UNIVERSITÀ DEGLI STUDI DI TORINO

SCUOLA DI DOTTORATO IN SCIENZE DELLA NATURA E TECNOLOGIE INNOVATIVE

DOTTORATO IN SCIENZE AGRARIE, FORESTALI ED AGROALIMENTARI

CICLO: XXX

CRISPR/Cas9 gene editing for functional genomics in Solanaceae species

Silvia Gianoglio

Relatore:
Prof.ssa Cinzia Comino

Coordinatore del Ciclo:
Prof.Aldo Ferrero

2014-2015-2016

There does not exist a category of science to which one can give the name applied science. There are science and the applications of science, bound together like the fruit to the tree which bears it.

Louis Pasteur

Summary

General Introduction ... 1

1 - Genome Editing ... 3

1.1 On genome modification and genome editing ... 3

1.2 Different classes of programmable nucleases ... 4

1.2.1 Zinc finger nucleases (ZFNs) ... 9

1.2.2 Transcription activator like effector nucleases (TALENs) ... 10

1.2.3 CRISPR: clustered regularly interspaced short palindromic repeats ... 11

1.3 Genome editing with CRISPR/Cas9 ... 11

1.3.1 CRISPR-Cas systems and adaptive immunity ... 12

1.3.2 CRISPR-Cas systems classification ... 14

1.3.3 Engineering of CRISPR-Cas systems ... 15

1.3.4 Development of novel synthetic CRISPR systems ... 18

1.4 Implications for the regulation of edited crops ... 19

2 - State of the art: gene editing of Solanaceae crops ... 22

2.1 CRISPR/Cas9 editing in tomato (*Solanum lycopersicum* L.) ... 22

2.2 CRISPR/Cas9 editing in potato (*Solanum tuberosum* L.) ... 25

2.3 Final remarks and perspectives ... 26

3 - The GoldenBraid cloning standard ... 28

3.1 GoldenBraid synthax ... 29

3.2 GoldenBraid domestication ... 33

3.3 GoldenBraid Multipartite Assemblies ... 34

3.4 GoldenBraid Binary Assemblies ... 35

3.5 The GoldenBraid CRISPR/Cas9 toolkit ... 36

3.6 Software tools for GoldenBraid cloning ... 38

Bibliography ... 40

Objectives ... 47

Chapter I : Generation of a tomato chlorophyll retaining *gf1* mutant using the CRISPR/Cas9 GoldenBraid toolkit ... 49

1 - Introduction ... 51

1.1 Chlorophyll catabolism ... 51

1.2 Staygreen proteins in plants ... 54

1.3 Staygreen phenotypes ... 56

1.4 Tomato *sgr* mutants ... 59

1.5 Experimental goals ... 60
2 - Materials and methods ... 62
2.1 Vector design and construction ... 62
2.2 Plant material ... 64
2.3 *Agrobacterium*-mediated transformation ... 65
2.4 Organogenesis and regeneration ... 65
2.5 DNA extraction and T7EI endonuclease assay ... 65
2.6 Genotyping of the T_0 generation ... 68
2.7 Evaluation of off-target effects ... 69
2.8 Genotyping of the T_1 generation ... 72
2.9 Phenotypic evaluation ... 72
3 - Results ... 73
3.1 Vector design and construction ... 73
3.2 Plant regeneration ... 74
3.3 Genetic characterization of the T_0 generation ... 74
3.4 Analysis of off-target effects ... 82
3.5 Genetic evaluation of the T_1 generation ... 84
3.6 Phenotype ... 85
4 - Discussion ... 88
4.1 Transformation and regeneration efficiency ... 88
4.2 Editing efficiency ... 89
4.3 Editing patterns ... 90
4.4 Continued activity of Cas9 during plant development ... 91
4.5 Effect of *in vitro* culture on polyploidization ... 92
4.6 Analysis of off-target effects ... 93
4.7 Heritability and segregation in T_1 plants ... 94
4.8 Effect of *gf1* on phenotype and shelf life ... 95
5 - Conclusions ... 96
Bibliography ... 97
Supplementary Material ... 103
Supplementary Sequence List ... 109
Chapter II : Engineering of tomato resistance to tomato yellow leaf curl virus (TYLCV) through a double CRISPR/Cas9 approach ... 111
1 - Introduction ... 113

1.1 The tomato yellow leaf curl disease ... 113
1.2 TYLCV resistance *loci* in tomato ... 115
1.3 CRISPR/Cas approaches to virus resistance ... 116
1.4 Experimental goals ... 118
2 - Materials and Methods ... 119
2.1 Vector design and construction ... 119
2.1.1 Editing of a plant endogenous gene ... 119
2.1.2 Transformation of the host with a CRISPR/Cas9 construct directed at viral sequences ... 119
2.2 Plant material ... 120
2.2.1 Stable transformation ... 120
2.2.2 Transient transformation ... 120
2.3 *Agrobacterium*-mediated transient transformation ... 120
2.4 *Agrobacterium*-mediated stable transformation ... 122
2.5 Organogenesis and regeneration of stable transformed plants ... 123
2.6 Genotyping of the T_0 generation ... 123
2.7 Evaluation of off-target effects ... 124
3 - Results ... 125
3.1 *Ty5/pelo* CRISPR/Cas9 editing ... 125
3.1.1 Target identification ... 125
3.1.2 *S*table transformation ... 125
3.1.3 Genetic and phenotypic characterization of the T_0 generation ... 125
3.1.4 Analysis of off-target effects ... 129
3.2 *TYLCV* CRISPR/Cas9 editing for plant resistance ... 131
3.2.1 Target identification ... 131
3.2.2 Transient expression assays ... 134
3.2.3 *TYLCSV* and *TYLCV* genotyping of stable transformants ... 139
4 - Discussion ... 140
4.1 *Ty5/pelo* CRISPR/Cas9 editing ... 140
4.1.1 Efficiency and patterns of mutation in *Ty5/pelo* lines ... 140
4.1.2 *Ty5* phenotype and alternative strategies for editing ... 142
4.1.3 Off-target activity ... 144
4.2 Transformation of the host with a CRISPR/Cas9 construct directed at viral sequences ... 144
4.2.1 Choice of targets for the tomato yellow leaf curl virus ... 144

4.2.2 Transient expression of TYLCV and TYLCSV CRISPR constructs in *N. benthamiana* ... 145
4.2.3 TYLCV/TYLCSV stable transformation ... 146
5 - Conclusions ... 147
Bibliography ... 148
Supplementary Material ... 152
Supplementary Sequence List ... 158
Chapter III : A CRISPR/Cas9 editing protocol for the polyphenol oxidase (*ppo*) gene family in eggplant ... 161
1 - Introduction ... 163
1.1 PPOs in plants ... 165
1.2 Roles of PPOs in plants ... 168
1.3 PPOs in Solanaceae ... 170
1.4 Genetic engineering of PPOs to reduce browning ... 172
1.5 Experimental goals ... 172
2 - Materials and methods ... 174
2.1 Target identification ... 174
2.2 Expression analysis ... 174
2.3 Vector design and construction ... 175
2.4 Plant material used for transformation ... 176
2.5 *Agrobacterium*-mediated stable transformation ... 176
2.6 Organogenesis and regeneration ... 176
2.7 Genotyping of target *loci* and evaluation of off-target effects in ‘Ecavi’ ... 177
2.8 Genotyping of target *loci* in ‘Black Beauty’ plantlets ... 178
3 - Results ... 180
3.1 PPO identification and phylogenetic analysis ... 180
3.2 Expression analysis ... 185
3.3 gRNA identification and vector design ... 186
3.4 Establishing an efficient regeneration protocol ... 187
3.5 Genetic characterization of the T_0 generation ... 189
3.5.1 ‘Ecavi’: genotyping of target and off-target *loci* ... 189
3.5.1.1 ‘Ecavi’: genotyping of target *loci* ... 189
3.5.1.2 ‘Ecavi’: genotyping of off-target *loci* ... 196
3.5.2 ‘Black Beauty’: genotyping of target *loci* ... 200
4 - Discussion ... 204

