as follows: plants accumulating a statistically similar level of viral CP to the wild-type as "fully susceptible", those accumulating significantly lower CP than the wild-type but higher than the accession PI24 as "partially resistant", and those accumulating undetectable level of viral CP as "fully resistant". According to this scale, the P69T line and G110E+A111T lines were generally fully susceptible to PVY-N605 at all three time points: 3-, 6-, and 9 wpi, with one exceptional G110E+A111T plant at 9wpi in Assay #2 (**Figure 4B-D, E-G**).

As for the two independent P69T/G110E+A111T lines, P69T/ET-1 and P69T/ET-3, the viral CP accumulation levels in the plants fluctuated over the three time points measured and showed resistance profiles ranging from partial to full resistance. At 3wpi, the two mutant lines showed partial resistance with varied viral CP levels between individuals within each line, in both assay #1 and #2 (**Figure 4B, 4E**). High viral accumulations in the P69T/G110E+A111T genotype observed in these assays were not

[14] Systemic leaves are leaves of plants that are not directly virus-inoculated and often younger or newly formed than the inoculated leaves.

observed previously in two independent experiments (Kuroiwa et al., 2023). At 6wpi, the viral CP levels lowered in most of the plants of independent P69T/G110E+A111T lines showing full resistance in assay #1 and stronger partial resistance in assay #2 (**Figure 4C, F**). Finally, at 9 wpi, the viral CP levels were elevated in most P69T/G110E+A111T plants, and the levels varied between individuals; however, they remained statistically significantly lower compared to the WT plants, showing partial resistance (**Figure 4D, G**). Throughout the long-term infection assays, the viral accumulation levels in most P69T/G110E+A111T plants were at detectable levels by DAS-ELISA; and these levels were higher than in the resistant accession PI24, but lower than in the WT plants (**Figure 4B-G**). Altogether, the results show that the previously reported resistance of P69T/G110E+A111T plants can be broken down by PVY-N605 isolate at as early as 3 wpi, and many plants continue to accumulate virus at least up to 9 wpi albeit at lower titer than the susceptible genotypes.

The method of virus inoculation, i.e. mechanical inoculation, used in this long-term infection test may apply stronger inoculum pressure to the resistance *SleIF4E1* allele than in the natural field setting, where the virus entry into plants depends on the vectors and other abiotic factors. Because the size of PVY-N605 population introduced in plants by inoculation may influence the probability of emergence of the resistance breaking isolates (Rousseau et al., 2018), the results obtained for the P69T/G110E+A111T genotype remain indicative of the true resistance durability of this edited *SleIF4E1* allele in the field. Meanwhile, the results demonstrate two things: one is that the edited P69T/G110E+A111T-coding *SleIF4E1* allele is not as durably resistant as the natural resistance allele *SheIF4E1*^{PI24}*-pot1* in the face of PVY-N605 mechanical inoculation; and another is that nevertheless, this allele is able to suppress significantly the viral loads compared to the susceptible alleles. Interestingly, the elevated PVY-N605 accumulation in the P69T/G110E+A111T genotype, which was initially reported as resistant in Kuroiwa et al. (2023), was reminiscent of the PVY-N605 accumulation observed in the *S. pimpinellifolium* accession LA0411, carrying *SpeIF4E1-pot1*^{2} allele (Lebaron et al., 2016). In the case of the *SpeIF4E1-pot1*^{2} allele, its resistance was broken by the emergence of resistance-breaking (RB) variants of PVY-N605. The elevated viral CP levels observed in many P69T/G110E+A111T plants in the long-term infection tests indeed hinted at emergence of such RB variants. Therefore, we proceeded to detection of potential RB variants among infected P69T/G110E+A111T plants in order to determine the cause of the resistance breakdown.

A

	Plant no.	Mutations in VPg	
		6 wpi	9 wpi
P69T/ET-1	62	R94G	R94G
	65	None	R94G
	67	R94G^{double}	R94G^{double} + L115M^{double}
	69	R94G	R94G^{double} + L115M^{double}
	72	None	R94G^{double} + S120I^{double}
	73	L115S	L115S^{double}
P69T/ET-3	76	None	No accumulation
	77	None	No accumulation
	79	R94G^{double}	N88 (silent) + R94G
	80	S101G	R94G^{double} + L115S^{double}
	84	None	No accumulation
	85	L115S	L115S + S120I^{double}
	90	None	K105R

B

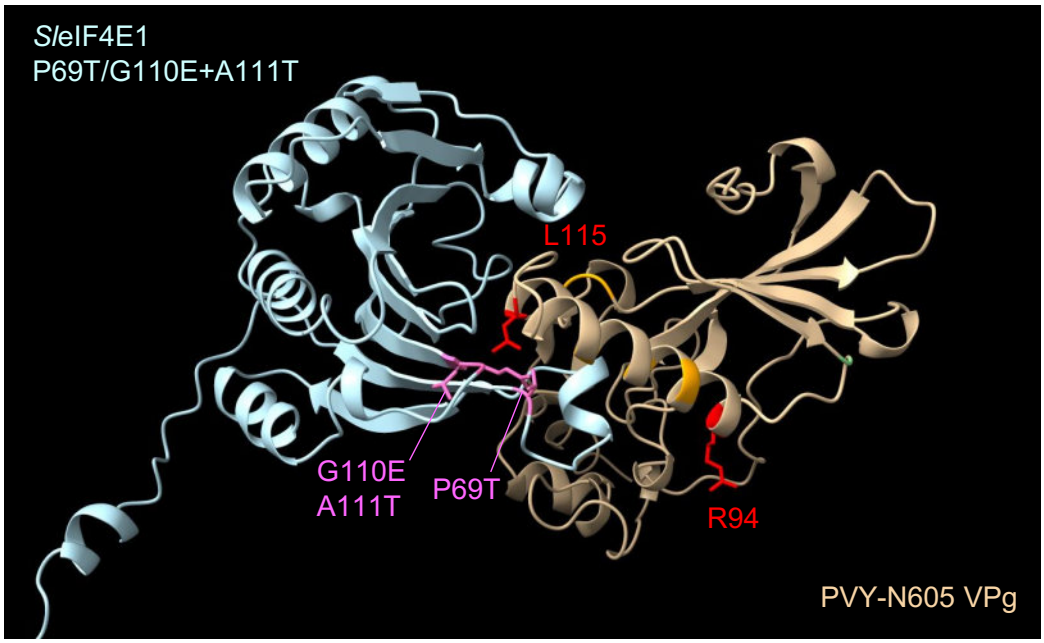


Figure 5 Amino acid substitutions were identified in the VPg sequence amplified from the infected P69T/G110E+A111T individual plants.

A List of amino acid substitutions in VPg sequence from samples of the plant individuals monitored at 6 and 9 wpi. Plant number corresponds to the individual plant numbers in Supplementary Data 4. The superscript « double » indicates that the both peaks of WT nucleotide and the substitution were chromatogram of the sequence. Mutations that were detected for the first time at each time point are colored. Colors correspond to the colored residues in B and Supplemtary Data 5A. In red are the most frequently found site of mutation. «No accumulation» was determined by DAS-ELISA and/or RT-PCR.

B Positions of the detected mutations on PVY-N605 VPg (gold) in complex with tomato eIF4E1 protein (light blue) harboring P69T and G110E+A111T substitutitons (pink). The WT VPg-*Sl*eIF4E1 complex structure was modeled using ColabFold. The unstructured N-terminus of *Sl*eIF4E1 is omitted from this figure.

The elevated CP accumulation in P69T/G110E+A111T plants at 6 and 9wpi was caused by resistance breaking PVY-N605 variants

Because many P69T/G110E+A111T plants accumulated detectable levels of viral CP in the long-term infection tests, the emergence of RB isolates was strongly suspected. The *eIF4E1*-based resistance is most often overcome by PVY due to nonsynonymous mutations in the central region of viral genome-linked protein (VPg) (Moury et al., 2004; Charron et al., 2008). Thus, the P69T/G110E+A111T plants that accumulated various CP levels were selected at 6- and 9 wpi, and their extracts used for DAS-ELISA were subjected to RNA extraction followed by RT-PCR to amplify and sequence the VPg cistron of the accumulated virus (**Figure 4A**). The analyses on the 3 wpi were not pursued because the preliminary results on the assay #1 samples showed that the amplified VPg sequences from the majority of plant samples with detectable virus titer (12/13 plants) did not harbor any mutation in VPg, except for one VPg sequence.

For each DAS-ELISA performed, one or more extracts from the susceptible genotypes, namely the WT WVa106, P69T, and G110E+A111T, were analyzed, and the results confirmed the accumulation of the WT VPg of PVY-N605 in accordance with their full susceptibility. RT-PCR on the samples from the resistant accession PI24 confirmed the lack of viral RNA accumulation, which is consistent with the DAS-ELISA results. (A representative RT-PCR result on the 6-wpi samples from the assay #1 is shown in **Supplementary Data 4C**.)

From the assay #1 (corresponding to **Figure 4C and D**), the ELISA extracts at 6 wpi from all 15 P69T/ET-1 plants and 15 P69T/ET-3 plants were analyzed (**Supplementary Data 4A**), and among those, the extracts at 9 wpi from six P69T/ET-1 plants and seven P69T/ET-3 plants were further monitored (**Supplementary Data 4B**). The VPg was not amplified by RT-PCR from the 6-wpi and 9-wpi P69T/G110E+A111T plant samples that accumulated undetectable level of CP by DAS-ELISA, like the accession PI24 (6-wpi results in **Supplementary Data 4C**). The amplified VPg sequences obtained from the P69T/G110E+A111T plant extracts at 6 wpi showed that total ten VPg sequences contained a single nonsynonymous mutation and twelve did not (**Figure 5A** column "6 wpi" and **Supplementary Data 4D**). At 9 wpi, all the VPg sequences detected from the 13 samples analyzed contained one or more amino acid substitution (**Figure 5A** column "9 wpi"). This included the samples in which mutant VPg was not detected at 6 wpi. Most of VPg sequences that obtained an amino acid substitution at 6 wpi retained the same

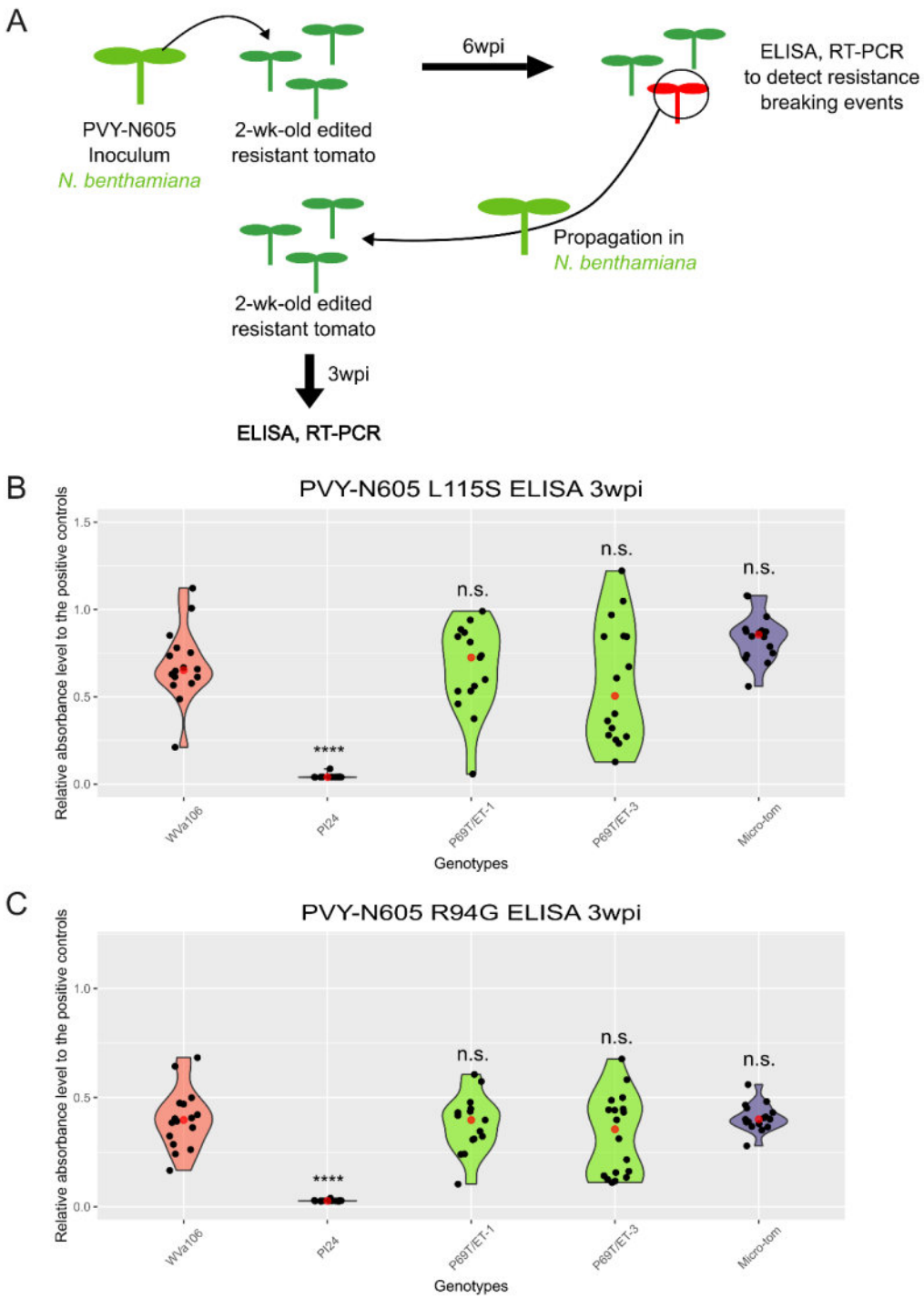


Figure 6 Backinoculation assay confirms the resistance breaking isolates.

A The schematics of the backinoculation assay. The 2-week-old tomato plants of resistance P69T/G110E+A111T genotype are inoculated with PVY-N605 propagated on the *N. benthamiana*. At 6wpi, viral accumulation is quantified by ELISA and the plants with elevated virus accumulation (compared to the resistant control PI24) are subjected to RT-PCR on the PVY-N605 VPg sequence. If mutation is identified in the VPg sequence from the plant extract, the accumulating virus is considered a potential resistance breaking isolate. The infected plant (red) is sampled, and the virus with VPg mutation is propagated on *N. benthmiana*. The propagated virus is inoculated on a new population of the same resistance P69T/G110E+A111T genotype. If a large portion of the population, if not all, is infected, then this isolate is a resistance breaking isolate. **B, C** The relative viral accumulations of RB suspect isolates (PVY-N605 L115S and R94G) measured by ELISA at 3 wpi. The values are corrected from the raw absorbance at 405nm, based on the positive controls in the ELISA plates (details in Materials and methods). The colors are according to the *eIF4E1* allele (Green: P69T/ G110E+A111T). WVa106 is the susceptible wild-type, PI24 the resistant accession PI247087 of *S. habrochaites.* Red point is median. Significance of difference compared with WVa106 based on Kruskal–Wallis and Dunn-test as a post-hoc test is indicated by asterisks. P-value $<10^{-4}$ ****. n.s.: no significant differences.

substitution, while some obtained additional substitution. Similar results were obtained for the long-term infection assay #2 and are summarized in **Supplementary Data 5A**.

While there was no single fixed VPg mutation in the population, three mutations: arginine to glycine at position 94 (R94G), leucine to serine at position 115 (L115S), and leucine to methionine at position 115 (L115M), were predominantly detected in independent assays (**Figure 5** and **Supplementary Data 5A**). R94G substitution is not located in the central region of VPg and has never been reported in PVY isolates that overcame *eIF4E1*-mediated resistance. On the contrary, L115 is located in the central region and both L115S and L115M substitutions in PVY isolates have been previously reported (Moury et al., 2004; Lebaron et al., 2016). Indeed, position 115 of VPg is one of the sites that are suggested to undergo positive selection in the PVY sequence along with position 105, 119, 123, 149, and 164 (Moury et al., 2004; Ayme et al., 2006). The predicted structure of the complex between the edited *Sl*eIF4E1 with P69T/G110E+A111T and the PVY-N605 VPg also shows the proximity of the edition and the L115 of VPg protein (**Figure 5B**). Importantly, L115S has been previously reported as a signature amino acid in the RB variants of PVY-N605 that overcame the low durability resistance allele *pot1*^{2} in tomato (Lebaron et al., 2016). Other mutations than these three predominantly found substitutions were found in both non-central and central regions of VPg; and those found in central region were either identical to previously reported mutations or located at the position associated for resistance breaking events in PVY isolates (**Supplementary Data 5B**).

To confirm that the observed viral accumulation in P69T/G110E+A111T plants in the long-term assays was due to the resistance breaking by PVY-N605 variants, back inoculation assays of RB suspect variants were carried out. The procedure of back inoculation assay is described in **Figure 6A**: briefly, when the PVY-N605 variant with mutation in VPg is detected in a plant of resistance genotype in the initial infection assay, this variant is recuperated and propagated on *N. benthamiana* plants to inoculate on a new population of the same resistance genotype. If this PVY variant accumulates in the resistance genotype at a level comparable to that in the susceptible wild-type, then this variant is considered a resistance breaking isolate.

Among the identified variants with different VPg mutations, we chose to test two distinct variants that were frequently observed: L115S and R94G. PVY-N605 L115S variant was chosen since it has been already characterized as an RB variant that

overcomes a weak resistance *eIF4E1* allele *pot1²* and hinted its association with susceptibility of edited resistance *eIF4E1* genotype P69T/G110E+A111T. PVY-N605 R94G variant was chosen because R94G-harboring VPg was found at intriguingly high frequency from many infected P69T/G110E+A111T individuals (7 out of 10 amplified VPg from 9-wpi samples in assay #1 and 6 out of 10 amplified VPg in assay #2) and was strongly suspected as a new RB signature of this PVY isolate. As the P69T/G110E+A111T-coding *eIF4E1* allele does not exist in the natural variation, it was possible that this R94G mutation emerged specifically to overcome this *de novo* resistance allele.

In the back inoculation assay, the WT WVa106, the resistant accession PI24, *S. lycopersicum* cv. Micro-tom, and two independent P69T/G110E+A111T lines were inoculated with each RB suspect variant. Micro-tom was included in this assay because this cultivar possesses a natural *eIF4E1* variant which has been shown to display low durability and quick RB to PVY N605 (Lebaron, C. unpublished data). Indeed, it resembles the weak resistance allele *pot1²*: the Microtom eIF4E1 harbors one amino acid substitution of aspartic acid to asparagine (D112N) in the region II, whereas the *pot1²*-coded eIF4E1 harbors D112G, plus V54K outside of the two regions. Therefore, the *SleIF4E1^{D112N}* genotype (Microtom) could serve as a proxy of resistance breaking by PVY-N605 variants. The ELISA at 3 wpi showed that for both PVY-N605 L115S and R94G variants, all of the genotypes except for the accession PI24 were susceptible, with statistically no significant differences between those genotypes (**Figure 6B** and **C**). This confirmed that both L115S and R94G variants were indeed RB variants, capable of accumulating in the tomatoes carrying the edited P69T/G110E+A111T-coding *SleIF4E1* allele as efficiently as in the susceptible WT tomato. The results also confirmed the low durability of resistance allele *SleIF4E1^{D112N}* against the PVY-N605 infection. Finally, the back inoculation showed that the neither L115S nor R94G mutation in VPg was not sufficient for PVY-N605 to overcome the *SheIF4E1^{PI24}-pot1*-mediated resistance which is consistent from the result for L115S variant reported in Lebaron et al. (2016).

By monitoring PVY-N605 infection in the edited resistance genotype P69T/G110E+A111T over period of nine weeks after inoculation, we show that the resistance conferred by P69T/G110E+A111T-coding *SleIF4E1* is overcome by RB variants after 3 wpi and no later than 9wpi. These RB variants possess one or more mutations in VPg, and such mutations are required for a successful proliferation of the virus in a long term, or more than 3 wpi. These include previously reported RB variant of PVY-N605 as

well as a new signature outside of the VPg central region, which may have evolved specifically to adapt to the P69T/G110E+A111T *Sl*eIF4E1. The reason that the virus could accumulate at 3 wpi without VPg mutation, i.e. a signature of resistance breaking, remains to be solved. It would be interesting to analyze the presence of mutations in other PVY cistrons than the VPg, such as the cylindrical inclusion or the P1, whose variants were associated with breaking of *eIF4E1*-mediated resistance in order to gain insights into the early infection phase in the resistance mutants (Abdul-Razzak et al., 2009; Nakahara et al., 2010).

Conclusion

This work provides complementary analyses of the edited resistance P69T/G110E+A111T-coding *SleIF4E1* allele which was generated in Kuroiwa et al. (2023) (*Chapter 2*). We demonstrate here that the precise editing of *SleIF4E1* gene in cherry tomato may reconcile the trade-off of growth and defense. The edited P69T/G110E+A111T genotype achieves a wider potyvirus resistance spectrum than the single KO mutant tomato *Sleif4e1*^{KO} or *Sleif4e2*^{KO} (Gauffier et al., 2016; Kuroiwa et al., 2023) without plant growth impairment as seen in the double KO *eif4e1*^{KO}*xeif4e2*^{KO} mutant.

However, we also report that the resistance conferred by the P69T/G110E+A111T-coding *SleIF4E1* allele against a stringent PVY-N605 isolate is not as durable as that by the natural resistance *SheIF4E1*^{PI24}*-pot1* allele coding for eight amino acid substitutions. The P69T/G110E+A111T-coding *SleIF4E1*-mediated resistance may be overcome by nine weeks after inoculation, as the virus acquires one or more mutations in the VPg. The low durability of resistance by the edited *SleIF4E1* allele compared to the natural resistance allele may at least be partially attributed to the lack of extensive nonsynonymous mutations in the region I: one in the edited allele versus three in the natural allele. Indeed, most of the naturally selected eIF4E1 variants, i.e. coded by *SheIF4E1*^{PI24}*-pot1* or various pepper *eIF4E1-pvr2* alleles, have more substitutions in the region I than the region II, with the exception of the tomato *pot1*^{2} allele as well as *SleIF4E1*^{D112N} allele (as attested in this manuscript), whose resistance were rapidly overcome. This difference points to the room for improvement of the edited resistance gene by introducing more mutations in eIF4E1 region I in order to mimic the natural resistance allele. Incidentally, the edited P69T/G110E+A111T-coding *SleIF4E1* genotype

would provide an interesting model to study PVY resistance breaking events in tomato, given the surprising dynamics of the viral accumulation level during the extended time of observation.

Materials and methods

Plant material

The wild-type genotype and the genetic background of the of *eIF4E1* (Solyc03g005870)-edited mutant lines were the cherry tomato (*Solanum lycopersicum* var. *cerasiforme*) cultivar (cv.) West Virginia 106 (WVa106). The wild tomato species used in the long-term virus infection test was the *S. habrochaites* accession PI247087 (referred in the text as PI24). For whole-plant phenotyping, the seedlings were grown in a growth chamber at 24°C/18°C in the 16 hr day/8 hr night cycle for 6-8 weeks until they were genotyped and transplanted in pots in the greenhouse environment. For virus testing, the plants were grown and maintained in a growth chamber at 24°C/18°C in the 16 hr day/8 hr night cycle throughout the experiments.

Genotyping by PCR and KASP (Kompetitive allele-specific PCR) assay

Genomic DNA was extracted from fresh tomato leaf samples according to (Fulton et al., 1995) and used for genotyping by PCR and KASP. PCR was performed on the *SleIF4E1, SlALS1,* and *SlALS2* genomic sequences using specific primers. *GAPDH* (*Glyceraldehyde-3-Phosphate Dehydrogenase*) gene was amplified as the technical control. PCR products were Sanger-sequenced by Genoscreen (Lille, France). KASP assay was performed to genotype the F2 individuals using specific KASP primers designed for *SleIF4E1* region I, *SleIF4E1* region II, and *SlALS1* at C-20 and C-14 from the PAM site (detail of the nucleotide position available in the main text). All the primer sequences and pairs used for detection are available in **Annex I** . Lecture of KASP fluorescent signal was performed on the Bio-Rad CFX-96 (California, USA).

Phenotyping in greenhouse

Representative parameters of vegetative and reproductive growth were measured to characterize the *SlALS*-edited plants (Vicente et al., 2015; Pons et al., 2022). Vegetative growth traits are the internode length between the stems of the first and third cluster ("c1" and "c3") of reproductive organs, the length of a terminal leaflet above the third cluster, and the average main stem diameter above the second cluster. These vegetative measurements were taken about 10-12 weeks after the sowing. Reproductive growth

traits included the number of fruits at cluster 3 and 5, the weight of five representative mature fruits from cluster 3 to 5, and the number of seeds from the same five fruits. The date of first flowering at cluster 3 and 5 was also recorded to compare the timing of inflorescence, an important transition from the vegetative to reproductive growth that influences the crop production. A non-destructive measurement of leaf content was performed using an optical sensor Dualex® (Pessl instruments, Austria). Four parameters related to plant photosynthetic activity and the nitrogen/carbon allocation were measured: chlorophyll content, anthocyanin content, flavonoid content, and the Nitrogen Balance Index® (NBI; the index of N/C allocation based on the chlorophyll and flavonoid ratio) (Fan et al., 2022).

Calculation of empirical recombination frequency

The free software R (https://www.R-project.org/) and rqtl "package" were used to calculate the empirical genetic distance of the *eIF4E1* and *ALS1* (Solyc03g044330) genes based on the observed ratios of segregation in the F2 populations of the co-edited plants. Mapping function Kosambi was used. The initial estimate of genetic distance r=0.293.

Virus isolates and infection assays using DAS-ELISA quantification

Potato virus Y (PVY) N605 isolate (PVY-N605) (Parrella et al., 2002) and its mutant progenies were used. Potyvirus long-term infection assay was performed using PVY-N605. Back inoculation assay was performed using the PVY-N605 progenies with mutation in viral genome-linked protein (VPg) obtained during the long-term infection test. The isolate was maintained in *Nicotiana benthamiana* plants 2 weeks before inoculation. Inocula were prepared from 1000 mg (fresh weight) of the infected tobacco leaves ground with 4 mL of grinding buffer: potassium phosphate buffer (0.03 M, pH = 7) with 0.2% of diethyldithiocarbamate, 80 mg active charcoal and 80 mg carborundum. 10-15 plants per genotype were mechanically inoculated on the cotyledons of 14-day-old tomato plants. At 3 weeks post-inoculation (wpi), 800mg – 1000mg of systemic leaves from each plant was collected and ground in 3.2-4 mL of grinding buffer (the sample:buffer weight ratio = 1:4). To measure the virus accumulation, DAS-ELISA was performed using anti-PVY (Sediag, Bretenière, France) according to the manufacturer's instructions. The absorbance was measured by a spectrophotometer at wavelength 405 nm. For long-term infection assay, the leaf sampling and DAS-ELISA was performed at 6 wpi and 9 wpi as well.

Resistance breaking suspect detection by total RNA extraction and cDNA analysis

Total RNA was extracted from 300 µL of the ELISA extracts of each plant of interest using 350 µL of TRI-Reagent (Sigma-Aldrich, St. Louis, USA). and quality-checked by concentration measurement on Nanodrop ND-1000 (ThermoFisher, Waltham, USA). Reverse transcription (RT) of the extracted RNA samples (1 µg) was performed using AMV (Avian Myeloblastosis Virus) Reverse Transcription Kit (Invitrogen, Waltham, USA) and oligo-dT18 (1 µM) to obtain coding DNA (cDNA). The cDNA for VPg sequence of PVY-N605 was amplified with a primer pair n605F/R and GAPDH (Glyceraldehyde-3-Phosphate Dehydrogenase) gene as the internal control with gapdhF/R. The PCR products were Sanger-sequenced by Genoscreen (Lille, France) and analyzed to detect mutations in VPg.

Protein-protein interaction modelling

Protein sequence of VPg sequence from the wild-type PVY-N605 polyprotein sequence (genbank ID: CAA66472.1) and *Sl*eIF4E1 sequence with P69T and G110E+A111T substitutions were used as input for multimer prediction by the Colabfold developed based on the AlphaFold2 (Jumper et al., 2021; Mirdita et al., 2022).

Data visualization and statistical analysis

Visualization of the results and the statistical tests were performed on the software R. One-way ANOVA test or Kruskal-Wallis non-parametric test ("pgirmess" package) were performed to evaluate the significance of differences between genotypes according to the sample profile. As the post-hoc test, Tukey-HSD test or Dunn test with p-value adjustment by Benjamini-Yekutieli method ("dunn.test" package) was performed to determine the significance. The type of statistical test to be used was determined based on the homogeneity of variances (Bartlett's test) and the normality of distribution (Shapiro-Wilk test) of a given dataset.

Acknowledgement

The authors thank Emmanuel Botton (INRAE GAFL) for taking care of tomato plants, Caroline Lebaron (INRAE GAFL) for helping with acquisition of the VPg sequences, Robin Thibert (INRAE GAFL) for helping with acquisition of the preliminary phenotyping data for this thesis.

References

Abdul-Razzak, A., Guiraud, T., Peypelut, M., Walter, J., Houvenaghel, M.-C., Candresse, T., LE Gall, O., and German-Retana, S. (2009). Involvement of the cylindrical inclusion (CI) protein in the overcoming of an eIF4E-mediated resistance against Lettuce mosaic potyvirus. Mol. Plant Pathol. 10:109–113.

Ayme, V., Souche, S., Caranta, C., Jacquemond, M., Chadoeuf, J., Palloix, A., and Moury, B. (2006). Different mutations in the genome-linked protein VPg of potato virus Y confer virulence on the pvr2(3) resistance in pepper. Mol. Plant. Microbe. Interact. 19:557–563.

Bastet, A., Robaglia, C., and Gallois, J. L. (2017). eIF4E Resistance: Natural Variation Should Guide Gene Editing. Trends Plant Sci. 22:411–419.

Bastet, A., Zafirov, D., Giovinazzo, N., Guyon-Debast, A., Nogué, F., Robaglia, C., and Gallois, J. L. (2019). Mimicking natural polymorphism in eIF4E by CRISPR-Cas9 base editing is associated with resistance to potyviruses. Plant Biotechnol. J. 17:1736–1750.

Bouzroud, S., Gasparini, K., Hu, G., Barbosa, M. A. M., Rosa, B. L., Fahr, M., Bendaou, N., Bouzayen, M., Zsögön, A., Smouni, A., et al. (2020). Down regulation and loss of auxin response factor 4 function using CRISPR/Cas9 alters plant growth, stomatal function and improves tomato tolerance to salinity and osmotic stress. Genes (Basel) 11:272.

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

Danilo, B., Perrot, L., Mara, K., Botton, E., Nogué, F., and Mazier, M. (2019). Efficient and transgene-free gene targeting using Agrobacterium-mediated delivery of the CRISPR/Cas9 system in tomato. Plant Cell Rep. 38:459–462.

Fan, K., Li, F., Chen, X., Li, Z., and Mulla, D. (2022). Nitrogen balance index prediction of winter wheat by canopy hyperspectral transformation and machine learning. Remote Sens. (Basel) 14:3504.

Fulton, T. M., Chunwongse, J., and Tanksley, S. D. (1995). Microprep protocol for extraction of DNA from tomato and other herbaceous plants. Plant Mol. Biol. Rep. 13:207–209.

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

Huang, X., Jia, H., Xu, J., Wang, Y., Wen, J., and Wang, N. (2023). Transgene-free genome editing of vegetatively propagated and perennial plant species in the T0 generation via a co-editing strategy. Nat. Plants Advance Access published September 18, 2023, doi:10.1038/s41477-023-01520-y.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature 596:583–589.

Kuroiwa, K., Danilo, B., Perrot, L., Thenault, C., Veillet, F., Delacote, F., Duchateau, P., Nogué, F., Mazier, M., and Gallois, J.-L. (2023). An iterative gene-editing strategy broadens eIF4E1 genetic diversity in Solanum lycopersicum and generates resistance to multiple potyvirus

isolates. Plant Biotechnol. J. Advance Access published January 30, 2023, doi:10.1111/pbi.14003.

Lebaron, C., Rosado, A., Sauvage, C., Gauffier, C., German-Retana, S., Moury, B., and Gallois, J. L. (2016). A new eIF4E1 allele characterized by RNAseq data mining is associated with resistance to potato virus Y in tomato albeit with a low durability. J. Gen. Virol. 97:3063–3072.

Michel, V., Julio, E., Candresse, T., Cotucheau, J., Decorps, C., Volpatti, R., Moury, B., Glais, L., Jacquot, E., de Borne, F. D., et al. (2019). A complex eIF4E locus impacts the durability of va resistance to Potato virus Y in tobacco. Mol. Plant Pathol. 20:1051–1066.

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. Nat. Methods 19:679–682.

Moury, B., Morel, C., Johansen, E., Guilbaud, L., Souche, S., Ayme, V., Caranta, C., Palloix, A., and Jacquemond, M. (2004). Mutations in potato virus Y genome-linked protein determine virulence toward recessive resistances in Capsicum annuum and Lycopersicon hirsutum. Mol. Plant. Microbe. Interact. 17:322–329.

Moury, B., Charron, C., Janzac, B., Simon, V., Gallois, J. L., Palloix, A., and Caranta, C. (2014). Evolution of plant eukaryotic initiation factor 4E (eIF4E) and potyvirus genome-linked protein (VPg): a game of mirrors impacting resistance spectrum and durability. Infect. Genet. Evol. 27:472–480.

Nakahara, K. S., Shimada, R., Choi, S.-H., Yamamoto, H., Shao, J., and Uyeda, I. (2010). Involvement of the P1 Cistron in Overcoming eIF4E-Mediated Recessive Resistance Against Clover yellow vein virus in Pea. Mol. Plant. Microbe. Interact. 23:1460–1469.

Parrella, G., Ruffel, S., Moretti, A., Morel, C., Palloix, A., and Caranta, C. (2002). Recessive resistance genes against potyviruses are localized in colinear genomic regions of the tomato (Lycopersicon spp.) and pepper (Capsicum spp.) genomes. Theor. Appl. Genet. 105:855–861.

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

Pons, C., Casals, J., Palombieri, S., Fontanet, L., Riccini, A., Rambla, J. L., Ruggiero, A., Figás, M. del R., Plazas, M., Koukounaras, A., et al. (2022). Atlas of phenotypic, genotypic and geographical diversity present in the European traditional tomato. Hortic. Res. 9.

Robaglia, C., and Caranta, C. (2006). Translation initiation factors: a weak link in plant RNA virus infection. Trends Plant Sci. 11:40–45.

Rousseau, E., Tamisier, L., Fabre, F., Simon, V., Szadkowski, M., Bouchez, O., Zanchetta, C., Girardot, G., Mailleret, L., Grognard, F., et al. (2018). Impact of genetic drift, selection and accumulation level on virus adaptation to its host plants. Mol. Plant Pathol. 19:2575–2589.

Sanfaçon, H. (2015). Plant Translation Factors and Virus Resistance. Viruses 7:3392–3419.

Sim, S.-C., Durstewitz, G., Plieske, J., Wieseke, R., Ganal, M. W., Van Deynze, A., Hamilton, J. P., Buell, C. R., Causse, M., Wijeratne, S., et al. (2012). Development of a large SNP genotyping array and generation of high-density genetic maps in tomato. PLoS One 7:e40563.

Tomato Genome Consortium (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485:635–641.

Veillet, F., Perrot, L., Chauvin, L., Kermarrec, M. P., Guyon-Debast, A., Chauvin, J. E., Nogué, F., and Mazier, M. (2019). Transgene-free genome editing in tomato and potato plants using Agrobacterium-mediated delivery of a CRISPR/Cas9 cytidine base editor. Int. J. Mol. Sci. 20.

Vicente, M. H., Zsögön, A., de Sá, A. F. L., Ribeiro, R. V., and Peres, L. E. P. (2015). Semi-determinate growth habit adjusts the vegetative-to-reproductive balance and increases productivity and water-use efficiency in tomato (Solanum lycopersicum). J. Plant Physiol. 177:11–19.

Supplementary Data 1 Genotypes of co-edited *eIF4E1*-mutant *ALS1*-mutant plants generated in Kuroiwa et al. (2023) in *eIF4E1*, *ALS1*, and *ALS2*.

	eIF4E1	***ALS1***				***ALS2***				**PVY-N605***
Plant lines from Kuroiwa et al.(2023)	Amino acid	C-20	C-14	C-13	Amino acid	C-20	C-14	C-13	Amino acid	3wpi
WVa106	WT	C	C	C	P186 (WT)	C	C	C	P184 (WT)	**S**
G110E+A111T#1	G110E+A111T	C	T	C	P186S	G	C	C	Q182E	**S**
G110E+A111T#2	G110E+A111T	C	T	C	P186S	Del9			Del184-186	**S**
P69T/G110E+A111T #1	P69T/ G110E+A111T	G	G	C	Q184E, P186A	Del5			189stp	**R**
P69T/G110E+A111T #2	P69T/ G110E+A111T	G	G	C	Q184E, P186A	G	C	C	Q182E	**R**
NIL-pot1	*pot-1*-coded 8 aa substitutions	-	-	-	-	-	-	-	-	**R**

eIF4E1 genetic mutation is abbreviated in this table. Genomic sequences of the *ALS1* and *ALS2* of NIL-pot1 line are not verified. Bolded were used as parental lines to generate F2 populations.

Supplementary Data 2 Segregation of the *ALS1* and *eIF4E1* alleles

A

Alleles:
A=*ALS1* WT
E=*eIF4E1* WT
a=*ALS1* mutant
e=*eIF4E1* mutant

Genotype:
(*ALS1*/*eIF4E1*)
w=WT
m=homozygous mutant
het=hetero-zygous

Parent: WVa106 (AAEE) x co-edited mutants (aaee)				
F1: heterozygous (AeEe)				
	0.35 AE	0.15 Ae	0.15 aE	0.35 ae
0.35 AE	w/w	w/het	het/w	het/het
	0.12	0.053	0.053	0.12
0.15 Ae	w/het	**w/m**	het/het	het/m
	0.053	0.023	0.023	0.053
0.15 aE	het/w	het/het	m/w	m/het
	0.053	0.023	0.023	0.053
0.35 ae	het/het	het/m	m/het	m/m
	0.12	0.050	0.050	0.12

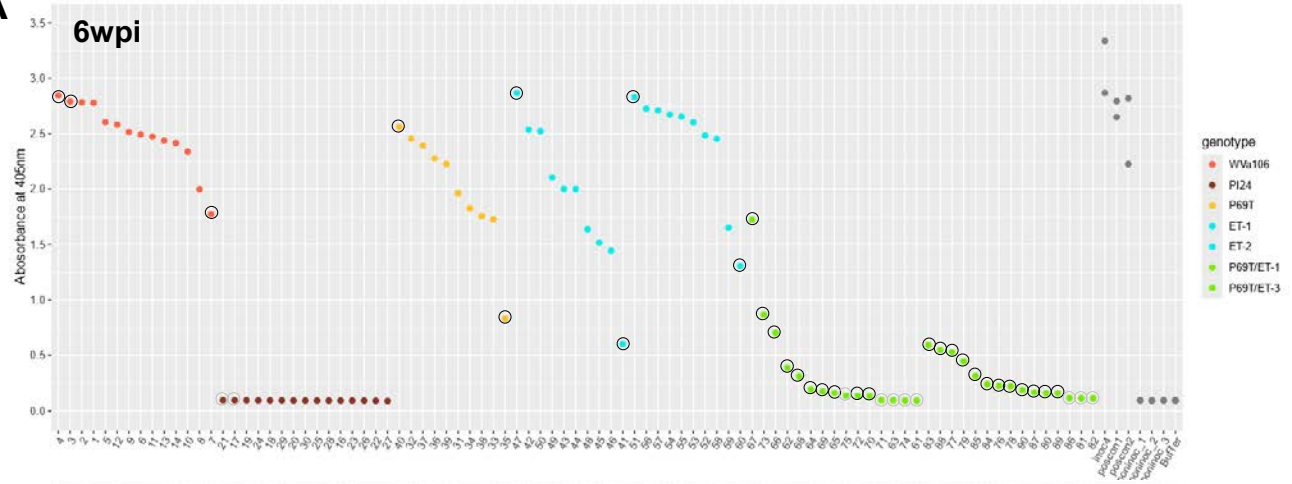
B

Parent of F2	ET-1		ET-2		P69T/ET-1		P69T/ET-2		Estimate in 88 individuals
No. of genotype	38	37	46	45	34	36	77	78	
het/w	2	7	6	2	15	12	12	7	9
m/het	11	4	4	7	9	12	3	11	9
w/het	1	5	2	8	8	10	10	10	9
het/m	5	7	7	4	9	14	11	14	9
w/m	**3**	**1**	**2**	**0**	**5**	**3**	**2**	**3**	**2**
m/w	4	1	1	2	5	3	5	7	2
w/w	3	4	3	4	3	2	7	9	11
m/m	6	3	3	5	5	5	5	5	11
het/het	13	8	11	8	22	26	29	22	26
unidentified	0	0	1	0	7	1	4	0	-
Total number	48	40	48	40	88	88	88	88	88
Percentage of w/m	6.3%	2.5%	4.2%	0.0%	5.7%	3.4%	2.3%	3.4%	2.3%

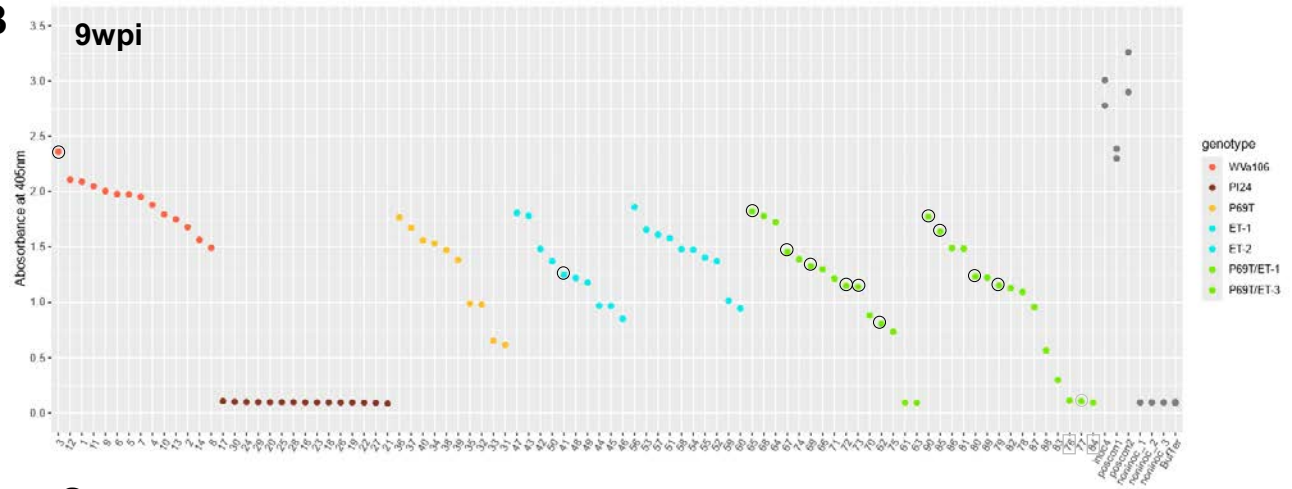
A Punnett square providing the theortical ratio of segregaation in the F2 progenies of WVa106 (wild-type) x co-edited mutant as parents. The possibility of obtaining the *ALS1* wild-type, *eIF4E1* homozygous mutant (w/m) in F2 population was calculated based on the estimated recombination frequency: 30% between the *eIF4E1* and *ALS1.* WT: wild-type. **B** Summary of *ALS1-eIF4E1* genotypes detected by KASP assay. Unid.: unidentified by KASP. The desired genotype of *ALS1* wild-type, *eIF4E1* homozygous mutant (w/m) is shaded in grey.