4.1 Identification of new PPO encoding *loci* 204
4.2 Functional specialization of PPOs in the Solanaceae family 205
4.3 Editing a conserved gene family: gRNA design, multiplexing and genotyping . 206
4.4 Editing patterns and efficiency in *calli* of the 'Ecavi' variety 208
4.5 Editing specificity and off-target analysis in *calli* of the 'Ecavi' variety............ 209
4.6 Editing patterns and efficiency in plantlets of the 'Black Beauty' variety.......... 210
4.7 Factors affecting *in vitro* regeneration efficiency of eggplant 211
5 - Conclusions 214
Bibliography 215
Supplementary Material.. 219
Supplementary Sequence List I 229
Supplementary Sequence List II 236
General Discussion and Conclusions.. 245
1. What to expect of the CRISPR era.. 247
2. The great bottleneck: regeneration and organogenesis. 249
3. Critical aspects of gene editing 251
4. Conclusive Remarks and Perspectives.. 253
Bibliography 256
Acknowledgements 259

General Introduction

1 - Genome Editing

1.1 On genome modification and genome editing

Genome modification is arguably as old an activity as agriculture itself. Thanks to the high adaptability and plasticity of plants and their genomes, along with the immense variability of traits and phenotypes, the very history of domestication and, later on, of deliberate breeding is a history of genome modifications. These range from whole genome duplications and polyploidization events to the selection of specific mutations responsible for relevant phenotypes such as growth habit, dwarfism, fertility, the content of nutrients and toxins, resistance to stress and pests and, not less importantly, the shape, size and color of leaves, flowers and fruits.

From the dawn of agriculture ten thousands of years ago to the XX century, when the laws of genetics published by Mendel in 1866 were rediscovered, plant breeding was necessarily a matter of identifying spontaneous mutations, followed by selection and crossing. An important innovation came from the discovery of the phenomenon of heterosis and the use of plant hybrids. In the course of the last century plant genetics has progressively integrated a wide array of technologies to serve its scope of improving crops, through both the identification and the induction of mutations. These technologies come from basic research in plant science as well as other branches of biology, with a particular obligation towards microbiology, and from physics and chemistry. For example, mutagenesis induced by chemical and physical agents, mainly based on nuclear techniques, especially ionizing radiations such as gamma- and X-rays, gave impulse to the creation of over 1,800 new plant varieties in the second half of the XX century (Ahloowalia and Maluszynski, 2001). Mutagenic techniques enabled the creation of genetic variability which was used to screen for desired phenotypes, but also potentially introduced important mutations and chromosomal rearrangements at undesired locations in the genome.

In the 1970s, new tools to recombine DNA from different sources were developed and, later in the 1980s, extended to allow the insertion of genes into a plant genome. In 1977, the discovery of the ability of *Agrobacterium* species (*i.e. Agrobacterium tumefaciens* and *A. rhizogenes*) to transfer and integrate bacterial DNA into plant genomes, in order to achieve expression of genes which are useful for the pathogen, drove the development of plant transformation. Beside *A. tumefaciens* and *A. rhizogenes*, various techniques for direct gene transfer were applied to plant transformation, especially particle bombardment, developed in the mid-Eighties, and polyethilen glycole (PEG)-

mediated transformation. Plants (or other organisms) carrying a gene inserted using recombinant DNA methods were referred to as genetically modified organisms (GMOs). Two categories of genetically modified plants can be distinguished: (i) transgenics, which contain genetic elements deriving from non-sexually compatible organisms (other plants, animals, fungi or Prokaryotes), and (ii) cisgenics, in which instead all new genetic elements (genes, but also regulatory sequences) derive from sexually compatible organisms. Transgenesis constituted a first step towards precise genome modifications, allowing the transfer of one or more selected traits and the silencing of target genes through RNA interference. Generally it does not, however, allow one to control the site of insertion of the transgene, nor the number of its copies, though both these traits can be subsequently evaluated. Transgenics, in addition, have been the object of an intense debate for over twenty years now, in relation to the concerns raised in different segments of society by the introduction of foreign DNA into an organism, its release into the environment and its effects on ecosystems and health. Cisgenesis partly escapes these concerns by exclusively employing genes and regulatory genetic elements (promoters, terminators) originating from the same species which is being transformed, or from a distinct but sexually compatible species (Espinoza *et al.*, 2013). This should not raise concerns about the introduction of DNA from an alien species and does not alter the gene pool of the transformed organism; ultimately, cisgenesis can be regarded as a means to accelerate a breeding process which could be obtained over long periods of time by backcrossing or by screening for spontaneous mutations. Still, there are some aspects of the transformation process, for which cisgenic plants would differ from spontaneous mutants: in particular, the number of copies and the site of insertion of the cisgene, and the possible presence of DNA fragments belonging not to the gene of interest, but to the bacterial vector used for transformation (Schouten *et al.*, 2006; Hunter, 2014). Some of the limitations associated to transgenesis and cisgenesis were overcome by the use of a class of enzymes called programmable nucleases, which allow to target and modify specific *loci* in a genome, in a way which is substantially more precise and predictable, and which leaves fewer traces of the introduction of alien genetic material into the target organism.

1.2 Different classes of programmable nucleases

Over the past twenty years, a group of techniques was developed, based on the activity of so-called programmable nucleases. These proteins share the ability to

perform what is known as targeted genome engineering or genome editing, that is the specific and precise modification of the genome at a predetermined *locus* (Kim & Kim, 2014; Weeks *et al.*, 2016). Programmable nucleases introduce double strand breaks (DSBs) at a specific, arbitrarily attributed *locus*, and therefore induce the activation of DNA repair systems (Fig. 1), which can act through non-homologous end joining (NHEJ) or through homology-directed repair (HDR). Non-homologous end joining is the simplest repair mechanism for DSBs, and is the most common in plants: the two broken ends are ligated back together, but in this process small mutations, especially insertions and deletions, are frequently introduced. Homology-directed repair is a more precise repair mechanism which uses homologous sequences (often sister chromatids during DNA replication) as templates to faithfully repair damaged DNA. This can result in the substitution of a particular allele with a different one.

The introduction of a DSB triggers the creation of mutations at the target site (Fig. 1). The simplest form of genome editing, and the most successful so far, is gene knock out through the introduction of indels, which is useful both for the creation of valuable phenotypes through the inactivation of specific genes and for the functional study of mutants. Gene knock out induced by mutations is significantly more effective than knockdown by RNA interference, which is often incomplete. Also, by coupling two gRNAs on the same target, deletions can be easily obtained. Genome editing may also be used for targeted insertions at specific *loci*, though with a low efficiency, especially in plants compared to other Eukaryotes (Hanin and Paszkowski, 2003). The creation of a double strand break can be exploited to promote homology-directed repair for the insertion or substitution of a DNA fragment. To this end, it is necessary to provide a suitable repair template together with the gene editing machinery. If a single DSB is induced, it is possible to insert a DNA fragment at the cut site. If two DSBs are induced, they can lead to the deletion of the DNA sequence between them, and the repair template serves to substitute the excised fragment. To increase the probability of the insertion of a DNA template, this must be flanked by sequences which are homologous to those found at the ends of the cut (Bortesi *et al.*, 2016; Ceasar *et al.*, 2016).