Supplementary Data 3 Viral CP accumulation in individual plants during the long-term infection assay #1 (corresponding to the assay in Figure 4)

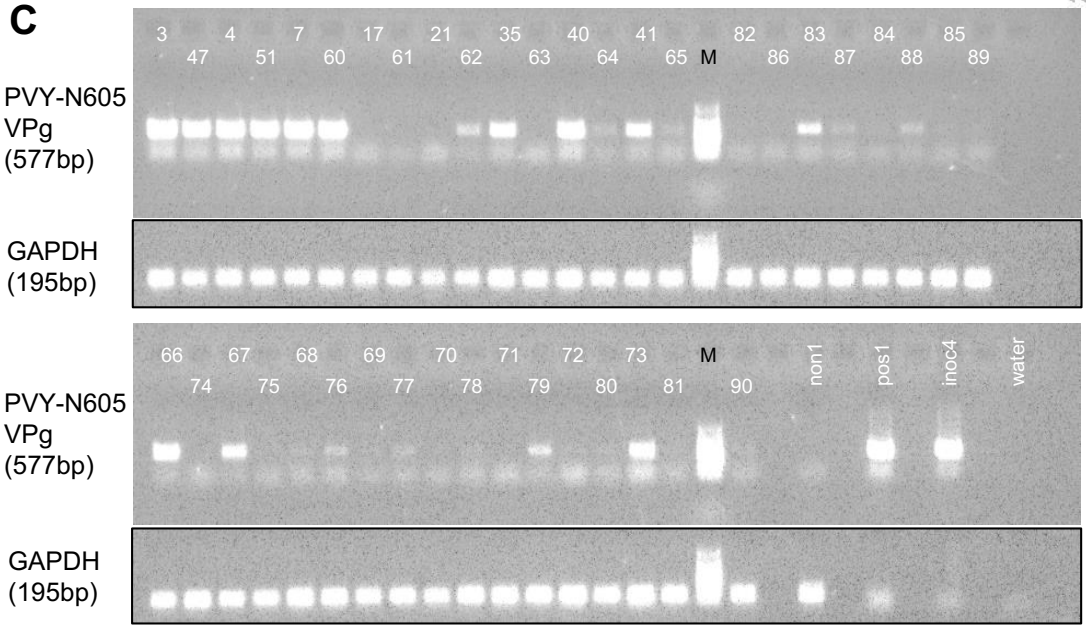
A



B



C

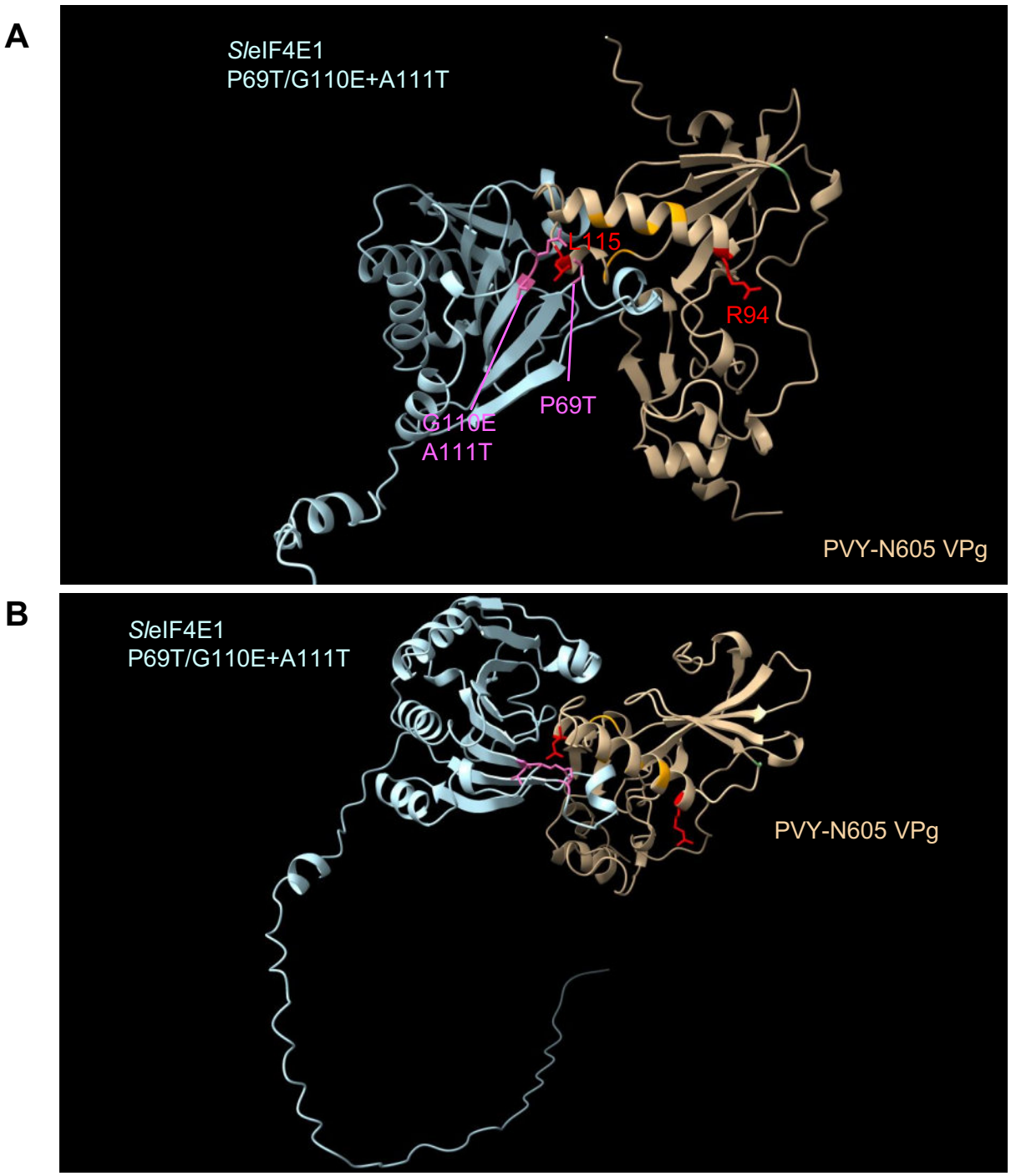


D

		Mutations in VPg
	Plant no.	**6 wpi**
P69T/ET-1	64	None
	66	R94G (double)
	68	R94G (double)
	70	L115S

		Mutations in VPg
	Plant no.	**6 wpi**
P69T/ET-3	78	None
	83	None
	87	None
	88	None
	89	None

A,B the viral coat protein (CP) accumulation in individual plants measured by ELISA at 6 and 9 wpi, respectively, are displayed in descending order by mutant lines. The colors are according to the *eIF4E1* allele. Blue: G110E+A111T and green: P69T/ G110E+A111T. WVa106 is the susceptible wild-type, PI24 the resistant accession PI247087 of *S. habrochaites.* Inoc4: inoculum at same wpi, poscon: positive control, noninoc: mock-inoculated plant, Buffer: ELISA buffer. The open circles indicate the samples that were subjected to RT-PCR and/or VPg sequencing. Black are those that amplified in RT-PCR to a detectable level and grey are those that did not. **B** #76 and #84 in grey square were analyzed by RT-PCR at 6 wpi but not at 9 wpi due to the lack of CP accumulation. **C** RT-PCR was performed on the RNA extraction from ELISA extracts of individual plants. The numbers correspond to the individual plants whose viral accumulation is shown in A. non1: mock-inoculated plant, pos1: positive control 1, inoc4: inoculum at 6wpi of the assay, water as the negative control. Amplification of partial tomato *GAPDH* (*glyceraldehyde 3-phosphate dehydrogenase*) gene is shown as the internal control. Note that this gene is not amplified from the « inoc » or « pos » as these samples are from *Nicotiana benthamiana*. **D** Amino acid substitutions observed in the VPg sequence detected from the infected P69T/G110E+A111T individual plants in assay #1. These plants were only analyzed at 6 wpi and not followed for 9 wpi. ELISA extracts used for viral CP detection in Figure 4 were subjected to RNA extraction and the PVY-N605 VPg sequence was amplified and sequenced by RT-PCR. Plant number corresponds to the individual plant numbers in Supplementary Data 5. Open circles are analyzed by RT-PCR. Broken-line circle did not amplify a complete cDNA sequence but a fragment. Grey circles were not amplified by RT-PCR.



Supplementary Data 4 Location of amino acid substitutions identified in the VPg sequence amplified from the infected P69T/G110E+A111T individual plants. Positions of the detected mutations on PVY-N605 VPg (gold) in complex with tomato eIF4E1 protein (light blue) harboring P69T and G110E+A111T substitutitons (pink). The WT VPg-*Sl*eIF4E1 complex structure was modeled using ColabFold. Colors correspond to the VPg mutations listed in Figure 5A and Supplementary Data 5A. **A**: another angle of the zoomed VPg-eIF4E1 mutant complex in Figure 5B. The unstructured N-terminus of *Sl*eIF4E1 is omitted from this figure. **B**: the entire VPg-*Sl*eIF4E1 P69T/G110E+A111T complex including the unstructured N-terminus of the *Sl*eIF4E1. The view from the same angle as the Figure 5B.

Supplementary Data 5 Amino acid substitutions in VPg detected from the long-term infection assay #2

A

		Mutations in VPg	
	Plant no.	**6 wpi**	**9 wpi**
P69T/ET-1	61	None	No accumulation
	63	S101G^{double}	No accumulation
	64	L115M^{double} + N121H^{double}	L115M^{double} + N121H^{double}
	65	N88 (silent)^{double}	R94G^{double}
	66	None	R94G
	67	R94G	R94G
	68	N88 (silent)^{double} + R94G	n.t.
	70	L115M	n.t.
	71	N88 (silent) + R94G	No accumulation
	73	L115M^{double}	No accumulation
	75	R94G^{double} + R99H^{double}	R94G^{double} + R99H^{double}
P69T/ET-3	76	L115S	n.t.
	78	L115S^{double}	No accumulation
	79	L115M	n.t.
	82	None	R94G^{double} + L115S^{double}
	83	L115S	L115S
	84	None	R94G
	85	K105M^{double}	K105M^{double} + N121H
	86	N88 (silent)^{double} + L115S^{double}	N88 (silent)^{double} + L115S/M^{double}
	88	L115S	n.t.
	90	L115S	n.t.

B

aa substitutions detected in this study	Previous reports
N88 silent	Not reported
R94G	Not reported
R99H	Not reported
S101G	PVY-SON41g (Ayme et al. 2006)
K105R	PVY-SON41(p) (Moury et al. 2004)
L115S	PVY-N605 RB variants (Lebaron et al. 2016)
L115M	PVY-LYE84 and LYE84.2 (Moury et al. 2004)
S120I	Other aa substitution at S120 (Michel et al. 2019)
N121H	PVY-SON41(p) (Moury et al. 2004)

A: Amino acid substitutions were identified in the VPg sequence amplified from the infected P69T/G110E+A111T individual plants in the assay #2. ELISA extracts used for viral CP detection in assay #2 were subjected to RNA extraction and the PVY-N605 VPg sequence was amplified and sequenced by RT-PCR. Plant number corresponds to the individual plant numbers in Supplementary Data 6. The amino acid substitutions in VPg detected Mutations that appeared after 6wpi is in red. **B**: Summary of all the amino acid substitutions in VPg identified in the long-term infection assays and the previous reports on the polymorphism itself or the site of polymorphism. Grey shaded are the substitutitons located in the central region of the VPg sequence and underlined are those detected from only one plant. PVY-SON41(p) was designated as PVY-SON41 in Moury et al. (2004) but later renamed as PVY-SON41p. N88 silent is a site of silent mutation not of amino acid substitution included in the table.

Supplementary Data 6: Table of edited *SlelF4E1* genotypes that are currently phenotyped at a greenhouse scale.

Mutant line	*Sl*eIF4E1	*Sl*ALS1	*Sl*ALS2	Number of plants	Group
WVa106 (WT)	WT	WT	WT	6 each	A & B
ET-1	G110**E**+A111**T**	WT	WT	6	A
ET-2	G110**E**+A111**T**	WT	WT	6	B
P69T/ET-1	**P69T**/G110**E**+A111**T**	WT	WT	6	A
P69T/ET-3	**P69T**/G110**E**+A111**T**	WT	WT	6	B

List of *SlelF4E1-edited*, *ALS*-edition-free mutant lines that are subjected to a whole-plant phenotyping at a greenhouse scale in spring 2024. The expected amino acid mutations due to the genetic mutations verified in each plant are listed. The plants are progenies of the mutants obtained in this work (cf. **Figure 1B,C**). Mutant line names correspond to the name of the mutants which were selfed to obtain these lines. Groups are according to the two distinct rows of double-gutter in the greenhouse which are watered separetely. Six plants are divided into two groups of three and randomly placed on both sides of the double-gutter.

Chapter 4: Assessment the impact of co-editing the *Acetolactate synthase* reporter gene on *Solanum lycopersicum* development

Potential list of authors for future publication:
Kyoka Kuroiwa[1], Anaë Laurens[1], Marie-Noëlle Corre[1], Raphaël Lugan[2], Marianne Mazier[1], and Jean-Luc Gallois[1]

[1]INRAE, GAFL, F-84143 Montfavet, France

[2]Avignon Université, 301 rue Baruch de Spinoza, 84140 Montfavet, France

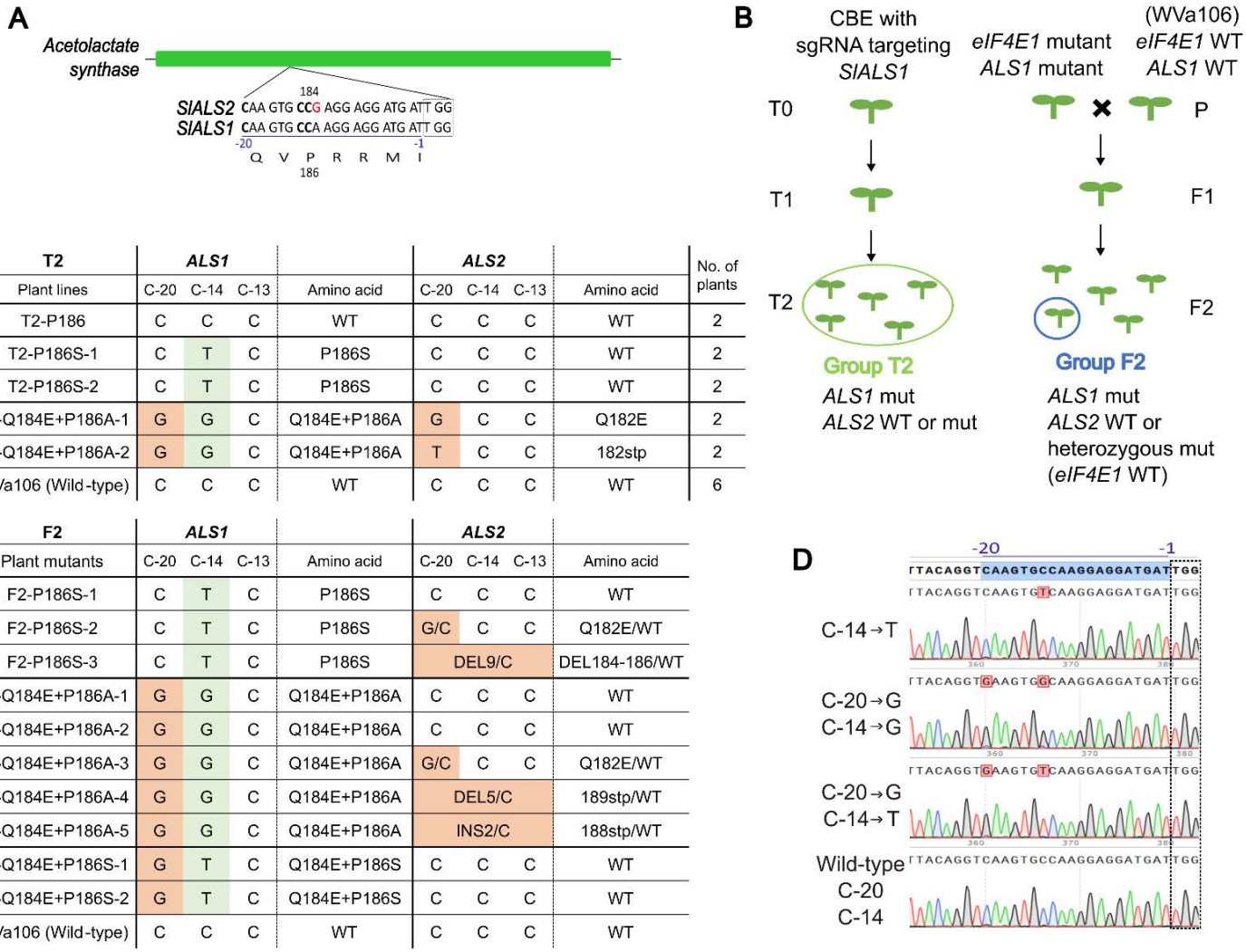
Introduction

New CRISPR/Cas-based technologies such as base-editing and prime-editing allow to introduce a larger range of mutations in genomes, but often involve a laborious selection step to find the desired mutants. One way to make these events more yielding is developing a co-editing strategy, where a reporter gene for the edition is targeted simultaneously with a given gene of interest. Recently, the *acetolactate synthase* (*ALS*; also known as a*cetohydroxyacid synthase* (AHAS)) genes have emerged as an efficient reporter gene of plant transformation and CRISPR activity, making an integral part of co-genome-editing strategy (Kuroiwa et al., 2023; Huang et al., 2023). In this strategy, a base editor, such as the cytosine base editor (CBE), is utilized to introduce in the *ALS* gene a point mutation that confers dominant herbicide resistance. This makes the *ALS* a suitable marker that allow tractable, positive selection of CBE-edited plants by the herbicide-containing media during regeneration of the plants.

ALS is a key enzyme in the first steps of branched-chain amino acid (BCAA) biosynthesis in chloroplast and is a target, or a site of action, of major herbicide groups to control weeds. In accordance with this, eight mutation sites associated with the herbicide resistance have been found in natural *ALS* alleles from weed species (Reviewed in Yu and Powles, 2014). Among them, namely the Proline-197 (amino acid position in Arabidopsis ALS; tomato ALS1 equivalent is P186) is one of the most commonly reported sites where at least 12 distinct substitutions are associated with resistance to an herbicide group sulfonylurea (SU); therefore, it has been suggested as a suitable base-editing target (Veillet et al., 2019) (online database http://www.weedscience.com/; accessed on 06 June 2024). Indeed, very recently, the efficiency of a genome-editing reporter system targeting the P197 equivalent codon by CBE has been demonstrated successfully in

tomato, tobacco, potato, and citrus *ALS*, implying the wide applicability of this strategy in different crop species (Huang et al., 2023). One important aspect of a reporter gene is that it should not have any side effect on the plant development. Thus far, the works on the co-editing strategy targeting the *ALS* gene and the gene of interest have focused on the efficiency of selection of the edited transformants. The potential side-effects of editing *ALS* on the plant growth have not been examined thoroughly despite the important role of the *ALS* gene in the plant physiology—BCAA synthesis—and the negative impacts reported on growth rate and development due to a P197 substitution in *Arabidopsis thaliana* and *Lolium rigidum* (Purrington and Bergelson, 1999; Yu et al., 2010).

Here, we evaluated the phenotypic consequences of the *ALS* editing strategy in cherry tomato (*Solanum lycopersicum* var. *cerasiforme*) cultivar (cv.) WVa106, taking advantage of the *ALS*-edited (Veillet et al., 2019) and the *ALS*-co-edited (Kuroiwa et al., 2023) plants that we have previously obtained. We characterized phenotypes of the different *SlALS1* and/or *SlALS2* genotypes generated by CBE. We show that the use of CBE generates a variety of *SlALS1* and *SlALS2* genotypes harboring mutation at the target P186 codon in *SlALS1* as well as at additional cytosine in Q184 codon within the edition window of CBE. Through whole-plant phenotyping, we demonstrate that the desired *Sl*ALS1-P186 mutant genotype grows as well as the wild-type WVa106, whereas the genotypes with additional mutations display a reduced vegetative growth. We also show that the genotype which is most negatively impacted on plant growth has indeed a metabolomic profile that is drastically altered from the wild-type plants. Altogether, this work verifies the advantage of the co-editing strategy targeting the P186 codon in *SlALS1* gene in cherry tomato without phenotypic effects. Simultaneously, this work addresses the potential drawback brought about by the mutations outside of the *Sl*ALS1-P186 site during the co-editing strategy that may lower the efficiency of the editing experiments as well as the vigor of the resulting crops, suggesting guidelines for the use of ALS.



C

T2	ALS1				ALS2				No. of plants
Plant lines	C-20	C-14	C-13	Amino acid	C-20	C-14	C-13	Amino acid	
T2-P186	C	C	C	WT	C	C	C	WT	2
T2-P186S-1	C	T	C	P186S	C	C	C	WT	2
T2-P186S-2	C	T	C	P186S	C	C	C	WT	2
T2-Q184E+P186A-1	G	G	C	Q184E+P186A	G	C	C	Q182E	2
T2-Q184E+P186A-2	G	G	C	Q184E+P186A	T	C	C	182stp	2
WVa106 (Wild-type)	C	C	C	WT	C	C	C	WT	6

F2	ALS1				ALS2			
Plant mutants	C-20	C-14	C-13	Amino acid	C-20	C-14	C-13	Amino acid
F2-P186S-1	C	T	C	P186S	C	C	C	WT
F2-P186S-2	C	T	C	P186S	G/C	C	C	Q182E/WT
F2-P186S-3	C	T	C	P186S	DEL9/C			DEL184-186/WT
F2-Q184E+P186A-1	G	G	C	Q184E+P186A	C	C	C	WT
F2-Q184E+P186A-2	G	G	C	Q184E+P186A	C	C	C	WT
F2-Q184E+P186A-3	G	G	C	Q184E+P186A	G/C	C	C	Q182E/WT
F2-Q184E+P186A-4	G	G	C	Q184E+P186A	DEL5/C			189stp/WT
F2-Q184E+P186A-5	G	G	C	Q184E+P186A	INS2/C			188stp/WT
F2-Q184E+P186S-1	G	T	C	Q184E+P186S	C	C	C	WT
F2-Q184E+P186S-2	G	T	C	Q184E+P186S	C	C	C	WT
WVa106 (Wild-type)	C	C	C	WT	C	C	C	WT

Figure 1 *SlALS1-edited* plants were obtained from the progenies of the simple base-editing and co-base-editing.

A The genomic region targeted for editing in the *SlALS1* sequence and the highly homologous sequence in *SlALS2*. Blue line indicates the nucleotide sequence of the single guide RNA (sgRNA) and the numbers in blue indicate the nucleotide positions from the protospacer adjacent motif (PAM) site. The PAM is in dotted line rectangle. The position of the target codon Pro-186 in *SlALS1* is indicated by "186" below and its equivalent Pro-184 in *SlALS2* by "184". The single mismatch in *SlALS2* compared to the *SlALS1* target sequence is in red. The three targeted cytidines (at position C-20, C-14, and C-13) are bolded. **B** The schematic of how the *SlALS1*-only-edited and *SlALS1* and *SlALS2* edited populations were obtained. Two methods were taken. One is by obtaining the T2 generation of the *SlALS*-edited population generated by CBE guided by the sgRNA in **A** (GroupT2). Another is by segregation in F2 populations of the co-edited population « *eIF4E1* mutant *ALS1* mutant » (Kuroiwa et al. (2023) and cf. *chapter 3*) (Group F2). The co-edited population was backcrossed with the wild-type WVa106 to obtain the F1 seeds. The F1 population was selfed to obtain the F2 population. **C** The tables of *SlALS1*-edited mutant genotypes obtained from the strategy depicted in **B**. (Top) Group T2. (Bottom) Group F2. The expected amino acid mutations due to the genetic mutations verified in each plant are listed. Heterozygosity of mutations in *SlALS2* gene is described as « mutation/C » and in protein as « mutation/WT ».

Results and Discussion

A single guide RNA designed to target SlALS1 gene generates various nucleotide substitutions in SlALS1 and SlALS2 in simple base-editing and co-base-editing.

Previous works on tomato *ALS* base-editing, both in simple and co-editing strategies, have been performed using a common single guide RNA (sgRNA) targeting the *SlALS1* P186-flanking sequence: CAAGTGCCAAGGAGGATGA, adjacent to a protospacer adjacent motif (PAM) (Veillet et al., 2019; Huang et al., 2023). The target sequence includes three cytosines within the edition window of the CRISPR/nCas9-fused CBE, namely -21 to -13 positions upstream of the PAM sequence (Nishida et al., 2016; Shimatani et al., 2017). Two out of three (at C-14 and C-13 positions) correspond to the P186 codon that is targeted, whereas the third cytosine at C-20 corresponds to the Q184 at which no herbicide resistance is associated (**Figure 1A**). It is to be noted that the target region is reported to be highly homologous with only a single mismatch to the *SlALS2* sequence flanking P184 codon, the *Sl*ALS1 P186 equivalent, making the *SlALS2* strong potential off-target site (Veillet et al., 2019).

Here, we re-analyzed two sets of T-DNA-free progeny populations of edited cherry tomato plants that have been previously produced, with a particular focus on the genomic sequences of *SlALS1* and *SlALS2*. One group consisted of the T2 progeny of the plants that were generated through the simple *ALS*-editing carried out using the same sgRNA in Veillet et al. (2019) (hereafter called "Group T2") (**Figure 1B "Group T2"**). To obtain another group of a distinct origin, we analyzed a set of independent plants resulting from the co-base-edition of *SlALS1* and *SleIF4E1*: we reasoned that these previously isolated *SleIF4E1*-edited plants (Kuroiwa et al., 2023) would also harbor mutations in *SlALS1* due to the linkage between the *SlALS1* locus and the *SleIF4E1* locus, both on the tomato chromosome 3. As expected, the co-edited plants contained various mutations in the *SlALS1* target region by the same sgRNA, which conferred chlorsulfuron resistance. In addition, some plants also contained either nucleotide substitution or an insertion or a deletion (indel) in the *SlALS2* in the homologous region. These plants were backcrossed to WT WVa106 and selfed to generate F2 populations, and the plants that were edited in *SlALS* gene(s) but not in the *SleIF4E1* (hence WT *SleIF4E1* allele) were selected for further experiments (hereafter referred to as "Group F2") (**Figure 1B "Group F2"**). In order to primarily characterize the mutant genotype of the main target *SlALS1* for their phenotypic aspects, homozygous *SlALS2* mutants were not included in the Group F2.

As attested by the genomic sequence amplification and sequencing, the different combinations of edited *SlALS1* and *SlALS2* alleles were obtained in the two groups (**Figure 1C**). Both groups included genotypes with the *SlALS1* alleles that were only edited at C-14 within the P186 codon and that were also edited at C-20 in the Q184 codon (**Figure 1C, D**). Some of the Group T2 and F2 mutants also contained substitution or deletion in *SlALS2* allele in the target region. These edition profiles in Group T2 are as previously reported in the simple *ALS*-editing experiments for the parental generation (T0) of the group (Veillet et al., 2019). Here, we similarly identified various edition profiles in the *SlALS1* and *SlALS2* genes in the Group F2, derived from the parental lines generated via *ALS*-co-editing strategy, allowing the evaluation of plant development based on two fully independent sets of plants.

In summary, three distinct *SlALS1* alleles were commonly identified from both the Group T2 and the Group F2, and an additional *SlALS1* allele was identified from the Group F2; and all of these included an amino acid substitution at the desired P186 codon sites. We also characterized more diverse, five distinct *SlALS2* alleles in total from two groups, and three of them resulted in *SlALS2* knock-out. Together, the genotyping results indicate that the common *SlALS1*-targeting sgRNA used with the CBE through stable T-DNA insertion can generate various *SlALS1* and *SlALS2* alleles in both *ALS* simple editing and co-editing experiments.

Q184E substitution in SlALS1 is associated with reduced vegetative growth but not with reproductive growth

In order to assess the potential effect of *SlALS* edition, we carried out a whole-plant analyses on both groups T2 and G2, albeit with a limited number of plants per genotype[o]. All the individuals in two groups were homozygous for the edited *SlALS1* allele. The T2 group plants were also homozygous for *SlALS2* allele, whereas the F2 group plants were heterozygous for the edition. The group T2 and F2 together allowed us to analyze the potential effect of four distinct *SlALS1* editions, with some of them in combination with the *SlALS2* edition. By using two groups of independent genome-editing experiments, we expected to obtain a robust data on the effect of *SlALS1* and *SlALS2* mutations on plant growth; especially we reckoned that in the F2 group, the potential effects on the

[o] Progenies of the T2 and F2 groups are currently being phenotyped at a larger number for each *SlALS1-SlALS2* combination (June 2024). The situation is explained at the end of the section.

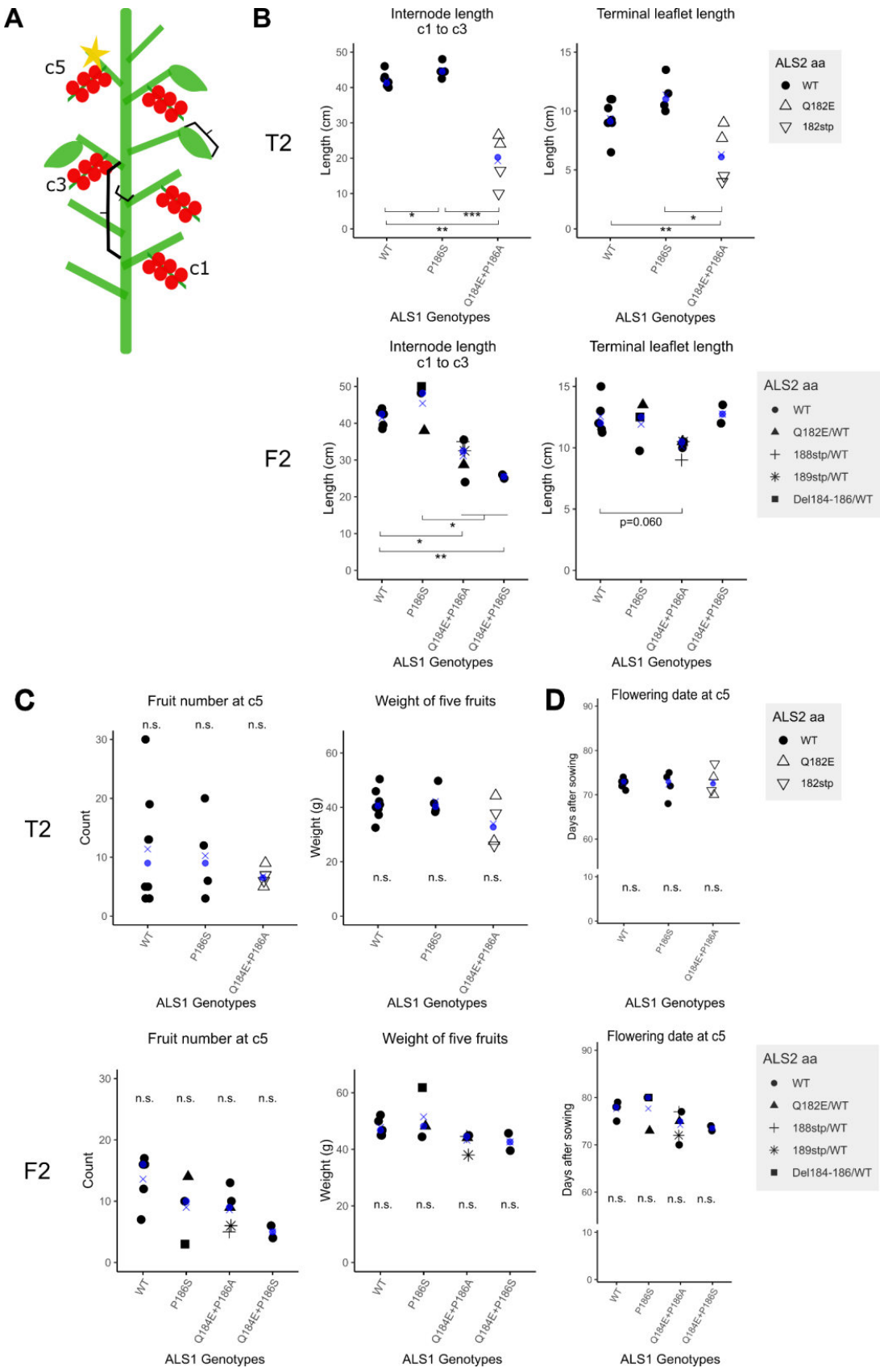


Figure 2 Substitution at Q184 in ALS1 impacts negatively the vegetative phenotypes but not reproductive phenotypes. **A** The schematic to show the positions of measurement of each parameter. c1, c3, and c5 indicate the cluster 1, 3 and 5 of flowers/fruits. **B** Measurement of vegetative parameters. Internode length between the cluster 1 and 3; and length of terminal leaflet, a leaf at the extremeties of a compound leaf, at one stem above cluster 3. **C** Measurement of reproductive parameters. Fruit number at cluster 5; and total weight of five representative fruits from each plant. **D** Flowering date recorded for the cluster 5 of each plant. One-way ANOVA test followed by Tukey HSD; or Kruskal-Wallis test was performed on each dataset. P-value $<10^{-4}$ ****;10^{-4} to 10^{-3} ***;10^{-3} to 10^{-2} **; 0.05 *. No significant difference is abbreviated. For **C-D,** n.s.: No statistically significant differences were observed. Shapes indicate the ALS2 amino acid substituttions. Blue dot is median, blue ack cross is the mean.

phenotypes from the plant transformation and regeneration would be alleviated due to the genetic recombination during backcrossing.

Representative parameters of vegetative and reproductive growth were measured to characterize the *SlALS*-edited plants, and the sites of measurement on the plant are displayed in **Figure 2A** (Vicente et al., 2015; Pons et al., 2022). As the *SlALS1* is the main target of the *ALS*-editing, the edited plants were divided into groups defined by the *SlALS1* genotype: three genotypes for T2 and four genotypes for F2.

The results show that the $ALS1^{P186S}$ genotype showed no significant differences from the wild-type in most aspects of vegetative and reproductive growth (**Figure 2,3, Supplementary Data 1**). As for the Q184E-containing genotype, i.e. $ALS1^{Q184E+P186A}$ and $ALS1^{Q184E+P186S}$, regardless of the *SlALS2* state, most vegetative growth parameters were generally reduced, while the reproductive growth parameters did not show significant differences, compared to the wild-type WVa106 and the edited P186S genotype (**Figure 2B-D, Supplementary Data 1A-C, 2**). Leaf content measurement showed decrease in the chlorophyll contents, while no obvious accumulation patterns or differences was observed for the other parameters between genotypes across the two groups (**Figure 3**, **Supplementary Data 1D**).

Overall, the phenotyping results indicate that the Q184E codon substitution in *SlALS1* due to the point mutation at C-20 position of the sgRNA is associated with reduction of vegetative growth. This may be at least partially explained by the lower chlorophyll contents, which reflect the nitrogen status and the productivity of the plants, in the same Q184E-coding *SlALS1* genotypes. In contrast, we did not observe any impacts of the ALS mutations studied here on the reproductive traits.

Due to time constraint, I am presenting the data from the preliminary phenotypic analyses carried out in spring 2023. Because the analyses of F2 were done directly on the plants after selection of different genotypes and the analyses of T2 were exploratory, the small number of plants analyzed makes it difficult to conclude on the phenotypic trends. At present (June 2024), new measurements are being carried out by an engineer school student, Anaë Laurens (INRAE GAFL). Data analyses are in progress, and I am hoping to present the complete data at my defense. The list of assayed genotypes in the ongoing assay is provided in Supplementary Data 4.

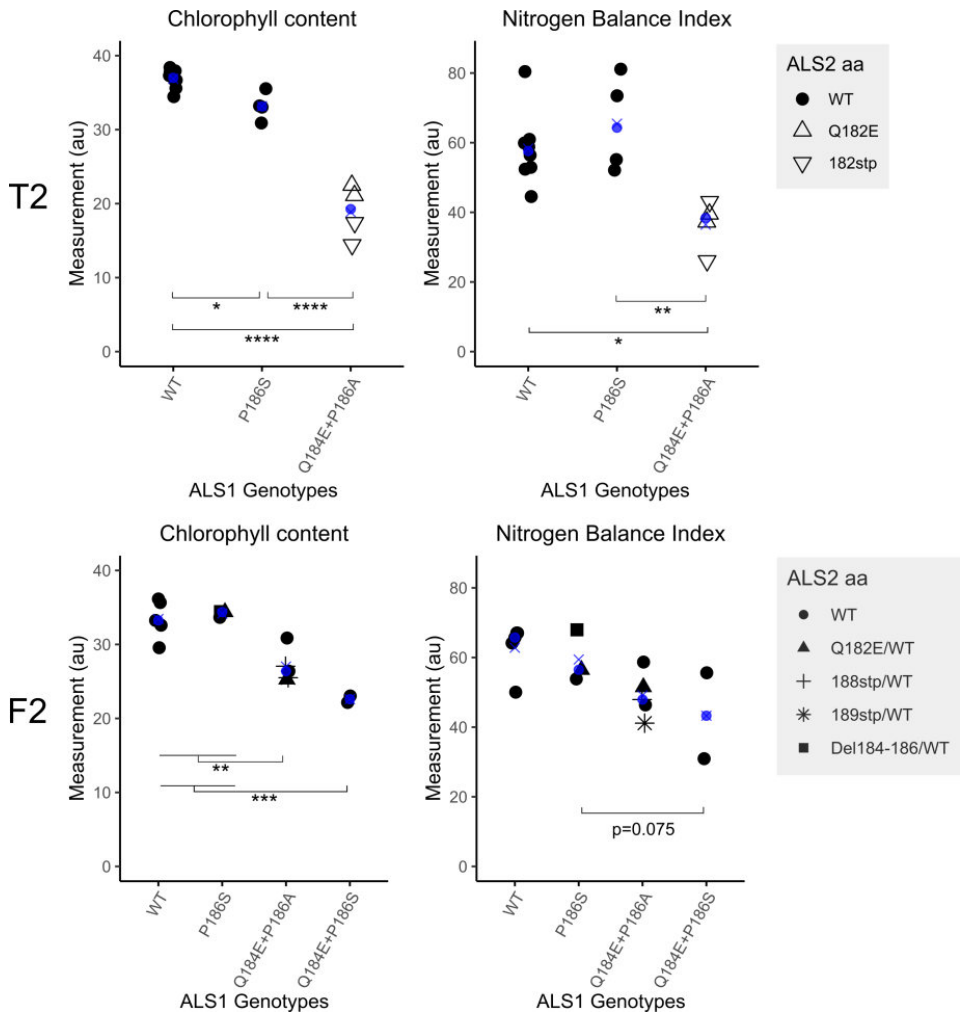


Figure 3 Substitution at Q184 in ALS1 impacts the leaf nitrogen contents negatively.
The indicators of leaf nitrogen contents: chloropyll content and Nitrogen Balance Index (NBI) were measured by an optical sensor Dualex®. au: arbitrary unit for NBI. One-way ANOVA test followed by Tukey HSD; or Kruskal-Wallis test followed by Conover-Iman test was performed on each dataset according to the data profile. P-value <10^{-4} ****;10^{-4} to 10^{-3} ***;10^{-3} to 10^{-2} **; 0.05 *. P-value < 0.1 is also indicated. No significant difference is abbreviated. Shapes indicate the ALS2 amino acid substituttions. The result is a representative of three measurements on the same leaf at a similar developmental stage (a terminal leaflet above the second to the top flower cluster). Blue dot is median, blue ack cross is the mean.

Combination of modification in SlALS1 and SlALS2 induces drastic change in the metabolomic status of plants

The phenotypic assessments indicated that specific modifications in the *SlALS* alleles reduced vegetative vigor. The different substitutions in *Sl*ALS1 have varying degrees of effects ranging from undetectable by morphological measurement ($ALS1^{P186S}$) to statistically significant reduction of growth ($ALS1^{Q184E+P186A}$ and $ALS1^{Q184E+P186S}$). As the ALS is a key enzyme in the BCAA biosynthesis pathway, there was a possibility that multiple physiological pathways were altered in the *Sl*ALS mutant genotypes, leading to the observed phenotypes in heavily edited genotypes (**Supplementary Data 3**).

Indeed, specific amino acid substitutions at P186 equivalent are reported to have impact on the ALS activity as well as on plant growth in different species, and the effects varied between plant species; for instance, the equivalent P197S mutation in ALS1 had no impact on plant growth of the monocot weed *L. rigidum*, while it reduced the growth of *Arabidopsis thaliana* (Purrington and Bergelson, 1999; Yu et al., 2010). Therefore, the $ALS1^{P186S}$ genotype that possessed minimal edition and showed similar growth to the wild-type tomato in greenhouse assay could also have some consequences at physiological level. On the other hand, there is no report on a substitution at Q184 equivalent.

Given this, we decided to assess the potential change in plant physiology of different genotypes. As a first step, we performed a preliminary metabolomics assay on the edited ALS genotype, which showed the strongest alteration in growth, to test whether the physiological effects of severe *ALS1* and *ALS2* editing could be pinpointed.

We performed the gas chromatography-electrospray ionization time-of-flight mass spectrometry (GC-EI-TOFMS) profiling on the primary metabolites in leaves from the wild-type WVa106 and the $ALS1^{Q184E+P186A}/ALS2^{182stp}$ genotype, whose vegetative growth was most affected. The $ALS1^{Q184E+P186A}/ALS2^{182stp}$ genotype corresponds to the T2-Q184E_P186A-2 line, possessing the *SlALS1* allele coding for Q184E and P186A substitutions and the *SlALS2* coding for a truncated protein at amino acid position 182 (**Figure 1C**). The GC-EI-TOFMS results show that the quantities of BCAA, namely isoleucine, leucine, and valine, were significantly modified as expected from the ALS's role in their biosynthesis: leucine and isoleucine were increased in the mutant, whereas valine was decreased (**Figure 4A**). Moreover, many other compounds showed significantly different accumulation levels between the wild-type and the $ALS1^{Q184E+P186A}/ALS2^{182stp}$

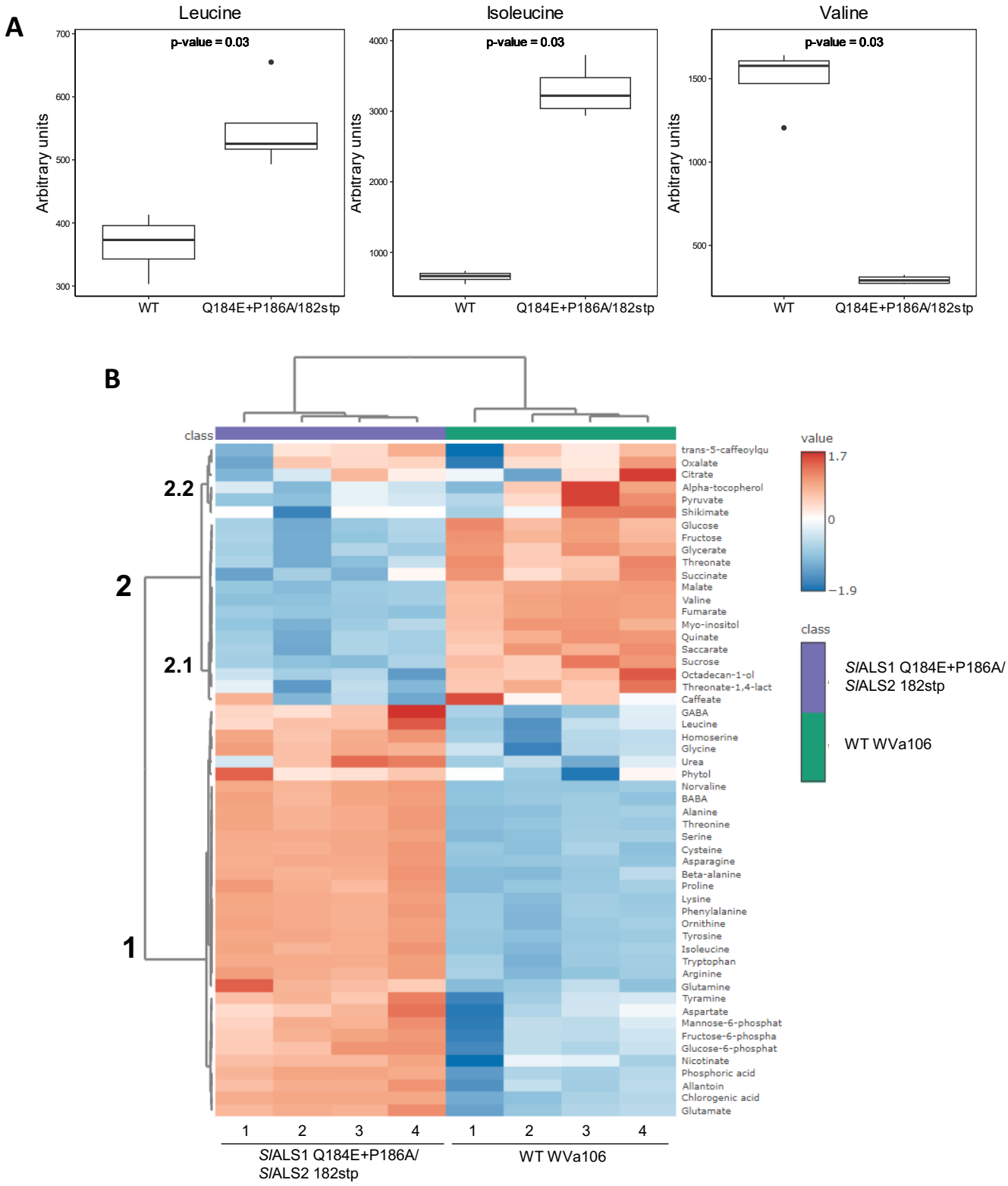


Figure 4 Metabolomic analyses show the drastic differences between the wild-type plants and the mutant genotype most affected for the plant growth (*Sl*ALS1 Q184E_P186A/*Sl*ALS2 182stp). *Sl*ALS1 Q184E_P186A/*Sl*ALS2 182stp corresponds to the T2_Q184E_P186A-2 in Figure 1C. **A** Accumulation levels of the branched-chain amino acids in each genotype. Calculated based on the peak area on the ion chromatogram, normalized by the sample mass. Four biological replicates per genotype was used. The statistical significance of difference between the two genotypes and the p-value was calculated by Mann-Whitney-Wilcoxon test. **B** Heatmap on the Log transformed/pareto-scaled data of the compound accumulation levels. Each biological replicate is displayed (number 1-4). Cluster 1 includes the compounds that were enriched and the cluster 2.1 includes the compounds that were depleted in the *Sl*ALS1 Q184E_P186A/ *Sl*ALS2 182stp genotype, compared to the wild-type WVa106. Cluster 2.2 includes the compounds that were unvaried between the two genotypes.

genotype. The mutant genotype displayed a tendency to accumulate higher amount of amino acids and lower amount of most organic acids compared to the wild-type plants (**Figure 4B**). These drastic alterations of the primary metabolite contents confirm an extensive impact of the *ALS*-editing beyond the disruption of BCAA synthesis that can explain the altered growth of the cherry tomato genotype, containing the *Sl*ALS1 with Q184E and P186A substitutions and truncated *Sl*ALS2 protein, at a greenhouse level. Further metabolomics assays on various *SlALS1/SlALS2* genotypes would determine the physiological effects of modification at each allele.

The metabolomic analyses presented in this chapter are exploratory. These results together with the preliminary phenotyping results have prompted us to carry out the same analyses on a larger panel of SlALS genotypes including those that are not visually impacted on their growth. The main question we want to address is to see whether the minimal edition on SlALS1 has any impact on metabolism. The metabolic analyses with a new set of genotypes is in progress by an assistant engineer Marie-Noëlle Corre (INRAE GAFL) and an Anae Laurens at present (June 2024). The list of assayed genotypes in the ongoing assay is provided in Supplementary Data 4.

Conclusion

This work positions itself as another complementary analysis to Kuroiwa et al. (2023) (*Chapter 2*), focusing on the co-base-editing strategy itself that was used to generate the cherry tomato genotypes in the publication. We confirm on a larger scale that the analogous sgRNA that has commonly been used for both the *ALS* simple editing and co-editing experiments in previous works can generate various *SlALS* genotypes in cherry tomato. These genotypes include those that are edited only in the *SlALS1* at the target, herbicide-resistance conferring P186 codon, and those that are additionally edited at another codon Q184 within the edition window and/or in the *SlALS2* at the equivalent, Q182 codon. We show through whole-plant phenotyping that the substitution at the target codon in *SlALS1* do not modify the plant growth behavior, while the additional substitutions in the *SlALS1*, often in combination with *SlALS2* mutation, can result in the undesirable phenotypes. However, these results will be soon completed by an assessment of potential effect of simple ALS1 edition (P186S) at the metabolomic level (currently underway). In summary, the *ALS*-co-editing is highly efficient with a risk of phenotypic

consequences due to the bystander and/or off-target mutations, which could eventually mask the true phenotypes of the edited alleles of gene of interest.

To avoid such risks, several alternatives to the experimental steps in the *ALS*-co-editing should be considered. One of them is to modulate the plant transformation/selection method. Interestingly, the variety of editing events represented in this work (based on the method in Veillet et al. (2019)) were not reported in the *SlALS*-co-editing experiments in Huang et al. (2023). In the latter, only the mutations at the *Sl*ALS1 P186 codon were reported.

Both works aimed to generate transgene-free plants by *Agrobacterium tumefaciens*-mediated transient expression of the T-DNA. Therefore, the differences in the resulting *ALS* edition profiles could be attributed to the choice of reporter gene for transfection, i.e. T-DNA transfer into the plant cell, or transformation, and hence the initial selection phase of edited plants. Veillet et al. (2019) as well as Kuroiwa et al. (2023) has first selected transfected plants by an antibiotic resistance gene and then the edited plants by herbicide chlorsulfuron; thus, the edited plants could retain the T-DNA after the first phase of selection. On the other hand, Huang et al. (2023) has selected the edited plants by chlorsulfuron and then the T-DNA-free plants by the absence of green fluorescence protein as a reporter. Inevitably, Veillet et al. (2019) analyzed the edited plants including the T-DNA-positive individuals; and the results showed that the 88% of the T-DNA positive *SlALS1*-edited plants were edited at the off-target *SlALS2* compared to the 16% in the T-DNA-free *SlALS1*-edited plants. Thus, enrichment of T-DNA-free, genome-edited events is important to minimize selection of plants edited extensively in the *ALS* genes with potential undesirable consequences. This is possible by selecting for edited events first, followed by selection based on an alternative reporter gene for T-DNA detection as demonstrated in (Huang et al., 2023).

Another method to avoid the potential negative impact of ALS co-editing is to prevent the bystander mutation on the *Sl*ALS1 by using a high-precision CBE. The phenotyping results demonstrate that the additional substitution of the Q184 codon within the edition window of the CBE used is associated with the decrease in the vegetative growth parameters. The CBE with a narrow activity window that can selectively edit a single nucleotide would be able to circumvent this unwanted mutation and only induce the target mutation (Tan et al., 2019). Moreover, a more precise tool like the prime editor—a search-and-replace strategy that allow to introduce very precisely

the desired mutation—can also bypass the bystander mutation issue and take advantage of *ALS* as a reporter without impact on plant growth (Anzalone et al., 2019; Vu et al., 2023). Prime editing experiments, specifically those targeting the *ALS* gene, have been mostly shown in monocots so far (with one exception of potato (Perroud et al., 2022)); and they are achieved at a very low efficiency so far (Vu et al., 2023). However, the work currently carried out on Arabidopsis in our laboratory shows that a very high precision of mutation can be achieved (Milkoff et Gallois, personal communication).

Finally, posterior reversion of the edited *ALS* allele into the wild-type allele is also a possibility (Gallois and Nogué, 2023). A proof of concept of such reversion strategy after co-editing has been demonstrated for the reporter gene *adenine phosphoribosyltransferase* (*APT*) in a model plant *Physcomitrella patens* (Guyon-Debast et al., 2021). A similar strategy of de-editing the reporter gene can be implemented for the *ALS* genes to restore the wild-type phenotypes.

The *ALS* is conserved across plant species, and an analogous target region flanking the tomato P186 equivalent codon has been used in both monocots and dicots by the base-editors to confer herbicide resistance (Tian et al., 2018; Zhang et al., 2019; Huang et al., 2023). Due to its high efficiency in enriching the editing events among the mutants selected by herbicide, we expect that the co-base-editing strategy will be applied in future genome-editing works by different groups, especially in the frame of low efficient strategies as PE. In this context, the results from this work warn that the co-editing strategy with ALS should be used with a careful choice of the plant transformation/selection protocol as well as of the CBE in order to avoid the risks of obtaining growth-defected plants. With this work, we hope to provide the guidelines and precautions for more efficient use of ALS co-editing strategy in the future experiments.

Materials and methods

Plant materials

The wild-type genotype and the genetic background of the of *eIF4E1* (Solyc03g005870)-edited mutant lines were the cherry tomato (*Solanum lycopersicum* var. *cerasiforme*) cultivar (cv.) West Virginia 106 (WVa106). For whole-plant phenotyping, the seedlings were grown in a growth chamber at 24°C/18°C in the 16 hr day/8 hr night cycle for 6-8 weeks until they were genotyped and transplanted in pots in the greenhouse

environment. For the metabolomic analyses, the seedlings were grown in a growth chamber at 24°C/18°C in the 16 hr day/8 hr night cycle.

Genotyping by PCR and KASP (Kompetitive allele-specific PCR) assay

Genomic DNA was extracted from fresh tomato leaf samples according to (Fulton et al., 1995) and used for genotyping by PCR and KASP. PCR was performed on the *SleIF4E1, SlALS1,* and *SlALS2* genomic sequences using specific primers. *GAPDH* (*Glyceraldehyde-3-Phosphate Dehydrogenase*) gene was amplified as the technical control. PCR products were Sanger-sequenced by Genoscreen (Lille, France). KASP assay was performed to genotype the F2 individuals using specific KASP primers designed for *SleIF4E1* region I, *SleIF4E1* region II, and *SlALS1* at C-20 and C-14 from the PAM site (detail of the nucleotide position available in the main text). All the primer sequences and pairs used for detection are available in **Annex I**. Lecture of KASP fluorescent signal was performed on the Bio-Rad CFX-96 (California, USA).

Phenotyping in greenhouse

Representative parameters of vegetative and reproductive growth were measured to characterize the *SlALS*-edited plants (Vicente et al., 2015; Pons et al., 2022). Vegetative growth traits are the internode length between the stems of the first and third cluster ("c1" and "c3") of reproductive organs, the length of a terminal leaflet above the third cluster, and the average main stem diameter above the second cluster. These vegetative measurements were taken about 10-12 weeks after the sowing. Reproductive growth traits included the number of fruits at cluster 3 and 5, the weight of five representative mature fruits from cluster 3 to 5, and the number of seeds from the same five fruits. The date of first flowering at cluster 3 and 5 was also recorded to compare the timing of inflorescence, an important transition from the vegetative to reproductive growth that influences the crop production. A non-destructive measurement of leaf content was performed using an optical sensor Dualex® (Pessl instruments, Austria). Four parameters related to plant photosynthetic activity and the nitrogen/carbon allocation were measured: chlorophyll content, anthocyanin content, flavonoid content, and the Nitrogen Balance Index® (NBI; the index of N/C allocation based on the chlorophyll and flavonoid ratio) (Fan et al., 2022). Plants were not randomized due to the small number used in this assay.

Untargeted GC-EI-TOFMS

Sampling: Seedlings were grown in a controlled growth chamber (condition explained in the "Plant materials") for three weeks. One or two developed true leaves were harvested from three to five plants, at three weeks after sowing date and immediately frozen in liquid nitrogen. Four replicates per genotype were collected.

Extraction: Leaf samples were ground into a fine powder in liquid nitrogen and freeze-dried. Then, around twenty mg of dried powder were weighed exactly and extracted for 15 min at 70°C under stirring (900rpm) with 2 ml of methanol/water (80/20, v/v). The extracts were centrifuged 5 min at 5100g and the supernatants were filtered through a 0.22µm membrane filter. Then, 50µl of the filtered extract were dried under vacuum and stored at -20°C until analysis.

GC-EI-TOFMS [gas chromatography-electrospray ionization time-of-flight mass spectrometry] : The method was adapted from (Roessner et al., 2000). Samples were derivatized online before injection with a MultiPurpose Sampler (Gerstel MPS, CTC Analytics AG, Mülheim an der Ruhr, Switzerland): dry extracts were incubated in 50 µL of a pyridine solution containing 20 mg/mL of methoxyamine hydrochloride under constant shaking at 900 rpm and 80°C for 90 min. Then, 80 µL of BSTFA containing a mixture of 9 n-alkanes were added before heating for 30 min at 80°C under constant shaking at 900 rpm. Data acquisition was performed with a gas chromatograph system (7890B GC, Agilent Technologies, Santa Clara, CA, USA) equipped with a capillary column (ZB-SemiVolatiles, 34.59 m, internal diameter 250 µm, film thickness 250 µm, Phenomenex, Torrance, USA) hyphenated to a time-of-flight (TOF) mass spectrometer (Pegasus BT, Leco, Saint Joseph, Benton Harbor, MI, USA). One microliter of sample was injected in split mode (1:50) at 230°C. Helium was used as carrier gas at 0.6 mL/min. The initial oven temperature was kept at 70°C for 1 min and then increased to 320°C (9°C/min) and maintained for 10 min. The m/z scan range was 70–600 with a cycle time of 20 scans/s. Source temperature and transfer line were set at 250°C. The MultiPurpose Sampler was controlled by Maestro Version 1.4.40.1. Gerstel and gas chromatography system with mass spectrometer were controlled by ChromaTOF Version 5.20.38.0.54864 (LECO, Saint Joseph, MI, USA).

Data analysis : GC–EI–TOFMS data were deconvoluted with the LECO NTD software (LECO, St.Joseph, MI, USA), then the peak list was curated manually to remove incorrectly deconvoluted peaks and contaminants. Peak annotation was based on spectral and retention index (RI) similarity using mass spectral libraries (Golm database, NIST 2014,

Leco-fiehn rtx5). Identification level of each metabolites was determined according to criteria inspired by (Schymanski et al., 2014): level 1, confirmed structure by comparison with authentic standard, reverse match > 850, difference between retention index (RI) < 1%; level 2, probable structure, reverse match > 800 and difference between RI < 1%; level 3, tentative candidate, 600 < reverse match < 1000 and difference between RI > 1%. A specific extracted ion chromatogram (XIC) was chosen for each molecule for integration; then peak areas were normalized against the sample mass so the final dataset consisted of semi-quantitative information. The list of metabolites detected and their analytical features can be found in Supplementary Data 4.

Data visualization and statistical analysis

Visualization of the results and the statistical tests were performed on the software R. One-way ANOVA test or Kruskal-Wallis non-parametric test ("pgirmess" package) were performed to evaluate the significance of differences between genotypes according to the sample profile. As the post-hoc test, Tukey-HSD test or Conover-Iman test with p-value adjustment by Benjamini-Yekutieli method ("conover.test" package) was performed to determine the significance. The type of statistical test to be used was determined based on the homogeneity of variances (Bartlett's test) and the normality of distribution (Shapiro-Wilk test) of a given dataset. Generation of the heatmap was done using the open-source web interface MetaboAnalyst 6.0 (https://www.metaboanalyst.ca/) (setting: Log transformation; pareto scaling; Euclidian distance; Ward algorithm).

Acknowledgement

The authors thank Emmanuel Botton (INRAE GAFL) for taking care of tomato plants and Robin Thibert (INRAE GAFL) for acquisition of the preliminary data for this thesis.

References

Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., Chen, P. J., Wilson, C., Newby, G. A., Raguram, A., et al. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. Nature 576:149–157.

Fan, K., Li, F., Chen, X., Li, Z., and Mulla, D. (2022). Nitrogen balance index prediction of winter wheat by canopy hyperspectral transformation and machine learning. Remote Sens. (Basel) 14:3504.

Fulton, T. M., Chunwongse, J., and Tanksley, S. D. (1995). Microprep protocol for extraction of DNA from tomato and other herbaceous plants. Plant Mol. Biol. Rep. 13:207–209.

Gallois, J.-L., and Nogué, F. (2023). Blueprint for non-transgenic edited plants. Nat. Plants 9:1579–1580.

Guyon-Debast, A., Alboresi, A., Terret, Z., Charlot, F., Berthier, F., Vendrell-Mir, P., Casacuberta, J. M., Veillet, F., Morosinotto, T., Gallois, J.-L., et al. (2021). A blueprint for gene function analysis through Base Editing in the model plant Physcomitrium (Physcomitrella) patens. New Phytol. 230:1258–1272.

Huang, X., Jia, H., Xu, J., Wang, Y., Wen, J., and Wang, N. (2023). Transgene-free genome editing of vegetatively propagated and perennial plant species in the T0 generation via a co-editing strategy. Nat. Plants Advance Access published September 18, 2023, doi:10.1038/s41477-023-01520-y.

Kuroiwa, K., Danilo, B., Perrot, L., Thenault, C., Veillet, F., Delacote, F., Duchateau, P., Nogué, F., Mazier, M., and Gallois, J.-L. (2023). An iterative gene-editing strategy broadens eIF4E1 genetic diversity in Solanum lycopersicum and generates resistance to multiple potyvirus isolates. Plant Biotechnol. J. Advance Access published January 30, 2023, doi:10.1111/pbi.14003.

Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K. Y., et al. (2016). Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science 353.

Perroud, P.-F., Guyon-Debast, A., Veillet, F., Kermarrec, M.-P., Chauvin, L., Chauvin, J.-E., Gallois, J.-L., and Nogué, F. (2022). Prime Editing in the model plant Physcomitrium patens and its potential in the tetraploid potato. Plant Sci. 316:111162.

Pons, C., Casals, J., Palombieri, S., Fontanet, L., Riccini, A., Rambla, J. L., Ruggiero, A., Figás, M. del R., Plazas, M., Koukounaras, A., et al. (2022). Atlas of phenotypic, genotypic and geographical diversity present in the European traditional tomato. Hortic. Res. 9.

Purrington, C. B., and Bergelson, J. (1999). Exploring the Physiological Basis of Costs of Herbicide Resistance in Arabidopsis thaliana. Am. Nat. 154:S82–S91.

Roessner, U., Wagner, C., Kopka, J., Trethewey, R. N., and Willmitzer, L. (2000). Simultaneous analysis of metabolites in potato tuber by gas chromatography–mass spectrometry. Plant J. 23:131–142.

Schymanski, E. L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H. P., and Hollender, J. (2014). Identifying small molecules via high resolution mass spectrometry: Communicating confidence. Environ. Sci. Technol. 48:2097–2098.

Shimatani, Z., Kashojiya, S., Takayama, M., Terada, R., Arazoe, T., Ishii, H., Teramura, H., Yamamoto, T., Komatsu, H., Miura, K., et al. (2017). Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. Nat. Biotechnol. 35:441–443.

Tan, J., Zhang, F., Karcher, D., and Bock, R. (2019). Engineering of high-precision base editors for site-specific single nucleotide replacement. Nat. Commun. 10:439.

Tian, S., Jiang, L., Cui, X., Zhang, J., Guo, S., Li, M., Zhang, H., Ren, Y., Gong, G., Zong, M., et al. (2018). Engineering herbicide-resistant watermelon variety through CRISPR/Cas9-mediated base-editing. Plant Cell Rep. 37:1353–1356.

Veillet, F., Perrot, L., Chauvin, L., Kermarrec, M. P., Guyon-Debast, A., Chauvin, J. E., Nogué, F., and Mazier, M. (2019). Transgene-free genome editing in tomato and potato plants using Agrobacterium-mediated delivery of a CRISPR/Cas9 cytidine base editor. Int. J. Mol. Sci. 20.

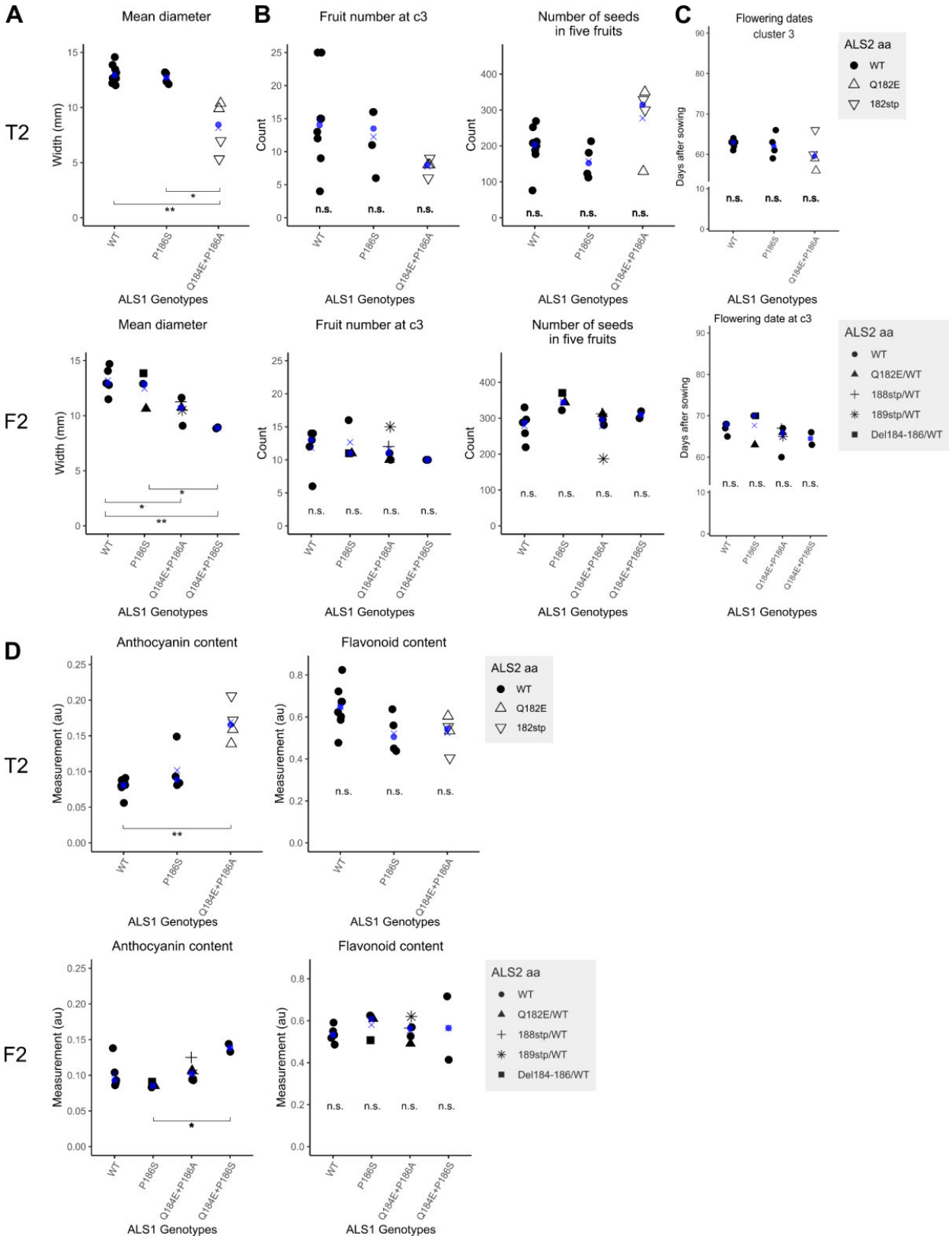
Vicente, M. H., Zsögön, A., de Sá, A. F. L., Ribeiro, R. V., and Peres, L. E. P. (2015). Semi-determinate growth habit adjusts the vegetative-to-reproductive balance and increases productivity and water-use efficiency in tomato (Solanum lycopersicum). J. Plant Physiol. 177:11–19.

Vu, T. V., Nguyen, N. T., Kim, J., Hong, J. C., and Kim, J.-Y. (2023). Prime editing: Mechanism insight and recent applications in plants. Plant Biotechnol. J. Advance Access published October 4, 2023, doi:10.1111/pbi.14188.

Yu, Q., and Powles, S. B. (2014). Resistance to AHAS inhibitor herbicides: current understanding. Pest Manag. Sci. 70:1340–1350.

Yu, Q., Han, H., Vila-Aiub, M. M., and Powles, S. B. (2010). AHAS herbicide resistance endowing mutations: effect on AHAS functionality and plant growth. J. Exp. Bot. 61:3925–3934.

Zhang, R., Liu, J., Chai, Z., Chen, S., Bai, Y., Zong, Y., Chen, K., Li, J., Jiang, L., and Gao, C. (2019). Generation of herbicide tolerance traits and a new selectable marker in wheat using base editing. Nat Plants 5:480–485.



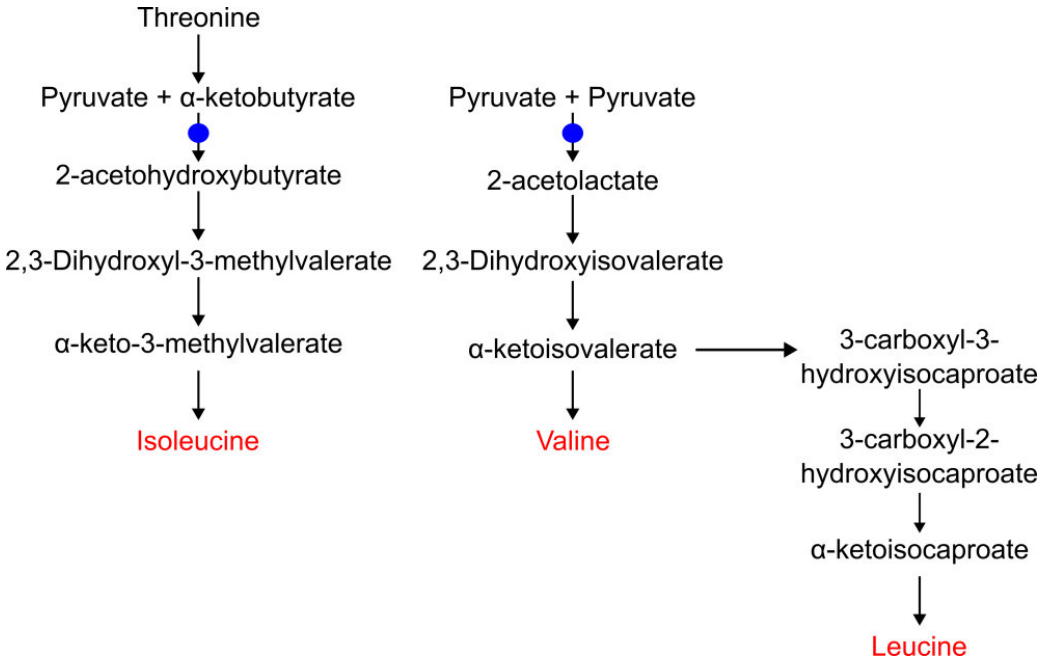
Supplementary Data 1 Additional phenotypes measured between the wild-type cv.WVa106 and the *SlALS1* and *SlALS2* edited mutants.

A Measurement of mean diameter of the two measurements of diagonal axis of main stem above the cluster 2. **B** Measurement of reproductive parameters. Fruit number at cluster 3 and the number of seeds in the same five representative fruits used for weight measurement for each plant. **C** Flowering date recorded for the cluster 3 and 5 of each plant. **D** anthocyanin content, flavonoid content measured by an optical sensor Dualex®. au: absorbance unit. The result is a representative of three measurements on the same leaf at a similar developmental stage (a terminal leaflet above the second to the top flower cluster). One-way ANOVA test followed by Tukey HSD; or Kruskal-Wallis test followed by Conover-Iman test was performed on each dataset. P-value $<10^{-4}$ ****;10^{-4} to 10^{-3} ***;10^{-3} to 10^{-2} **; 0.05 *. No significant difference is abbreviated. P-value < 0.1 is also indicated. For **B, C,** and **D** Flavonoid content, no statistically significant differences were observed in any of the parameters (denoted as n.s.) according to the one-way ANOVA test and the Kruskal-Wallis test. Shapes indicate the ALS2 amino acid substituttions. Blue dot is median, blue ack cross is the mean.



Plant lines	WVa106	F2-P186S-1	T2-P186S-1	WVa106	F2-Q184E+P186A-1	T2-Q184E+P186A-1	T2-Q184E+P186A-2	WVa106	F2-Q184E+P186S-1
ALS1	WT	P186S	P186S	WT	Q184E+P186A	Q184E+P186A	Q184E+P186A	WT	Q184E+P186S
ALS2	WT	WT	WT	WT	WT	Q184E	182stp	WT	WT

Supplementary Data 2 Substitution at Q184 in *Sl*ALS1 in combination with *Sl*ALS2 mutation impacts negatively the plant growth rate at 50 days after sowing. Below the mutations of each plant line are indicated. Red line indicates the 100 cm from the ground. These plants were not used for the phenotyping.



Supplementary Data 3 Acetolactate synthase (ALS) is involved in the first step of branched amino acid synthesis pathway. The step where ALS is the key enzyme is indicated by the blue circle. Figure created based on Kochvenko et al. (2012) and Singh and Shaner (1995)

	ALS1				*ALS2*					
Plant lines	C-20	C-14	C-13	Amino acid	C-20	C-14	C-13	Amino acid	No of plants	Group
F3-P186S-1	C	T	C	WT_P186S	C	C	C	WT	6	A
F3-P186S-2-ALS2wt	C	T	C	WT_P186S	C	C	C	WT	6	B
F3-P186S-2	C	T	C	WT_P186S	G	C	C	Q182E	6 & 6	A & B
F3-Q184E_P186A-1	G	G	C	Q184E_P186A	C	C	C	WT	6	B
F3-Q184E_P186A-2	G	G	C	Q184E_P186A	C	C	C	WT	6	A
F3-Q184E_P186A-3	G	G	C	Q184E_P186A	G	C	C	Q182E	6 & 4	A & B
F3-Q184E_P186S-1	G	T	C	Q184E_P186S	C	C	C	WT	6	A
F3-Q184E_P186S-2	G	T	C	Q184E_P186S	C	C	C	WT	6	B
F3_als2_Q184E	C	C	C	WT (P186)	G	C	C	Q182E	4 & 4	A & B
T2-Q184E_P186A-2	G	G	C	Q184E_P186A	T	C	C	182stp	6	B
WVa106 (Wildtype)	C	C	C	WT (P186)	C	C	C	WT (P184)	6 each	A & B

Supplementary Data 4 Table of edited *SlALS1* and *SlALS2* genotypes that are currently phenotyped at a greenhouse scale.
A list of *SlALS-edited*, *SleIF4E1*-edition-free mutant lines that are subjected to a whole-plant phenotyping at a greenhouse scale in spring 2024. The expected amino acid mutations due to the genetic mutations verified in each plant are listed. The plants are progenies of the mutants obtained in this work (cf. **Figure 1B,C**). Mutant line names correspond to the name of the mutants which were selfed to obtain these lines. Groups are according to the two distinct rows of double-gutter in the greenhouse which are watered separetely. Rows with red line marks the new set of genotypes selected for the metabolomic analyses. Six plants are divided into two groups of three and randomly placed on both sides of the double-gutter.

Normalized peak areas

		Peak area (test sample 15mg) (blue: low; red: high)								Mass spectrometry results								
		AHA1	AHA2	AHA3	AHA4	WA5	WA7	WA8	WA9	Id	ion	Proba	Reverse	RT sec	RT min	RI	RI theo.	Delta
Amino Acid	Valine	268	306	274	323	1205	1641	1595	1559	EINIST_122001-101-21_METB_1208.8_Valine, DL- (2TMS)	144	62,8	820	462	7,70	1210	1209	0,12%
	Leucine	493	525	526	655	356	390	413	303	EIQTMS_129002-101-9_METB_1266.2_Leucine, DL- (2TMS)	158	33	897	513	8,55	1269	1266	0,22%
	Isoleucine	3368	3073	2937	3799	639	690	739	551	EINIST_132002-101-19_METB_1288.6_Isoleucine, DL- (2TMS)	158	33,7	883	533	8,88	1291	1289	0,16%
	Proline	15537	10565	12360	15571	1249	1555	1737	1468	EIQTMS_132003-101_METB_1298_Proline, DL- (2TMS)	142	19,1	905	537	8,95	1296	1298	0,15%
	Serine	85696	89219	86062	95757	7630	10791	10791	8462	EIQTMS_138001-101-13_METB_1353.3_Serine, DL- (3TMS)	204	78,2	910	592	9,86	1359	1353	0,43%
	Threonine	13095	12166	11319	14363	1793	2173	2070	1954	EINIST_140001-101-20_METB_1379.4_Threonine, DL- (3TMS)	218	50,7	932	614	10,23	1384	1379	0,37%
	Phenylalanine	4491	3892	4161	5291	302	335	321	217	EIQTMS_164001-101_METB_1629.3_Phenylalanine, DL- (2TMS)	218	78,6	888	812	13,53	1627	1629	0,14%
	Ornithine	8490	6201	7080	9130	116	116	140	64	EITTMS_162001-101-1_MST| MST_1611.1_Ornithine, DL- (3TMS)	142	82,2	892	803	13,38	1615	1611	0,25%
	Glutamic acid	41611	42868	43771	49705	15529	21054	21819	18890	EIQTMS_163001-101-6_METB_1615.4_Glutamic acid, DL- (3TMS)	246	91,7	911	807	13,44	1620	1615	0,31%
	L Alanine	52254	49820	46261	56586	9664	11388	11839	9912	L-Alanine, 2TMS derivative	116	74	913	359	5,99	1093	1091	0,21%
	BABA	2786	2604	1728	2919	51	67	50	59	EIQTMS_117002-101-1_METB_1162,2_Butyric acid, 2-amino-, DL- (2TMS)	130	78,6	876	422	7,04	1165	1162	0,27%
	Glycine_2TMS	27726	25239	20803	23243	7897	6673	7378	2755	Glycine, 2TMS derivative	102	86,3	887	377	6,28	1113		
	Phosphoric acid	23419	28925	29502	27016	1071	2567	3196	2942	EIQTMS_129001-101_METB_1264,1_Phosphoric acid (3TMS)	299	97,5	912	512	8,53	1267	1264	0,24%
	Glycine_3TMS	1413	1117	773	1487	107	134	92	89	EIMOR_133001-101-17_METB_1304,5_Glycine (3TMS)	174	71,5	904	544	9,07	1304	1304	0,02%
	GABA	577	633	545	1157	305	280	393	237	EIQTMS_133002-101-3_METB_1298,8_Butyric acid, 4-amino- (2TMS)	102	89,1	822	541	9,02	1301	1299	0,17%
	Threonic acid lactone	230	210	164	180	290	287	359	314	EINIST_140005-101-5_METB_1373,4_Threonic acid-1,4-lactone (2TMS)	116	69,5	849	604	10,07	1373	1373	0,01%
	Pyroglutamic acid	446556	471383	454854	597146	19685	25803	28093	21772	EINIST_153002-101-19_METB| METB_1519,4_Pyroglutamic acid, DL- (2TMS)	156	39,1	937	728	12,14	1518	1519	0,09%
	Threonic acid	440	378	344	489	1294	942	1334	890	EIQTMS_156001-101-5_METB_1546,4_Threonic acid (4TMS)	205	47,5	893	758	12,64	1557	1546	0,69%
	Asparagine	571	583	572	831	15	20	31	17	L-Asparagine, 2TMS derivative	156	29,8	652	790	13,17	1598		
	B Alanine	337	327	314	398	66	74	96	64	EIQTMS_119004-101-1_METB_1183,1_Alanine, beta- (2TMS)	102	67,1	734	441	7,35	1187	1183	0,30%
	Homoserine	117	117	98	126	0	0	0	0	Homoserine (3TMS)_[2]_mz 218,2	103		897	669	11,15	1447		
	Norvaline	89	177	95	162	0	0	0	0	Norvaline (2TMS)_[2]_mz 144,1	100		855	484	8,07	1236		
	Tyrosine	10102	8912	10308	12496	130	139	155	103	EITTMS_194002-101-2_METB_1932,6_Tyrosine, DL- (3TMS)	218	52,8	911	1034	17,24	1937	1933	0,20%
	Asparagine-3TMS	1683	1932	1981	2981	0	0	0	0	EIQTMS_168001-101-6_METB_1665,8_Asparagine, DL- (3TMS)	218	92,8	902	844	14,07	1669	1666	0,19%
	Lysine	977	798	883	1151	0	0	0	0	EIMOR_170007-101_METB_1699,4_Lysine, DL- (3TMS)	156	95,9	869	873	14,55	1707	1699	0,46%
	Aspartic acid 3TMS	24317	26911	25451	30992	14188	19992	21621	18630	EIQTMS_152002-101_METB_1509.8_Aspartic acid, DL- (3TMS)	232	93	775	727	12,12	1516	1510	0,41%
Amine	Urea	78	174	125	164	64	53	81	72	EIFIEHN_127002-101-9_METB_1235,6_Urea (2TMS)	171	95,1	812	485	8,09	1237	1236	0,06%
Organic acid	Pyruvic acid	47	69	45	63	54	192	144	96	EISTR_104002-101-4_MST_1037.8_Pyruvic acid (1MEOX) (1TMS)	174	87	779	312	5,20	1040	1038	0,17%
	Malic acid	6966	7480	6053	7842	40615	53881	55394	49482	EITTMS_149001-101_METB_1477.3_Malic acid, DL- (3TMS)	233	89,5	889	701	11,68	1484	1477	0,47%
	Citric acid	3841	6539	4750	5549	5060	5781	8566	3623	EIQTMS_182004-101_METB| METB_1804.6_Citric acid (4TMS)	273	93,3	875	952	15,86	1811	1805	0,33%
	Succinic acid	108	116	133	186	273	231	281	206	EINIST_134001-101-25_METB_1310.2_Succinic acid (2TMS)	247	75,7	901	552	9,21	1314	1310	0,28%
	Shikimic acid	946	948	763	944	855	1107	1109	930	EITTMS_181002-101_METB_1792.5_Shikimic acid (4TMS)	204	38,8	780	946	15,76	1803	1792	0,59%
	Saccarid acid	80	86	47	78	315	433	600	562	EIGTMS_201001-101-13_METB_2002.2_Saccharic acid (6TMS)	333	58,8	651	1079	17,98	2015	2002	0,62%
	Fumaric acid	137	130	123	116	1137	1747	1666	1570	EITTMS_137001-101-4_METB_1346,2_Fumaric acid (2TMS)	245	86,2	839	585	9,75	1351	1346	0,37%
	Glyceric acid	66	69	54	64	153	157	143	121	EINIST_135003-101-17_METB_1321,7_Glyceric acid, DL- (3TMS)	189	67,2	752	565	9,41	1328	1322	0,46%
	Quinic acid	506	541	378	517	1342	1870	1817	1594	EITTMS_185001-101-1_METB_1843,2_Quinic acid, D(-)- (5TMS)	255	45,1	867	978	16,30	1850	1843	0,35%
Terpenoid	Phytol	372	274	268	294	249	153	256	198	EIQTMS_217006-101-1_METB_2169.1_Phytol (1TMS)	143	79	646	1170	19,50	2174	2169	0,24%
Polyol	Myo-inositol	6148	6579	5411	7409	16350	20451	17763	20631	EINIST_209002-101-15_METB_2083.9_Inositol, myo- (6TMS)	305	67,9	901	1125	18,75	2095	2084	0,55%
Alcool	Glycerol 3TMS	2168	2311	2422	3484	70	81	93	81	EINIST_129003-101-17_METB_1265,3_Glycerol (3TMS)	218	56,1	780	781	13,02	1586	1265	25,38%
Carbohydrate	Fructose-6-Phosphate	186	245	226	263	43	89	99	91	EIQTMS_232002-101_METB_2292.8_Fructose-6-phosphate (1MEOX) (6TMS)	315	72,9	819	1227	20,45	2315	2293	0,96%
	Glucose-6-Phosphate	290	417	312	418	68	126	147	138	EITTMS_233002-101_METB| METB_2307.1_Glucose-6-phosphate (1MEOX) (6TMS)	387	42,5	879	1232	20,53	2329	2307	0,94%
	Fructose 1	4617	3628	2734	4763	33083	29616	22646	23419	EITTMS_188004-101_METB_1865.4_Fructose, D (1MEOX) (5TMS)	307	15,3	904	987	16,45	1862	1865	0,14%
	Fructose 2	3219	2398	1811	3274	25080	22079	16670	17485	EITTMS_188004-101-3_METB_1865.4_Fructose, D (1MEOX) (5TMS)	307	19,4	885	994	16,56	1872	1865	0,37%
	D-Glucose	2073	1800	1208	2304	19569	16222	11491	11411	EITTMS_189002-101-3_METB_1881.6_Glucose, D- (1MEOX) (5TMS)	319	6,2	886	1004	16,73	1887	1882	0,27%
	D-Mannose	189	180	140	210	2179	1877	1088	1227	EITTMS_189001-101_METB_1887.5_Mannose, D- (1MEOX) (5TMS)	319	16,3	776	1017	16,96	1908	1887	1,09%
	Sucrose	39044	36659	37107	39909	57915	67902	63734	55424	Sucrose (8TMS)_[1]_mz 361.2	361		902	1348	22,46	2646		
	Glucose 5TMS	3914	2800	1933	1243	5834	5925	3725	2775	U_Glucopyranose (5TMS)_[3]_RM 907_mz 204,1	204		836	1448	24,14			
Phenylpropanoid	3CQ-cis	45	28	39	31	0	0	0	0	EIQTMS_299001-101-3_METB_2975.1_Quinic acid, 3-caffeoyl-, cis- (6TMS)	307	31	639	1454	24,24	2981	2975	0,21%
	1CQ-trans-1	106	861	702	1138	87	1025	2452	1146	EIQTMS_340001-101-3_METB_3390_Quinic acid, 1-caffeoyl-, trans- (6TMS)	307	19,4	707	1493	24,88	3120	3390	7,98%
	1CQ-trans-2	265	2033	1766	3682	98	1653	3099	2607	EIQTMS_340001-101-3_METB_3390_Quinic acid, 1-caffeoyl-, trans- (6TMS)	307	8	753	1508	25,14			
	Chlorogenic acid	4435	4731	4771	4736	350	630	706	464	Chlorogenic acid (6TMS)_[1]_mz 345.2	307		889	1489	24,81	3105		

Supplementary Data 5 The list of metabolites detected in the GC–EI–TOFMS assay and the peak area in each biological replicate from two genotypes. Red and blue for high and low values, respectively, when two genotypes were compared. RT: retention time. RI: retention index. (theo.: theoretical)

Primer name	Sequence 5′-3′	Amplicon size	Detection
ALS1_F	GTGGTTTTGCCGTTGCCAAT	761 bp	Genomic *SlALS1*
ALS1_R	CAGCTCCTCACTTGATTGCG		
ALS2_F	GCCGTTTCACTGTTTCAAAT	761 bp	Genomic *SlALS2*
ALS2_R	CAGCTCCTCACTCGACTGTG		
4er2_F	CATGCACCTTGTTTGGTGAG	526 bp	Genomic *SleIF4E1* Region II
4er2_R	CCCGGACACTAACAACTGCT		
gapdhF	GGTTACAGTTCCCGTGTGATTGA	190 /195 bp	Genomic or cDNA *GAPDH*
gapdhR	TCAAAGATAACAACAGGGTTCCGT		
n605F	CTCACCAGGGCAAGAACAAATCCAAAAGAATT	577 bp	cDNA PVY-N605 VPg
n605R	TTGGCTTCATGCTCCACTTCCTG		

Primer name	Sequence 5′-3′	Detection
KASP-SleIF4E1R1_F	AAGCATCCATTGGAGCATTC	*SleIF4E1*-Region I
KASP-SleIF4E1R1_R1	GAAGGTGACCAAGTTCATGCTGCAGTTTGTCGAGATTTAGTGGTA	
KASP-SleIF4E1R1_R2	GAAGGTCGGAGTCAACGGATTGCAGTTTGTCGAGATTTAGTGGTT	
KASP-SleIF4E1R2_F	ATCCATCACCCAAGCAAGTT	*SleIF4E1*-Region II
KASP-SleIF4E1R2_R1	GAAGGTGACCAAGTTCATGCTGTGCTTAAAACAATGAAAGTCTGC	
KASP-SleIF4E1R2_R2	GAAGGTCGGAGTCAACGGATTGTGCTTAAAACAATGAAAGTCTGT	
KASP-ALS1C14_F	GTTAGTGGTCTTGCGGATGC	*SlALS1*-C14
KASP-ALS1C14_R1	GAAGGTGACCAAGTTCATGCTCAGTACCAATCATCCTCCTTGG	
KASP-ALS1C14_R2	GAAGGTCGGAGTCAACGGATTCAGTACCAATCATCCTCCTTGA	
KASP-ALS1C14_R3	GAAGGTCGGAGTCAACGGATTCAGTACCAATCATCCTCCTTGC	
KASP-ALS1C20_F1	GAAGGTGACCAAGTTCATGCTTCCGATTGTTGCTATTACAGGTC	*SlALS1*-C20
KASP-ALS1C20_F2	GAAGGTCGGAGTCAACGGATTTCCGATTGTTGCTATTACAGGTG	
KASP-ALS1C20_R	CATCAGTACCAATCATCCTCCTT	

Annex I List of PCR primers and KASP primers used in C*hapter 3* and *4*

General Discussion

This PhD thesis work was carried out in an aim to demonstrate development of the *eIF4E1*-based genetic resistances against potyviruses using the genome-editing tools in a cherry tomato as a model crop. This was followed by evaluation of the practical consequences of the strategies taken—potential costs of editing (i) the susceptibility gene and (ii) the endogenous *ALS* as a selectable marker—in tomato plant growth. As the experimental results have been discussed in detail in each chapter, here I will provide the general discussion of the whole thesis and the perspective for future breeding strategy.