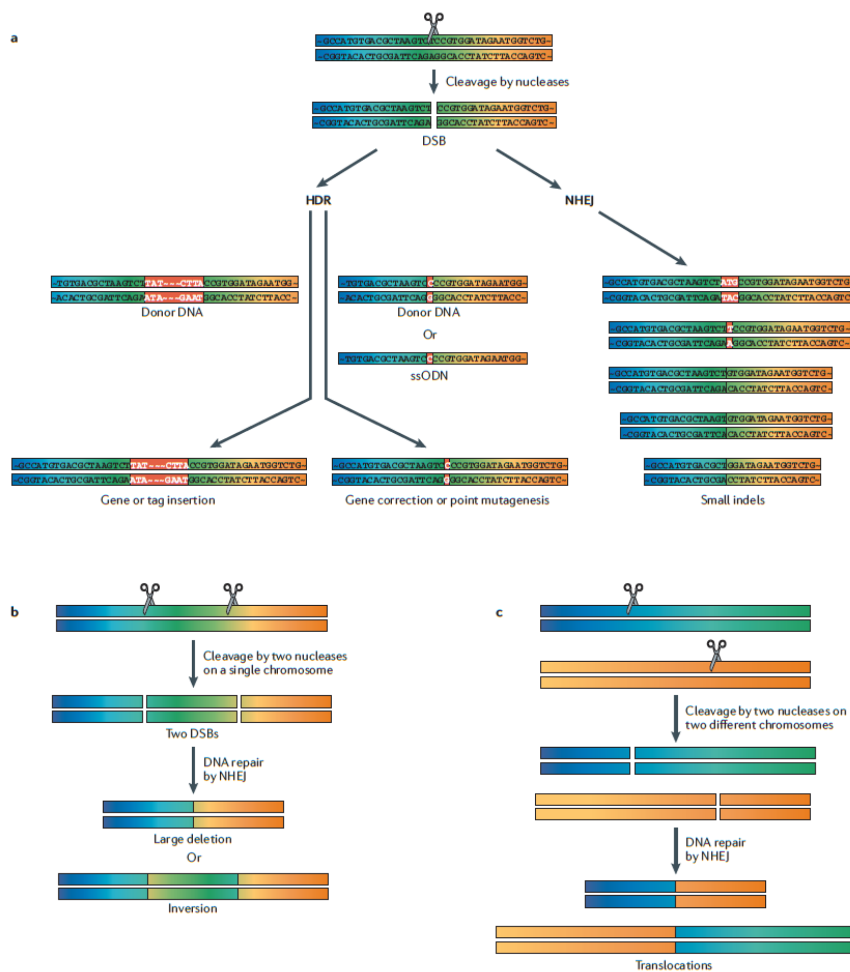


Figure 1. Possible outcomes of gene editing in plant cells. Introduction of double strand breaks at target sites induces the activation of repair mechanisms in the cell. (a) When a single DSB is induced, DNA repair can occur either through non-homologous end joining (NHEJ), which often results in inaccurate repair of the damaged site, with introduction of indels, or through homology-directed repair (HDR), a more precise but less frequent mechanism, in which the site is repaired using a homologous sequence as template. HDR can be exploited to induce the introduction of particular sequences at target sites, by providing them with homology arms to the sequence adjacent the breakpoint. (b) If two DSBs occur on the same DNA molecule, in the absence of a repair template, a deletion can occur or, alternatively, the excised fragments can be re-integrated in inverted orientation. (c) If DSBs occur simultaneously on two different chromosomes, translocations can occur through a NHEJ mechanism. Kim & Kim (2014).

Programmable nucleases essentially introduce three fundamental advances in the field of genome modifications:

1. it is possible to precisely determine the *locus* where the mutation will be introduced and, if the whole genome sequence is available, it is also possible to predict off-target effects based on sequence homology: by screening for mutations at the target and off-target *loci* one can assess the global impact of genome editing with unprecedented accuracy;
2. it is not required for programmable nucleases to be constitutively present and active in the cell to ensure their effect: once the DSB is created and repaired, the resulting mutation is passed on to daughter cells without further need of endonucleasic activity, which makes this technology substantially different from the expression of a transgene or the RNA interference strategy;
3. there are different possible methods for the delivery of programmable nucleases to plant cells, some of which are 'DNA-free' and do not imply the integration of a transgene into the host genome. Even when a transgene is involved, since its integration site and the edited *locus* are distinct, from transformed plants it is possible to obtain an offspring in which the two *loci* (the T-DNA and the edited target) are segregated, thereby retaining the mutation and discarding the transgene. These individuals are in fact undistinguishable from spontaneous mutants at the same genomic location.

Three main classes of programmable nucleases exist (Fig. 2), which differ for their specificities and the mechanisms responsible for sequence recognition: zinc finger nucleases (ZFNs), transcription activator like effector nucleases (TALENs), and clustered regularly interspaced short palindromic repeats (CRISPR)/Cas endonucleases, in particular Cas9. They have been used over the past twelve years to edit a number of important plant species (Fig. 3) and their general characteristics will be described in the following paragraph.

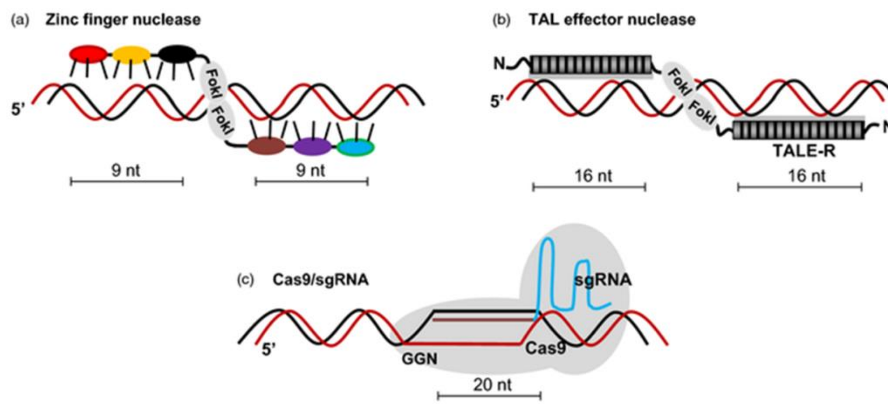


Figure 2. Three fundamental classes of customer nucleases. The identity of the recognized sequences depend either on the structure of the protein or on a RNA molecule. The length of the recognized sequence also varies. All induce double strand breaks which can be repaired through NHEJ or HDR. (a) Each ZFN recognizes a 3 nucleotide DNA sequence; 3-6 ZFNs can be combined to form zinc finger proteins. (b) TALENs are made of tandem arrays of 30-35 aminoacids, each array recognizing a 16 nucleotide sequence. (c) Cas9 is a RNA-guided endonuclease, which is targeted to a 20 nucleotide DNA sequence by a single guide RNA (sgRNA) partially complementary to the target. Weeks *et al.* (2016).

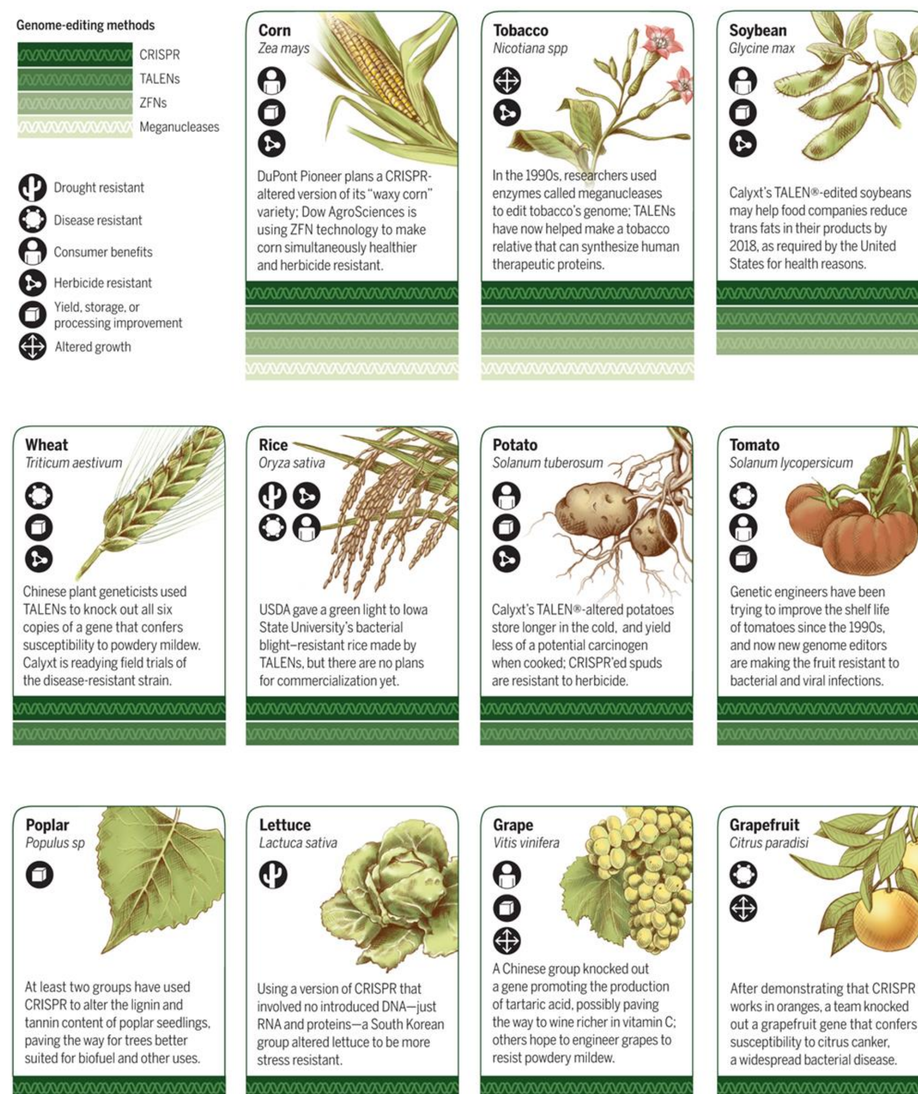


Figure 3. An overview of plant model species and crops that have been subject to gene editing with programmable nucleases. Pennisi, 2016.

1.2.1 Zinc finger nucleases (ZFNs)

The first programmable nucleases to be discovered were zinc-finger nucleases (ZFNs), which are composed of two domains: one domain is a DNA-binding zinc-finger protein (ZFP), the other is a nucleasic domain derived from the *Fok*I restriction enzyme (Klug, 2010; Urnov *et al.*, 2010). The DNA-binding domain

of *Fok*I is replaced by the ZFP to create a ZFN; in addition, the *Fok*I nucleasic domain must dimerize to cleave DNA, so that two ZFN monomers are needed to make a functional nuclease. The two ZFN monomers (called left ZFP and right ZFP) bind to adjacent half-sites on opposite strands, separated by a 5-7 nucleotide sequence. The need for dimerization and the recognition of two adjacent sequences increases the binding specificity of ZFNs. ZFP domains determine the specificity for a target sequence. Each ZFP domain is constituted of tandem arrays of C_2H_2 zinc-fingers, which represent the most common DNA-binding motifs in higher Eukaryotes. Each zinc-finger is capable of recognizing a 3 bp sequence; since a single ZFN subunit usually contains 3 to 6 zinc-fingers, the resulting target sequence is made of 9 to 18 bp. The fundamental feature of programmable nucleases is that they can be mutagenized to obtain different sequences specificities: different combinations of zinc-fingers can achieve customized target recognition, albeit with some restraints based on sequence characteristics and composition. ZFNs have been shown to function in human cells, mouse, zebrafish, frog and insects (Watanabe *et al.*, 2012; Chandrasegaran and Carroll, 2016); in plants, most experiments have been carried out in Arabidopsis (Petolino, 2015). Notable examples of the application of ZFN technology to some important crops are reported for tobacco (Wright *et al.*, 2005), maize (Shukla *et al.*, 2009; Ainley *et al.*, 2013) and soybean (Curtin *et al.*, 2011). In addition, Peer *et al.* (2015) successfully edited the genomes of apple and fig through a ZFN-based approach. The major disadvantage of this technology is the complex engineering of an appropriate DNA binding domain for each editing target.

1.2.2 Transcription activator like effector nucleases (TALENs)

Transcription activator like effector nucleases (TALENs) are a different class of programmable nucleases, and, like ZFNs, they contain the *Fok*I nucleasic domain but possess a different DNA-binding domain derived from *Xanthomonas* and *Ralstonia* bacteria (Joung and Sander, 2012; Sun and Zhao, 2013). In natural systems, bacteria transfer transcription activator like effectors (TALEs) to host cells through a type III secretion system and use these proteins to alter DNA transcription by binding to specific sites, thereby facilitating bacterial invasion.

These TALE domains are made of tandem arrays of 30-35 amino acid repeats, each recognizing a single base pair in the major groove of the DNA double helix. In particular, their binding specificity is due to two amino acids at positions 12 and 13, called repeat variable diresidues (RVDs); four RVDs (Asn-Asn, Asn-Ile, His-Asp and Asn-Gly) are usually employed to recognize guanine, adenine, cytosine and thymine, respectively. These features make the

design of TALENs easy compared to other types of nucleases and, most importantly, they can be targeted to any chosen sequence, with the only limitation of having a thymine at the 5' end. TALENs have been used to edit the genomes of several plant species, including Arabidopsis, tomato, potato, tobacco, soybean, maize, wheat and barley (Weeks *et al.*, 2016).

1.2.3 CRISPR: clustered regularly interspaced short palindromic repeats

The last genome editing technology derives from a Prokaryotic adaptive immune system that incorporates fragments of previously encountered invasive DNA sequences and, through the expression of clustered regularly interspaced short palindromic repeats (CRISPR), uses them to neutralize viral infections. CRISPR-mediated gene editing involves DNA targeting using guide RNAs (gRNAs), which direct the nuclease to its target site based on Watson-Crick base complementarity (Jinek *et al.*, 2012; Doudna and Charpentier, 2014). This means that the same protein can be directed to many different sites by providing it with the appropriate RNA guide. The most widely spread of the editing systems based on RNA-guided programmable nucleases is based on the *Streptococcus pyogenes* Cas9 endonuclease, called CRISPR/Cas9, and will be described in the next paragraph.