Precise genome-editing opens a way to balancing the trade-off of growth and defense

The work of *Chapter 1* provided an yet another example of how the CRISPR/Cas9 can efficiently induce gene inactivation in a targeted manner in the *SleIF4E2* in a crop species. This part also recapitulated the limitation of the single *eif4e* KO scheme in the context of redundant susceptibility genes, in line with former results obtained by EMS mutagenesis (Gauffier et al., 2016; Moury et al., 2020). The work of *Chapter 2* then showed that introducing base substitutions in the susceptibility gene *SleIF4E1,* without impairing its function, is sufficient to enlarge the spectrum of resistance compared to the single *Sleif4e1*^{KO} allele.

This work was built upon the proof-of-concept from our team carried out in Arabidopsis, where mimicking one or more signature amino acids of the pea natural resistance *eIF4E1* allele in Arabidopsis *eIF4E1* conferred multi-isolate virus resistance (Bastet et al., 2019). Previously, in tomato, only a double KO mutant *SleIF4E1*^{KO} x *SleIF4E2*^{KO} genotype or the natural resistance allele *SheIF4E1*^{PI24}*-pot1* from the wild tomato relative *S. habrochaites* accession PI247087 (PI24) could provide resistance to PVY isolates that can use either the eIF4E1 or eIF4E2 proteins (Gauffier et al., 2016; Lebaron et al., 2016). The former was associated with a severe growth defect, while the latter possessed a functional *eIF4E1* allele. The edited P69T/G110E+A111T-coding *SleIF4E1* allele, which we present in *Chapter 2* (Kuroiwa et al., 2023) marked the first example of an artificially generated resistance against one of such PVY isolates (PVY-N605). It is to be underlined that this edited resistance eIF4E1 protein is functional and exert resistance in the presence of the susceptible form of eIF4E2. Hence, we show that the precise modifications in the eIF4E1 guided by the natural resistance can mimic the

action of the natural *SheIF4E1*^{PI24}*-pot1*-coded protein, supporting both the growth and the defense of the plant at the same time.

Next challenge: achieving resistance of wider spectrum and higher durability

However, the edited P69T/G110E+A111T-coding *SleIF4E1* allele in the *Chapter 2* could neither suppress the viral accumulation at the same level nor present as large PVY/TEV resistance spectrum as the *SheIF4E1*^{PI24}*-pot1* allele. This resistance profile indicated that the edited allele could have less durable resistance to the potyviruses in comparison to the natural allele. We corroborated this hypothesis in the supplemental study carried out as *Chapter 3*: the edited *eIF4E1*-based resistance was indeed overcome by the emergence of resistance breaking PVY-N605 isolates after six to nine weeks. These results urge us to revisit the mechanisms of potyvirus resistance by the natural eIF4E1 variant and how we can implement it in the future editing for resistance by addressing the following questions:

i. Why does the tomato natural allele *pot-1* perform so well in potyvirus resistance?

Interestingly, despite the fact that the discovery of natural resistance *eIF4E1* alleles was initially made about two decades ago, the precise, underlying cellular or molecular mechanism by which the coded eIF4E1 proteins exert dominant resistance over the susceptible homologs remain unclear. Hypotheses for the resistance mechanism of *SheIF4E1*^{PI24}*-pot1* in tomato have been made previously during the thesis of Camille Gauffier from our laboratory, and so they will be revisited briefly in this section (Gauffier, 2015; later published in Gauffier et al., 2016).

One of the hypotheses is that the *SheIF4E1*^{PI24}*-pot1* gene or coded protein may have a regulatory role on the expression or the localization of the susceptible *Sl*eIF4E2. This was supported by the observation that the susceptible tomato *Sl*eIF4E2 form accumulated three times lower than in the wild-type *S. lycopersicum* cv. M82 (Gauffier et al., 2016). Additional hypothesis is that in general, *Sl*eIF4E1 is preferentially incorporated over *Sl*eIF4E2 into the eIF4F complex and this way, the resistance eIF4E1 form can not only avoid the hijacking of itself by the viruses but also protect the other translational machinery components such as *Sl*eIF4G from viruses by making them unavailable (Kang et al., 2007; Gauffier et al., 2016). The recent study in Arabidopsis implies an important role of interplay between eIF4E isoforms and eIF4G isoforms during TuMV infection,

encouraging the resistance mechanism in light of the homeostasis of the host translational machineries (Zafirov et al., 2023). (Approaches for exploring these hypotheses are extensively discussed in (Gauffier, 2015), so they will not be discussed here.)

These hypotheses, however, have been built upon the premise that the resistance conferred by the *SheIF4E1*^{PI24}*-pot1* mostly, if not entirely, relies on the amino acid substitutions in the protein and does not consider other factors due to the genetic background of the PI24. Indeed, whether or not the *SheIF4E1*^{PI24}*-pot1* exerts its resistance in other genetic background has only been explored in a cultivated tomato *S. lycopersicum* cv. Mospomorist which was developed with several other pathogen resistance traits (Lebaron et al., 2016). It would be interesting, therefore, to pursue the effect of *SheIF4E1*^{PI24}*-pot1*-based resistance in a more susceptible background (in progress in our laboratory).

In regards to this premise, the results from *Chapter 2* on the edited P69T/G110E+A111T genotype provide the evidence that indeed a combination of amino acid substitutions in the two *Sl*eIF4E1 regions important for the potyvirus interaction provide resistance to PVY-N605 in an otherwise susceptible cherry tomato background. This resistance also shows dominant effect over the susceptible *Sl*eIF4E2. Nevertheless, the reason that the edited resistance *SleIF4E1* allele did not perform as well as the natural resistance allele could be pointed to the fact that the profile of substituting amino acids were not appropriate and/or the number of substitutions were not sufficient. Interestingly, in the edited P69T/G110E+A111T genotype, the protein accumulation level of the susceptible homolog eIF4E2 was comparable to the wild-type plants, implying that the edited eIF4E1 protein may not play a regulatory role on the eIF4E2 and this may lead to the resistance breaking as shown in *Chapter 3*, by an unknown mechanism.

Now that the wild-type, the natural resistance, and the edited *eIF4E1* alleles with different virus resistance profiles are available, it would be interesting to investigate the behaviors of these eIF4E1 variants at non-infected and infected state through analyses of protein expression and localization and elucidate the underlying mechanism of resistance.

ii. How can we improve the resistance driven by the edited *eIF4E1* allele?

In *Chapter 3*, the edited resistance P69T/G110E+A111T-coding *SleIF4E1* allele was easily overcome during the long-term infection tests unlike *SheIF4E1^{PI24}-pot-1*, implying that in the agronomic point of view, further improvement is required to provide a durable resistance. Comparison of the natural and the edited resistance provide hints for improving the edited P69T/G110E+A111T-coding *SleIF4E1* allele based on the natural resistance alleles.

One of the most probable reason for the weaker resistance conferred by the edited-*SleIF4E1* is that the substitutions that were introduced, P69T in the region I and the G110E and A111T in the region II are not sufficient in number or as efficient or appropriate as those in the *SheIF4E1^{PI24}-pot1*-coded in avoiding hijacking by the virus. Generally, efficient *eIF4E* resistant alleles among natural diversity have been shown to be associated with several non-synonymous mutations within the region I, often leading to differences in electrostatic potential changes (Robaglia and Caranta, 2006b; Bastet et al., 2019; Coutinho de Oliveira et al., 2019). On the other hand, mutations in region II only are associated with limited resistance spectrum and low durability (Lebaron et al., 2016). Indeed, *SheIF4E1^{PI24}-pot1*-coded protein harbors eight amino acid substitutions, of which three are located in the region I and one in region II (Ruffel et al., 2005; Gauffier et al., 2016). It is possible that the differences in the durable resistance stem from these differences in the number and sites of mutations and effects on surface potential in the interaction site. Therefore, for instance, more base substitutions can be introduced in the region I of the P69T/G110E+A111T-coding *SleIF4E1* allele. The edition can be done by the base-editor like CBE even though it is limited by selectivity of base. Or it can also be achieved more precisely, by improved gene targeting, i.e. the CRISPR/Cas9-mediated homology-directed recombination (HDR) (Lu et al., 2020), or the prime-editor once the techniques are established (For review Veillet et al., 2020b).

Another possibility is that the genetic background of *SheIF4E1^{PI24}-pot1* played a role in the durability of resistance against PVY-N605, as the resistance associated with major genes can be influenced by the genetic background (Reviewed in Gallois et al., 2018). For example, genetic background of the accession or QTLs is known to modulate the durability of the pepper *eIF4E1-pvr*-mediated resistance (Quenouille et al., 2014; Quenouille et al., 2016). As mentioned above, the resistance of *SheIF4E1^{PI24}-pot1* allele has been demonstrated in genetic backgrounds that had already multiple pathogen resistance alleles (Gauffier et al., 2016; Lebaron et al., 2016). In fact, another isolate PVY-LYE84 was

shown to induce massive breakdown of the *SheIF4E1*^{PI24}*-pot1*-mediated resistance in a susceptible background cv. Mospomorist by obtaining resistance breaking mutations in VPg, implying that the durability of *eIF4E*-based resistance may indeed be supported by the other factors than the eIF4Es (Gauffier, 2015). Therefore, in order to improve the durability of the edited resistance *SleIF4E1* allele, identification of the genes that play role would be interesting and designing the stackable resistance QTLs for future improvement. Evidently, this strategy is not limited to stacking the genes involved in pathogen resistance but also extended to the other agronomic traits.

Precise genome-editing for transgene-free crop breeding

Finally, *Chapter 4* provides a precaution for the use of *acetolactate synthase* (*ALS*) gene as a reporter gene for selecting the genome-edited plants. The *ALS*-co-editing is an emerging method to generate genome-editing without stable integration of a transgene. Interestingly, the *ALS* gene was one of the first genes targeted to demonstrate the base-editing when the CBE was developed (Tian et al., 2018; Zhang et al., 2019b). This was inspired by the natural *ALS* variants in weed species that exhibit new (and selectable) phenotype, herbicide resistance, by obtaining nonsynonymous mutations in particular sites. Based on this, recently the *ALS* gene was proposed as a reporter gene in the co-editing strategy to enrich the target gene editing events by selecting the *ALS*-edited plants on herbicide-containing media (Veillet et al., 2019). This also meant that the plants edited in the endogenous *ALS* could be selected without a need of a foreign reporter gene, allowing the transgene-free editing; the implementation of the strategy was successfully demonstrated in multiple crop species (Huang et al., 2023). In this thesis, CBE-edited *SleIF4E1* mutants in *Chapter 3* were generated through the co-editing strategy proposed in Veillet et al., 2019.

In *Chapter 4*, we report for the first time that this co-base-editing scheme to facilitate the CRISPR-edited plant selection may cause growth retardation depending on the base substitutions introduced in the reporter *SlALS* genes in tomato. In the previous works, the phenotypic outcomes of the *ALS*-edited plants have not been fully discussed despite the essential role of *ALS* genes in the branched-chain amino acid synthesis. The results of *Chapter 4* strongly suggest that the common or orthologous sgRNA used to edit the *ALS* gene should be used with care. While a substitution at the herbicide resistance conferring site in *SlALS1* does not impact the plant phenotype, other modification at *SlALS1* as well as off-target modification on the *SlALS2* gene are associated with a negative

impact on the plant's growth. As we provide in the Chapter 4, several methods of circumventing such consequences are available or in progress of development, and these should be considered for the future *ALS*-co-editing strategy.

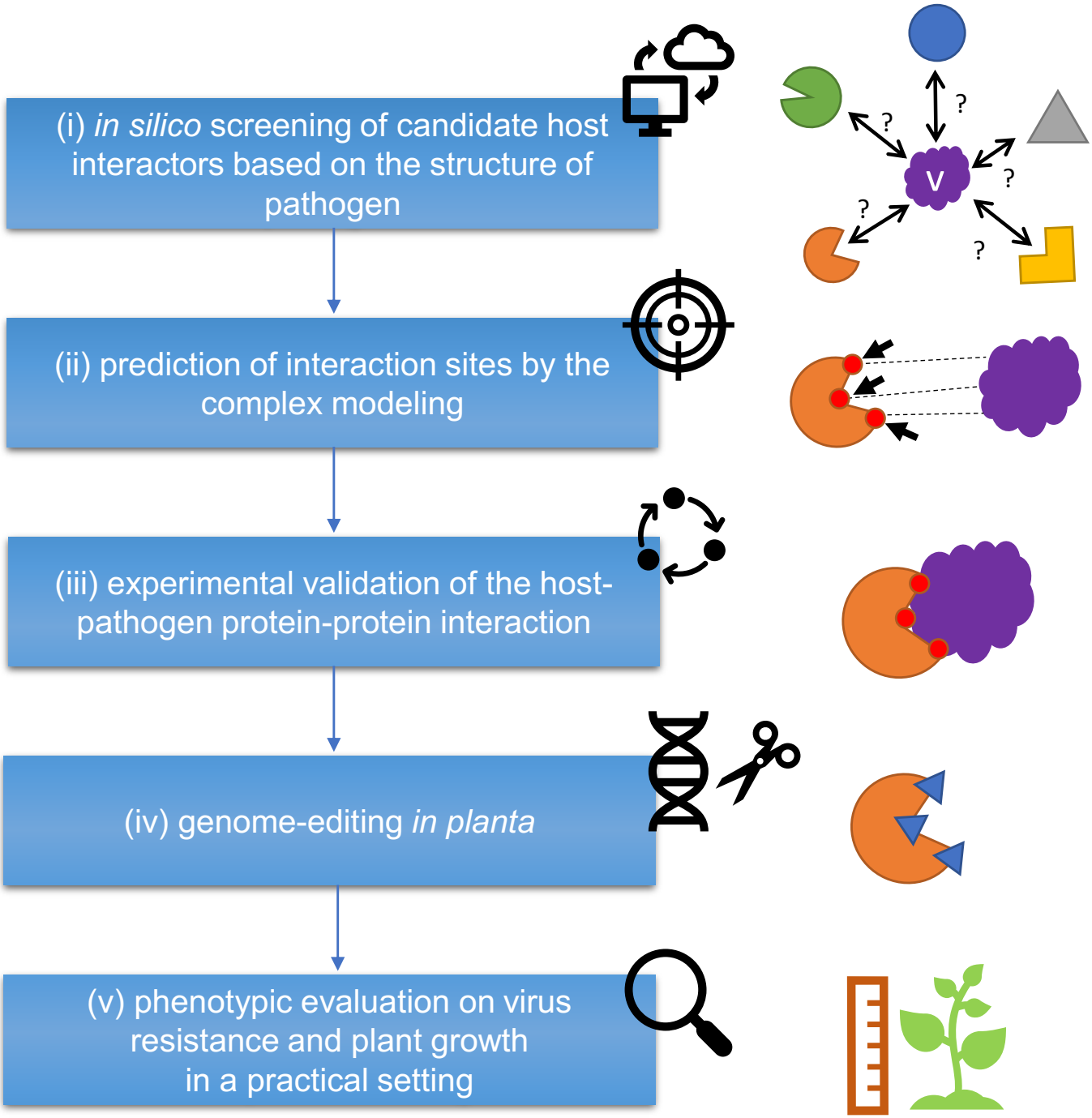


Figure D1: Pipeline for designing *de novo* resistance alleles based on loss-of-susceptibility and implementing them in species of interest by genome-editing.
From the *in silico* screening of candidate host interactors to the phenotyping of plants edited in the susceptibility gene to confer resistance. Schematics of each step are depicted by pictograms. This pipeline can be applied to other pathogens, but here it is described for viruses. v.: viral factor. (i) *in silico* screening of host interactors to a viral factor using protein structure prediction tools like AlphaFold2 and ColabFold. Once the candidate host interactor(s) are identified and the protein-protein interactions are modeled, (ii) the interaction sites on the viral and host factors can also be predicted. The prediction of interaction sites on host factors should be supported by experimental evidences and comparative studies of natural resistance alleles if they are available. (iii) The host-pathogen protein-protein interactions and the disruption of the interactions are to be verified experimentally. (iv) Based on the prediction and the experimental results, the genome-editing strategy to introduce base edition in the crop of interest can be implemented. (v) Finally, succefully edited plants are to be checked for their phenotypes including the virus resistance profile and the plant growth.

Perspectives

This thesis work represents how the fundamental studies of *eIF4E* genes as a key susceptibility factor have paved the way to designing resistance eIF4E variant in a crop model, cherry tomato. The significance of achieving artificial resistance by precise genome-editing in tomato is that this can be transferred to other crop species that are targeted by the potyviruses and are difficultly improved by the natural resistance alleles through the conventional crossing, including economically important crops like potatoes and cassava.

The success of the resistance by the *SleIF4E1*-editing partially, if not largely, owes to the compilation of the discovered natural resistance *eIF4E* genes from different species and the structural biological studies that allowed to theorize the involvement of two eIF4E regions in the virus susceptibility/resistance (Robaglia and Caranta, 2006b). This was later confirmed by the resolution of the eIF4E-VPg (PVY) complex structure (Coutinho de Oliveira et al., 2019).

Today, tools like AlphaFold2 or its derivative tools (e.g. ColabFold) to predict structure of proteins and protein complexes can allow us to skip the above steps or accelerate the designing of new resistance alleles based on susceptibility factors (Jumper et al., 2021; McLellan et al., 2023). McLellan et al. (2023) has proposed a pathway for converting a host plant into a non-host plant by predicting key mutations of host factors that perceive the effectors from fungal and oomycete pathogens with the help of protein structure prediction. Indeed, based on these tools, a pipeline for engineering resistant alleles by genome-editing can be proposed for preventing viruses to which the susceptibility genes are unknown or not well studied due to lack of natural varieties. Briefly, the pipeline consists of (i) *in silico* screening of candidate host interactors based on the structure of pathogen (Bryant et al., 2022; e.g. Ibrahim et al., 2023), (ii) prediction of interaction sites by the same complex modeling, (iii) experimental validation of the host-pathogen protein-protein interaction, (iv) editing *in planta*, and (v) phenotypic evaluation on the pathogen resistance as well as the plant growth in a practical setting (**Figure D1**). The room of improvement remains in the accurate prediction and efficiency of edition to be introduced; and these can be improved by the deeper understanding of the mechanism of resistance as well as the technical advancements. Such pipeline is a promising avenue to the vast possibility to improve crop species that have been

marginalized in the global breeding programs as well as to promote ecological agriculture with lower cost of chemicals for cultivating such crops.

References

Abdul-Razzak, A., Guiraud, T., Peypelut, M., Walter, J., Houvenaghel, M.-C., Candresse, T., LE Gall, O., and German-Retana, S. (2009). Involvement of the cylindrical inclusion (CI) protein in the overcoming of an eIF4E-mediated resistance against Lettuce mosaic potyvirus. *Mol. Plant Pathol.* **10**:109–113.

Altpeter, F., Springer, N. M., Bartley, L. E., Blechl, A., Brutnell, T. P., Citovsky, V., Conrad, L., Gelvin, S. B., Jackson, D., Kausch, A. P., et al. (2016). Advancing crop transformation in the era of genome editing. *Plant Cell* Advance Access published June 22, 2016.

Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., Chen, P. J., Wilson, C., Newby, G. A., Raguram, A., et al. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**:149–157.

Ashby, J. A., Stevenson, C. E. M., Jarvis, G. E., Lawson, D. M., and Maule, A. J. (2011). Structure-based mutational analysis of eIF4E in relation to sbm1 resistance to pea seed-borne mosaic virus in pea. *PLoS One* **6**:e15873.

Atarashi, H., Jayasinghe, W. H., Kwon, J., Kim, H., Taninaka, Y., Igarashi, M., Ito, K., Yamada, T., Masuta, C., and Nakahara, K. S. (2020). Artificially edited alleles of the eukaryotic translation initiation factor 4E1 gene differentially reduce susceptibility to cucumber mosaic virus and potato virus Y in tomato. *Front. Microbiol.* **11**:564310.

Atkins, P. A., and Voytas, D. F. (2020). Overcoming bottlenecks in plant gene editing. *Curr. Opin. Plant Biol.* **54**:79–84.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**:1709–1712.

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

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

Bastet, A., Zafirov, D., Giovinazzo, N., Guyon-Debast, A., Nogué, F., Robaglia, C., and Gallois, J. L. (2019). Mimicking natural polymorphism in eIF4E by CRISPR-Cas9 base editing is associated with resistance to potyviruses. *Plant Biotechnol. J.* **17**:1736–1750.

Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, S. D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**:2551–2561.

Borden, K. L. B. (2023). The eukaryotic translation initiation factor eIF4E unexpectedly acts in splicing thereby coupling mRNA processing with translation: eIF4E induces widescale splicing reprogramming providing system-wide connectivity between splicing, nuclear mRNA export and translation. *Bioessays* Advance Access published November 5, 2023.

Boutrot, F., and Zipfel, C. (2017). Function, discovery, and exploitation of plant pattern recognition receptors for broad-spectrum disease resistance. *Annu. Rev. Phytopathol.* **55**:257–286.

Brown, J. W. S., and Simpson, C. G. (1998). SPLICE SITE SELECTION IN PLANT PRE-mRNA SPLICING. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49**:77–95.

Browning, K. S., and Bailey-Serres, J. (2015). Mechanism of Cytoplasmic mRNA Translation. *Arabidopsis Book* **13**:e0176.

Bryant, P., Pozzati, G., and Elofsson, A. (2022). Improved prediction of protein-protein interactions using AlphaFold2. *Nat. Commun.* **13**:1265.

Büschges, R., Hollricher, K., Panstruga, R., Simons, G., Wolter, M., Frijters, A., van Daelen, R., van der Lee, T., Diergaarde, P., Groenendijk, J., et al. (1997). The Barley Mlo Gene: A Novel Control Element of Plant Pathogen Resistance. *Cell* **88**:695–705.

Bush, M. S., Hutchins, A. P., Jones, A. M. E., Naldrett, M. J., Jarmolowski, A., Lloyd, C. W., and Doonan, J. H. (2009). Selective recruitment of proteins to 5' cap complexes during the growth cycle in Arabidopsis. *Plant J.* **59**:400–412.

Callot, C., and Gallois, J. L. (2014). Pyramiding resistances based on translation initiation factors in Arabidopsis is impaired by male gametophyte lethality. *Plant Signal. Behav.* **9**:e27940.

Caranta, C. (1996). A complementation of two genes originating from susceptible*capsicum annuum*lines confers a new and complete resistance to pepper veinal mottle virus. *Phytopathology* **86**:739.

Castellano, M. M., and Merchante, C. (2021). Peculiarities of the regulation of translation initiation in plants. *Curr. Opin. Plant Biol.* **63**:102073.

Cavatorta, J., Perez, K. W., Gray, S. M., Van Eck, J., Yeam, I., and Jahn, M. (2011). Engineering virus resistance using a modified potato gene. *Plant Biotechnol. J.* **9**:1014–1021.

Cebrian-Serrano, A., and Davies, B. (2017). CRISPR-Cas orthologues and variants: optimizing the repertoire, specificity and delivery of genome engineering tools. *Mamm. Genome* **28**:247–261.

Cermak, T., Doyle, E. L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J. A., Somia, N. V., Bogdanove, A. J., and Voytas, D. F. (2011). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.* **39**:7879–7879.

Chandrasekaran, J., Brumin, M., Wolf, D., Leibman, D., Klap, C., Pearlsman, M., Sherman, A., Arazi, T., and Gal-On, A. (2016). Development of broad virus

resistance in non-transgenic cucumber using CRISPR/Cas9 technology. *Mol. Plant Pathol.* **17**:1140–1153.

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

Chat, V., Ferguson, R., Morales, L., and Kirchhoff, T. (2022). Ultra low-coverage whole-genome sequencing as an alternative to genotyping arrays in genome-wide association studies. *Front. Genet.* **12**.

Che, P., Wu, E., Simon, M. K., Anand, A., Lowe, K., Gao, H., Sigmund, A. L., Yang, M., Albertsen, M. C., Gordon-Kamm, W., et al. (2022). Wuschel2 enables highly efficient CRISPR/Cas-targeted genome editing during rapid de novo shoot regeneration in sorghum. *Commun. Biol.* **5**.

Chen, K., Wang, Y., Zhang, R., Zhang, H., and Gao, C. (2019). CRISPR/Cas Genome Editing and Precision Plant Breeding in Agriculture. *Annu. Rev. Plant Biol.* **70**:667–697.

Chen, Q., Chuai, G., Zhang, H., Tang, J., Duan, L., Guan, H., Li, W., Li, W., Wen, J., Zuo, E., et al. (2023). Genome-wide CRISPR off-target prediction and optimization using RNA-DNA interaction fingerprints. *Nat. Commun.* **14**.

Cheng, X., and Wang, A. (2017). The Potyvirus Silencing Suppressor Protein VPg Mediates Degradation of SGS3 via Ubiquitination and Autophagy Pathways. *J. Virol.* **91**.

Christie, M., and Igreja, C. (2021). eIF4E-homologous protein (4EHP): a multifarious cap-binding protein. *FEBS J.* Advance Access published November 10, 2021, doi:10.1111/febs.16275.

Chung, B. Y.-W., Miller, W. A., Atkins, J. F., and Firth, A. E. (2008). An overlapping essential gene in the Potyviridae. *Proc. Natl. Acad. Sci. U. S. A.* **105**:5897–5902.

Collard, B. C. Y., and Mackill, D. J. (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **363**:557–572.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**:819–823.

Coutinho de Oliveira, L., Volpon, L., Rahardjo, A. K., Osborne, M. J., Culjkovic-Kraljacic, B., Trahan, C., Oeffinger, M., Kwok, B. H., and Borden, K. L. B. (2019). Structural studies of the eIF4E-VPg complex reveal a direct competition for capped RNA: Implications for translation. *Proc. Natl. Acad. Sci. U. S. A.* **116**:24056–24065.

Culjkovic, B., Topisirovic, I., and Borden, K. L. B. (2007). Controlling gene expression through RNA regulons: the role of the eukaryotic translation initiation factor eIF4E. *Cell Cycle* **6**:65–69.

Danilo, B., Perrot, L., Botton, E., Nogué, F., and Mazier, M. (2018). The DFR locus: A smart landing pad for targeted transgene insertion in tomato. *PLoS One* **13**:e0208395.

Danilo, B., Perrot, L., Mara, K., Botton, E., Nogué, F., and Mazier, M. (2019). Efficient and transgene-free gene targeting using Agrobacterium-mediated delivery of the CRISPR/Cas9 system in tomato. *Plant Cell Rep.* **38**:459–462.

de Ronde, D., Butterbach, P., Lohuis, D., Hedil, M., van Lent, J. W. M., and Kormelink, R. (2013). Tsw gene-based resistance is triggered by a functional RNA silencing suppressor protein of the Tomato spotted wilt virus. *Mol. Plant Pathol.* **14**:405–415.

Dempewolf, H., Baute, G., Anderson, J., Kilian, B., Smith, C., and Guarino, L. (2017). Past and future use of wild relatives in crop breeding. *Crop Sci.* **57**:1070–1082.

Desaint, H., Aoun, N., Deslandes, L., Vailleau, F., Roux, F., and Berthomé, R. (2021). Fight hard or die trying: when plants face pathogens under heat stress. *New Phytol.* **229**:712–734.

Diamond, J. (2017). *Guns, Germs, and Steel: The Fates of Human Societies (20th Anniversary Edition)*. W. W. Norton & Company.

Doebley, J. F., Gaut, B. S., and Smith, B. D. (2006). The molecular genetics of crop domestication. *Cell* **127**:1309–1321.

Duan, H., Richael, C., and Rommens, C. M. (2012). Overexpression of the wild potato eIF4E-1 variant Eva1 elicits Potato virus Y resistance in plants silenced for native eIF4E-1. *Transgenic Res.* **21**:929–938.

Dubin, M. J., Mittelsten Scheid, O., and Becker, C. (2018). Transposons: a blessing curse. *Curr. Opin. Plant Biol.* **42**:23–29.

Duprat, A., Caranta, C., Revers, F., Menand, B., Browning, K. S., and Robaglia, C. (2002). The Arabidopsis eukaryotic initiation factor (iso)4E is dispensable for plant growth but required for susceptibility to potyviruses. *Plant J.* **32**:927–934.

Eisenstein, M. (2022). Base edit your way to better crops. *Nature* **604**:790–792.

Erickson, F. L., Holzberg, S., Calderon-Urrea, A., Handley, V., Axtell, M., Corr, C., and Baker, B. (1999). The helicase domain of the TMV replicase proteins induces the N-mediated defence response in tobacco. *Plant J.* **18**:67–75.

Eskelin, K., Hafrén, A., Rantalainen, K. I., and Mäkinen, K. (2011). Potyviral VPg enhances viral RNA Translation and inhibits reporter mRNA translation in planta. *J. Virol.* **85**:9210–9221.

Estevan, J., Maréna, A., Callot, C., Lacombe, S., Moretti, A., Caranta, C., and Gallois, J.-L. (2014). Specific requirement for translation initiation factor 4E or its isoform drives plant host susceptibility to Tobacco etch virus. *BMC Plant Biol.* **14**:67.

Ezura, H. (2022). Letter to the Editor: The World's First CRISPR Tomato Launched to a Japanese Market: The Social-Economic Impact of its Implementation on Crop Genome Editing. *Plant Cell Physiol.* **63**:731–733.

Flagel, L. E., and Wendel, J. F. (2009). Gene duplication and evolutionary novelty in plants. *New Phytol.* **183**:557–564.

Fulton, T. M., Chunwongse, J., and Tanksley, S. D. (1995). Microprep protocol for extraction of DNA from tomato and other herbaceous plants. *Plant Mol. Biol. Rep.* **13**:207–209.

Gadhave, K. R., Gautam, S., Rasmussen, D. A., and Srinivasan, R. (2020). Aphid Transmission of Potyvirus: The Largest Plant-Infecting RNA Virus Genus. *Viruses* **12**.

Gaillochet, C., Develtere, W., and Jacobs, T. B. (2021). CRISPR screens in plants: approaches, guidelines, and future prospects. *Plant Cell* **33**:794–813.

Gallie, D. R., and Browning, K. S. (2001). eIF4G Functionally Differs from eIFiso4G in Promoting Internal Initiation, Cap-independent Translation, and Translation of Structured mRNAs*. *J. Biol. Chem.* **276**:36951–36960.

Gallois, J. L., Charron, C., Sanchez, F., Pagny, G., Houvenaghel, M. C., Moretti, A., Ponz, F., Revers, F., Caranta, C., and German-Retana, S. (2010). Single amino acid changes in the turnip mosaic virus viral genome-linked protein (VPg) confer virulence towards Arabidopsis thaliana mutants knocked out for eukaryotic initiation factors eIF(iso)4E and eIF(iso)4G. *J. Gen. Virol.* **91**:288–293.

Gallois, J.-L., Moury, B., and German-Retana, S. (2018). Role of the Genetic Background in Resistance to Plant Viruses. *Int. J. Mol. Sci.* **19**.

Gao, C. (2021). Genome engineering for crop improvement and future agriculture. *Cell* **184**:1621–1635.

Gao, Z., Johansen, E., Eyers, S., Thomas, C. L., Noel Ellis, T. H., and Maule, A. J. (2004). The potyvirus recessive resistance gene, sbm1, identifies a novel role for translation initiation factor eIF4E in cell-to-cell trafficking. *Plant J.* **40**:376–385.

Garcia, D., Garcia, S., and Voinnet, O. (2014). Nonsense-mediated decay serves as a general viral restriction mechanism in plants. *Cell Host Microbe* **16**:391–402.

Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., and Liu, D. R. (2017). Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**:464–471.

Gauffier, C. (2015). Recherche et mise en place de résistances génétiques aux begomovirus et aux potyvirus chez la tomate Advance Access published November 24, 2015.

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

Ge, L., Cao, B., Qiao, R., Cui, H., Li, S., Shan, H., Gong, P., Zhang, M., Li, H., Wang, A., et al. (2023). SUMOylation-modified Pelota-Hbs1 RNA surveillance complex restricts the infection of potyvirids in plants. *Mol. Plant* **16**:632–642.

German-Retana, S., Walter, J., Doublet, B., Roudet-Tavert, G., Nicaise, V., Lecampion, C., Houvenaghel, M.-C., Robaglia, C., Michon, T., and Le Gall, O. (2008).

Mutational analysis of plant cap-binding protein eIF4E reveals key amino acids involved in biochemical functions and potyvirus infection. *J. Virol.* **82**:7601–7612.

Gibbs, A. J., Hajizadeh, M., Ohshima, K., and Jones, R. A. C. (2020). The potyviruses: An evolutionary synthesis is emerging. *Viruses* **12**:132.

Gomez, M. A., Lin, Z. D., Moll, T., Chauhan, R. D., Hayden, L., Renninger, K., Beyene, G., Taylor, N. J., Carrington, J. C., Staskawicz, B. J., et al. (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 Biotechnol. J.* **17**:421–434.

Goodfellow, I. (2011). The genome-linked protein VPg of vertebrate viruses - a multifaceted protein. *Curr. Opin. Virol.* **1**:355–362.

Greenleaf, W. H. (1986). Pepper breeding. In *Breeding Vegetable Crops*, pp. 67–134.

Gupta, P. K., Balyan, H. S., and Gautam, T. (2021). SWEET genes and TAL effectors for disease resistance in plants: Present status and future prospects. *Mol. Plant Pathol.* **22**:1014–1026.

Guyon-Debast, A., Alboresi, A., Terret, Z., Charlot, F., Berthier, F., Vendrell-Mir, P., Casacuberta, J. M., Veillet, F., Morosinotto, T., Gallois, J.-L., et al. (2021). A blueprint for gene function analysis through Base Editing in the model plant *Physcomitrium (Physcomitrella) patens*. *New Phytol.* **230**:1258–1272.

Hafrén, A., Lõhmus, A., and Mäkinen, K. (2015). Formation of Potato Virus A-Induced RNA Granules and Viral Translation Are Interrelated Processes Required for Optimal Virus Accumulation. *PLoS Pathog.* **11**:e1005314.

Hafrén, A., Üstün, S., Hochmuth, A., Svenning, S., Johansen, T., and Hofius, D. (2018). Turnip Mosaic Virus Counteracts Selective Autophagy of the Viral Silencing Suppressor HCpro. *Plant Physiol.* **176**:649–662.

Hagiwara-Komoda, Y., Choi, S. H., Sato, M., Atsumi, G., Abe, J., Fukuda, J., Honjo, M. N., Nagano, A. J., Komoda, K., Nakahara, K. S., et al. (2016). Truncated yet functional viral protein produced via RNA polymerase slippage implies underestimated coding capacity of RNA viruses. *Sci. Rep.* **6**:21411.

Hamann, E., Blevins, C., Franks, S. J., Jameel, M. I., and Anderson, J. T. (2021). Climate change alters plant-herbivore interactions. *New Phytol.* **229**:1894–1910.

Hansjakob, A., Riederer, M., and Hildebrandt, U. (2011). Wax matters: absence of very-long-chain aldehydes from the leaf cuticular wax of the glossy11 mutant of maize compromises the prepenetration processes of Blumeria graminis. *Plant Pathol.* **60**:1151–1161.

Haxim, Y., Ismayil, A., Jia, Q., Wang, Y., Zheng, X., Chen, T., Qian, L., Liu, N., Wang, Y., Han, S., et al. (2017). Autophagy functions as an antiviral mechanism against geminiviruses in plants. *Elife* **6**:e23897.

Heidmann, I., de Lange, B., Lambalk, J., Angenent, G. C., and Boutilier, K. (2011). Efficient sweet pepper transformation mediated by the BABY BOOM transcription factor. *Plant Cell Rep.* **30**:1107–1115.

Hess, G. T., Tycko, J., Yao, D., and Bassik, M. C. (2017). Methods and applications of CRISPR-mediated base editing in eukaryotic genomes. *Mol. Cell* **68**:26–43.

Hickey, L. T., N Hafeez, A., Robinson, H., Jackson, S. A., Leal-Bertioli, S. C. M., Tester, M., Gao, C., Godwin, I. D., Hayes, B. J., and Wulff, B. B. H. (2019). Breeding crops to feed 10 billion. *Nat. Biotechnol.* **37**:744–754.

Hickey, K. L., Dickson, K., Cogan, J. Z., Replogle, J. M., Schoof, M., D'Orazio, K. N., Sinha, N. K., Hussmann, J. A., Jost, M., Frost, A., et al. (2020). GIGYF2 and 4EHP Inhibit Translation Initiation of Defective Messenger RNAs to Assist Ribosome-Associated Quality Control. *Mol. Cell* **79**:950-962.e6.

Hipper, C., Brault, V., Ziegler-Graff, V., and Revers, F. (2013). Viral and cellular factors involved in phloem transport of plant viruses. *Front. Plant Sci.* **4**.

Hohn, T., and Vazquez, F. (2011). RNA silencing pathways of plants: silencing and its suppression by plant DNA viruses. *Biochim. Biophys. Acta* **1809**:588–600.

Holme, I. B., Gregersen, P. L., and Brinch-Pedersen, H. (2019). Induced genetic variation in crop plants by random or targeted Mutagenesis: Convergence and differences. *Front. Plant Sci.* **10**:1468.

Hu, X., Wang, C., Fu, Y., Liu, Q., Jiao, X., and Wang, K. (2016). Expanding the range of CRISPR/Cas9 genome editing in rice. *Mol. Plant* **9**:943–945.

Hu, J. H., Miller, S. M., Geurts, M. H., Tang, W., Chen, L., Sun, N., Zeina, C. M., Gao, X., Rees, H. A., Lin, Z., et al. (2018). Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**:57–63.

Hua, K., Zhang, J., Botella, J. R., Ma, C., Kong, F., Liu, B., and Zhu, J.-K. (2019). Perspectives on the Application of Genome-Editing Technologies in Crop Breeding. *Mol. Plant* **12**:1047–1059.

Huang, Z., and van der Knaap, E. (2011). Tomato fruit weight 11.3 maps close to fasciated on the bottom of chromosome 11. *Züchter Genet. Breed. Res.* **123**:465–474.

Huang, X., Jia, H., Xu, J., Wang, Y., Wen, J., and Wang, N. (2023). Transgene-free genome editing of vegetatively propagated and perennial plant species in the T0 generation via a co-editing strategy. *Nat. Plants* Advance Access published September 18, 2023, doi:10.1038/s41477-023-01520-y.

Ibrahim, T., Khandare, V., Mirkin, F. G., Tumtas, Y., Bubeck, D., and Bozkurt, T. O. (2023). AlphaFold2-multimer guided high-accuracy prediction of typical and atypical ATG8-binding motifs. *PLoS Biol.* **21**:e3001962.

Inoue-Nagata, A. K., Jordan, R., Kreuze, J., Li, F., López-Moya, J. J., Mäkinen, K., Ohshima, K., Wylie, S. J., and Ictv Report Consortium (2022). ICTV virus Taxonomy profile: Potyviridae 2022. *J. Gen. Virol.* **103**.

Iqbal, Z., Iqbal, M. S., Ahmad, A., Memon, A. G., and Ansari, M. I. (2020). New prospects on the horizon: Genome editing to engineer plants for desirable traits. *Curr. Plant Biol.* **24**:100171.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., and Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J. Bacteriol.* **169**:5429–5433.

Ishino, Y., Krupovic, M., and Forterre, P. (2018). History of CRISPR-Cas from encounter with a mysterious repeated sequence to genome editing technology. *J. Bacteriol.* **200**.

Ivanov, K. I., Eskelin, K., Lõhmus, A., and Mäkinen, K. (2014). Molecular and cellular mechanisms underlying potyvirus infection. *J. Gen. Virol.* **95**:1415–1429.

Jacob, P., Avni, A., and Bendahmane, A. (2018). Translational Research: Exploring and Creating Genetic Diversity. *Trends Plant Sci.* **23**:42–52.

Jansen, R., van Embden, J. D. A., Gaastra, W., and Schouls, L. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43**:1565–1575.

Jayasinghe, W. H., Kim, H., Nakada, Y., and Masuta, C. (2021). A plant virus satellite RNA directly accelerates wing formation in its insect vector for spread. *Nat. Commun.* **12**.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**:816–821.

Jones, J. D. G., and Dangl, J. L. (2006). The plant immune system. *Nature* **444**:323–329.

Joshi, B., Lee, K., Maeder, D. L., and Jagus, R. (2005). Phylogenetic analysis of eIF4E-family members. *BMC Evol. Biol.* **5**:48.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* **596**:583–589.

Kang, B.-C., Yeam, I., and Jahn, M. M. (2005a). Genetics of plant virus resistance. *Annu. Rev. Phytopathol.* **43**:581–621.

Kang, B.-C., Yeam, I., Frantz, J. D., Murphy, J. F., and Jahn, M. M. (2005b). The pvr1 locus in Capsicum encodes a translation initiation factor eIF4E that interacts with Tobacco etch virus VPg. *Plant J.* **42**:392–405.

Kang, B.-C., Yeam, I., Li, H., Perez, K. W., and Jahn, M. M. (2007). Ectopic expression of a recessive resistance gene generates dominant potyvirus resistance in plants. *Plant Biotechnol. J.* **5**:526–536.

Karlsson Green, K., Stenberg, J. A., and Lankinen, Å. (2020). Making sense of Integrated Pest Management (IPM) in the light of evolution. *Evol. Appl.* **13**:1791–1805.

Khoury, W. E., and Makkouk, K. (2010). Integrated plant disease management in developing countries. *J. Plant Pathol.* **92**:S35–S42.

Kim, S.-B., Lee, H.-Y., Seo, S., Lee, J. H., and Choi, D. (2015). RNA-dependent RNA polymerase (NIb) of the potyviruses is an avirulence factor for the broad-spectrum resistance gene Pvr4 in Capsicum annuum cv. CM334. *PLoS One* **10**:e0119639.

Kim, S.-B., Kang, W.-H., Huy, H. N., Yeom, S.-I., An, J.-T., Kim, S., Kang, M.-Y., Kim, H. J., Jo, Y. D., Ha, Y., et al. (2017). Divergent evolution of multiple virus-resistance genes from a progenitor in Capsicum spp. *New Phytol.* **213**:886–899.

Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Topkar, V. V., Nguyen, N. T., Zheng, Z., Gonzales, A. P. W., Li, Z., Peterson, R. T., Yeh, J.-R. J., et al. (2015). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**:481–485.

Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., and Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**:420–424.

Koonin, E. V., Dolja, V. V., Krupovic, M., Varsani, A., Wolf, Y. I., Yutin, N., Zerbini, F. M., and Kuhn, J. H. (2020). Global organization and proposed megataxonomy of the virus world. *Microbiol. Mol. Biol. Rev.* **84**.

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

Kuroiwa, K., Thenault, C., Nogué, F., Perrot, L., Mazier, M., and Gallois, J.-L. (2022). CRISPR-based knock-out of eIF4E2 in a cherry tomato background successfully recapitulates resistance to pepper veinal mottle virus. *Plant Sci.* **316**:111160.

Lapidot, M., Karniel, U., Gelbart, D., Fogel, D., Evenor, D., Kutsher, Y., Makhbash, Z., Nahon, S., Shlomo, H., Chen, L., et al. (2015). A Novel Route Controlling Begomovirus Resistance by the Messenger RNA Surveillance Factor Pelota. *PLoS Genet.* **11**:e1005538.

Lebaron, C., Rosado, A., Sauvage, C., Gauffier, C., German-Retana, S., Moury, B., and Gallois, J. L. (2016). A new eIF4E1 allele characterized by RNAseq data mining is associated with resistance to potato virus Y in tomato albeit with a low durability. *J. Gen. Virol.* **97**:3063–3072.

Lellis, A. D., Kasschau, K. D., Whitham, S. A., and Carrington, J. C. (2002a). Loss-of-susceptibility mutants of Arabidopsis thaliana reveal an essential role for eIF(iso)4E during potyvirus infection. *Current Biology* **12**.

Lellis, A. D., Kasschau, K. D., Whitham, S. A., and Carrington, J. C. (2002b). Loss-of-susceptibility mutants of Arabidopsis thaliana reveal an essential role for eIF(iso)4E during potyvirus infection. *Curr. Biol.* **12**:1046–1051.

Lemmon, Z. H., Reem, N. T., Dalrymple, J., Soyk, S., Swartwood, K. E., Rodriguez-Leal, D., Van Eck, J., and Lippman, Z. B. (2018). Rapid improvement of domestication traits in an orphan crop by genome editing. *Nat Plants* **4**:766–770.

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

Levins, E., Tseng, C.-Y., Patrick, R. M., Mayberry, L. K., Cole, N., and Browning, K. S. (2016). Fusion proteins of Arabidopsis cap-binding proteins: Cautionary "tails" of woe. *Translation* **4**.

Li, F., and Wang, A. (2018). RNA decay is an antiviral defense in plants that is counteracted by viral RNA silencing suppressors. *PLoS Pathog.* **14**:e1007228.

Li, T., Yang, X., Yu, Y., Si, X., Zhai, X., Zhang, H., Dong, W., Gao, C., and Xu, C. (2018). Domestication of wild tomato is accelerated by genome editing. *Nat. Biotechnol.* **36**:1160–1163.

Li, Z., Xiong, X., Wang, F., Liang, J., and Li, J.-F. (2019). Gene disruption through base editing-induced messenger RNA missplicing in plants. *New Phytol.* **222**:1139–1148.

Lin, T., Zhu, G., Zhang, J., Xu, X., Yu, Q., Zheng, Z., Zhang, Z., Lun, Y., Li, S., Wang, X., et al. (2014). Genomic analyses provide insights into the history of tomato breeding. *Nat. Genet.* **46**:1220–1226.

Lin, Q., Zong, Y., Xue, C., Wang, S., Jin, S., Zhu, Z., Wang, Y., Anzalone, A. V., Raguram, A., Doman, J. L., et al. (2020). Prime genome editing in rice and wheat. *Nat. Biotechnol.* **38**:582–585.

Lin, Q., Jin, S., Zong, Y., Yu, H., Zhu, Z., Liu, G., Kou, L., Wang, Y., Qiu, J.-L., Li, J., et al. (2021). High-efficiency prime editing with optimized, paired pegRNAs in plants. *Nat. Biotechnol.* **39**:923–927.

Lippman, Z., and Tanksley, S. D. (2001). Dissecting the genetic pathway to extreme fruit size in tomato using a cross between the small-fruited wild species Lycopersicon pimpinellifolium and L. esculentum var. Giant Heirloom. *Genetics* **158**:413–422.

Lopez-Gomollon, S., and Baulcombe, D. C. (2022). Roles of RNA silencing in viral and non-viral plant immunity and in the crosstalk between disease resistance systems. *Nat. Rev. Mol. Cell Biol.* **23**:645–662.

Lowe, K., Wu, E., Wang, N., Hoerster, G., Hastings, C., Cho, M.-J., Scelonge, C., Lenderts, B., Chamberlin, M., Cushatt, J., et al. (2016). Morphogenic regulators Baby boom and wuschel improve monocot transformation. *Plant Cell* **28**:1998–2015.

Lu, Y., Tian, Y., Shen, R., Yao, Q., Wang, M., Chen, M., Dong, J., Zhang, T., Li, F., Lei, M., et al. (2020). Targeted, efficient sequence insertion and replacement in rice. *Nat. Biotechnol.* **38**:1402–1407.

Maher, M. F., Nasti, R. A., Vollbrecht, M., Starker, C. G., Clark, M. D., and Voytas, D. F. (2020). Plant gene editing through de novo induction of meristems. *Nat. Biotechnol.* **38**:84–89.

Majeed, A., Johar, P., Raina, A., Salgotra, R. K., Feng, X., and Bhat, J. A. (2022). Harnessing the potential of bulk segregant analysis sequencing and its related approaches in crop breeding. *Front. Genet.* **13**:944501.

Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I., and Koonin, E. V. (2006). A putative RNA-interference-based immune system in prokaryotes: computational

analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct* **1**.

Mäkinen, K., Lõhmus, A., and Pollari, M. (2017). Plant RNA Regulatory Network and RNA Granules in Virus Infection. *Front. Plant Sci.* **8**:2093.

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., and Church, G. M. (2013). RNA-guided human genome engineering via Cas9. *Science* **339**:823–826.

Manghwar, H., Lindsey, K., Zhang, X., and Jin, S. (2019). CRISPR/Cas system: Recent advances and future prospects for genome editing. *Trends Plant Sci.* **24**:1102–1125.

Mao, L., Begum, D., Chuang, H. W., Budiman, M. A., Szymkowiak, E. J., Irish, E. E., and Wing, R. A. (2000). JOINTLESS is a MADS-box gene controlling tomato flower abscission zone development. *Nature* **406**:910–913.

Marks, R. A., Hotaling, S., Frandsen, P. B., and VanBuren, R. (2021). Representation and participation across 20 years of plant genome sequencing. *Nat Plants* **7**:1571–1578.

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

McLellan, H., Boevink, P. C., and Birch, P. R. J. (2023). How to convert host plants into nonhosts. *Trends Plant Sci.* **28**:876–879.

Meyer, R. S., and Purugganan, M. D. (2013). Evolution of crop species: genetics of domestication and diversification. *Nat. Rev. Genet.* **14**:840–852.

Mickelbart, M. V., Hasegawa, P. M., and Bailey-Serres, J. (2015). Genetic mechanisms of abiotic stress tolerance that translate to crop yield stability. *Nat. Rev. Genet.* **16**:237–251.

Miller, S. M., Wang, T., Randolph, P. B., Arbab, M., Shen, M. W., Huang, T. P., Matuszek, Z., Newby, G. A., Rees, H. A., and Liu, D. R. (2020). Continuous evolution of SpCas9 variants compatible with non-G PAMs. *Nat. Biotechnol.* **38**:471–481.

Miras, M., Truniger, V., Silva, C., Verdaguer, N., Aranda, M. A., and Querol-Audí, J. (2017). Structure of eIF4E in Complex with an eIF4G Peptide Supports a Universal Bipartite Binding Mode for Protein Translation. *Plant Physiol.* **174**:1476–1491.

Miyoshi, H., Suehiro, N., Tomoo, K., Muto, S., Takahashi, T., Tsukamoto, T., Ohmori, T., and Natsuaki, T. (2006). Binding analyses for the interaction between plant virus genome-linked protein (VPg) and plant translational initiation factors. *Biochimie* **88**:329–340.

Miyoshi, H., Okade, H., Muto, S., Suehiro, N., Nakashima, H., Tomoo, K., and Natsuaki, T. (2008). Turnip mosaic virus VPg interacts with Arabidopsis thaliana eIF(iso)4E and inhibits in vitro translation. *Biochimie* **90**:1427–1434.

Mojica, F. J. M., Juez, G., and Rodriguez-Valera, F. (1993). Transcription at different salinities of *Haloferax mediterranei* sequences adjacent to partially modified *Pst*I sites. *Mol. Microbiol.* **9**:613–621.

Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J., and Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60**:174–182.

Monzingo, A. F., Dhaliwal, S., Dutt-Chaudhuri, A., Lyon, A., Sadow, J. H., Hoffman, D. W., Robertus, J. D., and Browning, K. S. (2007). The structure of eukaryotic translation initiation factor-4E from wheat reveals a novel disulfide bond. *Plant Physiol.* **143**:1504–1518.

Moon, S. L., and Wilusz, J. (2013). Cytoplasmic viruses: rage against the (cellular RNA decay) machine. *PLoS Pathog.* **9**:e1003762.

Moscou, M. J., and Bogdanove, A. J. (2009). A simple cipher governs DNA recognition by TAL effectors. *Science* **326**:1501.

Moury, B., and Desbiez, C. (2020). Host Range Evolution of Potyviruses: A Global Phylogenetic Analysis. *Viruses* **12**.

Moury, B., Morel, C., Johansen, E., Guilbaud, L., Souche, S., Ayme, V., Caranta, C., Palloix, A., and Jacquemond, M. (2004). Mutations in potato virus Y genome-linked protein determine virulence toward recessive resistances in Capsicum annuum and Lycopersicon hirsutum. *Mol. Plant. Microbe. Interact.* **17**:322–329.

Moury, B., Fabre, F., and Senoussi, R. (2007). Estimation of the number of virus particles transmitted by an insect vector. *Proc. Natl. Acad. Sci. U. S. A.* **104**:17891–17896.

Moury, B., Charron, C., Janzac, B., Simon, V., Gallois, J. L., Palloix, A., and Caranta, C. (2014). Evolution of plant eukaryotic initiation factor 4E (eIF4E) and potyvirus genome-linked protein (VPg): a game of mirrors impacting resistance spectrum and durability. *Infect. Genet. Evol.* **27**:472–480.

Moury, B., Lebaron, C., Szadkowski, M., Ben Khalifa, M., Girardot, G., Bolou Bi, B. A., Koné, D., Nitiema, L. W., Fakhfakh, H., and Gallois, J.-L. (2020). Knock-out mutation of eukaryotic initiation factor 4E2 (eIF4E2) confers resistance to pepper veinal mottle virus in tomato. *Virology* **539**:11–17.

Muños, S., Ranc, N., Botton, E., Bérard, A., Rolland, S., Duffé, P., Carretero, Y., Le Paslier, M.-C., Delalande, C., Bouzayen, M., et al. (2011). Increase in tomato locule number is controlled by two single-nucleotide polymorphisms located near WUSCHEL. *Plant Physiol.* **156**:2244–2254.

Nagy, E. D., Stevens, J. L., Yu, N., Hubmeier, C. S., LaFaver, N., Gillespie, M., Gardunia, B., Cheng, Q., Johnson, S., Vaughn, A. L., et al. (2021). Novel disease resistance gene paralogs created by CRISPR/Cas9 in soy. *Plant Cell Rep.* **40**:1047–1058.

Nakahara, K. S., Shimada, R., Choi, S.-H., Yamamoto, H., Shao, J., and Uyeda, I. (2010). Involvement of the P1 Cistron in Overcoming eIF4E-Mediated Recessive Resistance Against Clover yellow vein virus in Pea. *Mol. Plant. Microbe. Interact.* **23**:1460–1469.

Nakahara, K. S., Masuta, C., Yamada, S., Shimura, H., Kashihara, Y., Wada, T. S., Meguro, A., Goto, K., Tadamura, K., Sueda, K., et al. (2012). Tobacco calmodulin-like protein provides secondary defense by binding to and directing degradation of virus RNA silencing suppressors. *Proc. Natl. Acad. Sci. U. S. A.* **109**:10113–10118.

Nazarov, P. A., Baleev, D. N., Ivanova, M. I., Sokolova, L. M., and Karakozova, M. V. (2020). Infectious Plant Diseases: Etiology, Current Status, Problems and Prospects in Plant Protection. *Acta Naturae* **12**:46–59.

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

Nelson, R., Wiesner-Hanks, T., Wisser, R., and Balint-Kurti, P. (2018). Navigating complexity to breed disease-resistant crops. *Nature Reviews Genetics* **19**:21–33.

Ng, J. C. K., and Falk, B. W. (2006). Virus-vector interactions mediating nonpersistent and semipersistent transmission of plant viruses. *Annu. Rev. Phytopathol.* **44**:183–212.

Ngou, B. P. M., Ding, P., and Jones, J. D. G. (2022). Thirty years of resistance: Zig-zag through the plant immune system. *Plant Cell* **34**:1447–1478.

Nicaise, V. (2014). Crop immunity against viruses: Outcomes and future challenges. *Frontiers in Plant Science* **5**.

Nicaise, V., German-Retana, S., Sanjuán, R., Dubrana, M.-P., Mazier, M., Maisonneuve, B., Candresse, T., Caranta, C., and LeGall, O. (2003). The eukaryotic translation initiation factor 4E controls lettuce susceptibility to the Potyvirus Lettuce mosaic virus. *Plant Physiol.* **132**:1272–1282.

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

Nicolas, O., and Laliberte, J.-F. (1992). The complete nucleotide sequence of turnip mosaic potyvirus RNA. *J. Gen. Virol.* **73**:2785–2793.

Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K. Y., et al. (2016). Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**.

Nishimasu, H., Shi, X., Ishiguro, S., Gao, L., Hirano, S., Okazaki, S., Noda, T., Abudayyeh, O. O., Gootenberg, J. S., Mori, H., et al. (2018a). Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* **361**:1259–1262.

Nishimasu, H., Shi, X., Ishiguro, S., Gao, L., Hirano, S., Okazaki, S., Noda, T., Abudayyeh, O. O., Gootenberg, J. S., Mori, H., et al. (2018b). Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* **361**:1259–1262.

Nonaka, S., Arai, C., Takayama, M., Matsukura, C., and Ezura, H. (2017). Efficient increase of ɣ-aminobutyric acid (GABA) content in tomato fruits by targeted mutagenesis. *Sci. Rep.* **7**:7057.

Olsen, K. M., and Wendel, J. F. (2013). A bountiful harvest: genomic insights into crop domestication phenotypes. *Annu. Rev. Plant Biol.* **64**:47–70.

Olspert, A., Chung, B. Y.-W., Atkins, J. F., Carr, J. P., and Firth, A. E. (2015). Transcriptional slippage in the positive-sense RNA virus family Potyviridae. *EMBO Rep.* **16**:995–1004.

Østerberg, J. T., Xiang, W., Olsen, L. I., Edenbrandt, A. K., Vedel, S. E., Christiansen, A., Landes, X., Andersen, M. M., Pagh, P., Sandøe, P., et al. (2017). Accelerating the domestication of new crops: Feasibility and approaches. *Trends Plant Sci.* **22**:373–384.

Pallas, V., and García, J. A. (2011). How do plant viruses induce disease? Interactions and interference with host components. *J. Gen. Virol.* **92**:2691–2705.

Parrella, G., Ruffel, S., Moretti, A., Morel, C., Palloix, A., and Caranta, C. (2002). Recessive resistance genes against potyviruses are localized in colinear genomic regions of the tomato (Lycopersicon spp.) and pepper (Capsicum spp.) genomes. *Theor. Appl. Genet.* **105**:855–861.

Pasin, F., Daròs, J.-A., and Tzanetakis, I. E. (2022). Proteome expansion in the *Potyviridae* evolutionary radiation. *FEMS Microbiol. Rev.* **46**.

Patrick, R. M., and Browning, K. S. (2012). The eIF4F and eIFiso4F Complexes of Plants: An Evolutionary Perspective. *Comp. Funct. Genomics* **2012**:287814.

Patrick, R. M., Mayberry, L. K., Choy, G., Woodard, L. E., Liu, J. S., White, A., Mullen, R. A., Tanavin, T. M., Latz, C. A., and Browning, K. S. (2014). Two Arabidopsis loci encode novel Eukaryotic Initiation Factor 4E isoforms that are functionally distinct from the conserved plant Eukaryotic Initiation Factor 4E. *Plant Physiol.* **164**:1820–1830.

Pechar, G. S., Donaire, L., Gosalvez, B., García-Almodovar, C., Sánchez-Pina, M. A., Truniger, V., and Aranda, M. A. (2022). Editing melon eIF4E associates with virus resistance and male sterility. *Plant Biotechnol. J.* Advance Access published July 1, 2022, doi:10.1111/pbi.13885.

Perring, T. M., Gruenhagen, N. M., and Farrar, C. A. (1999). Management of plant viral diseases through chemical control of insect vectors. *Annu. Rev. Entomol.* **44**:457–481.

Perroud, P.-F., Guyon-Debast, A., Casacuberta, J. M., Paul, W., Pichon, J.-P., Comeau, D., and Nogué, F. (2023). Improved prime editing allows for routine predictable gene editing in *Physcomitrium patens*. *J. Exp. Bot.* **74**:6176–6187.

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

Pouclet, A., Gagliardi, D., and Garcia, D. (2023). No-go decay as a novel route to restrict viral infection in plants. *Mol. Plant* **16**:509–510.

Poulicard, N., Pacios, L. F., Gallois, J.-L., Piñero, D., and García-Arenal, F. (2016). Human Management of a Wild Plant Modulates the Evolutionary Dynamics of a Gene Determining Recessive Resistance to Virus Infection. *PLoS Genet.* **12**:e1006214.

Pourcel, C., Salvignol, G., and Vergnaud, G. (2005). CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151**:653–663.

Puchta, H. (2005). The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. *J. Exp. Bot.* **56**:1–14.

Pumplin, N., and Voinnet, O. (2013). RNA silencing suppression by plant pathogens: defence, counter-defence and counter-counter-defence. *Nat. Rev. Microbiol.* **11**:745–760.

Pyott, D. E., Sheehan, E., and Molnar, A. (2016). Engineering of CRISPR/Cas9-mediated potyvirus resistance in transgene-free Arabidopsis plants. *Mol. Plant Pathol.* **17**:1276–1288.

Qin, R., Li, J., Li, H., Zhang, Y., Liu, X., Miao, Y., Zhang, X., and Wei, P. (2019). Developing a highly efficient and wildly adaptive CRISPR-SaCas9 toolset for plant genome editing. *Plant Biotechnol. J.* **17**:706–708.

Quenouille, J., Montarry, J., Palloix, A., and Moury, B. (2013). Farther, slower, stronger: how the plant genetic background protects a major resistance gene from breakdown. *Mol. Plant Pathol.* **14**:109–118.

Quenouille, J., Paulhiac, E., Moury, B., and Palloix, A. (2014). Quantitative trait loci from the host genetic background modulate the durability of a resistance gene: a rational basis for sustainable resistance breeding in plants. *Heredity* **112**:579–587.

Quenouille, J., Saint-Felix, L., Moury, B., and Palloix, A. (2016). Diversity of genetic backgrounds modulating the durability of a major resistance gene. Analysis of a core collection of pepper landraces resistant to Potato virus Y. *Mol. Plant Pathol.* **17**:296–302.

Raja, P., Sanville, B. C., Buchmann, R. C., and Bisaro, D. M. (2008). Viral genome methylation as an epigenetic defense against geminiviruses. *J. Virol.* **82**:8997–9007.

Raman, R. (2017). The impact of Genetically Modified (GM) crops in modern agriculture: A review. *GM Crops Food* **8**:195–208.

Ramanathan, A., Robb, G. B., and Chan, S.-H. (2016). mRNA capping: biological functions and applications. *Nucleic Acids Res.* **44**:7511–7526.

Ramu, P., Esuma, W., Kawuki, R., Rabbi, I. Y., Egesi, C., Bredeson, J. V., Bart, R. S., Verma, J., Buckler, E. S., and Lu, F. (2017). Cassava haplotype map highlights fixation of deleterious mutations during clonal propagation. *Nat. Genet.* **49**:959–963.

Ran, F. A., Cong, L., Yan, W. X., Scott, D. A., Gootenberg, J. S., Kriz, A. J., Zetsche, B., Shalem, O., Wu, X., Makarova, K. S., et al. (2015). In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520**:186–191.

Ran, Y., Liang, Z., and Gao, C. (2017). Current and future editing reagent delivery systems for plant genome editing. *Sci. China Life Sci.* **60**:490–505.

Ranc, N., Mũos, S., Santoni, S., and Causse, M. (2008). A clarified position for solanum lycopersicum var. cerasiforme in the evolutionary history of tomatoes (solanaceae). *BMC Plant Biology* **8**.

Rantalainen, K. I., Eskelin, K., Tompa, P., and Mäkinen, K. (2011). Structural flexibility allows the functional diversity of potyvirus genome-linked protein VPg. *J. Virol.* **85**:2449–2457.

Ren, Q., Sretenovic, S., Liu, S., Tang, X., Huang, L., He, Y., Liu, L., Guo, Y., Zhong, Z., Liu, G., et al. (2021). PAM-less plant genome editing using a CRISPR-SpRY toolbox. *Nat Plants* **7**:25–33.

Ren, Y., Tao, X., Li, D., Yang, X., and Zhou, X. (2022). Ty-5 confers broad-spectrum resistance to geminiviruses. *Viruses* **14**:1804.

Revers, F., and García, J. A. (2015). Molecular biology of potyviruses. *Adv. Virus Res.* **92**:101–199.

Riechmann, J. L., Lain, S., and Garcia, J. A. (1992). Highlights and prospects of potyvirus molecular biology. *J. Gen. Virol.* **73**:1–16.

Robaglia, C., and Caranta, C. (2006a). Translation initiation factors: a weak link in plant RNA virus infection. *Trends in plant science* **11**:40–45.

Robaglia, C., and Caranta, C. (2006b). Translation initiation factors: a weak link in plant RNA virus infection. *Trends Plant Sci.* **11**:40–45.

Rodamilans, B., Valli, A., Mingot, A., San León, D., Baulcombe, D., López-Moya, J. J., and García, J. A. (2015). RNA polymerase slippage as a mechanism for the production of frameshift gene products in plant viruses of the potyviridae family. *J. Virol.* **89**:6965–6967.

Rodriguez, C. M., Freire, M. A., Camilleri, C., and Robaglia, C. (1998). The Arabidopsis thaliana cDNAs coding for eIF4E and eIF(iso)4E are not functionally equivalent for yeast complementation and are differentially expressed during plant development. *Plant J.* **13**:465–473.

Rodríguez-Leal, D., Lemmon, Z. H., Man, J., Bartlett, M. E., and Lippman, Z. B. (2017). Engineering Quantitative Trait Variation for Crop Improvement by Genome Editing. *Cell* **171**:470-480.e8.

Rodriguez-Moreno, L., Song, Y., and Thomma, B. P. (2017). Transfer and engineering of immune receptors to improve recognition capacities in crops. *Curr. Opin. Plant Biol.* **38**:42–49.

Rojas, M. R., Macedo, M. A., Maliano, M. R., Soto-Aguilar, M., Souza, J. O., Briddon, R. W., Kenyon, L., Rivera Bustamante, R. F., Zerbini, F. M., Adkins, S., et al. (2018). World Management of Geminiviruses. *Annu. Rev. Phytopathol.* **56**:637–677.

Roldan, M. V. G., Périlleux, C., Morin, H., Huerga-Fernandez, S., Latrasse, D., Benhamed, M., and Bendahmane, A. (2017). Natural and induced loss of function mutations in SlMBP21 MADS-box gene led to jointless-2 phenotype in tomato. *Sci. Rep.* **7**:4402.

Rothan, C., Diouf, I., and Causse, M. (2019). Trait discovery and editing in tomato. *Plant J.* **97**:73–90.

Ruffel, S., Dussault, M. H., Palloix, A., Moury, B., Bendahmane, A., Robaglia, C., and Caranta, C. (2002a). A natural recessive resistance gene against potato virus Y in pepper corresponds to the eukaryotic initiation factor 4E (elF4E). *Plant Journal* **32**.

Ruffel, S., Dussault, M.-H., Palloix, A., Moury, B., Bendahmane, A., Robaglia, C., and Caranta, C. (2002b). A natural recessive resistance gene against potato virus Y in pepper corresponds to the eukaryotic initiation factor 4E (eIF4E). *Plant J.* **32**:1067–1075.

Ruffel, S., Gallois, J. L., Lesage, M. L., and Caranta, C. (2005). The recessive potyvirus resistance gene pot-1 is the tomato orthologue of the pepper pvr2-eIF4E gene. *Mol. Genet. Genomics* **274**:346–353.

Ruffel, S., Gallois, J.-L., Moury, B., Robaglia, C., Palloix, A., and Caranta, C. (2006). Simultaneous mutations in translation initiation factors eIF4E and eIF(iso)4E are required to prevent pepper veinal mottle virus infection of pepper. *J. Gen. Virol.* **87**:2089–2098.

Ruud, K. A., Kuhlow, C., Goss, D. J., and Browning, K. S. (1998). Identification and characterization of a novel cap-binding protein from Arabidopsis thaliana. *J. Biol. Chem.* **273**:10325–10330.

Saha, S., and Mäkinen, K. (2020). Insights into the Functions of eIF4E-Binding Motif of VPg in Potato Virus A Infection. *Viruses* **12**.

Salazar-Díaz, K., Aquino-Luna, M., Hernández-Lucero, E., Nieto-Rivera, B., Pulido-Torres, M. A., Jorge-Pérez, J. H., Gavilanes-Ruiz, M., and Dinkova, T. D. (2021). Arabidopsis thaliana eIF4E1 and eIF(iso)4E participate in cold response and promote translation of some stress-related mRNAs. *Front. Plant Sci.* **12**.

Sanfaçon, H. (2015). Plant Translation Factors and Virus Resistance. *Viruses* **7**:3392–3419.

Scheben, A., Wolter, F., Batley, J., Puchta, H., and Edwards, D. (2017). TowardsCRISPR/Cas crops – bringing together genomics and genome editing. *New Phytol.* **216**:682–698.

Schmitt-Keichinger, C. (2019). Manipulating Cellular Factors to Combat Viruses: A Case Study From the Plant Eukaryotic Translation Initiation Factors eIF4. *Front. Microbiol.* **10**:17.

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., Zhang, K., Liu, J., Xi, J. J., Qiu, J.-L., et al. (2013). Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat. Biotechnol.* **31**:686–688.

Shimura, H., Pantaleo, V., Ishihara, T., Myojo, N., Inaba, J.-I., Sueda, K., Burgyán, J., and Masuta, C. (2011). A viral satellite RNA induces yellow symptoms on tobacco by targeting a gene involved in chlorophyll biosynthesis using the RNA silencing machinery. *PLoS Pathog.* **7**:e1002021.

Simpson, C. G., Clark, G., Davidson, D., Smith, P., and Brown, J. W. (1996). Mutation of putative branchpoint consensus sequences in plant introns reduces splicing efficiency. *Plant J.* **9**:369–380.

Somssich, M., Je, B. I., Simon, R., and Jackson, D. (2016). CLAVATA-WUSCHEL signaling in the shoot meristem. *Development* **143**:3238–3248.

Soyk, S., Lemmon, Z. H., Oved, M., Fisher, J., Liberatore, K. L., Park, S. J., Goren, A., Jiang, K., Ramos, A., van der Knaap, E., et al. (2017). Bypassing Negative Epistasis on Yield in Tomato Imposed by a Domestication Gene. *Cell* **169**:1142-1155.e12.

Stadler, L. J. (1928). Genetic effects of X-rays in maize. *Proc. Natl. Acad. Sci. U. S. A.* **14**:69–75.

Stenberg, J. A. (2017). A conceptual framework for integrated pest management. *Trends Plant Sci.* **22**:759–769.

Stuttmann, J., Barthel, K., Martin, P., Ordon, J., Erickson, J. L., Herr, R., Ferik, F., Kretschmer, C., Berner, T., Keilwagen, J., et al. (2021). Highly efficient multiplex editing: one-shot generation of 8× Nicotiana benthamiana and 12× Arabidopsis mutants. *Plant J.* **106**:8–22.

Tadele, Z. (2019). Orphan crops: their importance and the urgency of improvement. *Planta* **250**:677–694.

Tamisier, L., Lacombe, S., Caranta, C., Gallois, J. L., and Moury, B. (2023). *Viral Fitness and Evolution*. Springer International Publishing.

Tan, J., Zhang, F., Karcher, D., and Bock, R. (2019). Engineering of high-precision base editors for site-specific single nucleotide replacement. *Nat. Commun.* **10**:439.

Tang, Y., Abdelrahman, M., Li, J., Wang, F., Ji, Z., Qi, H., Wang, C., and Zhao, K. (2021). CRISPR/Cas9 induces exon skipping that facilitates development of fragrant rice. *Plant Biotechnol. J.* **19**:642–644.

Tanksley, S. D. (2004). The genetic, developmental, and molecular bases of fruit size and shape variation in tomato. *Plant Cell* **16 Suppl**:S181-9.

Tavert-Roudet, G., Anne, A., Barra, A., Chovin, A., Demaille, C., and Michon, T. (2017). The Potyvirus Particle Recruits the Plant Translation Initiation Factor eIF4E by Means of the VPg covalently Linked to the Viral RNA. *Mol. Plant. Microbe. Interact.* **30**:754–762.

Tian, S., Jiang, L., Cui, X., Zhang, J., Guo, S., Li, M., Zhang, H., Ren, Y., Gong, G., Zong, M., et al. (2018). Engineering herbicide-resistant watermelon variety through CRISPR/Cas9-mediated base-editing. *Plant Cell Rep.* **37**:1353–1356.

Tomato brown rugose fruit virus (2020).*Bull. OEPP* **50**:529–534.

Truniger, V., and Aranda, M. A. (2009). Recessive resistance to plant viruses. *Adv. Virus Res.* **75**:119–159.

Turnbull, C., Lillemo, M., and Hvoslef-Eide, T. A. K. (2021). Global Regulation of Genetically Modified Crops Amid the Gene Edited Crop Boom - A Review. *Front. Plant Sci.* **12**:630396.

Uppalapati, S. R., Ishiga, Y., Doraiswamy, V., Bedair, M., Mittal, S., Chen, J., Nakashima, J., Tang, Y., Tadege, M., Ratet, P., et al. (2012). Loss of abaxial leaf epicuticular wax in Medicago truncatula irg1/palm1 mutants results in reduced spore differentiation of anthracnose and nonhost rust pathogens. *Plant Cell* **24**:353–370.

Valli, A. A., Gallo, A., Rodamilans, B., López-Moya, J. J., and García, J. A. (2018). The HCPro from the*Potyviridae*family: an enviable multitasking Helper Component that every virus would like to have. *Mol. Plant Pathol.* **19**:744–763.

van de Wouw, M., Kik, C., van Hintum, T., van Treuren, R., and Visser, B. (2010). Genetic erosion in crops: concept, research results and challenges. *Plant Genet. Resour.* **8**:1–15.

Van Schie, C. C. N., and Takken, F. L. W. (2014). Susceptibility genes 101: How to be a good host. *Annu. Rev. Phytopathol.* **52**:551–581.

Vaucheret, H., and Voinnet, O. (2024). The plant siRNA landscape. *Plant Cell* **36**:246–275.

Veillet, F., Gaillard, C., Coutos-Thévenot, P., and La Camera, S. (2016). Targeting the AtCWIN1 Gene to Explore the Role of Invertases in Sucrose Transport in Roots and during Botrytis cinerea Infection. *Front. Plant Sci.* **7**:1899.

Veillet, F., Perrot, L., Chauvin, L., Kermarrec, M. P., Guyon-Debast, A., Chauvin, J. E., Nogué, F., and Mazier, M. (2019). Transgene-free genome editing in tomato and potato plants using Agrobacterium-mediated delivery of a CRISPR/Cas9 cytidine base editor. *Int. J. Mol. Sci.* **20**.

Veillet, F., Perrot, L., Guyon-Debast, A., Kermarrec, M.-P., Chauvin, L., Chauvin, J.-E., Gallois, J.-L., Mazier, M., and Nogué, F. (2020a). Expanding the CRISPR Toolbox in P. patens Using SpCas9-NG Variant and Application for Gene and Base Editing in Solanaceae Crops. *Int. J. Mol. Sci.* **21**.

Veillet, F., Durand, M., Kroj, T., Cesari, S., and Gallois, J. L. (2020b). Precision Breeding Made Real with CRISPR: Illustration through Genetic Resistance to Pathogens. *Plant Communications* **1**.

Venbrux, M., Crauwels, S., and Rediers, H. (2023). Current and emerging trends in techniques for plant pathogen detection. *Front. Plant Sci.* **14**:1120968.

Venezia, M., and Creasey Krainer, K. M. (2021). Current advancements and limitations of gene editing in orphan crops. *Front. Plant Sci.* **12**:742932.

Voloudakis, A. E., Kaldis, A., and Patil, B. L. (2022). RNA-based vaccination of plants for control of viruses. *Annu. Rev. Virol.* **9**:521–548.

Vu, T. V., Nguyen, N. T., Kim, J., Hong, J. C., and Kim, J.-Y. (2023). Prime editing: Mechanism insight and recent applications in plants. *Plant Biotechnol. J.* Advance Access published October 4, 2023, doi:10.1111/pbi.14188.

Wallace, J. G., Rodgers-Melnick, E., and Buckler, E. S. (2018). On the Road to Breeding 4.0: Unraveling the Good, the Bad, and the Boring of Crop Quantitative Genomics. *Annu. Rev. Genet.* **52**:421–444.

Walton, R. T., Christie, K. A., Whittaker, M. N., and Kleinstiver, B. P. (2020). Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* **368**:290–296.

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

Wang, M.-B., Masuta, C., Smith, N. A., and Shimura, H. (2012a). RNA silencing and plant viral diseases. *Mol. Plant. Microbe. Interact.* **25**:1275–1285.

Wang, Y., Ma, M., Xiao, X., and Wang, Z. (2012b). Intronic splicing enhancers, cognate splicing factors and context-dependent regulation rules. *Nat. Struct. Mol. Biol.* **19**:1044–1052.

Weber, R., Chung, M.-Y., Keskeny, C., Zinnall, U., Landthaler, M., Valkov, E., Izaurralde, E., and Igreja, C. (2020). 4EHP and GIGYF1/2 Mediate Translation-Coupled Messenger RNA Decay. *Cell Rep.* **33**:108262.

Whitham, S., Dinesh-Kumar, S. P., Choi, D., Hehl, R., Corr, C., and Baker, B. (1994). The product of the tobacco mosaic virus resistance gene N: similarity to toll and the interleukin-1 receptor. *Cell* **78**:1101–1115.

Wittmann, S., Chatel, H., Fortin, M. G., and Laliberté, J.-F. (1997). Interaction of the viral protein genome linked of turnip mosaic Potyvirus with the translational eukaryotic initiation factor (iso) 4E ofArabidopsis thalianaUsing the yeast two-hybrid system. *Virology* **234**:84–92.

Wolter, M., Hollricher, K., Salamini, F., and Schulze-Lefert, P. (1993). The mlo resistance alleles to powdery mildew infection in barley trigger a developmentally controlled defence mimic phenotype. *Mol. Gen. Genet.* **239**:122–128.

Xiang, Q., Lott, A. A., Assmann, S. M., and Chen, S. (2021). Advances and perspectives in the metabolomics of stomatal movement and the disease triangle. *Plant Sci.* **302**:110697.

Xu, Y.-C., and Guo, Y.-L. (2020). Less Is More, Natural Loss-of-Function Mutation Is a Strategy for Adaptation. *Plant Commun* **1**:100103.

Xu, C., Liberatore, K. L., MacAlister, C. A., Huang, Z., Chu, Y.-H., Jiang, K., Brooks, C., Ogawa-Ohnishi, M., Xiong, G., Pauly, M., et al. (2015). A cascade of arabinosyltransferases controls shoot meristem size in tomato. *Nat. Genet.* **47**:784–792.

Xue, C., and Greene, E. C. (2021). DNA Repair Pathway Choices in CRISPR-Cas9-Mediated Genome Editing. *Trends Genet.* **37**:639–656.

Yang, X., Li, Y., and Wang, A. (2021). Research Advances in Potyviruses: From the Laboratory Bench to the Field. *Annu. Rev. Phytopathol.* Advance Access published April 23, 2021, doi:10.1146/annurev-phyto-020620-114550.

Yaschenko, A. E., Alonso, J. M., and Stepanova, A. N. (2024). Arabidopsis as a model for translational research. *Plant Cell* Advance Access published February 27, 2024, doi:10.1093/plcell/koae065.

Yeam, I., Cavatorta, J. R., Ripoll, D. R., Kang, B.-C., and Jahn, M. M. (2007). Functional dissection of naturally occurring amino acid substitutions in eIF4E that confers recessive potyvirus resistance in plants. *Plant Cell* **19**:2913–2928.

Yoon, Y.-J., Venkatesh, J., Lee, J.-H., Kim, J., Lee, H.-E., Kim, D.-S., and Kang, B.-C. (2020). Genome editing of eIF4E1 in tomato confers resistance to pepper mottle virus. *Front. Plant Sci.* **11**:1098.

Zafirov, D., Giovinazzo, N., Bastet, A., and Gallois, J.-L. (2020). When a knockout is an Achilles' heel: Resistance to one potyvirus species triggers hypersusceptibility to another one in Arabidopsis thaliana. *Mol. Plant Pathol.* Advance Access published December 29, 2020, doi:10.1111/mpp.13031.

Zafirov, D., Giovinazzo, N., Bastet, A., and Gallois, J.-L. (2021). When a knockout is an Achilles' heel: Resistance to one potyvirus species triggers hypersusceptibility to another one in Arabidopsis thaliana. *Molecular Plant Pathology* **22**:334–347.

Zafirov, D., Giovinazzo, N., Lecampion, C., Field, B., Ducassou, J. N., Couté, Y., Browning, K. S., Robaglia, C., and Gallois, J.-L. (2023). Arabidopsis eIF4E1 protects the translational machinery during TuMV infection and restricts virus accumulation. *PLoS Pathog.* **19**.

Zandalinas, S. I., Fritschi, F. B., and Mittler, R. (2021). Global Warming, Climate Change, and Environmental Pollution: Recipe for a Multifactorial Stress Combination Disaster. *Trends Plant Sci.* **26**:588–599.

Zdrzałek, R., Stone, C., De la Concepcion, J. C., Banfield, M. J., and Bentham, A. R. (2023). Pathways to engineering plant intracellular NLR immune receptors. *Curr. Opin. Plant Biol.* **74**:102380.

Zhang, Y., Liang, Z., Zong, Y., Wang, Y., Liu, J., Chen, K., Qiu, J.-L., and Gao, C. (2016). Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA. *Nat. Commun.* **7**:12617.

Zhang, H., Mittal, N., Leamy, L. J., Barazani, O., and Song, B.-H. (2017). Back into the wild—Apply untapped genetic diversity of wild relatives for crop improvement. *Evol. Appl.* **10**:5–24.

Zhang, C., Wang, P., Tang, D., Yang, Z., Lu, F., Qi, J., Tawari, N. R., Shang, Y., Li, C., and Huang, S. (2019a). The genetic basis of inbreeding depression in potato. *Nat. Genet.* **51**:374–378.

Zhang, R., Liu, J., Chai, Z., Chen, S., Bai, Y., Zong, Y., Chen, K., Li, J., Jiang, L., and Gao, C. (2019b). Generation of herbicide tolerance traits and a new selectable marker in wheat using base editing. *Nat Plants* **5**:480–485.

Zhang, X., Zhang, Y., Wang, T., Li, Z., Cheng, J., Ge, H., Tang, Q., Chen, K., Liu, L., Lu, C., et al. (2019c). A Comprehensive Map of Intron Branchpoints and Lariat RNAs in Plants. *Plant Cell* **31**:956–973.

Zhang, J., Yu, H., and Li, J. (2023). De novo domestication: retrace the history of agriculture to design future crops. *Curr. Opin. Biotechnol.* **81**:102946.

Zohary, D. (2004). Unconscious selection and the evolution of domesticated plants. *Econ. Bot.* **58**:5–10.

Zong, Y., Song, Q., Li, C., Jin, S., Zhang, D., Wang, Y., Qiu, J.-L., and Gao, C. (2018). Efficient C-to-T base editing in plants using a fusion of nCas9 and human APOBEC3A. *Nat. Biotechnol.* **36**:950–953.

Zsögön, A., Čermák, T., Naves, E. R., Notini, M. M., Edel, K. H., Weinl, S., Freschi, L., Voytas, D. F., Kudla, J., and Peres, L. E. P. (2018). De novo domestication of wild tomato using genome editing. *Nat. Biotechnol.* Advance Access published October 1, 2018, doi:10.1038/nbt.4272.

Citation for figures

Belfort, M., and Bonocora, R. P. (2014). Homing endonucleases: From genetic anomalies to programmable genomic clippers. In *Methods in Molecular Biology*, pp. 1–26. Totowa, NJ: Humana Press.

Jaramillo-Mesa, H., and Rakotondrafara, A. M. (2023). All eggs in one basket: How potyvirus infection is controlled at a single cap-independent translation event. *Semin. Cell Dev. Biol.* **148–149**:51–61.(Belfort and Bonocora, 2014; Jaramillo-Mesa and Rakotondrafara, 2023)

Annexes

Plant Science 316 (2022) 111160

Contents lists available at ScienceDirect

Plant Science

journal homepage: www.elsevier.com/locate/plantsci

CRISPR-based knock-out of *eIF4E2* in a cherry tomato background successfully recapitulates resistance to pepper veinal mottle virus

Kyoka Kuroiwa [a,1], Christina Thenault [a,1], Fabien Nogué [b], Laura Perrot [c], Marianne Mazier [a], Jean-Luc Gallois [a,*]

[a] INRAE, GAFL, 84143 Montfavet, France
[b] Institut Jean-Pierre Bourgin, INRAE, AgroParisTech, Université Paris-Saclay, 78000 Versailles, France
[c] Toulouse Biotechnology Institute, Université de Toulouse, 135 avenue de Rangueil, 31077 Toulouse CEDEX 04, France

ARTICLE INFO

Keywords:
eIF4E
CRISPR-Cas9
Genome editing
Virus resistance
Tomato
Cas9-NG
Potyvirus

ABSTRACT

The host susceptibility factors are important targets to develop genetic resistances in crops. Genome editing tools offer exciting prospects to develop resistances based on these susceptibility factors, directly in the cultivar of choice.

Translation initiation factors 4E have long been known to be a susceptibility factor to the main genus of *Potyviridae*, potyviruses, but the inactivation of the *eIF4E2* gene has only recently been shown to provide resistance to some isolates of pepper veinal mottle virus (PVMV) in big-fruit tomato plants. Here, using CRISPR-Cas9-NG, we show how *eIF4E2* can be targeted and inactivated in cherry tomato plants. Three independent knockout alleles caused by indel in the first exon of *eIF4E2*, resulted in the complete absence of the eIF4E2 protein. All three lines displayed a narrow resistance spectrum to potyvirus, similar to the one described earlier for an *eIF4E2* EMS mutant of M82, a big-fruit tomato cultivar; the plants were fully resistant to PVMV-Ca31, partially to PVMV-IC and were fully susceptible to two isolates of PVY assayed: N605 and LYE84.

These results show how easily a resistance based on *eIF4E2* can be transferred across tomato cultivar, but also confirm that gene redundancy can narrow the resistances based on eIF4E knockout.

1. Introduction

The recently developed techniques of genome editing have already proved themselves to be essential tools for crop breeding. Among the various traits that can be modified through these techniques, resistance to pathogens appears as being particularly promising. One of the main ways to do so and generate resistance is by the ability to suppress or modify in the plant host genome the genes encoding the susceptibility factors that the pathogens do require to infect the plant. The modification of these susceptibility factors makes the plant resistant to the pathogen thereby helping reduce the use of pesticide during culture [1].

Among those susceptibility factors, the translation initiation factors eIF4E, which are encoded by a small multigene family, appear as very meaningful targets for crop breeding [2]. Initially, a translation initiation factor eIF4E was characterized among the natural variability of pepper (*Capsicum* species) as being responsible for resistance to potyviruses, the largest group of viruses affecting plants, including several isolates from potato virus Y (PVY) and tobacco etch virus (TEV) [3]. Indeed, many alleles of the *eIF4E1* gene have been selected during pepper domestication: they are characterized by non-synonymous mutations that lead to resistance to potyviruses. It was interesting to note that those mutations do not affect the role of eIF4E1 in translation initiation [4]. Translation initiation is essential for the plant physiology as it is an essential step to produce new proteins, so the selection of functional eIF4E factors could be a way of maintaining a trade-off between resistance and plant development during plant domestication. Following the characterization of *eIF4E* as a resistance gene to potyvirus in pepper, translation initiation factors have been shown to play a pivotal role in many crops by providing genetic resistance to potyviruses and related viruses [2,5].