1.3 Genome editing with CRISPR/Cas9

The first observation of what would have become known as clustered regularly interspaced short palindromic repeats (CRISPR) dates back to the work of Atsuo Nakata's group (Ishino *et al.*, 1987; Nakata *et al.*, 1989), who identified the presence of spaced repeated motifs in Gram-negative bacteria (*E. coli*, *Shigella* spp. and *Salmonella* spp.). In the early 1990s, such direct repeats were identified also in Gram positive bacteria by Hermans *et al.* (1991) and in Archaea by Mojica *et al.* (1993). The architecture of CRISPR sites shows some common features in different microbial species (Sorek *et al.*, 2013): each CRISPR *locus* displays a series of short, 20-50 bp repeat sequences, separated by spacers of similar length; the spacers have unique sequences. In each individual CRISPR *locus*, repeat sequences are conserved, but they vary between CRISPR *loci*. CRISPR arrays are present in about 90% of bacterial and 50% of archaeal analyzed species, and there can also be more than one CRISPR *locus* in a single prokaryotic genome. Repeats can have different types of sequences: some are palindromic and are predicted to form hairpin structures, while others are not. In addition, CRISPR sites often have what is referred to as a leader sequence (an A-T rich region) at their 5'. The size and number of CRISPR *loci* are not correlated to genome size, meaning that they can occupy a substantial share of the small-sized prokaryotic genomes (over 1%). This,

together with the presence of a variable cassette of CRISPR-associated genes (known as *cas*), led to the hypothesis that such repeats and spacers could actually have an important biological function. What this function was, however, remained obscure, also due to the fact that the bioinformatics tools and sequence databases which were available at the time were not sufficient to associate these sequences to any known function based on phylogeny (Mojica *et al.*, 2000; Mojica and Rodriguez-Valera, 2016). When genomics tools became advanced enough to sequence the entire genomes of many microorganisms and sequence information became more abundant, it was finally possible to find a match for the spacers, which were surprisingly homologous to sequences belonging to bacteriophages or to a conjugative plasmid of *E. coli* (Mojica *et al.*, 2005; Mojica *et al.*, 2009). This allowed to finally formulate the hypothesis that CRISPR sequences constituted an immune system of Bacteria and Archaea, which were thought to use these spacers as templates for RNA-mediated silencing of the corresponding viruses. The role of CRISPR in defense against viruses and as a way to limit horizontal gene transfer in Bacteria was substantiated by a series of studies carried out in the course of the following years (Lillestøl *et al.*, 2006; Barrangou *et al.*, 2007; Brouns *et al.*, 2008; Marraffini and Sontheimer, 2008; Marraffini, 2015). In addition to immunity, other functions were proposed for CRISPR/Cas systems (Westra *et al.*, 2014) in regulating gene expression, virulence, DNA repair and group behavior.

1.3.1 CRISPR-Cas systems and adaptive immunity

Approximately 40% of bacterial genomes and 70% of sequenced archaeal species employ CRISPR-Cas systems as an endogenous adaptive immunity against viruses and plasmids (Donohoue *et al.*, 2017). This defense system involves a conserved sequence of events: adaptation, expression and interference. Upon viral infections, Prokaryotes can recognize short stretches of the invading viral DNA (usually 20 nucleotides in length), called protospacers, and integrate them in their genome by inserting them between direct repeats of a CRISPR array, in a process called adaptation. A viral DNA sequence is recognized as a protospacer if it is adjacent to a specific motif called protospacer-adjacent motif (PAM), which has been proved to be crucial to target recognition and binding. Once the protospacers are integrated, they become known as spacers. Subsequently, transcription of the CRISPR array leads to "guide RNA biogenesis", that is the production of a short RNA molecule complementary to a target site of the virus it was acquired from. Upon a new infection by the same virus, the guide RNA (or CRISPR RNA, crRNA) leads Cas endonucleases to the target site, causing its cleavage and virus inactivation. CRISPR/Cas systems are classified according to the type of protein

effector complexes; different Cas proteins recognize different PAM sequences, so that the sequences recognized as protospacers in viral genomes vary according to the CRISPR/Cas system which is in place. Once a CRISPR array is transcribed, it is then processed, freeing individual spacer-repeat RNAs, each recognizing a particular target sequence (Fig. 4). A fascinating feature of CRISPR-mediated immunity, as compared to other prokaryotic defenses, is that it is adaptive. Since a limited number of spacers are present in a cell at a given time, and given that they are more diverse in the region adjacent to the leader sequence (while spacers tend to be degenerate at the opposite, trailer end), it was proposed that spacers are integrated progressively at the 5' end of the CRISPR array upon encounter with viruses; CRISPR *loci* thus represent a repository which dynamically keeps track of recent infections. This aspect is particularly interesting, because it demonstrates that Prokaryotes possess environmentally tuned adaptive and heritable immune systems. CRISPR systems are a way in which Prokaryotes respond to environmental threats and pass their defenses on to offspring cells.

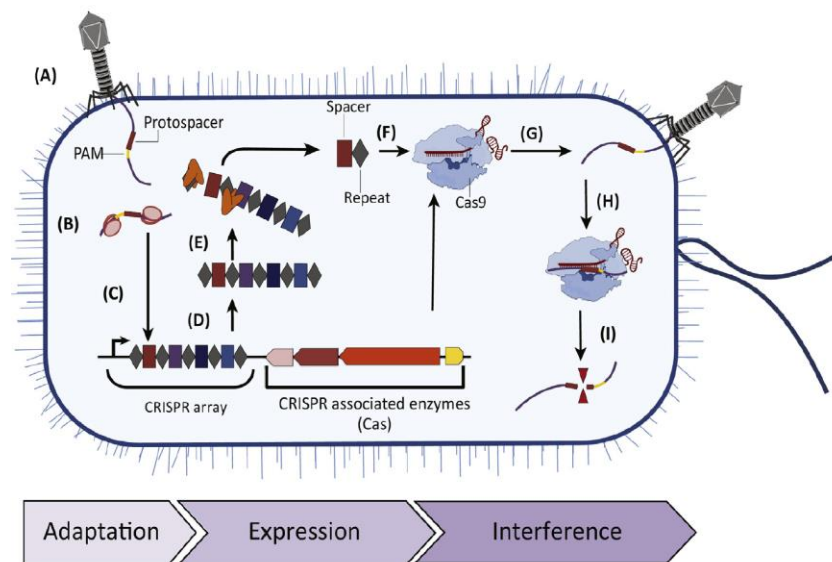


Figure 4. Phases of CRISPR-mediated immunity in Prokaryotes. Viral sequences recognized as protospacers (A) are acquired during infection and stored as spacers in *loci* known as CRISPR arrays (B-C-D), which are associated to genes encoding CRISPR-associated enzymes (Cas), usually endonucleases; this process is called **adaptation**. **Expression** of CRISPR arrays leads to the synthesis of RNA molecules known as CRISPR RNA (crRNA, E). Finally, if another infection by the same virus occurs, **interference** is mediated by crRNAs guiding Cas endonucleases (F, G) to recognize (H) and cleave (I) the viral sequence complementary to the spacer sequence specified by the crRNA. Donohoue *et al.* (2017).

1.3.2 CRISPR-Cas systems classification

CRISPR-associated systems are divided into three main categories based on the phylogeny of repeats and *cas* genes (Makarova *et al.*, 2011; Sorek *et al.*, 2013).

1. Type I CRISPR-associated systems include 6 distinct subtypes (A-F) and are characterized by the Cas3 protein, which has a N-terminal HD phosphohydrolase domain and a C-terminal DExH helicase domain: these, together, cleave and unwind dsDNA at target sequences. Cas3 is not capable, alone, to identify target DNA and crRNA-guided surveillance complexes constituted of many subtype-specific Cas proteins are in place.
2. Type II CRISPR-associated systems are unique to Bacteria and comprise four Cas proteins, namely Cas1, Cas2, Csn2/Cas4 (depending on the subtype, A or B) and Cas9. Cas9 is the defining Cas protein of this type of CRISPR-associated system and is a large multifunctional endonuclease which participates both in crRNA biogenesis and in recognition and cleavage of target DNA. It possesses two domains, a HNH nuclease domain which cleaves the DNA strand complementary to the crRNA guide, and a RuvC-like domain, which in turn cleaves the opposite, non-complementary DNA strand (Fig. 5). Cleavage thus results in a double-strand break with blunt ends. Another distinctive feature of type II systems is the presence of a second RNA molecule, called tracrRNA (trans-activating crRNA), which hybridizes with the crRNA and forms a dsRNA molecule which is processed by the cellular RNAse III. Mutations in the *Cas9* gene are known to impact tracrRNA synthesis and thus crRNA biogenesis.
3. Type III CRISPR-associated systems are a hallmark of Archaea and include Cas6 and Cas10 proteins.