The ubiquitous role of eIF4E in resistance to potyvirus makes it a very exciting model for translational biology across species, or how a function that has been revealed in one species can be transferred to another one [6,7]. While pepper was instrumental in studying the

* Corresponding author.
E-mail address: jean-luc.gallois@inrae.fr (J.-L. Gallois).
[1] These authors contributed equally.

https://doi.org/10.1016/j.plantsci.2021.111160
Received 20 September 2021; Received in revised form 22 November 2021; Accepted 17 December 2021
Available online 20 December 2021
0168-9452/© 2021 Elsevier B.V. All rights reserved.

co-evolution between natural eIF4E-based resistances and potyviruses, it is not well suited for biotechnology studies. On the other hand, the closely related Tomato (*Solanum lycopersicum*) was ideally suited and presents the same redundancy in the small multigene family encoding eIF4E as pepper does, with two genes encoding eIF4E proteins—eIF4E1 (Solyc03g005870) and eIF4E2 (Solyc02g021550)—and one gene encoding a smaller protein named eIFiso4E (Solyc09g090580) [8–10]. Besides being a model plant, it is also a crop of major economical importance at the forefront of biotechnological innovations [11]. Thanks to the numerous genetic tools available in this species, it was possible to generate eIF4E-based resistance in cultivated tomato accessions and compare them with the natural resistance alleles that had been characterized in wild accessions [10,12]. In addition to providing a better understanding of the virus/eIF4E interaction, the knowledge obtained can then be transferred to other cultures for which natural resistance to potyviruses is not available. Also, given the specifics of market requirements by the consumer, tomato posits essential challenges for the breeders, such as combining new resistances with other traits. Of those are fruit shape and size that make immediate appeal to the consumer [11].

Initial works to generate resistances based on eIF4E in tomato were carried out using gene inactivation, through gene silencing or by chemical mutagenesis through TILLING [8,13]. Interestingly, those studies revealed important limitation of the gene knock-out strategies. Natural resistances to the two important potyviruses affecting tomato, PVY and TEV, consist in non-synonymous mutations in *eIF4E1* that can be associated with a large resistance spectrum to several virus isolates, whereas the *eIF4E1* knock-out (KO) tomato plants did not mimic the same resistance spectrum. These KO plants displayed a very narrow resistance spectrum to one isolate of PVY (LYE90), and to pepper mottle virus (PepMoV) while being fully susceptible to other PVY and TEV isolates tested [8,9]. In addition, it was shown that, when *eIF4E1* is knocked-down or knocked-out, most PVY and TEV isolates can recruit eIF4E2, a very close homolog of eIF4E1. The emerging picture shows that in tomato, some potyvirus isolates can selectively hijack either eIF4E1 or eIF4E2 for their infectious cycle, whereas most can recruit both. Unfortunately, knocking out both *eIF4E1* and *eIF4E2* genes cannot be an option for tomato improvement because of detrimental effects on plant growth and development [9,13]. Interestingly, it was shown recently, that some isolates of pepper veinal mottle virus (PVMV), which is an important constraint on tomato production, did rely on eIF4E2 alone and big-fruit tomato plants knocked out for *eIF4E2* were resistant to several isolates of PVMV, making this gene an important target to develop resistance trait in tomato [14].

The rapid development of genome editing technologies, based on the use of the CRISPR-Cas9 system soon appeared as an ideal tool to knock out genes in a given genotype. The expression *in planta* of a specific guide RNA and the Cas9 nuclease can lead to insertion or deletion of nucleotides at the targeted site in the genome. This system has been widely used to target gene coding sequence: the eventual indel leading to frameshift or inaccurate splicing results in incorrect protein translation, creating loss-of-function alleles. Interestingly those edited lines can be transgene-free by segregating away or not integrating the transgene harboring the Cas9 and sgRNA machinery [15,16]. Such approaches have already been used to target translation initiation factors 4E. CRISPR-Cas9 inactivation of *eIFiso4E* in *Arabidopsis thaliana*, allowed to generate a resistance to turnip mosaic virus similar to the one that had been previously characterized using chemical or insertional mutation of the same gene [17–19]. In cucumber (*Cucumis sativa*), inactivation of eIF4E resulted in resistance to zucchini yellow mosaic virus and papaya ring spot mosaic virus-W [20]. Likewise, inactivation of the two genes encoding the atypical translation initiation factors nCBP1 and nCBP2 resulted in reduced susceptibility and reduced symptoms severity to cassava brown streak disease (CBSV) in cassava [21].

Recently, two studies in tomato showed that inactivation through a CRISPR-Cas9 strategy of *eIF4E1* leads to a narrow resistance spectrum to potyvirus [22,23]. In order to complete these studies, we present here the inactivation through CRISPR-Cas9 of *eIF4E2*. We could demonstrate that tomato plants knocked out for the *eIF4E2* gene are fully resistant to one isolate of PVMV but only partially to another one and fully susceptible to the two tested PVY isolates. These results confirm and complete what has been observed in big-fruit tomato plants in which the *eIF4E2* gene was inactivated through EMS [9,14]. Finally, by developing the *eIF4E2* KO mutant in a cherry fruit accession, we show how resistance to PVMV can be easily translated into a genome background of choice using a CRISPR-Cas9 strategy.

2. Materials and methods

2.1. Plant material

West Virginia 106 (WVa106), a cherry fruit tomato (*Solanum lycopersicum* var. *cerasiforme*) which is highly susceptible to potyvirus, was used as background for the editing of *eIF4E2* (Solyc02g021550). The construct and methods to edit *eIF4E2* has been described previously [24]. Briefly, two closely located loci of *eIF4E2* first exon were targeted through *Agrobacterium*-mediated transformation, using a construct expressing two guide RNAs and a *Streptococcus pyogenes* Cas9 NG (SpCas9-NG). Primary transformants edited in *eIF4E2* were selected through High Resolution Melting (HRM) analysis. The TILLING tomato lines $eIF4E1^{KO}$ and $eIF4E2^{KO}$ are knocked out for *eIF4E1* (Solyc03g005870) and *eIF4E2*, respectively. They have been obtained through Ethyl Methyl Sulfonate (EMS) mutagenesis in the *Solanum lycopersicum* M82 background and their characterization has been described earlier [8,9]. All plants were grown in a growth chamber at 22 °C during the day, 20 °C during the night, 70 % relative humidity and 16 h day length.

2.2. Genotyping and selection of edited tomato lines

Genomic DNA was extracted from fresh leaf samples according to [25].

Detection of the stable integration of the T-DNA was detected by PCR analysis from genomic DNA using the following 6 primers: AHAF, AHAR, Cas9F, Cas9R, kanaF and KanaR in the same 20 µL reaction containing the GoTaq® G2 Flexi DNA Polymerase (Promega, Madison, USA). All primer sequences are available in Supplementary data (Appendix A).

The High Resolution Melting-curve analysis was mostly performed as described in [26]. Briefly, PCR amplification of small amplifications products (131 bp) was carried out from genomic DNA using 4e2HRMF and 4e2HRMR primers and the Precision Melt Supermix (BioRad, Hercules, USA) with the CFX96™ Real-Time PCR Detection System (BioRad, Hercules, USA), according to the manufacturer's instructions. Melting-curves analysis was performed using the Precision Melt Analysis software (BioRad, Hercules, USA). Sequence information was obtained by amplifying the target locus with 4e2F and 4e2R primers and GoTaq® G2 Flexi DNA Polymerase (Promega, Madison, USA). PCR products were purified and Sanger sequenced (Genoscreen, Lille, France).

2.3. Total RNA extraction and cDNA synthesis

Total RNA was extracted from young tomato leaf tissue (< 100 mg) of each plant in 500 µL of TRI-Reagent (Sigma-Aldrich, St. Louis, USA) and quality-checked by concentration measurement on Nanodrop ND-1000 (ThermoFisher, Waltham, USA). Reverse transcription (RT) of

the extracted RNA sample (1 μg) was performed using AMV (Avian Myeloblastosis Virus) Reverse Transcription Kit (Invitrogen, Waltham, USA) and oligo-dT18 (1 μM) to obtain coding DNA (cDNA).

The cDNA for *eIF4E2* and control *GAPDH* (Glyceraldehyde-3-Phosphate Dehydrogenase) were amplified by PCR using primer pairs Z1379F/Z1380F and Q3F/Q3R, respectively, and Sanger sequenced (Genoscreen, Lille, France).

2.4. Protein extraction and western blot analysis

Total protein was extracted from 100 mg of young tomato leaves of each plant, and 150 μL of Laemmli 2x DTT buffer was added to the crushed leaves. The homogenized solution was boiled for 5 min. The supernatant after centrifugation at 4 °C was collected and conserved at −20 °C.

The extracted total protein samples were run on the 12 % polyacrylamide gel in the SDS-PAGE system at 80 V for 2.5 h to separate proteins by their sizes. Then, the separated proteins were transferred from the gel on to the nitrocellulose membrane (GE Healthcare, Chicago, USA) using semi-dry transfer method in 30−45 min. The membranes were blocked with TBS 1x + 5 % milk overnight and then incubated with primary antibody for 2 h and after with secondary antibody for 1 h. The anti-Sl-eIF4E2 [9] serum was diluted 1/500, and combined with secondary goat horseradish peroxidase-labelled anti-rabbit serum (Sigma-Aldrich, St. Louis, USA) diluted 1/2000. As loading controls, monoclonal anti-plant actin antibodies (1/10,000 dilution) (Sigma-Aldrich, St. Louis, USA) were used with horseradish peroxidase-labelled rabbit anti-mouse serum (1/10,000 dilution) (Sigma-Aldrich, St. Louis, USA). The results of immunoblotting were revealed on autoradiography films Blu-Lite (MTC Bio, Sayreville, USA) using LumiGLO Reserve® Chemiluminescent Substrate Kit (LGC Seracare, Milford, USA).

2.5. 3D modelization of eIF4E2 structure

Homology modelling of the eIF4E2 protein were carried out as previously [14] using the YASARA software (http://www.yasara.org/), using structural data from pea (*Pisum sativum*) eIF4E (GenBank ID: AY423375, PDB ID: 2WMC-C) as the template. Protein structure was visualized using PyMol software (https://pymol.org/).

2.6. Virus isolates and infectious assays

Potyvirus infection assays were performed using PVY-N605 [27], PVY-LYE84 [28], PVMV isolates Cameroon 31 (PVMV-Ca31) and Ivory Coast (PVMV-IC) [14]. Isolates were maintained in *Nicotiana benthamiana* plants 12 days before inoculation.

Inocula were prepared from 1 g (fresh weight) of the infected tobacco leaves ground with 4 mL of grinding buffer: potassium phosphate buffer (0.03 M, pH = 7) with 0.2 % of diethyldithiocarbamate, 80 mg active charcoal and 80 mg carborundum. Six plants per genotype were mechanically inoculated by manually rubbing cotyledons of 14 day-old tomato plants with inoculum extract and rinsed with water 5 min later. Infected tomato plants were grown in a growth chamber in the same condition of photoperiod and temperature as before inoculation. At 21 days post-inoculation (dpi), 1 g of leaves from each plant was collected and ground in 4 mL of grinding buffer. For the PVMV tests only, the extracts were subjected to a serial dilution (1:25) with the same grinding buffer. To evaluate the accumulation for each virus, DAS-ELISA were performed using anti-PVY and anti-PVMV antisera (Sediag, www.sediag.com) according to the manufacturer's instructions. A pool of non-inoculated plants was used as healthy control. The absorbance was measured by a spectrophotometer at wavelength 405 nm.

2.7. Statistical analysis

Kruskal–Wallis non-parametric tests were performed to identify significant differences between genotypes using the pgirmess package on the free software R (http://www.r-project.org/).

3. Results

3.1. Confirmation of efficient editing at two targets in eIF4E2 in the cherry tomato WVa106 background

We previously reported the editing at two closely located sequences of the tomato *SleIF4E2* gene, using a rationally engineered SpCas9 variant, the SpCas9-NG, that recognizes NGN protospacer adjacent motifs [24,29]. Based on the sequencing of both targeted regions in *eIF4E2* on primary tomato plant transformants [24], three independent T0 regenerated individuals were selected with potential mutations (indel) at both sites. The plants were allowed to self in greenhouse for seed production. Two successive screenings on T1 and T2 individualized progenies, monitored by HRM and sequencing, resulted in selecting T2 genotypes homozygous for the modifications in *eIF4E2* gene (Supplementary Fig. 1). A third selfing produced T3 seeds in which the T-DNA–harboring the SpCas9-NG construct—has been eliminated by segregation, as verified by PCR (Supplementary Fig. 2). The resulting plants are therefore edited but do not contain the transgene.

The three independent edited lines are hereafter named as 2A, 2B and 2C, in which genome modifications (insertion or deletion of bases) had occurred at both target regions. The respective alleles will be named $eIF4E2^{2A}$, $eIF4E2^{2B}$ and $eIF4E2^{2C}$. Sequence data from the screening was re-analyzed closely to determine the indels caused by genome editing (Fig. 1). Finally, cDNAs were amplified from all three lines to check 1/ that the expression of the *eIF4E2* was not significantly affected, and 2/ that the mutations were indeed present on the expressed mRNA (Supplementary Fig. 3). The *eIF4E2* cDNA from the $eif4e2^{KO}$ EMS mutant, that had been isolated in the M82 big-fruit tomato by TILLING, was sequenced and showed accordingly a G254A mutation resulting in a tryptophan to stop mutation at position 85 [8,9].

3.2. eIF4E2 encoded-protein may suggest different outcomes for the three edited alleles

Based on the cDNA sequences obtained, the resulting translation product was predicted for each allele $eIF4E2^{2A}$, $eIF4E2^{2B}$ and $eIF4E2^{2C}$, respectively (Fig. 2). For the three edited alleles, a single-base insertion at the first targeted site leads to a frameshift in translation resulting in AA changes starting at the 35th residue. However, based on modification occurring at the second target, two cases could be considered.

For $eIF4E2^{2A}$ and $eIF4E2^{2C}$, the modifications occurring at the second targeted region did not restore the frame and the frameshift resulted in a premature stop codon at AA position 127 and 126 in the exon 2 out of total 221 amino acids, respectively. As a result, only the 34 first AA of the wild-type eIF4E2 are conserved in plants 2A and 2C, and more than 40 % of the C-terminus of the protein is missing. Based on that, *eIF4E2* alleles of 2A and 2C can be expected to be loss-of-function.

On the contrary, in plant 2B, the frameshift caused by the insertion of an A in target 1 was shifted back to the correct reading frame due to the deletion of another A in target 2. Thus, $eIF4E2^{2B}$ encodes a predicted protein with modification of 14 amino acids within exon1-encoded region of the eIF4E2 protein, in the N-term region, while the rest of the protein–spanning AA 49 to 221—is similar to the wild-type eIF4E2 protein. The wild-type *eIF4E2* cDNA encodes a 221 Amino Acid (AA) protein, whose structure can be modelized based on its strong homologies with crystallized eIF4E proteins (Supplementary Fig. 4, [14]). As the structure of the 54 first amino acids of eIF4E protein have not been resolved by protein 3D crystallization, it is not possible to foresee whether this limited peptide change (highlighted in yellow on Supplementary Fig. 4) will have any effect on the protein expression or function.

Overall, protein predictions suggest that out of the three selected

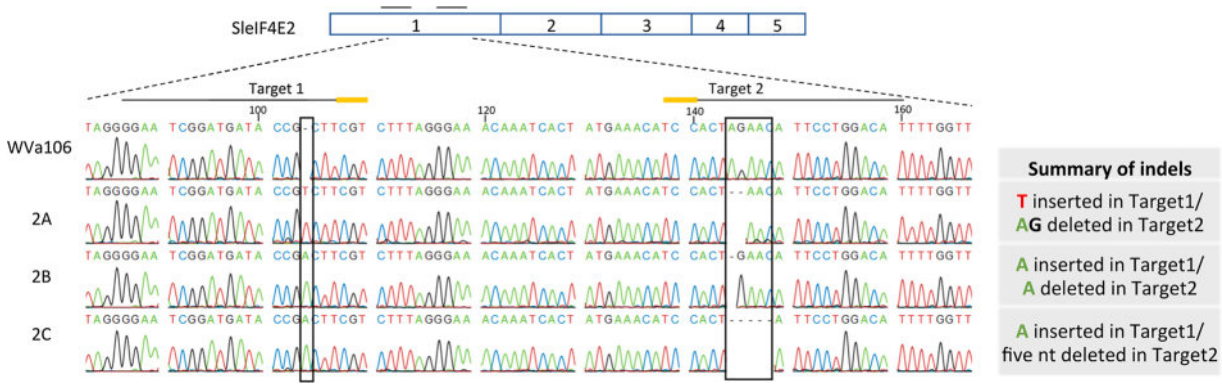


Fig. 1. Editing at two targets in *eIF4E2* exon1 caused nucleotide insertion or deletion in the genomic *eIF4E2*.
Genomic sequence of alleles of *eIF4E2* CRISPR mutants from lines 2A, 2B, and 2C. Target 1 and Target 2 represent the sequences of the two guide RNAs used in [24] and yellow bar indicates the PAM sequence for each target. The blue numbered boxes indicate the five exons of *eIF4E2*. Nucleotide position in respect to the WVa106 sequence is indicated at the top of the alignment. Summary of the modifications in each line is indicated on the right. nt: nucleotide.

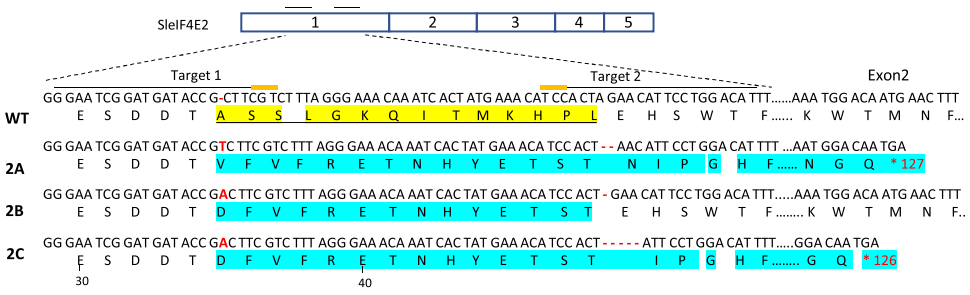


Fig. 2. Indels in *eIF4E2* exon1 induced by CRISPR/Cas9 cause extensive amino acid changes and premature stop codons.
Genomic sequence of alleles of *eIF4E2* CRISPR mutants from lines 2A, 2B, and 2C. Target 1 and Target 2 represent the sequences of the two guide RNAs used in [24] and yellow bar indicates the PAM sequence for each target. The blue numbered boxes indicate the five exons of *eIF4E2*. Red letters or hyphens are the insertions or deletions (edited sites) in respect to WT sequence in the 2A, 2B, and 2C genomic sequences. Red asterisks represent the premature stop codons in plant 2A and 2C. Highlighted in blue are amino acid substitutions in respect to WT amino acid sequence in 2A, 2B and 2C. Underlined (highlighted in yellow) is the amino acid sequence of WT that corresponds to the 14 AA modified in 2B. Amino acid position number is indicated at the bottom.

alleles, at least two are full loss-of-function alleles, while the question remains open on the third one.

3.3. *Protein expressed from the three edited eIF4E2 alleles do not accumulate in planta*

To confirm the loss-of-function of alleles *eIF4E2*^{2A}and *eIF4E2*^{2C,} as well as to investigate the potential effect of the stretch of 14 AA changes in the N-term part of eIF4E2^{2B}, the accumulation of the eIF4E2 protein was investigated in tomato leaf tissues from the three genotypes. To this aim, we used a specific polyclonal antibody that had been raised against the tomato eIF4E2 full protein. The serum has been shown to be highly specific for the tomato eIF4E2 protein, as attested by analysis of protein accumulation in wild type plants and in the EMS *eif4e2*^{KO} EMS mutant ([9] and Fig. 3, first two lanes). Analysis of eIF4E2 protein accumulation in leaf tissues showed that none of the three edited alleles mutants accumulate any detectable eIF4E2 protein (Fig. 3). While this could be expected from plants carrying the *eIF4E2*^{2A}or *eIF4E2*^{2C} alleles, the same holds true for *eIF4E2*^{2B}. This suggests that the 14 AA changes in the N-term part of the protein is sufficient to suppress the eIF4E2 protein accumulation, probably by triggering its degradation. In conclusion, all three edited alleles can be considered as independent loss-of-function alleles for *eIF4E2* in the cherry tomato WVa106.

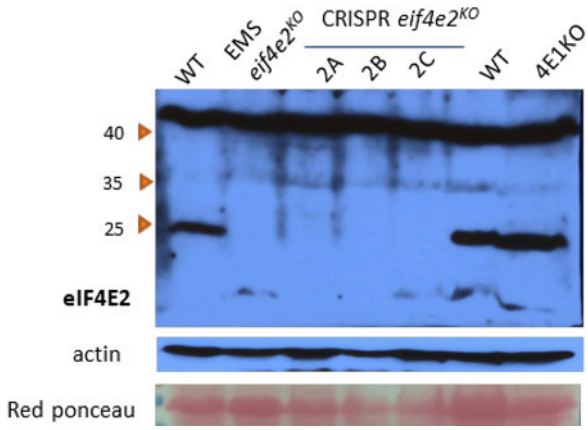


Fig. 3. All three *eIF4E2* CRISPR-edited alleles are loss-of function alleles that do not accumulate eIF4E2.
Analysis of eIF4E2 protein accumulation in tomato leaves by western blot. WT, the wildtype WVa106 tomato and the EMS *eif4e2*^{KO} mutant, are used as positive and negative controls, respectively, for eIF4E2 protein detection. The band relative to eIF4E2 stands at 25 kDa. Actin detection and red ponceau are shown as loading controls.

3.4. *All three edited eIF4E2 alleles are associated with resistance to some PVMV isolates*

We next assessed whether the inactivation of *eIF4E2* in the cherry tomato WVa106 was associated with resistance to PVMV as it has been shown in big-fruit tomato. Previously, we showed that the *eif4e2*^{KO} EMS plants showed very contrasted response to different isolates of PVMV, ranging from full resistance, intermediate resistance, to full susceptibility [14]. Out of those isolates we selected two, Cameroon 31 (Ca31) and Ivory Coast (IC), towards which the *eif4e2*^{KO} EMS plants responded with full and intermediate resistance, respectively. The edited plants were challenged with these isolates and the virus accumulation was

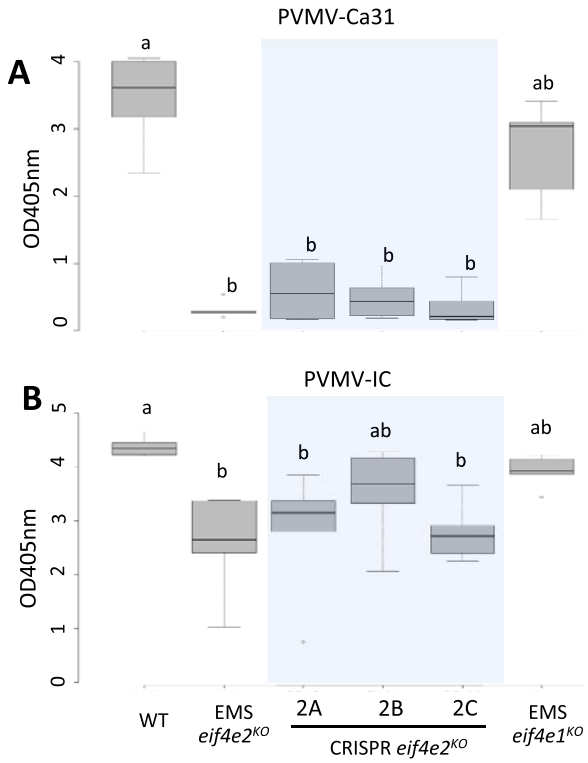


Fig. 4. *eIF4E2* CRISPR-edited loss-of function alleles are associated with resistance to PVMV-Ca31 but not to PVMV-CI.
(A, B) Resistance assay to PVMV isolates in tomato plants. The accumulation of viral coat proteins was detected in tomato plant tissues by DAS-ELISA from plants challenged with PVMV-Ca31 (A) or PVMV-IC (B). Samples were collected at 21 dpi. Different alphabets indicate the statistically significant difference ($p < 0.05$) after Kruskal-Wallis analysis. Six plants were tested per genotype.

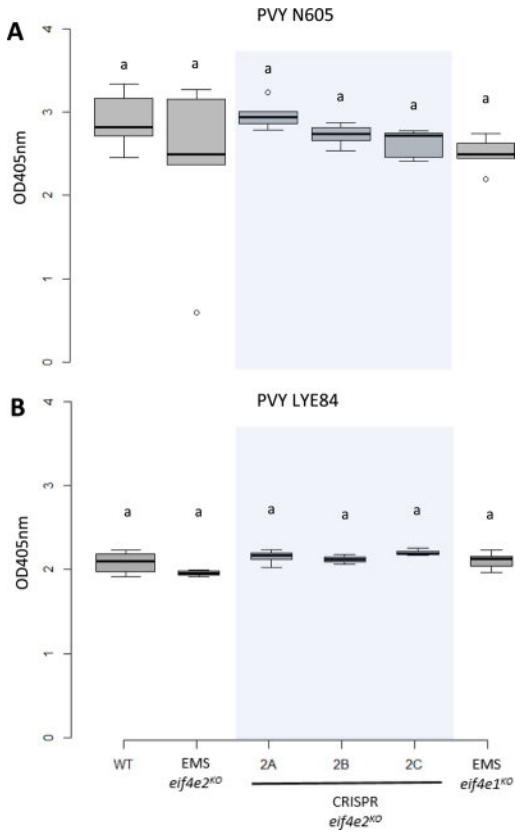


Fig. 5. *eIF4E2* CRISPR-edited loss-of function alleles are fully susceptible to PVY-N605 and PVY-LYE84.
(A, B) Resistance assay to PVY isolates in tomato plants. The accumulation of viral coat proteins was detected by DAS-ELISA in tomato plant tissues from plants challenged with PVY-N605 (A) or PVY-LYE84 (B). Samples were collected at 21 dpi. Different alphabets indicate the statistically significant difference ($p < 0.05$) after Kruskal-Wallis analysis. Six plants were tested per genotype.

assayed by DAS-ELISA after 21 days (Fig. 4). Results showed that the three edited lines, carrying either *eif4e2*^{2A}, *eif4e2*^{2B}, or *eif4e2*^{2C}, behaved similarly to the *eif4e2*^{KO} EMS plants, by being fully resistant to PVMV-Ca31 while accumulating significantly less PVMV-IC virus. By directly editing *eIF4E2* using CRISPR tools, we show therefore that we can reproduce the specific resistance to PVMV isolates in another cultivar of choice.

In contrast to PVMV, Moury et al. [14] showed that the plants carrying the EMS knock out in *eIF4E2* are fully susceptible to all other isolates of potyvirus tested, including PVY. We therefore challenged the three *eIF4E2*-edited lines with two isolates of PVY, N605 and LYE84. We found all three lines to be fully susceptible to the two PVY isolates (Fig. 5). As for the *eif4e2*^{KO} EMS plants, it suggests that knocking out *eIF4E2* alone by CRISPR-Cas9 in the WVa106 background results in a very narrow resistance spectrum.

4. Discussion

In this work, using genome editing of the susceptibility allele *eIF4E2* in tomato, we show quick and efficient development of genetic resistance to PVMV based on CRISPR-Cas gene knockout.

Here, we used a very straightforward approach by targeting two regions of eIF4E2, using the recently developed SpCas9-NG. The use of a Cas9 with a broader PAM selection (NGN instead of NGG), makes it easier to directly target the first exon of *eIF4E2*, in order to avoid the production of truncated protein with remnant activities. Indeed, we could select three independent tomato lines harboring loss-of-function in *eIF4E2*, as attested by the absence of the encoded protein accumulation in leaf tissues. At the same time, by selecting efficient editing at both targeted regions, we could isolate two types of mutants. In the first type, editing resulted in frameshifts and subsequent premature stop codons, leading to loss-of function alleles. For the other type, we isolated one allele (*eIF4E2*^{2B}) where the frameshift caused at the first target is canceled by a nucleotide insertion at the second target, leading to a short stretch of AA changes (spanning AA 35–48). We show that this modification is sufficient to obliterate the accumulation of the edited *eIF4E2* in tomato leaves although its mRNA is correctly produced. Little is known about the N-term part of eIF4E proteins. The first N-term part of eIF4E1 has not been crystallized yet, as its presence is detrimental to the crystallization process [30–32]. In addition, linking other proteins to the N-term or C-term parts of eIF4E is detrimental to its function [33]. This could be caused either by destabilizing the protein or impairing protein-protein association. Here, our results show that modifying a short region of the N-term part of the eIF4E2 protein can result in its destabilization. Incidentally, this work here shows how combining several guides using low complexity PAM sequences could give opportunities to selectively mutate specific regions coding for domains, which could help characterizing their function.

One of the advantages of genome editing is the possibility to introduce mutations directly in the genetic background of choice. For the breeder, it implies the possibility to modify directly pre-breeding or breeding material and avoid reshuffling of traits that have already been carefully selected. Previous reports of edition of *eIF4E1*, the homolog of *eIF4E2*, have been carried out in the model *S. lycopersicum* cultivar Micro-Tom [22] and in a big-fruit tomato inbred line [23]. Here, we implement a similar strategy in a cherry-tomato fruit, WVa106, belonging to the *Solanum lycopersicum* var. *cerasiforme,* and very closely

related to the *Solanum pimpinellifolium* species [34]. Instead of eIF4E1, we decided to target eIF4E2, that has been previously characterized as a susceptibility factor to PVMV in a big-fruit tomato mutant issued from an EMS mutagenesis. We show here that we can easily translate such resistance from a big-fruit tomato to this new background. It has been documented that the genetic background of an accession can be instrumental for the efficiency of the genetic resistance trait or its durability (for review [35],). For recessive resistances such as the one relying on eIF4E, such effects are generally caused by modification of gene redundancy among the *eIF4E* gene family which can provide or remove alternative host factors that the virus hijacks. We show here that there is no genetic background effect between M82 and WVa106 affecting resistance to PVMV and thus, that there is not major reshuffling of redundancy between *eIF4E1* and *eIF4E2* between M82 and WVa106. This is in line with the very close relationship between *S. lycopersicum* and cherry tomatoes, such as WVa106 which results from the admixture of the *S. lycopersicum* and *S. pimpinellifolium* genomes [34]. However, it will be interesting to see whether these two genes, *eIF4E1* and *eIF4E2*, are fully redundant or not during the plant development: it has been shown that an EMS loss-of-function of *eIF4E2* is associated with a slight developmental delay [9], and it will be interesting to see whether this effect is conserved in WVa106 background.

Potyviruses have long been known to selectively target specific translation initiation factors eIF4E or eIFiso4E [2,5]. In tomato, such knowledge has been acquired through different biotechnological approaches, ranging from RNA interference, TILLING to now genome editing. These studies showed that most potyvirus isolates are able to recruit either eIF4E1 or eIF4E2 in tomato. A smaller subset of isolates does rely on eIF4E1 alone (PVY-LYE90 and PepMoV) while it was shown recently that some isolates of PVMV did rely on eIF4E2 alone [8,9,13,14]. These virus/host factor relationships are confirmed by the recent analysis of *eIF4E1* and *eIF4E2* knock-out alleles generated by genome editing ([22,23] ; this study). Concerning the *eIF4E1* CRISPR-Cas9 knock-out Micro-Tom lines produced by Yoon et al. [22], although the eIF4E1 protein accumulation was not confirmed in these lines, the deletions caused by the edition presumably induced premature stop codon in the edited alleles. The resulting plants were found to be resistant to PepMoV, but not to TEV. This is in accordance with results obtained previously on tomato lines generated through TILLING and knocked-out for the same *eIF4E1* gene [8]. In those plants, a splicing defect caused by an EMS mutation resulted in the absence of eIF4E1 protein accumulation. Those plants were fully resistant to PepMoV as well. From the more extensive resistance spectrum assayed with the TILLING lines, it can be expected that the plants edited for eIF4E1 will be also resistant to one isolate of PVY-LYE90, but susceptible to most PVY isolates and all TEV isolates tested so far [8,9]. Concerning the *eIF4E2* CRISPR-Cas9 knock-out, we could confirm in this study the very limited resistance spectrum observed in a *eIF4E2* knock-out obtained by EMS [14]. The three independent lines we generated showed full resistance to one PVMV isolate (PVMV-Ca31), intermediate resistance to another one (PVMV-IC) and full susceptibility to two PVY isolates -N605 and LYE84 that are known to rely on eIF4E1 and eIF4E2. As previously shown for other plant/potyviruses pathosystems [9,14,36], it is very likely that differences between the PVMV genome sequences, and variability in the proteins they encode, are the basis for these differences in resistance by modulating the virus selectivity for one or several eIF4E factors.

Our results, together with those from Yoon et al. [22] on tomato, highlight the fact that CRISPR-based genome editing is not going to be the silver bullet to achieve genetic resistances if we limit ourselves to gene knock-out. Beyond the very exciting opportunities provided by the CRISPR technology, what will matter is the mechanism by which the virus or other pathogens do subvert the host machinery. In the case of eIF4E-based resistance, the redundancy among 4E factors, as well as the virus's ability to break down resistance by recruiting other 4E homologs should be taken into considerations. As we previously stated, by taking advantage of the coevolution that happened during crop domestication, *eIF4E* natural resistance alleles could help provide a blueprint for generating resistance alleles while avoiding the drawback associated with loss-of-functions such as limited resistance spectrum, higher resistance breaking, or over-susceptibility mechanism [2,9,37]. It is all the more because pyramiding KO in 4E genes, that could extend resistance, is likely to be associated with plant developmental default due to the importance of translation initiation in the plant cell physiology [2,38,39]. On the other hand, base editing techniques, allowing precise modification of the gene coding sequence could help reproduce or mimic natural resistance without impairing the eIF4E translation initiation. A proof of concept for such an approach has been done in *Arabidopsis*, where a single non-synonymous mutation in the *eIF4E1* gene was found to be sufficient to trigger resistance to clover yellow vein virus while not impairing the plant yield [40]. A recent study provides exciting prospect along this line: by targeting the genomic sequence of *eIF4E1* in a tomato breeding line, Atarashi et al., [23] achieved targeted deletion of one or two codon in a specific region of the encoded eIF4E1 protein. Strikingly, the targeted mutation managed to reduce significantly the plant susceptibility to cucumber mosaic virus, a cucumovirus. Likewise, it is tempting to expect that a combination of multiplexed base editing, prime editing, or codon deletion are likely to provide new alleles of interest by generating a high level of diversity in the genetic backgrounds of choice. There again, by constituting a master plan for designing genes, it is to be expected that such strategy can be extended to other susceptibility factors to pathogens. This is of major importance as other host susceptibility factors targeted by pathogens may have very important functions for the plant development as well [41].

Overall, the feasibility of transferring resistances by loss-of-susceptibility based on CRISPR-Cas9 technologies is well acted, but idiosyncratic properties of these susceptibility factors can jeopardize their efficiency in the field. A major challenge for the years to come will be to assess on a case-by-case the efficiency to deliver efficient resistance traits for plant breeding. To do so, it will be worth providing more extensive resistance spectrum studies to characterize resistances, and at term one should start tackling the durability of those resistances.

Data availability

Raw data and other supplementary material are available at the following repository: https://osf.io/7zpcj.

Author contributions

FN, MM and JLG conceived and designed the experiments. KK, CT, LP, MM performed and interpreted the experiments. KK and CT designed the figures. JLG wrote the manuscript with inputs from all authors. All authors have approved the final manuscript.

Funding

This work was supported by the French National Research Agency (ANR11-BTBR-0001-GENIUS). The IJPB benefits from the support of the LabEx Saclay Plant Sciences-SPS (ANR-10-LABX-0040-SPS). KK was supported by the Ito foundation (Japan, N0257) and Bourses du Gouvernement Français 2020–2021 (B-20001 JCA/NH).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors thank Benoît Moury for advising and supplying virus isolates as well as for helpful advices for ELISA, Mathilde Causse for

discussions, Delyan Zafirov for advices on western blots, Emmanuel Botton for taking care of tomato plants and Nathalie Truglio for supplying tobacco plants.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.plantsci.2021.111160.

References

[1] F. Veillet, M. Durand, T. Kroj, S. Cesari, J.-L. Gallois, Precision breeding made real with CRISPR: illustration through genetic resistance to pathogens, Plant Commun. 1 (2020), https://doi.org/10.1016/j.xplc.2020.100102.

[2] A. Bastet, C. Robaglia, J.L. Gallois, eIF4E resistance: natural variation should guide gene editing, Trends Plant Sci. 22 (2017) 411–419, https://doi.org/10.1016/j.tplants.2017.01.008.

[3] S. Ruffel, M.H. Dussault, A. Palloix, B. Moury, A. Bendahmane, C. Robaglia, C. Caranta, A natural recessive resistance gene against potato virus Y in pepper corresponds to the eukaryotic initiation factor 4E (elF4E), Plant J. 32 (2002), https://doi.org/10.1046/j.1365-313X.2002.01499.x.

[4] C. Charron, M. Nicolaï, J.L. Gallois, C. Robaglia, B. Moury, A. Palloix, C. Caranta, Natural variation and functional analyses provide evidence for co-evolution between plant eIF4E and potyviral VPg, Plant J. 54 (2008), https://doi.org/10.1111/j.1365-313X.2008.03407.x.

[5] C. Robaglia, C. Caranta, Translation initiation factors: a weak link in plant RNA virus infection, Trends Plant Sci. 11 (2006) 40–45, https://doi.org/S1360-1385(05)00292-X [pii], https://doi.org/10.1016/j.tplants.2005.11.004.

[6] C. Schmitt-Keichinger, Manipulating cellular factors to combat viruses: a case study from the plant eukaryotic translation initiation factors eIF4, Front. Microbiol. 10 (2019), https://doi.org/10.3389/fmicb.2019.00017.

[7] P. Jacob, A. Avni, A. Bendahmane, Translational research: exploring and creating genetic diversity, Trends Plant Sci. 23 (2018), https://doi.org/10.1016/j.tplants.2017.10.002.

[8] F. Piron, M. Nicolaï, S. Minoïa, E. Piednoir, A. Moretti, A. Salgues, D. Zamir, C. Caranta, A. Bendahmane, An induced mutation in tomato eiF4E leads to immunity to two potyviruses, PLoS One 5 (2010), https://doi.org/10.1371/journal.pone.0011313.

[9] C. Gauffier, C. Lebaron, A. Moretti, C. Constant, F. Moquet, G. Bonnet, C. Caranta, J.L. Gallois, A TILLING approach to generate broad-spectrum resistance to potyviruses in tomato is hampered by eIF4E gene redundancy, Plant J. 85 (2016) 717–729, https://doi.org/10.1111/tpj.13136.

[10] C. Lebaron, A. Rosado, C. Sauvage, C. Gauffier, S. German-Retana, B. Moury, J.-L. Gallois, A new eIF4E1 allele characterized by RNAseq data mining is associated with resistance to potato virus Y in tomato albeit with a low durability, J. Gen. Virol. 97 (2016) 3063–3072, https://doi.org/10.1099/jgv.0.000609.

[11] C. Rothan, I. Diouf, M. Causse, Trait discovery and editing in tomato, Plant J. 97 (2019), https://doi.org/10.1111/tpj.14152.

[12] S. Ruffel, J.L. Gallois, M.L. Lesage, C. Caranta, The recessive potyvirus resistance gene pot-1 is the tomato orthologue of the pepper pvr2-eIF4E gene, Mol. Genet. Genomics 274 (2005), https://doi.org/10.1007/s00438-005-0003-x.

[13] M. Mazier, F. Flamain, M. Nicolaï, V. Sarnette, C. Caranta, Knock-down of both eIF4E1 and eIF4E2 genes confers broad-spectrum resistance against potyviruses in tomato, PLoS One 6 (2011), https://doi.org/10.1371/journal.pone.0029595.

[14] B. Moury, C. Lebaron, M. Szadkowski, M. Ben Khalifa, G. Girardot, B.A. Bolou Bi, D. Koné, L.W. Nitiema, H. Fakhfakh, J.L. Gallois, Knock-out mutation of eukaryotic initiation factor 4E2 (eIF4E2) confers resistance to pepper veinal mottle virus in tomato, Virology 539 (2020), https://doi.org/10.1016/j.virol.2019.09.015.

[15] B. Danilo, L. Perrot, K. Mara, E. Botton, F. Nogué, M. Mazier, Efficient and transgene-free gene targeting using Agrobacterium-mediated delivery of the CRISPR/Cas9 system in tomato, Plant Cell Rep. 38 (2019), https://doi.org/10.1007/s00299-019-02373-6.

[16] F. Veillet, L. Perrot, L. Chauvin, M.-P. Kermarrec, A. Guyon-Debast, J.-E. Chauvin, F. Nogué, M. Mazier, Transgene-free genome editing in tomato and potato plants using agrobacterium-mediated delivery of a CRISPR/Cas9 cytidine base editor, Int. J. Mol. Sci. 20 (2019) 402, https://doi.org/10.3390/ijms20020402.

[17] A.D. Lellis, K.D. Kasschau, S.A. Whitham, J.C. Carrington, Loss-of-susceptibility mutants of Arabidopsis thaliana reveal an essential role for eIF(iso)4E during potyvirus infection, Curr. Biol. 12 (2002), https://doi.org/10.1016/S0960-9822(02)00898-9.

[18] A. Duprat, C. Caranta, F. Revers, B. Menand, K.S. Browning, C. Robaglia, The Arabidopsis eukaryotic initiation factor (iso)4E is dispensable for plant growth but required for susceptibility to potyviruses, Plant J. 32 (2002), https://doi.org/10.1046/j.1365-313X.2002.01481.x.

[19] D.E. Pyott, E. Sheehan, A. Molnar, Engineering of CRISPR/Cas9-mediated potyvirus resistance in transgene-free Arabidopsis plants, Mol. Plant Pathol. 17 (2016) 1276–1288, https://doi.org/10.1111/mpp.12417.

[20] J. Chandrasekaran, M. Brumin, D. Wolf, D. Leibman, C. Klap, M. Pearlsman, A. Sherman, T. Arazi, A. Gal-On, Development of broad virus resistance in non-transgenic cucumber using CRISPR/Cas9 technology, Mol. Plant Pathol. 17 (2016) 1140–1153, https://doi.org/10.1111/mpp.12375.

[21] M.A. Gomez, Z.D. Lin, T. Moll, R.D. Chauhan, L. Hayden, K. Renninger, G. Beyene, N.J. Taylor, J.C. Carrington, B.J. Staskawicz, R.S. Bart, Simultaneous CRISPR/Cas9-mediated editing of cassava eIF4E isoforms nCBP-1 and nCBP-2 reduces cassava brown streak disease symptom severity and incidence, Plant Biotechnol. J. 17 (2019), https://doi.org/10.1111/pbi.12987.

[22] Y.-J. Yoon, J. Venkatesh, J.-H. Lee, J. Kim, H.-E. Lee, D.-S. Kim, B.-C. Kang, Genome editing of eIF4E1 in tomato confers resistance to pepper mottle virus, Front. Plant Sci. 11 (2020) 1098, https://doi.org/10.3389/fpls.2020.01098.

[23] H. Atarashi, W.H. Jayasinghe, J. Kwon, H. Kim, Y. Taninaka, M. Igarashi, K. Ito, T. Yamada, C. Masuta, K.S. Nakahara, Artificially edited alleles of the eukaryotic translation initiation factor 4E1 gene differentially reduce susceptibility to cucumber mosaic virus and potato virus Y in tomato, Front. Microbiol. 11 (2020) 3075, https://doi.org/10.3389/fmicb.2020.564310.