Among Cas proteins, only Cas1 and Cas2, which participate in the integration of new protospacers, are universally conserved in CRISPR-containing organisms. Overall, a pool of up to 45 proteins has been identified to belong to CRISPR systems.

Makarova *et al.* (2015) proposed a new classification of CRISPR-Cas systems which recognizes only two major classes, whose main difference lies in the structure of the crRNA-effector complexes: class 1 systems have multisubunit crRNA-effector complexes, while class 2 systems are defined by a single protein (Cas9, for example) carrying out all the functions of the effector complex.

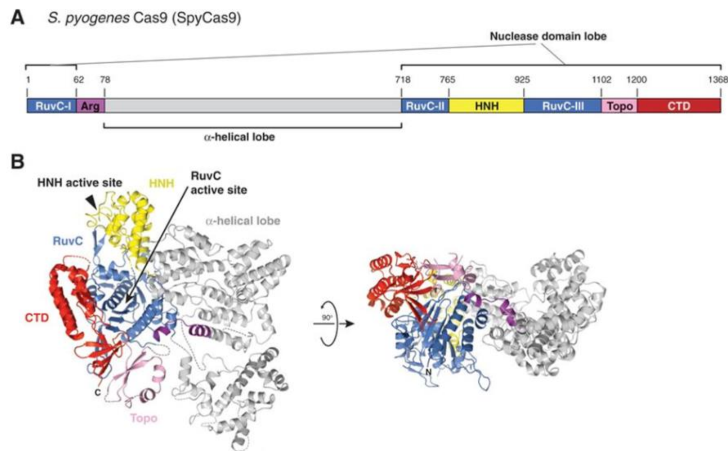


Figure 5. Structure of the Cas9 endonuclease. (A) Protein domains of the S. pyogenes Cas9 endonuclease, and (B) tridimensional structure of the protein, with the RuvC domain highlighted in blue and the HNH domain highlighted in yellow. Jinek *et al.* (2014).

1.3.3 Engineering of CRISPR-Cas systems

Jinek *et al.* (2012) demonstrated it was possible to reassemble *in vitro* some of the elements of the naturally occurring Type II CRISPR system from *Streptococcus pyogenes*, namely the tracrRNA, the crRNA and the Cas9 endonuclease. They also proved that the tracrRNA and the crRNA could be synthetically fused into a single element, named a single guide RNA (sgRNA), which would become one of the central and most interesting features of CRISPR gene editing (Fig. 6).

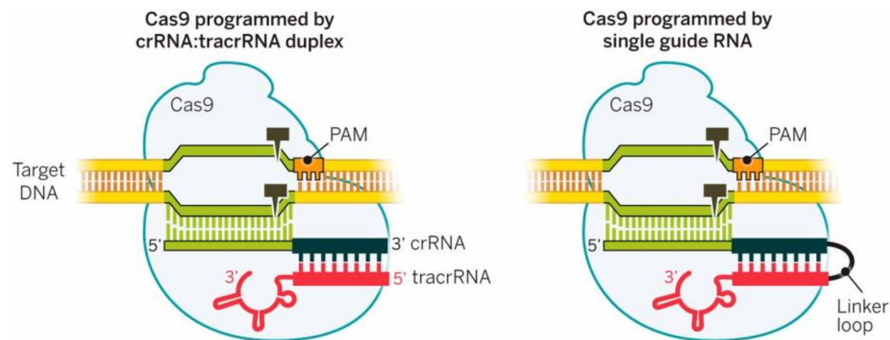


Figure 6. Comparison of natural and synthetic CRISPR/Cas9 systems. Synthetic CRISPR/Cas9 systems (right side of the figure) are composed of the Cas9 endonuclease and one synthetic single guide RNA, which is sufficient to guide the endonuclease to its target *locus*. In natural systems (to the left side of the figure), 2 RNA molecules are instead required, called a crRNA (CRISPR RNA) and a tracrRNA (trans-activating CRISPR RNA), which perform the same targeting functions. Doudna & Charpentier (2014).

The single guide RNA can be split into two components: one is specific for each target, is 20 bp in size and must be chosen by the user specifically for every gene editing experiment; the other is an invariant 80 bp sequence, called scaffold, which has a structural role and mediates the interaction of the sgRNA with the Cas9 endonuclease. It follows that the whole molecular machinery, except the 20 bp recognition sequence, can be standardized and does not need adaptation. Synthetic CRISPR/Cas systems developed for gene editing in Eukaryotes thus include fundamentally two expression units: one for the expression of the sgRNA, and the other for the expression of the Cas9 endonuclease. The single guide RNA, made of the user-designed target sequence, normally 20 bp long, and of the scaffold RNA, is placed under the control of a RNA PolIII promoter; in plants, this is usually the Arabidopsis U6 promoter for Dicots, and the U3 promoter of rice for Monocots. This particular choice of promoters imposes some constraints on the design of the gRNA, since the U6 promoter requires a G as the first nucleotide to be transcribed, and the U3 promoter requires an A (Belhaj *et al.*, 2013). The other important requirement for a target sequence is that, in analogy to the natural system, it possesses a PAM sequence at its 3' terminal. PAM sequences vary according to the endonuclease which is being used: for Cas9, the PAM sequence is 5'-NGG-3'. A series of software tools exist for assisting in target selection; along with putative target sequences and PAMs, they also usually provide scores to express the quality and the strength of each target. On-target scores represent the strength of the binding of the gRNA/Cas9 complex to the target, and depend on GC content and other sequence features of the gRNA (Doench *et al.*, 2014; Liang *et al.*, 2016; Pan *et al.*, 2016). If a reference genome is available, it is also possible to calculate an off-target score, which expresses the likelihood of the gene editing system being active in other *loci* other than the target. CRISPR/Cas9 has been known to induce some off-target effects under specific conditions (Hsu *et al.*, 2013; Doudna and Charpentier, 2014). An off-target *locus* is a sequence which presents some mismatches compared to the target; the likelihood of the CRISPR/Cas system to bind an off-target depends on the number of mismatches and on their position relative to the PAM. The nearest a mismatch is to the PAM, the more it affects recognition and binding. In fact, a seed region was identified in gRNAs, spanning about 10 nucleotides upstream from the PAM, which is thought to strongly affect CRISPR/Cas specificity (Hsu *et al.*, 2013). The C*as9* gene is usually under the control of a strong constitutive promoter, like the CaMV 35S promoter; however, tissue-specific promoters can also be used to modulate the expression of Cas9 according to specific experimental settings. Some codon optimized versions of the C*as9* gene exist for expression in Eukaryotes; in particular, a human codon optimized version of

the gene is widely used and was proved effective also for expression and activity in plants (Pan *et al.*, 2016; Vazquez-Vilar, *et al.*, 2016). Mutated versions of Cas9 are also available, in which one or both protein subunits are inactivated. If only one of the two domains (the HNH nuclease domain or the RuvC-like domain) is mutated, Cas9 acts as a nickase, cutting only the DNA strand recognized by the functional subunit. By using paired nickases, it is possible to introduce a pair of single strand breaks, leaving sticky ends. If both subunits are inactivated, Cas9 retains its ability to recognize the target sequence, but it loses its endonucleasic activity. This dead Cas9 (dCas9) can be used for transcriptional modulation of target *loci*: the dCas9 itself acts as a repressor by binding to target sites and making them inaccessible for transcription. dCas9 can also be fused to specific activator or repressor domains, or to chromatin-modifying domains (Fig. 7) (Doudna and Charpentier, 2014; Didovyk *et al.*, 2016).