[24] F. Veillet, L. Perrot, A. Guyon-Debast, M.P. Kermarrec, L. Chauvin, J.E. Chauvin, J. L. Gallois, M. Mazier, F. Nogué, Expanding the CRISPR toolbox in P. Patens using SpCas9-NG variant and application for gene and base editing in solanaceae crops, Int. J. Mol. Sci. 21 (2020), https://doi.org/10.3390/ijms21031024.

[25] T.M. Fulton, J. Chunwongse, S.D. Tanksley, Microprep protocol for extraction of DNA from tomato and other herbaceous plants, Plant Mol. Biol. Report 13 (1995), https://doi.org/10.1007/BF02670897.

[26] F. Veillet, C. Gaillard, P. Coutos-Thévenot, S. La Camera, Targeting the AtCWIN1 gene to explore the role of invertases in sucrose transport in roots and during Botrytis cinerea infection, Front. Plant Sci. 7 (2016), https://doi.org/10.3389/fpls.2016.01899.

[27] G. Parrella, S. Ruffel, A. Moretti, C. Morel, A. Palloix, C. Caranta, Recessive resistance genes against potyviruses are localized in colinear genomic regions of the tomato (Lycopersicon spp.) and pepper (Capsicum spp.) genomes, Theor. Appl. Genet. 105 (2002), https://doi.org/10.1007/s00122-002-1005-2.

[28] B. Moury, C. Morel, E. Johansen, L. Guilbaud, S. Souche, V. Ayme, C. Caranta, A. Palloix, M. Jacquemond, Mutations in Potato virus Y genome-linked protein determine virulence toward recessive resistances in Capsicum annuum and Lycopersicon hirsutum, Mol. Plant Microbe Interact. 17 (2004), https://doi.org/10.1094/MPMI.2004.17.3.322.

[29] H. Nishimasu, X. Shi, S. Ishiguro, L. Gao, S. Hirano, S. Okazaki, T. Noda, O. O. Abudayyeh, J.S. Gootenberg, H. Mori, S. Oura, B. Holmes, M. Tanaka, M. Seki, H. Hirano, H. Aburatani, R. Ishitani, M. Ikawa, N. Yachie, F. Zhang, O. Nureki, Engineered CRISPR-Cas9 nuclease with expanded targeting space, Science (80-.) 361 (2018) 1259–1262, https://doi.org/10.1126/SCIENCE.AAS9129.

[30] A.F. Monzingo, S. Dhaliwal, A. Dutt-Chaudhuri, A. Lyon, J.H. Sadow, D. W. Hoffman, J.D. Robertas, K.S. Browning, The structure of eukaryotic translation initiation factor-4E from wheat reveals a novel disulfide bond, Plant Physiol. 143 (2007), https://doi.org/10.1104/pp.106.093146.

[31] M. Miras, V. Truniger, C. Silva, N. Verdaguer, M.A. Aranda, J. Querol-Audí, Structure of eiF4E in complex with an eiF4G peptide supports a universal bipartite binding mode for protein translation, Plant Physiol. 174 (2017), https://doi.org/10.1104/pp.17.00193.

[32] J.A. Ashby, C.E.M. Stevenson, G.E. Jarvis, D.M. Lawson, A.J. Maule, Structure-based mutational analysis of eIF4E in relation to sbm1 resistance to Pea seed-borne mosaic virus in Pea, PLoS One 6 (2011), https://doi.org/10.1371/journal.pone.0015873.

[33] E. Levins, C.-Y. Tseng, R.M. Patrick, L.K. Mayberry, N. Cole, K.S. Browning, Fusion proteins of Arabidopsis cap-binding proteins: cautionary "tails" of woe, Translation 4 (2016), https://doi.org/10.1080/21690731.2016.1257408.

[34] N. Ranc, S. Mūos, S. Santoni, M. Causse, A clarified position for solanum lycopersicum var. cerasiforme in the evolutionary history of tomatoes (solanaceae), BMC Plant Biol. 8 (2008), https://doi.org/10.1186/1471-2229-8-130.

[35] J.L. Gallois, B. Moury, S. German-Retana, Role of the genetic background in resistance to plant viruses, Int. J. Mol. Sci. 19 (2018), https://doi.org/10.3390/ijms19102856.

[36] A. Bastet, B. Lederer, N. Giovinazzo, X. Arnoux, S. German-Retana, C. Reinbold, V. Brault, D. Garcia, S. Djennane, S. Gersch, O. Lemaire, C. Robaglia, J.L. Gallois, Trans-species synthetic gene design allows resistance pyramiding and broad-spectrum engineering of virus resistance in plants, Plant Biotechnol. J. 16 (2018) 1569–1581, https://doi.org/10.1111/pbi.12896.

[37] D. Zafirov, N. Giovinazzo, A. Bastet, J.-L. Gallois, When a knockout is an Achilles' heel: Resistance to one potyvirus species triggers hypersusceptibility to another one in Arabidopsis thaliana, Mol. Plant Pathol. 22 (2021) 334–347, https://doi.org/10.1111/mpp.13031.

[38] C. Callot, J.L. Gallois, Pyramiding resistances based on translation initiation factors in Arabidopsis is impaired by male gametophyte lethality, Plant Signal. Behav. 9 (2014), https://doi.org/10.4161/psb.27940.

[39] R.M. Patrick, L.K. Mayberry, G. Choy, L.E. Woodard, J.S. Liu, A. White, R. A. Mullen, T.M. Tanavin, C.A. Latz, K.S. Browning, Two arabidopsis loci encode novel eukaryotic initiation factor 4E isoforms that are functionally distinct from the conserved plant eukaryotic initiation factor 4E, Plant Physiol. 164 (2014), https://doi.org/10.1104/pp.113.227785.

[40] A. Bastet, D. Zafirov, N. Giovinazzo, A. Guyon-Debast, F. Nogue, C. Robaglia, J. L. Gallois, Mimicking natural polymorphism in eIF4E by CRISPR-Cas9 base editing is associated with resistance to potyviruses, Plant Biotechnol. J. 17 (2019) 1736–1750, https://doi.org/10.1111/pbi.13096.

[41] C.C.N. van Schie, F.L.W. Takken, Susceptibility genes 101: how to be a good host, Annu. Rev. Phytopathol. 52 (2014) 551–581, https://doi.org/10.1146/annurev-phyto-102313-045854.

Plant Biotechnology Journal
Open Access

Plant Biotechnology Journal (2023), pp. 1–13 doi: 10.1111/pbi.14003

An iterative gene-editing strategy broadens *eIF4E1* genetic diversity in *Solanum lycopersicum* and generates resistance to multiple potyvirus isolates

Kyoka Kuroiwa[1], Benoit Danilo[1], Laura Perrot[2], Christina Thenault[1], Florian Veillet[3], Fabien Delacote[4], Philippe Duchateau[4], Fabien Nogué[5], Marianne Mazier[1] and Jean-Luc Gallois[1,*]

[1]*INRAE, GAFL, Montfavet, France*
[2]*Toulouse Biotechnology Institute, Université de Toulouse, Toulouse, France*
[3]*INRAE, Agrocampus Ouest, Université de Rennes, IGEPP, Ploudaniel, France*
[4]*CELLECTIS S.A., Paris, France*
[5]*Université Paris-Saclay, INRAE, AgroParisTech, Institut Jean-Pierre Bourgin (IJPB), Versailles, France*

Received 5 September 2022;
revised 9 December 2022;
accepted 17 December 2022.
**Correspondence* (Tel +33 (0)4 32 72 27 96 fax +33 (0)4 32 72 27 02; email jean-luc.gallois@inrae.fr)

Keywords: genome editing, base editing, tomato, eIF4E, potyvirus, genetic resistance.

Summary

Resistance to potyviruses in plants has been largely provided by the selection of natural variant alleles of *eukaryotic translation initiation factors* (*eIF*) *4E* in many crops. However, the sources of such variability for breeding can be limited for certain crop species, while new virus isolates continue to emerge. Different methods of mutagenesis have been applied to inactivate the *eIF4E* genes to generate virus resistance, but with limited success due to the physiological importance of translation factors and their redundancy. Here, we employed genome editing approaches at the base level to induce non-synonymous mutations in the e*IF4E1* gene and create genetic diversity in cherry tomato (*Solanum lycopersicum* var. *cerasiforme*). We sequentially edited the genomic sequences coding for two regions of eIF4E1 protein, located around the cap-binding pocket and known to be important for susceptibility to potyviruses. We show that the editing of only one of the two regions, by gene knock-in and base editing, respectively, is not sufficient to provide resistance. However, combining amino acid mutations in both regions resulted in resistance to multiple potyviruses without affecting the functionality in translation initiation. Meanwhile, we report that extensive base editing in exonic region can alter RNA splicing pattern, resulting in gene knockout. Altogether our work demonstrates that precision editing allows to design plant factors based on the knowledge on evolutionarily selected alleles and enlarge the gene pool to potentially provide advantageous phenotypes such as pathogen resistance.

Introduction

Recent genome-editing techniques enable to directly inactivate or modify genomes at specific loci, thereby opening the possibility to control specific agronomic traits (Eisenstein, 2022). These techniques, including transcription activator-like effector nucleases (TALEN) and CRISPR (clustered, regularly interspaced, short palindromic repeat)-associated (Cas) endonucleases, allow the targeting of genomic regions of interest and cut the nucleic acid strand(s) to introduce random mutations or to precisely customize target loci with elements like donor sequences and deaminases (Gao, 2021; Veillet *et al.*, 2020). Such precise targeting is crucial to minimize side effects especially for genes that code for proteins involved in several physiological activities through different domains.

Pathogen resistance in certain accessions is a result of evolution of the interaction between plants and pathogens. This is often represented by the natural diversity in resistance alleles of susceptibility genes. These genes can be hijacked by viruses in susceptible plant accessions and are often essential for plant physiology: plants have to develop efficient resistance without impairing their growth. *Eukaryotic translation initiation factor* (*eIF*) *4E* is one of such genes co-evolving with the viral-genome linked protein (VPg)-coding gene of potyvirus, the largest plant RNA virus group. *eIF4E*s make a small multi-gene family of mRNA 5′-cap binding proteins that are essential for translation initiation in eukaryotes. The interaction of the potyviruses' VPg with host eIF4E(s) is important to establish the infection although the precise mechanism is still unknown (Coutinho de Oliveira *et al.*, 2019; Miras *et al.*, 2017; Saha and Mäkinen, 2020; Tavert-Roudet *et al.*, 2017). Hence, the mutant forms of *eIF4E* have been found as resistant alleles in several crop species including lettuce (*Lactuca sativum*), pea (*Pisum sativum*), and tomato (Gao *et al.*, 2004; Nicaise *et al.*, 2003; Ruffel *et al.*, 2002, 2005) though they are not available in all species (Bastet *et al.*, 2017). In pepper (*Capsicum* spp.), 25 allelic forms of *eIF4E1* have been reported and associated with various spectra and durabilities of potyvirus resistance (Charron *et al.*, 2008;

Please cite this article as: Kuroiwa, K., Danilo, B., Perrot, L., Thenault, C., Veillet, F., Delacote, F., Duchateau, P., Nogué, F., Mazier, M. and Gallois, J-L. (2023) An iterative gene-editing strategy broadens *eIF4E1* genetic diversity in *Solanum lycopersicum* and generates resistance to multiple potyvirus isolates. *Plant Biotechnol J.*., https://doi.org/10.1111/pbi.14003.

© 2023 The Authors. *Plant Biotechnology Journal* published by Society for Experimental Biology and The Association of Applied Biologists and John Wiley & Sons Ltd.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

14677652, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/pbi.14003 by Inrae - Dipso, Wiley Online Library on [16/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Moury *et al.*, 2014; Poulicard *et al.*, 2016). Interestingly, those natural mutants harbour only substitutions in their eIF4E1 amino acid sequences compared with their susceptible counterparts. These substitutions are highly clustered in two domains—named Region I and II—of the eIF4E1 protein surrounding the cap-binding pocket (Charron *et al.*, 2008; Robaglia and Caranta, 2006). Despite the importance of this domain for the protein's original function, many resistance genes have been proven functional reflecting the natural selection of the resulting amino acids (Charron *et al.*, 2008; Moury *et al.*, 2014). This can provide a blueprint to be implemented in resistance breeding.

Similar resistance is found among the ancestral species of tomato. The recessive resistance gene *pot1*, isolated from the wild tomato relative *S. habrochaites* PI247087 accession, encodes an eIF4E1 protein which possesses eight amino acid mutations (L48F, N68K, P69S, A77D, V85L, M109I, K123Q, and N224S) compared with the susceptible cultivated tomato, including five within and in the proximity of Regions I and II. The *pot1*-encoded eIF4E1 protein is associated with a broad resistance spectrum to multiple potyvirus species and strains, while it retains its function to initiate translation (Gauffier *et al.*, 2016; Ruffel *et al.*, 2005).

Various eIF4E knock-out mutants targeting eIF4E1 and/or eIF4E2 have been generated by methods such as TILLING, RNAi and CRISPR/Cas9 to generate potyvirus resistance in tomato as a model crop (Gauffier *et al.*, 2016; Kuroiwa *et al.*, 2022; Mazier *et al.*, 2011; Piron *et al.*, 2010; Schmitt-Keichinger, 2019; Yoon *et al.*, 2020). Although the mutant genotypes exhibited resistance to potyviruses, the resistance spectrum associated with single-gene knock-out was limited. Only the severely stunted, double *eif4e1*; *eif4e2* knock-out (KO) mutants had a resistance spectrum as large as the one provided by the natural allele *pot1*. This is because most potyviruses can use interchangeably both tomato *Sl*eIF4E1 and *Sl*eIF4E2 (Gauffier *et al.*, 2016; Lebaron *et al.*, 2016). Consequently, KO strategies were achievable, but unlike the natural resistance eIF4Es, neither the single nor double KO of tomato eIF4Es fully met the agronomic demand due to the functional redundancy and the importance of this protein. These results have prompted us to study how more precise genome-editing techniques can help designing resistant phenotypes based on the positively selected non-synonymous mutations rather than producing KO alleles in crops like tomatoes (Bastet *et al.*, 2017).

Previously, we provided a proof-of-concept by showing how six mutations selected in the pea (*Pisum sativum*) *eIF4E1*: *sbm1* alleles could be transferred through transgenesis in the *Arabidopsis eIF4E1* susceptible allele and provide resistance to a potyvirus, clover yellow vein virus (Bastet *et al.*, 2018). Moreover, this resistance could be pinpointed to discrete mutations. And when one of those—N176K—was introduced in the susceptible wild-type *Arabidopsis* through base editing, the resulting plants showed the same resistance. This was to our knowledge the first example of using precise genome editing to copy a non-synonymous mutation associated with resistance to viruses (Bastet *et al.*, 2019).

In the present work, we decided to introduce in cultivated tomato a potential amino acid mutation by gene knock-in in the *eIF4E1* Region I, which is known to be important for resistance. Facing the difficulty of such approach, we next decided to randomly tinker another genomic domain in *eIF4E1*—Region II, also known to be involved in resistance, using cytosine base editor (CBE). We show that by iteratively generating diversity in *eIF4E1* genomic regions known to be involved in resistance across plant species, we succeeded in developing a functional potyvirus resistance allele. Incidentally, we observed that heavily mutating a coding region with the CBE can impair the proper gene splicing leading to knock out. Altogether, this work presents the promising potential and a caveat in using genome editing for improvement of crops such as for virus resistance by generating targeted variability in alleles.

Results

A codon change in *SleIF4E1* associated with a P69T mutation is not sufficient to confer resistance to potyviruses

The broad-spectrum resistance allele *eIF4E1-pot1* differs from the wild-type susceptible tomato allele by eight non-synonymous codon substitutions, and the role of individual substitutions in the resistance process has not been fully elucidated (Gauffier *et al.*, 2016). Because a related *Solanaceae*, pepper, displays a wide range of resistant alleles, we looked at this *eIF4E1* gene pool to identify the causative mutations for the resistance. In many pepper resistant varieties, one or more mutations at amino acid position 66 to 68, such as P66T, V67E, and A68E within the eIF4E1 Region I, are present (Charron *et al.*, 2008; Kang *et al.*, 2005). As P69 (pepper P66 equivalent) is also mutated in the *pot1*-encoded protein, we decided to introduce a P69T mutation in the susceptible tomato *eIF4E1* allele to generate a new resistant tomato (Figure S1).

In order to mimic the pepper P66T by mutating codon #69 from CCT to ACA, we attempted a gene knock-in strategy, using a pair of TALEN® to generate a double-strand break in the intron 1 of the *SleIF4E1* gene and to recombine with a donor DNA template. The template contained the *eIF4E1* exon 1 with substituted codon ACA as well as a kanamycin resistance *NptII* gene for selection, flanked by left and right homology arms to promote homologous recombination (HR) (Figure 1a). Both the TALEN® T-DNA vectors and the donor DNA template were introduced into cells of cherry tomato cultivar (cv.) WVa106 by the biolistic method. Thirty-two regenerating buds were assessed by PCR and sequencing, but none of them harboured an integration of the DNA template via HR (Figure S2). Nevertheless, one plant had integrated the donor template through non-homologous end joining (NHEJ) and was named P69T-NHEJ-npt. (Figure 1a P69T-NHEJ-npt, Figure S2, Table S1, Data S1). This NHEJ-based insertion in the native intron 1 has resulted in the expected insertion of *NptII* gene as well as an undesirable, partial duplication within the *eIF4E1* gene, namely the entire exon 1 deriving from the native and recombined sequences.

The P69T-NHEJ-npt was selfed to produce TALEN-free plants, homozygous for the modified *eIF4E1* locus. As expected, eIF4E1 protein accumulation was not detected in the progenies using specific polyclonal antibodies (Figure 1b, P69T-NHEJ-npt): this was consistent with the presence of an approximately 1.4 kb *NptII* expression cassette in the native exon 1. Thus, in order to restore a functional *eIF4E1* gene, two successive CRISPR/Cas9-mediated editing strategies were used to eliminate undesirable sequences of the recombined *eIF4E1* gene at the 3′ and the 5′ of the insert, respectively (Figure 1a, alleles P69T-NHEJ and P69T; Figure S3). Several independent tomato lines were recovered in which both the *NptII* cassette and the endogenous exon 1 of *eIF4E1* were removed. In these plants, the resulting edited *eIF4E1* gene harboured the T69 codon, and the eIF4E1 protein accumulation was restored to wild-type level (Figure 1b, P69T lanes). These plant lines are hereafter referred to as *eIF4E1*^{P69T} lines. This

© 2023 The Authors. *Plant Biotechnology Journal* published by Society for Experimental Biology and The Association of Applied Biologists and John Wiley & Sons Ltd.,

14677652, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/pbi.14003 by Inrae - Dipso, Wiley Online Library on [16/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

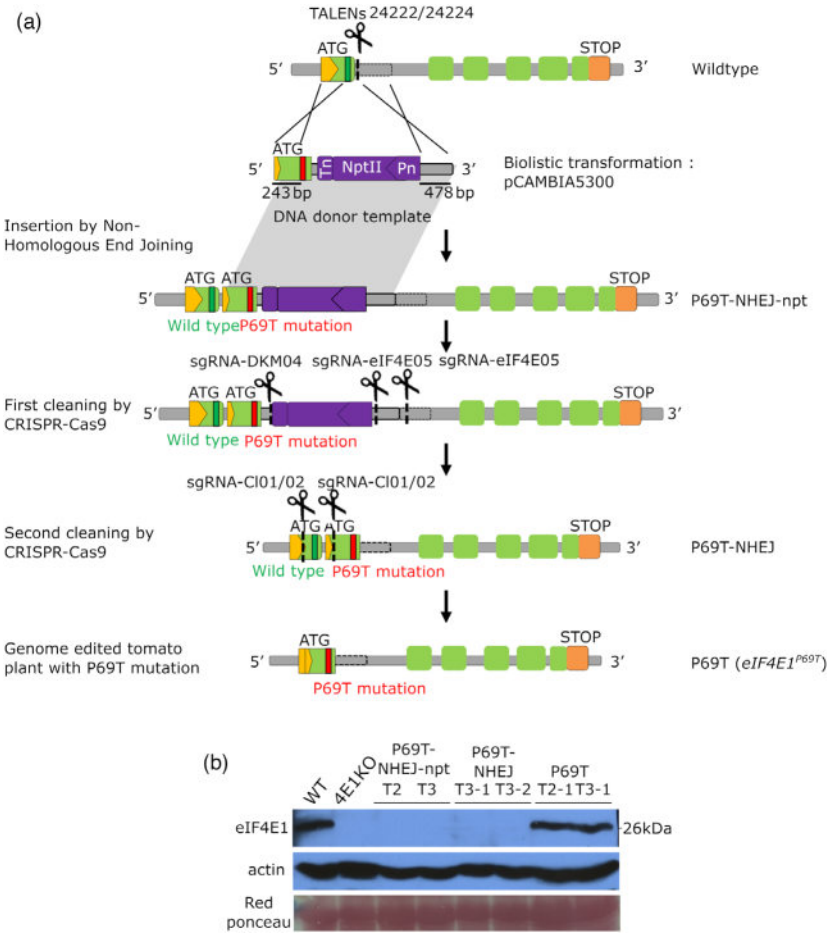


Figure 1 A gene knock-in strategy using TALEN® followed by the targeted elimination of undesirable inserted parts using CRISPR/Cas9 generates a single AA change in *Sl*eIF4E1 exon 1. (a) Schematic diagram of the targeted gene Knock-In and the successive targeted deletion of genomic region. Dotted line under the scissors indicate each targeted site of TALEN or CRISPR sgRNA indicated above. The codon position for P69T mutation is indicated by the strips in green or red (wildtype or mutant). Green boxes represent the exons and yellow pentagon the 5′-UTR. The crossed lines between the genome and the donor DNA template indicate recombined region. Pn, Nos promoter; Tn, nos terminator. (b) Western blot analysis to detect the *Sl*eIF4E1 accumulation in the homozygous mutant plants from successive genetic modification steps. T2 and T3 indicate the transformant generation. Actin accumulation and red ponceau are shown as the loading control. 4E1KO, TILLING *eif4e1*^{KO}; WT, wildtype.

suggests that the targeted modification of an exon in *eIF4E1* through knock-in with a DNA donor template is achievable but might be labour-intensive due to low frequency of homologous recombination and possible rearrangements of the targeted locus.

To evaluate the effect of the eIF4E1 Region I modification on the virus resistance, we used four potyviruses—pepper mottle virus (PepMoV) and three different isolates of potato virus Y (PVY)—to challenge the plants homozygous at the three editing stages (Figure 1a) and quantified the viral accumulation by ELISA 21 days after infection (Table 1). In accordance with the lack of eIF4E1 expression in the first steps of this editing scheme, intermediate lines (P69T-NHEJ-npt and P69T-NHEJ in Figure 1a and b) were found to be fully resistant to PepMoV and PVY LYE90 isolate but not to other PVY isolates. They therefore present, as expected, the same resistance spectrum as a TILLING *eif4e1*^{KO} line (Gauffier *et al.*, 2016; Piron *et al.*, 2010). Interestingly, the *eIF4E1*^{P69T} lines in which the eIF4E1 expression was restored were fully susceptible to all four potyvirus isolates tested (Table 1). This implied that introducing a single P69T mutation in the tomato eIF4E1 is not sufficient to induce resistance to the potyvirus infection. However, based on the natural *eIF4E* resistance alleles, which often possess multiple non-synonymous mutations, we reasoned that additional mutations could confer resistance. We thus proceeded to the editing of another genomic region of *eIF4E1*, Region II, where non-synonymous mutations are documented in resistance alleles from multiple species in combination with the Region I.

A base-editing strategy efficiently generates diverse alleles with multiple substitutions in *eIF4E1* region II

Since many natural potyvirus resistance alleles possess amino acid substitutions in both Region I and II of eIF4E1, we decided to target the Region II genomic sequence in *eIF4E1*^{P69T} plants. The same editing strategy was carried out in parallel on wild-type WVa106 plants. In order to bypass the labour-intensive knock-in strategy, we chose a base-editing strategy using nCas9-cytidine deaminase enzyme (or cytosine base editor; CBE). Because a CBE can introduce mutations at one or more cytosines, provided they are located at appropriate position related to a PAM sequence (Nishida *et al.*, 2016), it can trigger varied amino acid changes in this region instrumental for potyvirus resistance. An *SleIF4E1*

© 2023 The Authors. *Plant Biotechnology Journal* published by Society for Experimental Biology and The Association of Applied Biologists and John Wiley & Sons Ltd.,

14677652, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/pbi.14003 by Inrae - Dipso, Wiley Online Library on [16/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 1 Resistance phenotypes against four potyvirus isolates of plants at different stages during the knock-in strategy

	Plant genotypes						
				P69T-NHEJ T3		P69T	
Viruses tested	WVa106 (wildtype)	TILLING *eif4e1*^{KO}	P69T-NHEJ-npt T3-1	1	2	1	2
PepMoV	S	R	R	R	R	S	S
PVY-LYE90	S	PR	PR/S	PR	PR	S	S
PVY-N605	S	S	S	S	S	S	S
PVY-SON41	S	S	S	S	S	S	S

PR, partially resistant (CP level lower than S but higher than R; coloured in orange); R, resistant (coloured in yellow); S, susceptible.

The accumulation of viral coat protein (CP) 21 days post inoculation (dpi) was measured by ELISA. Six plants are used for each mutant line represented by column (two lines for the P69T-NHEJ and P69T genotypes) and 12 for wildtype.

guide RNA was designed to modify a maximum of four cytosines situated in the editing window at 16–20 nt relative to the PAM on the reverse strand, with potential to induce amino acid changes for M109, G110 and A111 (Figure S4a). Moreover, to facilitate the selection, we implemented a co-editing strategy which can simultaneously mutate cytosines in the *acetolactate synthase SlALS1*-targeted genomic regions (Figure S4c). As shown previously, plants edited for *SlALS1* can be selected on ALS inhibitors such as chlorsuluron, allowing a selection of potentially edited plantlets by CRISPR activity (Veillet *et al*., 2019).

Explants that grew on the chlorsulphuron selection media were recovered from both WVa106 and *eIF4E1*^{P69T} backgrounds. Based on the sequencing of the genomic DNA covering *eIF4E1* exon 2 that includes the targeted Region II, two edited profiles were selected (i) a double amino acid change of G110E + A111T (hereafter referred to as ET) and (ii) a triple change of M109I + G110K + A111T (hereafter referred to as IKT) (Figure 2a). We therefore obtained four original combinations of mutations in eIF4E1 that have never been described among the natural diversity of eIF4Es. Moreover, this setup allows to potentially monitor the effect of mutations in the two regions, independently or together. Two independent lines for each of the four new alleles (*eIF4E1*^{ET}, *eIF4E1*$^{P69T/ET}$, *eIF4E1*^{IKT} and *eIF4E1*$^{P69T/IKT}$) were selected and after three selfings, transgene-free homozygous lines for the *eIF4E1* modification were obtained (Figure 2b–d).

Extensive base editing in exon 2 is associated with mis-splicing of *eIF4E1* mRNA

To characterize the effects of modifications on *eIF4E1* gene function in each mutant line, we performed RT-PCR on the *eIF4E1* mRNA using specific primers. In the *eIF4E1*^{P69T}, *eIF4E1*^{ET}, and *eIF4E1*$^{P69T/ET}$-edited lines, cDNA of the same size as wildtype (696 bp) was amplified (Figure 3a). A shorter cDNA of 404 bp was amplified for *eif4e1*^{KO} mutant which is a product of an aberrant splicing due to a mutation in a 5′-splice site flanking the exon 3 (Piron *et al*., 2010). In contrast, in the *eIF4E1*^{IKT} and *eIF4E1*$^{P69T/IKT}$ lines, cDNA with a size smaller than wildtype but larger than that of *eif4e1*^{KO} mutant was amplified (Figure 3b). Based on the Sanger-sequencing and alignment with the wild-type sequence, all but the *eIF4E1* cDNA derived from *eIF4E1*^{IKT} and *eIF4E1*$^{P69T/IKT}$ contained the predicted sequences including the correct nucleotide substitutions already verified in genomic DNA (Figures 2d, 3c). In contrast, the shorter cDNAs amplified from all of the IKT modification mutants were missing a 166-bp-long segment which entails Region II of *eIF4E1*, rendering the cDNA shorter (530 bp-long) (Figure 3e). In fact, this missing segment perfectly matched the entire exon 2.

To gain insight into this event, we analysed closely the amplified gDNA sequences surrounding the Region I and II and the aberrant cDNA sequence obtained from the edited tomato lines in comparison with previous studies on regulatory elements. In the gDNA fragment that spans the 526-bp genomic region from the last 100 bp of the intron 1 to the first 30 bp of the exon 3 (amplified by 4e2F/R in Figure 2b), we did not observe any other mutations than our predicted target codon modifications that code for the IKT substitutions. The sequencing also revealed that the 3′-splice sites of exons 1 and 2 and the 5′-splice sites of exons 2 and 3 were unchanged in the edited plant compared with the wild-type, with the conserved dinucleotides known as the donor and acceptor sites for the pre-mRNA splicing (Brown and Simpson, 1998; Li *et al*., 2019) (Figure S5a). Thus, we concluded that the modification of a wild-type sequence ATGGGAGCA into an IKT-coding sequence ATAAAACA in the exonic region has altered the splicing pattern and induced an exon skipping in *eIF4E1* (Figure S5b).

At translational level, this exon deletion would cause a frameshift at position 97 in amino acid sequence, resulting in a premature termination codon corresponding to position 100 in the wild-type sequence (Figure 3d). This possibly leads to truncated eIF4E1 production and/or the degradation of such aberrant form in *eIF4E1*^{IKT} and *eIF4E1*$^{P69T/IKT}$ lines.

Mis-splicing resulting from heavy editing at *eIF4E1* region II affects the eIF4E1 protein accumulation while a moderate editing maintains accumulation and functionality of eIF4E1

Not only the mis-splicing but also the introduction of amino acid modifications itself can destabilize the protein structure affecting the accumulation and/or functionality of the protein. To examine the protein accumulation, we performed western blot analysis on the proteins extracted from leaves for each edited line. All independent lines of plants with the *eIF4E1*^{P69T}, *eIF4E1*^{ET}, and *eIF4E1*$^{P69T/ET}$ alleles expressed eIF4E1 protein (26 kDa) at a level comparable to that of the wildtype (Figure 4a). In contrast, lines harbouring three mutations in eIF4E1 Region II (IKT) showed no to very low accumulation of eIF4E1 at the expected size (Figure 4b). This suggests that the amino acid substitutions in the eIF4E1^{P69T}, eIF4E1^{ET}, and eIF4E1$^{P69T/ET}$ do not affect the protein accumulation. The results also implied that the mis-splicing due to the IKT modification in Region II has led to lack of accumulation or

© 2023 The Authors. *Plant Biotechnology Journal* published by Society for Experimental Biology and The Association of Applied Biologists and John Wiley & Sons Ltd.,

14677652, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/pbi.14003 by Inrae - Dipso, Wiley Online Library on [16/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

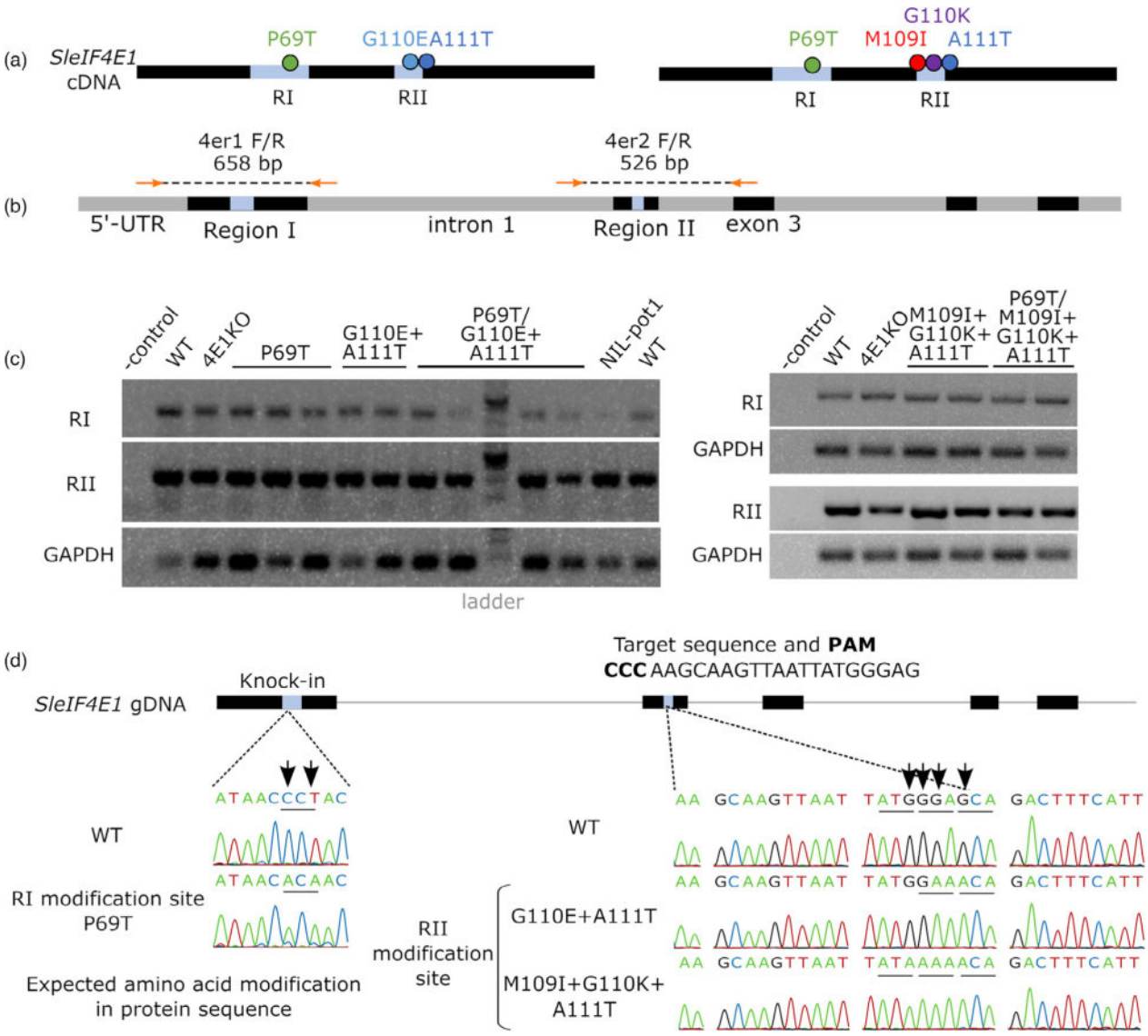


Figure 2 Base editing in *eIF4E1* Region II generates two combinations of amino acid substitutions in different Region I backgrounds. (a) Schematic diagram of the relative positions of amino acid substitutions induced in this work. Light blue boxes represent Region I and II. (b) Primer positions used in the genomic sequence analyses of *SleIF4E1* Region I and II. Orange arrows are primers. Grey and black box represent introns and exons, respectively. Light blue boxes represent Region I and II. (c) Amplification of genomic sequence of *SleIF4E1* Region I and II. Bands of correct lengths were verified. Two to four independent lines, including those not used in the further analyses, were used per allele. Mutant lines are indicated by their expected amino acid modifications in the *Sle*IF4E1 protein. -control is water. GAPDH is used as internal control. 4E1KO, TILLING *eif4e1*^{KO}; WT, wildtype. (d) Sanger-sequencing results of the amplified genomic regions confirm the expected modifications in the edited plants. Black boxes indicate the exons in *SleIF4E1* genomic DNA. Representative chromatogram from each modified region (light blue boxes) is shown. Analyses were performed for at least two plants of independent lines of each genotype. Arrows indicate the positions of nucleotide substitutions. The sgRNA in Reverse Complementary sequence used in the base-editing of the Region is shown.

expression of the eIF4E1 protein. As a consequence, the *eIF4E1*^{IKT} and *eIF4E1*$^{P69T/IKT}$ alleles could be considered *eIF4E1* knock-out alleles.

Given the importance of eIF4E1 protein in plant translation, we further checked, in addition to the protein accumulation, whether the functionality of the edited eIF4E1 proteins was not affected. To verify the ability of the edited proteins to bind the 5′-cap structure and initiate translation, we performed both *in vitro* and *in vivo* methods. We first assessed the ability of the eIF4E1 proteins to bind the m^{7}GTP-cap analogue. After the incubation of total protein extract with cap analogue, eIF4E1 protein was detected for the wildtype as well as for all three edited lines, namely *eIF4E1*^{P69T}, *eIF4E1*^{ET} and, *eIF4E1*$^{P69T/ET}$, showing that the edited alleles code for eIF4E1 with cap-binding ability (Figure 4c). Additionally, we used a yeast complementation assay, which relies on the conditional complementation of a yeast strain lacking its native *eIF4E* gene. In accordance with the cap affinity purification results, the heterologous expression of the three edited tomato alleles complemented the lack of a native yeast *eIF4E* (Figure 4d). The serial dilution suggests that the complementation efficiency with the edited version of eIF4E1 is comparable to that achieved with the wild-type eIF4E1 as well

© 2023 The Authors. *Plant Biotechnology Journal* published by Society for Experimental Biology and The Association of Applied Biologists and John Wiley & Sons Ltd., 1–13

14677652, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/pbi.14003 by Inrae - Dipso, Wiley Online Library on [16/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

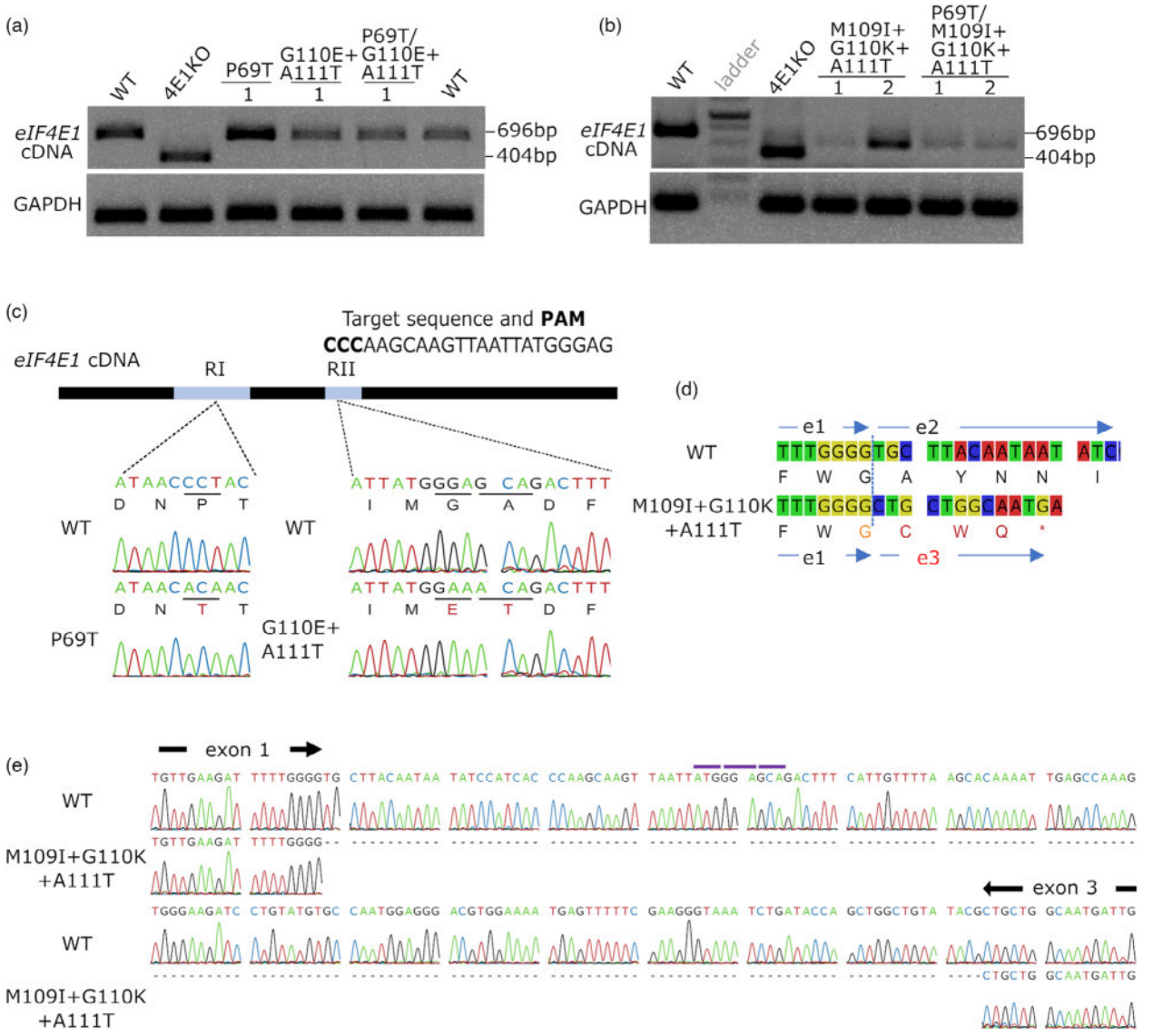


Figure 3 *eIF4E1* cDNA analysis confirms correct mRNA expression for moderately edited plants but reveals abnormal splicing for the triple-codon mutants in Region II. (a, b) Amplification of *eIF4E1* cDNA following reverse transcription; the expected size of the PCR fragment is 696 bp for WT and 404 bp for the TILLING *eif4e1*^{KO}. The GAPDH cDNA is amplified as internal control. Representative sample from two independent mutant lines is shown. 4E1KO, TILLING *eif4e1*^{KO}; WT: wildtype. (c) Sanger-sequencing of the amplified *Sl*e*IF4E1* cDNA confirms the expected modifications in the P69T- and G110E + A111T-coding mutant lines. Representative chromatogram from each modified region is shown. Analyses were performed for at least two plants of independent lines of each genotype. Mutant lines are indicated by their expected amino acid modifications in the *Sl*eIF4E1 protein. Substitution targets are underlined. Predicted translation of each sequence is shown under the nucleotide sequence. Amino acid substitution relative to WT sequence is indicated in red. (d) Predicted protein translation of cDNAs from the WT plant and from an IKT-containing edited plant. The premature stop codon in the eIF4E1 cDNA from IKT plants is represented by a star. e1, e2 and e3 are exons 1, 2 and 3. (e) Chromatogram of the representative cDNA sequence of IKT group shows an unexpected deletion corresponding to the wild-type exon 2. Shown from the 3′ end of the exon 1 to the 5′ end of the exon 3. Purple bars indicate the positions of the codons expected for mutations based on the genomic DNA sequencing.

as the *pot1*-encoded eIF4E1 (Figure 4d). Therefore, the *in vitro* and *in vivo* assays strongly suggest that the editing carried out in *eIF4E1*^{P69T}, *eIF4E1*^{ET}, and *eIF4E1*$^{P69T/ET}$ alleles did not affect their functionality in translation initiation processes, in a similar fashion to the natural *eIF4E* varieties.

Taken together, these results show that the P69T editing in Region I does not affect eIF4E1 expression or function. In contrast, while the editing of two codons in Region II (G110E + A111T) does not result in any apparent change, a more extensive, three-codon mutation (M109I + G110K + A111T) results in a lack of wild-type level eIF4E1 protein expression due to the altered splicing in the genomic sequence.

A combination of mutations in region I and II of *eIF4E1* associates with resistance to several isolates of potyviruses

Finally, we evaluated the resistance spectrum to potyviruses of all the 10 edited tomato lines—two for each allele, along with plants of wild-type WVa106, TILLING *eif4e1*^{KO} mutant, and a *pot1* near-isogenic line (NIL-pot1) (Lebaron *et al.*, 2016). Plants were

© 2023 The Authors. *Plant Biotechnology Journal* published by Society for Experimental Biology and The Association of Applied Biologists and John Wiley & Sons Ltd., 1–13

14677652, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/pbi.14003 by Inrae - Dipso, Wiley Online Library on [16/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

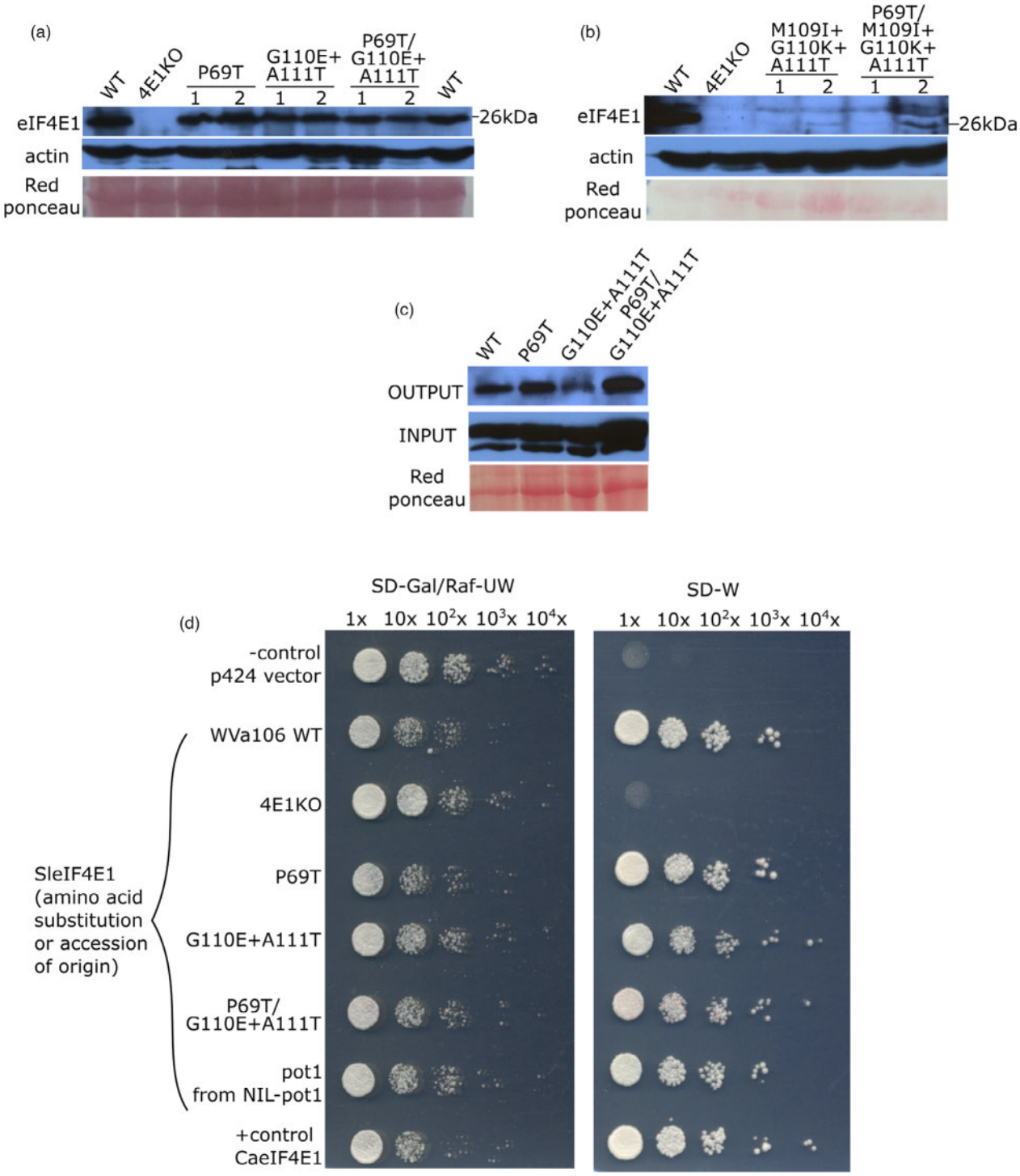


Figure 4 Moderate editing in eIF4E1 region II does not affect the protein accumulation level *in planta* and the protein functionality *in vitro* and *in vivo*. (a, b) Western blot analysis to detect the *Sl*eIF4E1 accumulation in the edited lines. Actin accumulation and red ponceau are shown as the loading control. Mutant lines are indicated by their expected amino acid modifications in the *Sl*eIF4E1 protein. Two independent edited lines were used per genotype. The expected molecular weight of *Sl*eIF4E1 is 26 kDa. 4E1KO, TILLING *eif4e1*^{KO}; WT, wildtype. (c) Cap-affinity purification and following western blot shows the cap-binding ability of the edited eIF4E1 proteins at wild-type level. OUTPUT is immunoblot using anti-SleIF4E1 specific antibody on the eluted samples after incubation with m^{7}GTP-cap analogue and repeated washing. INPUT is immunoblot using anti-actin antibody on the total soluble protein extracts as the loading control. Red ponceau staining of the INPUT loading is also shown as the loading control. (d) Yeast complementation assay demonstrates that the edited *SleIF4E1* proteins correctly expressed *in planta* retain the functionality in translation initiation. Gal/Raf-UW is the control media and SD-W is the selective media for the eIF4E function. J055 transformed with p424 empty vector is shown as a negative control. *Ca*eIF4E1 is the positive control. All *Sl*eIF4E1 are the transformants with the cDNA obtained from the edited plants. X indicates the dilution level of the yeast culture. 4E1KO, TILLING *eif4e1*^{KO}; Ca, *Capsicum annuum*; WT, wildtype.

© 2023 The Authors. *Plant Biotechnology Journal* published by Society for Experimental Biology and The Association of Applied Biologists and John Wiley & Sons Ltd., 1–13

14677652, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/pbi.14003 by Inrae - Dipso, Wiley Online Library on [16/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

challenged with three isolates of PVY and two other potyviruses, PepMoV and tobacco etch virus (TEV) CAA10 isolate, previously tested (Gauffier *et al.*, 2016). Resistance of each genotype was based on viral accumulation assessed by ELISA after 21 days.

All eight lines harbouring Region II mutations ET or IKT, regardless of the presence of P69T in Region I, showed complete resistance to PepMoV and complete or partial resistance to PVY-LYE90 (Table 2). This is significant but not any better than the limited resistance spectrum associated with knocking out *eIF4E1*. Indeed, the limited resistance spectra of the *eIF4E1*^{IKT} and *eIF4E1*$^{P69T/IKT}$ mutants were consistent with their loss of function of the *eIF4E1* gene just like in the TILLING *eif4e1*^{KO}. In contrast, while *eif4e1*^{KO} plants are fully susceptible to PVY-N605 and PVY-SON41, we show that the combination of mutations P69T and ET in *eIF4E1* (eIF4E1$^{P69T/ET}$) triggered resistance to PVY-N605, as well as partial resistance to the PVY-SON41 isolate (Table 2, Figure 5a). The infection of PVY-N605 was further assessed by inoculating a GFP-tagged infectious clone of this isolate and monitoring the GFP fluorescence. GFP fluorescence could be observed on the inoculated and systemic leaves of the wild-type WVa106 and of the *eIF4E1*^{ET} line, mutated only in Region II. On the contrary, the fluorescence was observed neither on the inoculated nor the systemic leaves of the *eIF4E1*$^{P69T/ET}$ lines and NIL-pot1, thereby confirming the resistance of the specific Region I + II edited tomato (Figure 5b). Nevertheless, the *eIF4E1*$^{P69T/ET}$ allele was not associated with resistance to TEV-CAA10 unlike the natural resistance *pot1* allele in which a larger set of amino acid substitutions has been selected (Table 2).

Discussion

In this work, we have used TALEN-based gene knock-in and CRISPR-mediated base editing techniques on the tomato *eIF4E1* allele in an attempt to generate genetic resistance against a broad range of potyviruses. Our targeted approach to the specific regions in *eIF4E1* for multiple base substitutions has brought about a *de novo* functional resistance tomato *eIF4E1* allele. Edited plants with this allele showed the broadest potyvirus resistance spectrum achieved by genetic engineering in tomato so far. Meanwhile, we have observed a previously unreported effect on RNA splicing by certain base modifications in a target exon resulting in abnormal expression.

Since the identification of pepper PVY-resistance alleles *pvr1*/*pvr2* as alleles of the *eIF4E1* gene in early 2000s, there has been a number of potyvirus resistance *eIF4E* alleles found among genetic resources, including *pot1* in tomato (Kang *et al.*, 2005; Ruffel *et al.*, 2002, 2006). Apart from using transgenic expression, direct implementation of these genes to other species in resistance development was restricted because they are found in a limited number of species and the conventional breeding is often bound by the interspecies barriers. Moreover, the functional significance and redundancy of the *eIF4E* genes in plants make the gene KO or knock-down strategies insufficient (Chandrasekaran *et al.*, 2016; Gauffier *et al.*, 2016; Kuroiwa *et al.*, 2022; Mazier *et al.*, 2011; Pechar *et al.*, 2022; Piron *et al.*, 2010). One intriguing feature of natural *eIF4E1*-based resistance is how the presence of an eIF4E1 protein variant is sufficient to bar the potyvirus from hijacking eIF4E2 although the latter can be expressed. Several hypothesis have been put forward, such as the fact that the functional eIF4E1 protein encoded by the *pot1* resistance allele might titrate other host factors which are also required by the virus (Bastet *et al.*, 2017; Gauffier *et al.*, 2016; Kang *et al.*, 2007). These results have suggested that an alternate strategy, based on the generation of resistant functional alleles, similar to the ones selected among the natural variability, should be more efficient than knocking out *eIF4E* genes. Thus, more delicate engineering at the base level is clearly needed to achieve resistance and keep eIF4E function at the same time (Bastet *et al.*, 2017).

Though this work confirms the high efficiency of base editing to generate diversity in a specific region of a gene, we also highlight a potential drawback of base editing following the mutation of four bases in the *eIF4E1* Region II. While the genomic modification in the IKT-coding mutants was done correctly, the expression of the resulting full-length mRNA and protein was impaired because the genomic modification induced an exon skipping. So far, there are no reported cases of such aberrant splicing due to CRISPR-mediated editing in exons of a gene although there are some examples after editing in introns or inducing indels in exons (Li *et al.*, 2019; Tang *et al.*, 2021). It is possible that this set of modifications has converted the exon 2 into an intron-like sequence. Introns are often defined by UA-richness of the sequence (Simpson *et al.*, 1996). The modified sequence does not completely correspond to known consensus of branchpoint or intronic splicing enhancer sequences which are important intronic *cis*-elements in splicing events (Wang

Table 2 Schematic diagram of the selected edited alleles and resistance phenotypes associated

	Region I	Region II		Virus isolates tested				
Genotypes	P69T	G110E + A111T	M109I + G110K + A111T	PepMoV	PVY-LYE90	PVY-N605	PVY-SON41	TEV-CAA10
WVa106 (wildtype)				S	S	S	S	S
TILLING *eif4e1*^{KO}				R	PR	S	S	S
NIL-pot1				R	R	R	R	R
P69T	X			S	S	S	S	S
ET		X		R	PR	S	S	S
P69T/ET	X	X		R	R	R	PR	S
IKT			X	R	PR	S	S	S
P69T/IKT	X		X	R	PR	S	S	S

PR, partially resistant (CP level lower than S but higher than R; coloured in orange); R, resistant (coloured in yellow); S, susceptible.

Summary of virus infection testing results of each mutant against five potyvirus isolates. The accumulation of viral coat protein (CP) after 21 dpi was measured by ELISA. Six to ten plants were used for each mutant line and 12–15 plants for wildtype. Two independent lines were tested for the five edited genotypes.

© 2023 The Authors. *Plant Biotechnology Journal* published by Society for Experimental Biology and The Association of Applied Biologists and John Wiley & Sons Ltd.,

14677652, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/pbi.14003 by Inrae - Dipso, Wiley Online Library on [16/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

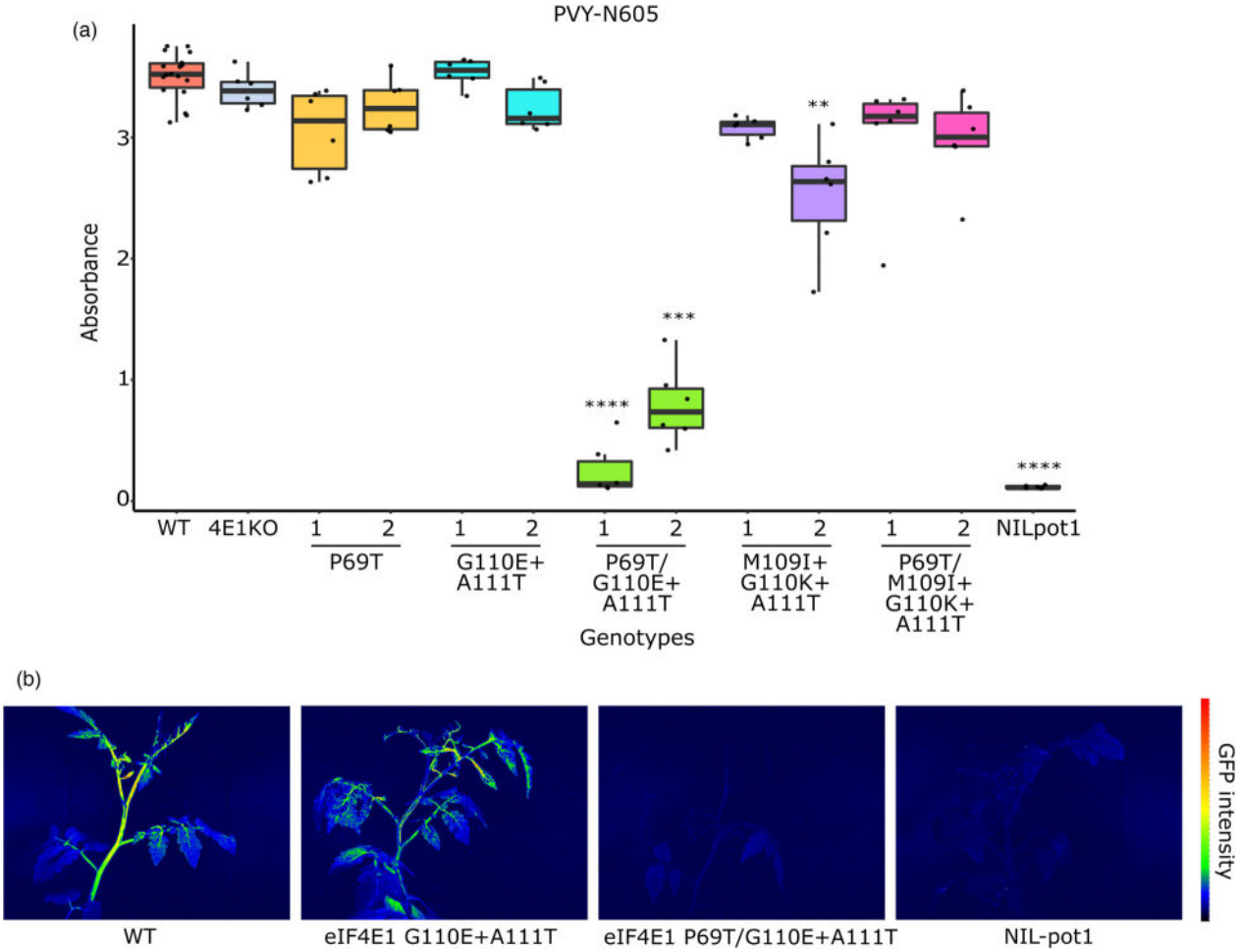


Figure 5 A combination of mutations at eIF4E1 region I and II triggers resistance to PVY N605. (a) DAS-ELISA quantification of the accumulation of PVY-N605 in tomato plants at 21 dpi. Points overlaid on boxplots represent the individual plants. Wild-type WVa106 and KO4E1 (TILLING *eif4e1*^{KO}) lines were used as positive controls, while NIL-pot1 plants are used as controls for resistance. Two independent lines were tested for the five edited genotypes. Coloured by alleles. Six plants were tested per mutant and 12 for wildtype. Significance of difference compared with WVa106 based on Kruskal–Wallis and Dunn-test as a *post-hoc* test is indicated by asterisks. <10^{-4} ****; 10^{-4} to 10^{-3} ***; 10^{-3} to 10^{-2} **. 4E1KO, TILLING *eif4e1*^{KO}. (b) Monitoring of the virus infection using a GFP-tagged PVY-N605 clone corroborates the resistance of the edited *eIF4E1*$^{P69T/ET}$ line demonstrated by ELISA results. Wild-type WVa106 was used as the positive control, while NIL-pot1 was used as the resistant control. GFP signal was observed at 15 dpi under blue light (420 nm) by using GFP camera imaging. The bar next to the photos indicates the intensity of GFP perceived by the device, from blue (no GFP) to red (high intensity). Each photo is a representative of 10–15 plants observed for each genotype. Two independent lines were tested for the edited alleles. WT, wildtype.

et al., 2012; Zhang *et al*., 2019). Nevertheless, it is tempting to associate the increased A-richness in an otherwise exonic region, with the false recognition of the intron. On a practical point, our results stress that the sequencing of the genomic target is not sufficient for inferring the effect of a base-editing strategy on a gene. This holds an important message that (i) the results of the genome editing should be confirmed by assessing cDNA and protein expressions and that (ii) codon choices should not be neglected.

Here, by inducing multiple amino acid substitutions in eIF4E1, we have designed cherry tomato mutant lines which present resistance against a wider range of potyviruses than the single *eif4e1*^{KO} mutant. To our knowledge, this is the first example of virus resistance conferred to a crop species via base-editing technology by generating a *de novo* functional resistance *eIF4E1* allele. The amino acid changes induced in the resistance eIF4E1 include side-chain-lacking Glycine to negatively charged Glutamic Acid at position 110 (G110E). Since potyviral VPg is negatively charged, such substitution may repel the VPg protein from the two regions and reduce the interaction. Nevertheless, this substitution does not abolish the functionality of the protein as shown in the cap-affinity purification and yeast complementation assay. Remarkably, tomatoes possessing mutant eIF4E1$^{P69T/ET}$ have accumulated a very low level of PVY-N605 which is known to use either *Sl*eIF4E1 or *Sl*eIF4E2 in the absence of one of them in susceptible wild-type cultivars. The resistance against this isolate had been achieved only when the two eIF4Es were knocked-out or when the *SheIF4E1-pot1* allele from the wild relative *S. habrochaites*, was expressed (Gauffier *et al*., 2016; Lebaron *et al*., 2016). Our results show that with the exception of TEV-CAA10, the eIF4E1$^{P69T/ET}$ protein behaves similarly to the natural resistant *pot1*-encoded protein that appears to make the *Sl*eIF4E2 unavailable to the potyvirus infection by a yet unknown mechanism. Indeed, we confirm that editing eIF4E1 does not affect the

© 2023 The Authors. *Plant Biotechnology Journal* published by Society for Experimental Biology and The Association of Applied Biologists and John Wiley & Sons Ltd., 1–13

14677652, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/pbi.14003 by Inrae - Dipso, Wiley Online Library on [16/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

accumulation of eIF4E2 mRNA or protein in leaves (Figure S6) suggesting a same mechanism.

In all, our results prove that it is possible to pyramid modifications in two separate regions in one gene to establish resistance while maintaining the functionality and that indeed a combination of mutations in the two regions is necessary to establish resistance. To induce targeted genetic variation in the tomato *eIF4E1*, we have focused on two separate domains in the genome that are in the proximity in the predicted 3D protein structure (Figure S7) and achieved virus resistance.

Our work summarizes the progress of base editing in the recent years: from gene knock-in using TALENs to the CRISPR-based base editors. In the coming years, we expect that a similar strategy could be designed and carried on in a faster and more accurate manner, by using prime editor and/or CRISPR/Cas proteins with relaxed PAM requirement (Ren *et al*., 2021; Veillet *et al*., 2020). These techniques should allow us to tinker the protein function as we design and to minimize the unfavourable effects brought about by the insertions/deletions or lack of precision. By taking the susceptibility factor eIF4E as an example, we exemplify how genome editing can be used to design complex alleles and generate useful traits that could boost plant breeding.

Material and methods

Plant material

Cherry tomato (*Solanum lycopersicum* var. *cerasiforme*) cv. West Virginia 106 (WVa106) was used for the co-editing of *eIF4E1* (Solyc03g005870) and *ALS1* (Solyc03g044330) and as the wildtype. The TILLING line *eif4e1*^{KO} was previously obtained through ethyl methanesulphonate (EMS) mutagenesis in the *S. lycopersicum* cv. M82 background (Piron *et al*., 2010). NIL-pot1 is a near-isogenic line of the *eIF4E1* allele *pot1* from *S. habrochaites* PI247087 introgressed in *S. lycopersicum* cv. Mospomorist (Lebaron *et al*., 2016). All plants were grown in a growth chamber at 24 °C/18 °C in the 16 h day/8 h night cycle.

Constructs for genome editing using TALEN®, CRISPR/Cas9 and Cytosine Base editors (CBE)

TALEN® designed to target 5′-TTTCCTCTTTCAAATTggtgat agtgtagtgTAAGGGAAAACAGGGA (nucleotides recognized by the TAL effectors are capitalized) in the *SleIF4E1* genomic sequence were cloned into vectors, pCLS23222 and pCLS23224, and provided by Cellectis (Paris, France) (Data S2 and S3). The DNA donor template for gene knock-in was synthesized by GenScript (Piscataway, NJ) and cloned into pBGWD0 vector (Gateway vector VIB) through LR Gateway recombination. The linear DNA template designed for HR was generated by digestion with PacI and PmeI (Figure 1b, Data S4). For the two targeted deletion steps using CRISPR/Cas9 after the NHEJ event, different pairs of single guide RNAs: pair 1 [sgRNA-DKM04 and sgRNA-eIF4E05] and pair 2 [sgRNA-Cl-01 and sgRNA- Cl-02] were cloned into pDe-Cas9-Hpt binary vector and pDe-Cas9-NptII binary vector, respectively (Danilo *et al*., 2018). The sgRNA cassettes were synthesized by Integrated DNA Technologies (Coralville, IA). For co-base-editing of the *SleIF4E1* and *SlALS1*, an *SleIF4E1* sgRNA cassette with *StU6* promoter was cloned into pDicAID_nCas9-PmCDA1_NptII binary vector inserted with *AtU6p-SlALS1* guide (Veillet *et al*., 2019) (Figure S4a,b, Data S5) and named CBE-4E1-RII. The *SleIF4E1* sgRNA cassette was synthesized by TwistBioscience (San Francisco, CA). TALEN targets and CRISPR guide RNA sequences are listed in Table S2.

Genetic transformation of tomato

TALEN® plasmids and the linear DNA template were delivered into leaves from 4-week-old tomato plants by biolistic method. Other transformations with CRISPR/Cas9 or CBE vectors were performed on cotyledon pieces from 8 to 12-day-old tomato seedlings by Agrobacterium-mediated transformation using *Agrobacterium tumefaciens* C58pGV2260 strain. Plant regeneration and transformant selection following both transformations were performed as previously described (Danilo *et al*., 2018 for plant regeneration and Veillet *et al*., 2019 for transformant selection). *In vitro* culture was carried out in a controlled environment at the temperature of 22 °C/18 °C with 16 h/8 h (day/night) photoperiod.

Genotyping by PCR and high-resolution melting analysis

Genomic DNA was extracted from fresh tomato leaf samples according to Fulton *et al*. (1995). T0 plantlets after editing with TALEN and CRISPR/Cas9 were screened by PCR using specific primers for each transformation. T0 plantlets after co-base editing using CBE were screened by HRM analysis using specific primers and Precision Melt Supermix (BioRad, Hercules) with the CFX96™ Real-Time PCR Detection System (BioRad), as previously described (Veillet *et al*., 2019). Melting-curve analysis was performed using the Precision Melt Analysis software (BioRad, Hercules). PCR was also performed on the *SleIF4E1* genomic sequence to further verify. *GAPDH* (*Glyceraldehyde-3-Phosphate Dehydrogenase*) gene was amplified as the technical control with gapdhF/R. PCR products were Sanger-sequenced by Genoscreen (Lille, France). T-DNA detection was done by PCR. All the primer sequences and pairs used for detection are available in Table S3.

Total RNA extraction and cDNA analysis

Total RNA was extracted from young tomato leaf tissue (<100 mg) of each plant in 500 µL of TRI-Reagent (Sigma-Aldrich, St. Louis) and quality-checked by concentration measurement on Nanodrop ND-1000 (ThermoFisher, Waltham). Reverse transcription (RT) of the extracted RNA samples (1 µg) was performed using AMV (Avian Myeloblastosis Virus) Reverse Transcription Kit (Invitrogen, Waltham) and oligo-dT18 (1 µM) to obtain coding DNA (cDNA). The cDNA for *eIF4E1* was amplified with a primer pair c4e1F/R and *GAPDH* (*Glyceraldehyde-3-Phosphate Dehydrogenase*) gene as the internal control with gapdhF/R.

Protein extraction and western blot analysis

Total protein was extracted from 100 mg of young tomato leaves, and 150 µL of Laemmli 2× DTT buffer was added to the crushed leaves. The homogenized solution was boiled for 5 min. The supernatant after centrifugation at 4 °C was collected and conserved at −20 °C. Western blot was performed as previously described using specific antibodies (Kuroiwa *et al*., 2022). The anti-*Sl*eIF4E1 polyclonal serum (Gauffier *et al*., 2016) was diluted 1/2000 and combined with secondary goat horseradish peroxidase-labelled anti-rabbit serum (Sigma-Aldrich, St. Louis) diluted 1/5000. As loading controls, monoclonal anti-plant actin antibodies (1/10 000 dilution) (Sigma-Aldrich) were used with horseradish peroxidase-labelled anti-mouse serum (1/10 000 dilution) (Sigma-Aldrich). Dilution was done in TBS 1x + 5% milk (Tween 0.05% + Triton X-100 0.2% for cap-affinity purification).

© 2023 The Authors. *Plant Biotechnology Journal* published by Society for Experimental Biology and The Association of Applied Biologists and John Wiley & Sons Ltd.,

14677652, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/pbi.14003 by Inrae - Dipso, Wiley Online Library on [16/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Virus isolates and infection assays using DAS-ELISA quantification

Potyvirus infection assays were performed using PepMoV (Mazier *et al*., 2011), PVY-N605 (Parrella *et al*., 2002), PVY-LYE84, PVY-SON41 (Moury *et al*., 2004) and TEV-CAA10 (Charron *et al*., 2008). Isolates were maintained in *Nicotiana benthamiana* plants 12 days before inoculation. For the PVY infection monitoring, the infectious clone of PVY-N605 tagged with GFP (hereafter, PVY-N605-GFP) incorporated in pCambia vector was multiplied on *N. tabacum* cv. Xanthi plants before inoculation of 21-day-old tomato plants. Inocula were prepared from 1 g (fresh weight) of the infected tobacco leaves ground with 4 mL of grinding buffer: potassium phosphate buffer (0.03 M, pH = 7) with 0.2% of diethyldithiocarbamate, 80 mg active charcoal and 80 mg carborundum. Six to 12 plants per genotype were mechanically inoculated on the cotyledons of 14-day-old tomato plants. At 21 days post-inoculation (dpi), 1 g of leaves from each plant was collected and ground in 4 mL of grinding buffer. To measure the virus accumulation, DAS-ELISA was performed using anti-PVY or anti-TEV antiserum (Sediag, Bretenière, France) or anti-polypoty antiserum (Agdia, Grigny, France) for PepMoV according to the manufacturer's instructions. The absorbance was measured by a spectrophotometer at wavelength 405 nm.

Virus detection by fluorometric camera analysis

Monitoring of PVY-N605-GFP infection was carried out with a closed fluorometric camera FluorCam FC 800-C/1010- GFP (Photon System Instruments, Drasov, Czech Republic) equipped with a GFP filter. For each genotype, 10–15 plants were used. GFP fluorescence was captured by the camera upon excitation at 420 nm and the intensity of the GFP signal was indicated by the colour scale.

m^7GTP cap-affinity purification

Total soluble proteins were extracted from 300 mg leaf tissue in 1200 µL binding buffer with protease inhibitor (+PI) (performed in three sets 100 mg:400 µL each). After centrifugation at 15 000 *g* for 10 min at 4 °C, 50 µL of the supernatant (INPUT) was recovered, and the rest was incubated with 150 µL of γ-aminophenyl-m^7GTP (C10-spacer)-agarose beads (Jena Bioscience, Jena, Germany) pre-equilibrated with binding buffer without protease inhibitor at 4 °C for 1.5 h. The beads were pelleted for 1 min at 15 000 *g* and washed three times with binding buffer +PI at 4 °C to remove unbound proteins. The INPUT was diluted two-fold with Laemmli DTT 2× buffer. The proteins linked to the cap analogue were eluted using 100 µL Laemmli DTT 2× buffer (OUTPUT) and analysed by western blot assay as described above.

Yeast complementation assay

Amplified cDNA from the different *eIF4E1* alleles was subcloned using the gateway technology (ThermoFisher) in a converted p424-GDP-DEST plasmids. Yeast complementation method was adapted from Gauffier *et al*. (2016). Briefly, attB sequence was attached to cDNA for BP recombination reaction to generate the entry clones. The entry clone was then used for LR recombination reaction with the destination vector p424-GPD to generate the tryptophan-selectable expression vector with a constitutive promoter pGAD. Then, yeast (*Saccharomyces cerevisiae*) strain J055 (*cdc33-Δ*: *LEU2 leu2 ura3 his3 trp1 ade2* [YCp33supex-h4E URA3]) that lacks its native SceIF4E and expresses human eIF4E (H eIF4E) under galactose-inducible, glucose-repressible promoter pGAL was transformed with each expression vector. The transformants were selected on galactose-containing, uracil/tryptophan-deficient media (−UW); then, the transformed colonies were spotted on glucose-containing, tryptophan-deficient media (−W) to see if the eIF4E expressed from the expression vector complements the lack of its native eIF4E or repressed H eIF4E. Incubation for the complementation assay was done at 20 °C/16 °C with 16 h/8 h cycle. The empty vector is p424-RfA.

3D modelization of eIF4E structures

Homology modelling of the eIF4E1 protein were carried out as previously (Moury *et al*., 2020) using the YASARA software (http://www.yasara.org/), using structural data from pea (*Pisum sativum*) eIF4E (GenBank ID: AY423375, PDB ID: 2WMC-C) as the template. Protein structure was visualized using PyMol software (https://pymol.org/).

Statistical analysis

All the statistical tests were performed on the free software R (https://www.R-project.org/). Kruskal–Wallis non-parametric tests were performed to evaluate the significance of differences between genotypes using the 'pgirmess' package. As the *post-hoc* test, Dunn test with *P*-value adjustment by Benjamini–Yekutieli method or Benjamini–Krieger–Yekutieli method was performed using 'dunn.test' package to determine the significance.

Acknowledgements

The authors thank Emmanuel Botton (INRAE GAFL) for taking care of tomato plants, Nathalie Truglio (INRAE Pathologie Vegetale) for supplying tobacco plants, Céline Valeille for providing the p424-GDP-DEST plasmid, Pierre Barret, Caroline Tassy and Anne Partier (INRAE GDEC) for helpful advices to set up biolistic experiments, and Luc Mathis (Cellectis) for TALEN® design and production.

Competing interests

TALEN® is a registered trademark owned by Cellectis and a patented technology.

Funding

This work was supported by the French national Research Agency (ANR11-BTBR-0001-GENIUS). The IJPB benefits from the support of Saclay Plant Sciences-SPS (ANR-17-EUR-0007).

Author contributions

KK, FV, FN, MM and JLG designed the experiments. KK, BD, LP, CT, FV, FD and MM carried out the experiments. FD and PD contributed new reagents. KK and JLG wrote the manuscript with contributions from all authors. All authors read and approved the final manuscript.

References

Bastet, A., Robaglia, C. and Gallois, J.L. (2017) eIF4E resistance: natural variation should guide gene editing. *Trends Plant Sci.* **22**, 411–419.

Bastet, A., Lederer, B., Giovinazzo, N., Arnoux, X., German-Retana, S., Reinbold, C., Brault, V. *et al.* (2018) Trans-species synthetic gene design

14677652, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/pbi.14003 by Inrae - Dipso, Wiley Online Library on [16/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

allows resistance pyramiding and broad-spectrum engineering of virus resistance in plants. *Plant Biotechnol. J.* **16**, 1569–1581.

Bastet, A., Zafirov, D., Giovinazzo, N., Guyon-Debast, A., Nogué, F., Robaglia, C. and Gallois, J.L. (2019) Mimicking natural polymorphism in eIF4E by CRISPR-Cas9 base editing is associated with resistance to potyviruses. *Plant Biotechnol. J.* **17**, 1736–1750.

Brown, J.W.S. and Simpson, C.G. (1998) SPLICE SITE SELECTION IN PLANT PRE-mRNA SPLICING. *Annu. Rev. Plant. Physiol. Plant. Mol. Biol.* **49**, 77–95.

Chandrasekaran, J., Brumin, M., Wolf, D., Leibman, D., Klap, C., Pearlsman, M., Sherman, A. *et al*. (2016) Development of broad virus resistance in non-transgenic cucumber using CRISPR/Cas9 technology. *Mol. Plant Pathol.* **17**, 1140–1153.

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

Coutinho de Oliveira, L., Volpon, L., Rahardjo, A.K., Osborne, M.J., Culjkovic-Kraljacic, B., Trahan, C., Oeffinger, M. *et al*. (2019) Structural studies of the eIF4E-VPg complex reveal a direct competition for capped RNA: implications for translation. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 24056–24065.

Danilo, B., Perrot, L., Botton, E., Nogué, F. and Mazier, M. (2018) The DFR locus: a smart landing pad for targeted transgene insertion in tomato. *PLoS ONE*, **13**, e0208395.

Eisenstein, M. (2022) Base edit your way to better crops. *Nature*, **604**, 790–792.

Fulton, T.M., Chunwongse, J. and Tanksley, S.D. (1995) Microprep protocol for extraction of DNA from tomato and other herbaceous plants. *Plant Mol. Biol. Rep.* **13**, 207–209.

Gao, C. (2021) Genome engineering for crop improvement and future agriculture. *Cell*, **184**, 1621–1635.

Gao, Z., Johansen, E., Eyers, S., Thomas, C.L., Noel Ellis, T.H. and Maule, A.J. (2004) The potyvirus recessive resistance gene, sbm1, identifies a novel role for translation initiation factor eIF4E in cell-to-cell trafficking. *Plant J.* **40**, 376–385.

Gauffier, C., Lebaron, C., Moretti, A., Constant, C., Moquet, F., Bonnet, G., Caranta, C. *et al*. (2016) A TILLING approach to generate broad-spectrum resistance to potyviruses in tomato is hampered by eIF4E gene redundancy. *Plant J.* **85**, 717–729.

Kang, B.-C., Yeam, I., Frantz, J.D., Murphy, J.F. and Jahn, M.M. (2005) The pvr1 locus in Capsicum encodes a translation initiation factor eIF4E that interacts with Tobacco etch virus VPg. *Plant J.* **42**, 392–405.

Kang, B.-C., Yeam, I., Li, H., Perez, K.W. and Jahn, M.M. (2007) Ectopic expression of a recessive resistance gene generates dominant potyvirus resistance in plants. *Plant Biotechnol. J.* **5**, 526–536.

Kuroiwa, K., Thenault, C., Nogué, F., Perrot, L., Mazier, M. and Gallois, J.-L. (2022) CRISPR-based knock-out of eIF4E2 in a cherry tomato background successfully recapitulates resistance to pepper veinal mottle virus. *Plant Sci.* **316**, 111160.

Lebaron, C., Rosado, A., Sauvage, C., Gauffier, C., German-Retana, S., Moury, B. and Gallois, J.L. (2016) A new eIF4E1 allele characterized by RNAseq data mining is associated with resistance to potato virus Y in tomato albeit with a low durability. *J. Gen. Virol.* **97**, 3063–3072.

Li, Z., Xiong, X., Wang, F., Liang, J. and Li, J.-F. (2019) Gene disruption through base editing-induced messenger RNA missplicing in plants. *New Phytol.* **222**, 1139–1148.

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

Miras, M., Truniger, V., Silva, C., Verdaguer, N., Aranda, M.A. and Querol-Audí, J. (2017) Structure of eIF4E in complex with an eIF4G peptide supports a universal bipartite binding mode for protein translation. *Plant Physiol.* **174**, 1476–1491.

Moury, B., Morel, C., Johansen, E., Guilbaud, L., Souche, S., Ayme, V., Caranta, C. *et al*. (2004) Mutations in potato virus Y genome-linked protein determine virulence toward recessive resistances in Capsicum annuum and Lycopersicon hirsutum. *Mol. Plant Microbe Interact.* **17**, 322–329.

Moury, B., Charron, C., Janzac, B., Simon, V., Gallois, J.L., Palloix, A. and Caranta, C. (2014) Evolution of plant eukaryotic initiation factor 4E (eIF4E) and potyvirus genome-linked protein (VPg): a game of mirrors impacting resistance spectrum and durability. *Infect. Genet. Evol.* **27**, 472–480.

Moury, B., Lebaron, C., Szadkowski, M., Ben Khalifa, M., Girardot, G., Bolou Bi, B.A., Koné, D. *et al*. (2020) Knock-out mutation of eukaryotic initiation factor 4E2 (eIF4E2) confers resistance to pepper veinal mottle virus in tomato. *Virology*, **539**, 11–17.

Nicaise, V., German-Retana, S., Sanjuán, R., Dubrana, M.-P., Mazier, M., Maisonneuve, B., Candresse, T. *et al*. (2003) The eukaryotic translation initiation factor 4E controls lettuce susceptibility to the Potyvirus Lettuce mosaic virus. *Plant Physiol.* **132**, 1272–1282.

Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M. *et al*. (2016) Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science*, **353**, aaf8729.

Parrella, G., Ruffel, S., Moretti, A., Morel, C., Palloix, A. and Caranta, C. (2002) Recessive resistance genes against potyviruses are localized in colinear genomic regions of the tomato (Lycopersicon spp.) and pepper (Capsicum spp.) genomes. *Theor. Appl. Genet.* **105**, 855–861.

Pechar, G.S., Donaire, L., Gosalvez, B., García-Almodovar, C., Sánchez-Pina, M.A., Truniger, V. and Aranda, M.A. (2022) Editing melon eIF4E associates with virus resistance and male sterility. *Plant Biotechnol. J.* **20**, 2006–2022.

Piron, F., Nicolaï, M., Minoïa, S., Piednoir, E., Moretti, A., Salgues, A., Zamir, D. *et al*. (2010) An induced mutation in tomato eIF4E leads to immunity to two potyviruses. *PLoS ONE*, **5**, e11313.

Poulicard, N., Pacios, L.F., Gallois, J.-L., Piñero, D. and García-Arenal, F. (2016) Human management of a wild plant modulates the evolutionary dynamics of a gene determining recessive resistance to virus infection. *PLoS Genet.* **12**, e1006214.

Ren, Q., Sretenovic, S., Liu, S., Tang, X., Huang, L., He, Y., Liu, L. *et al*. (2021) PAM-less plant genome editing using a CRISPR-SpRY toolbox. *Nat. Plants*, **7**, 25–33.

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

Ruffel, S., Dussault, M.-H., Palloix, A., Moury, B., Bendahmane, A., Robaglia, C. and Caranta, C. (2002) A natural recessive resistance gene against potato virus Y in pepper corresponds to the eukaryotic initiation factor 4E (eIF4E). *Plant J.* **32**, 1067–1075.

Ruffel, S., Gallois, J.L., Lesage, M.L. and Caranta, C. (2005) The recessive potyvirus resistance gene pot-1 is the tomato orthologue of the pepper pvr2-eIF4E gene. *Mol. Genet. Genomics*, **274**, 346–353.

Ruffel, S., Gallois, J.-L., Moury, B., Robaglia, C., Palloix, A. and Caranta, C. (2006) Simultaneous mutations in translation initiation factors eIF4E and eIF (iso)4E are required to prevent pepper veinal mottle virus infection of pepper. *J. Gen. Virol.* **87**, 2089–2098.

Saha, S. and Mäkinen, K. (2020) Insights into the functions of eIF4E-binding motif of VPg in potato virus A infection. *Viruses*, **12**, 197.

Schmitt-Keichinger, C. (2019) Manipulating cellular factors to Combat viruses: a case study from the plant eukaryotic translation initiation factors eIF4. *Front. Microbiol.* **10**, 17.

Simpson, C.G., Clark, G., Davidson, D., Smith, P. and Brown, J.W. (1996) Mutation of putative branchpoint consensus sequences in plant introns reduces splicing efficiency. *Plant J.* **9**, 369–380.

Tang, Y., Abdelrahman, M., Li, J., Wang, F., Ji, Z., Qi, H., Wang, C. *et al*. (2021) CRISPR/Cas9 induces exon skipping that facilitates development of fragrant rice. *Plant Biotechnol. J.* **19**, 642–644.

Tavert-Roudet, G., Anne, A., Barra, A., Chovin, A., Demaille, C. and Michon, T. (2017) The Potyvirus particle recruits the plant translation initiation factor eIF4E by means of the VPg covalently linked to the viral RNA. *Mol. Plant Microbe Interact.* **30**, 754–762.

Veillet, F., Perrot, L., Chauvin, L., Kermarrec, M.P., Guyon-Debast, A., Chauvin, J.E., Nogué, F. *et al*. (2019) Transgene-free genome editing in tomato and potato plants using Agrobacterium-mediated delivery of a CRISPR/Cas9 cytidine base editor. *Int. J. Mol. Sci.* **20**, 402.

Veillet, F., Durand, M., Kroj, T., Cesari, S. and Gallois, J.L. (2020) Precision breeding made real with CRISPR: illustration through genetic resistance to pathogens. *Plant Commun.* **1**, 100102.

Wang, Y., Ma, M., Xiao, X. and Wang, Z. (2012) Intronic splicing enhancers, cognate splicing factors and context-dependent regulation rules. *Nat. Struct. Mol. Biol.* **19**, 1044–1052.

© 2023 The Authors. *Plant Biotechnology Journal* published by Society for Experimental Biology and The Association of Applied Biologists and John Wiley & Sons Ltd.,

14677652, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/pbi.14003 by Inrae - Dipso, Wiley Online Library on [16/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Yoon, Y.-J., Venkatesh, J., Lee, J.-H., Kim, J., Lee, H.-E., Kim, D.-S. and Kang, B.-C. (2020) Genome editing of eIF4E1 in tomato confers resistance to pepper mottle virus. *Front. Plant Sci.* **11**, 1098.

Zhang, X., Zhang, Y., Wang, T., Li, Z., Cheng, J., Ge, H., Tang, Q. *et al*. (2019) A comprehensive map of intron Branchpoints and lariat RNAs in plants. *Plant Cell*, **31**, 956–973.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1 Alignment of the amino acid sequences from potyvirus susceptible and resistant tomato and pepper accessions.
Figure S2 Detection of insertion of the DNA donor template after the biolistic experiment with TALEN.
Figure S3 Details on the target deletion using CRISPR-Cas9 to produce eIF4E1^{P69T}-expressing plants after the TALEN method.
Figure S4 The targets, constructs and scheme for the co-base editing of *SlALS1* and *SleIF4E1* genes.
Figure S5 Genomic and cDNA sequencing of P69T/M109I + G110K + A111T indicate the abnormal splicing of the *eIF4E1* gene.
Figure S6 eIF4E editing does not affect the *eIF4E2* mRNA expression and eIF4E2 protein accumulation level *in planta*.
Figure S7 Three-dimensional homology modelling of the tomato eIF4E1 protein, based on crystallography data from the *Pisum sativum* eIF4E 3D structure (PDB ID: 2WMC-C).
Data S1 Sequence analysis of plant #266 with insertion of the donor template.
Data S2 *SleIF4E*1 genomic sequence and TALEN® target sequences pCLS24222 and 24224.
Data S3 The left and the right TALEN® cassette sequences in pCLS24222 and pCLS24224, respectively.
Data S4 DNA donor template with P69T mutation.
Data S5 The single guide RNA cassette sequence for targeting *SleIF4E1*.
Table S1 Number of the individuals used and obtained for the Region I codon insertion.
Table S2 Synthesis of all nucleases and target sites used.
Table S3 List of primers used in this study.

© 2023 The Authors. *Plant Biotechnology Journal* published by Society for Experimental Biology and The Association of Applied Biologists and John Wiley & Sons Ltd.

14677652, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/pbi.14003 by Inrae - Dipso, Wiley Online Library on [16/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License