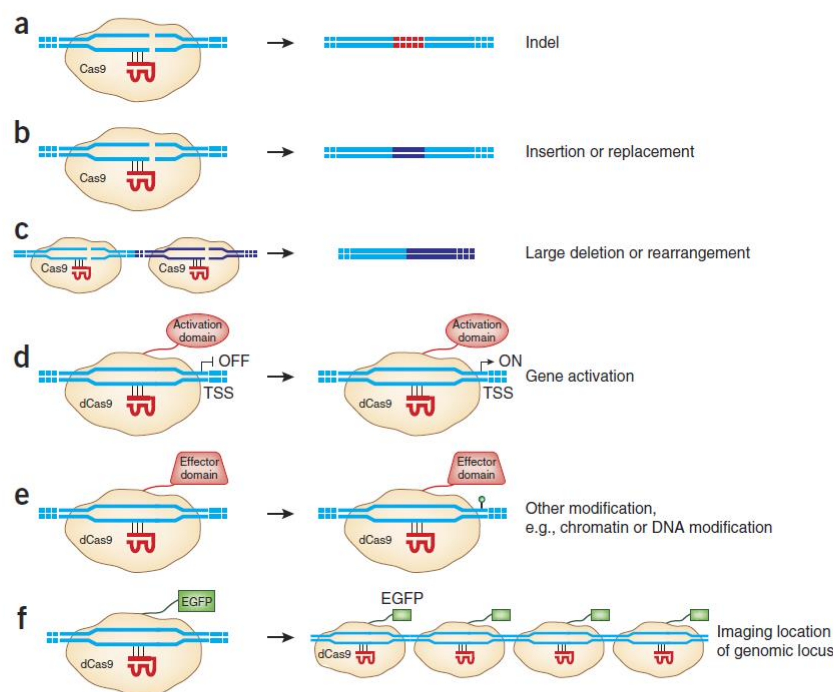


Figure 7. Cas9 variant proteins for different applications. Cas9 targeted to a single *locus* can be used to introduce a single DSB, which leads to indels, or to insertion or replacement (a, b); if Cas9 is targeted to two *loci*, large deletions or rearrangements can occur (c); an inactivated form of Cas9, called dCas9 can be fused to activation domains for gene activation (d), or to effector domains to introduce other modifications, such as chromatin remodelling and DNA modification (e); finally, dCas9 can be fused to a fluorescent tag to allow imaging of genomic locations (f). Sander and Joung (2014).

1.3.4 Development of novel synthetic CRISPR systems

Cas9 was the first CRISPR-associated endonuclease to be extensively characterized and exploited for gene editing. Figure 8 summarizes the milestones in the evolution of gene editing techniques in parallel with the uncovering of the mechanisms responsible for CRISPR-mediated defenses in Prokaryotes. Additional nucleases are starting to emerge as powerful gene editing tools. Since different nucleases recognize different target sequences and different PAMs, cleaving DNA in a specific fashion, expanding the endonuclease toolbox for CRISPR leads to expanding the possibilities of gene editing. Cpf1 is a class II type V endonuclease (Zetsche *et al.*, 2015; Zaidi *et al.*, 2017) which generates cohesive ends with 4- or 5-nt overhangs, a feature predicted to improve the efficiency of HDR mechanisms. The PAM site is different from that of Cas9, is located at the 5' of the target sequence and is a T-rich region. An interesting feature is that Cpf1, in addition to its endonucleasic activity, also has a RNAse III activity for crRNA processing, suggesting that it could be possible to express arrays of gRNAs, which would then be processed by the endonuclease itself. These features make Cpf1 an attractive alternative to Cas9 and hint to greater efficiency for some applications. Cpf1 endonucleases from *Lachnospiraceae* bacterium ND 2006 (LbCpf1) and from *Acidaminococcus* sp. BV3L6 (AsCpf1) have been shown to function effectively (Kleinstiver *et al.*, 2016). Also, the engineering of the inactivated dCas9 has led to useful applications for base editing, that is for the substitution, rather than the deletion or insertion, of a given nucleotide; this is done mainly by fusing dCas9 with a cytidine deaminase (Li *et al.*, 2017; Zong *et al.*, 2017).

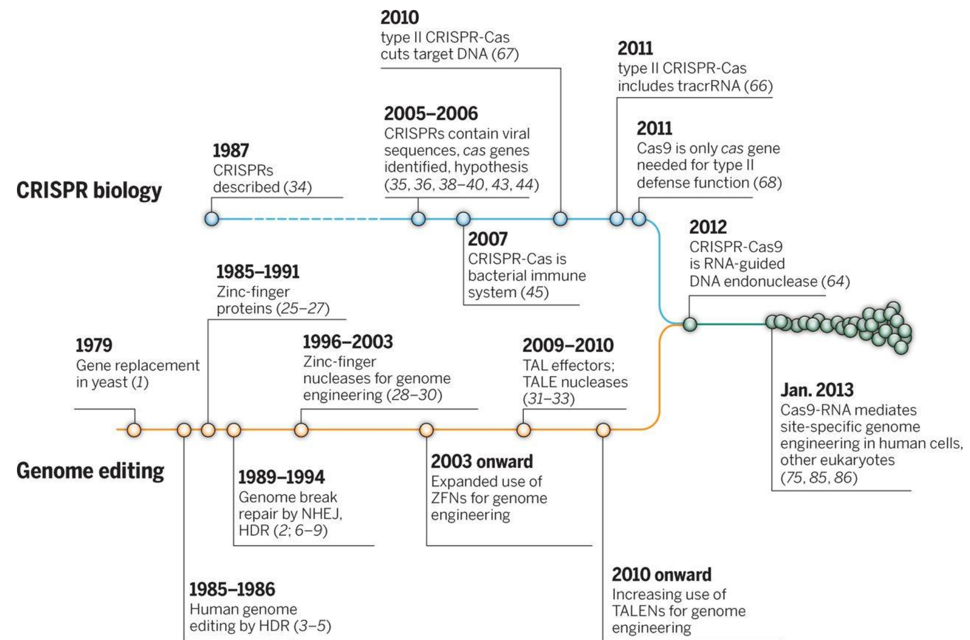


Figure 8. Evolution and milestones of gene editing techniques and of CRISPR biology. Doudna & Charpentier (2014).

1.4 Implications for the regulation of edited crops

The peculiar technological characteristics of gene editing tools leave much room for debate when considering whether they should be regulated and whether their derivatives actually fall under the definition of genetically modified organisms (Huang *et al.*, 2016; Wolt *et al.*, 2016). Different countries regulate biotechnological crops in different ways: while the European Union bases its regulatory framework on the precaution principle and thus requires thorough risk assessment and governmental authorization to release and market genetically modified crops, North American countries like the USA and Canada have a less stringent approach. The USA do not have a specific legislation on GMOs, but regulate them under the legislation on pests, plant protection and food safety, and in general refer to the substantial equivalence principle. While, according to the European precaution principle, it is necessary to prove that a novel product does not pose any foreseeable risk, the substantial equivalence principle requires novel products to be considered equivalent, except for the novel trait, to corresponding traditional products whose use has proven safe over time. Canada has a specific regulation for "plants with novel traits" (PNTs) which is based on the evaluation of novel traits themselves and not on the technology used to acquire them. As Custers (2017) clearly discusses in his

review of the European legal framework for the regulation of genetically modified crops, it is essential to consider that the EU Directive 2001/18/EC (EU, 2001), in defining the terms for regulation, explicitly takes into account *both* the technique used to obtain the modification and the characteristics of the end product. Currently, the definition of a genetically modified organism can be found in the Annex IA Part I of the Directive, stating that:

Techniques of genetic modification referred to in Article 2(2)(a) are inter alia*:*

1. *recombinant nucleic acid techniques involving the formation of new combinations of genetic material by the insertion of nucleic acid molecules produced by whatever means outside an organism, into any virus, bacterial plasmid or other vector system and their incorporation into a host organism in which they do not naturally occur but in which they are capable of continued propagation;*
2. *techniques involving the direct introduction into an organism of heritable material prepared outside the organism including micro-injection, macro-injection and micro-encapsulation;*
3. *cell fusion (including protoplast fusion) or hybridisation techniques where live cells with new combinations of heritable genetic material are formed through the fusion of two or more cells by means of methods that do not occur naturally.*

It derives that if the technique used to perform gene editing is analogous to those used to obtain any genetically modified organism, it is undoubtedly covered by this definition (in the case of crops, mainly *Agrobacterium*-mediated transformation and biolistics). The use of ribonucleoproteins constituted of Cas9 and gRNA used to transfect protoplasts, without integration in the genome, might fall outside of this definition. But what about the characteristics of the product? Even if integration of a transgene occurs, if transgene and mutation are segregated, it is possible to obtain an edited individual in which it is in fact impossible to retrieve any trace of the molecular machinery used to perform editing, and only the effect persists. The crucial point and the regulatory difficulty seem to lie in the fact that CRISPR-edited crops (after segregation of the transgene carrying the editing machinery) do not meet the requirements to be considered as genetically modified organisms, but necessarily *derive* from a genetically modified organism. The European Court of Justice is expected to rule on the regulation of edited crops in the spring of 2018. In his Opinion on Case C-528/16, Advocate General of the European Court of Justice Michal Bobek concluded that organisms obtained by mutagenesis are, in principle, exempted from the obligations in the Genetically

Modified Organisms Directive of the European Parliament and that crops obtained through gene editing might be regarded as products of mutagenesis rather than GMOs, if they have not integrated foreign DNA.

2 - State of the art: gene editing of Solanaceae crops

One of the greatest impacts of CRISPR/Cas9 mediated gene editing will probably be seen in their application to food and industrial crops. For this reason, many early reports of CRISPR/Cas9 editing in plants, aside from widely used model species such as Arabidopsis and *N. benthamiana*, regarded important crops such as rice (Miao *et al.*, 2013; Mikami *et al.*, 2015; Sun *et al.*, 2016), maize (Svitashev *et al.*, 2015; Svitashev *et al.*, 2016) and soybean (Jacobs *et al.*, 2017; Li *et al.*, 2015). Among horticultural crops, the Solanaceae family occupies an important position for its wide distribution and the economic relevance of species such as potato, tomato, eggplant and pepper. To date, among these, reports have been published for gene editing in tomato and potato. Aside from the introduction of agronomically relevant traits, CRISPR/Cas9 has also proven to be a valuable instrument for functional genetics in tomato, especially because of its role as a model species and thanks to the availability of efficient regeneration methods.

CRISPR/Cas9 gene editing has been demonstrated also in tobacco (Gao *et al.*, 2015) and petunia (Zhang *et al.*, 2016).

2.1 CRISPR/Cas9 editing in tomato (*Solanum lycopersicum* L.)

The first report of the use of CRISPR/Cas9 in tomato is from Brooks *et al.* (2014). Here, the authors used two guide RNAs directed at the argonaute 7 (*AGO7*) gene, involved in leaf development and morphogenesis, to induce a deletion; the target was chosen because it causes an early, clearly recognizable phenotype characterized by needle-like leaves. The genotyping of mutants carrying the T-DNA revealed that a series of edited alleles were present. The edited individuals were characterized as homozygous, heterozygous, biallelic or chimeric based on whether both alleles are edited in the same fashion (homozygous), only one allele was edited and the other was wild type (heterozygous), both alleles were edited, but in different fashions (biallelic) or a number of different alleles were present (chimeric). Overall, 48% of the regenerated plants showed a phenotype comparable to that of known strong mutants; being the mutation recessive, this implies that both alleles were edited. Notably, only 1 in 29 plants was homozygous for the expected deletion between the two gRNAs, while 2 others were chimeric. The remaining individuals had a range of smaller deletions. The authors hypothesize that the low frequency of

the expected deletion can be attributed to the fact that the two gRNAs should cut simultaneously. The breakpoints were usually at 3 bp upstream of the PAM, with some exceptions. The heritability of the mutation was assessed in the following generation, and in some plants it was possible to segregate the transgene.

A thorough analysis of the efficiency, mutation patterns and heritability of CRISPR in tomato was performed by Pan *et al.* (2016). Here, two gRNAs in independent expression cassettes were directed to the phytoene desaturase (*PDS)* gene, while two other gRNAs were targeted to the phytochrome interacting factor *(PIF4)* gene. *Solanum lycopersicum PIF4 (SlPIF4)* is involved in the transduction of light signals and is a bHLH transcriptional factor, which offers insight into CRISPR specificity when targeting members of a multigenic conserved gene family. *SlPDS* knock-outs showed an albino phenotype, while no obvious phenotype as associated to *SlPIF4* mutations. Editing efficiency proved to be very high (72-100%) and no off-target effects were found by analyzing putative off-target sites with 1-3 mismatches in the gRNA seed region.

Čermák *et al.* (2015) used TALENs and CRISPR/Cas9 to induce homologous recombination in tomato plants, where they inserted a strong constitutive promoter (CaMV 35S) upstream of the *ANT1* gene, encoding a Myb transcriptional factor involved in anthocyanin biosynthesis, to increase the production and ectopic accumulation of purple pigments. The activity of the two classes of endonucleases, TALENs and Cas9, were compared, but extensive characterization of mutants and heritability was performed only for TALEN-edited plants. CRISPR-induced gene editing was shown to happen at a comparable or lower rate than editing by TALENs but, in contrast, the rate of NHEJ-induced mutagenesis was shown to be higher. This is consistent with the different kind of DSB induced by the two nucleases: TALENs leave 5' overhangs, while Cas9 leaves blunt ends. This supports the idea that new CRISPR endonucleases, like Cpf1, with the ability to cut and leave overhangs will be crucial to achieve efficient recombination in plants. The strategy used to deliver DNA to the cells relied on the development of a viral vector based on the single-component bean yellow dwarf virus (BeYDV). This vector was used both for gene editing cassettes (the endonucleases under the control of a 35S promoter, and the gRNAs under the control of the AtU6 promoter) and for donor templates, containing the 35S promoter for *ANT1* overexpression and the *nptII* gene as a selection marker for kanamycin resistance. The donor template was flanked by 987- and 719-bp homology arms. The use of viral vectors makes it possible to obtain large amounts (in the order of hundreds of thousands of

copies per cell) of the constructs, greatly increasing the probability of the desired event. In addition, viral replication is expected to happen via rolling circle replication, avoiding integration of the T-DNA and making it possible to retrieve edited but T-DNA free cells: this was verified by the authors in this paper both in T_0 and in T_1 plants.

An interesting example of the application of CRISPR/Cas9 to engineer disease resistance is reported by de Toledo Thomazella *et al.* (2016), who induced broad-spectrum disease resistance in tomato through the knock-out of the downy mildew resistance 6 (*DMR6*) gene. In Arabidopsis, *AtDRM6,* which is a member of the 2-oxoglutarate Fe(II) dependent oxygenases, is known to be involved in the response to several pathogens and is required for susceptibility to downy mildew; loss-of-function *drm6* mutants show increased levels of salicylic acid and increased disease resistance. The tomato homolog for *AtDRM6*, identified by phylogenetic analysis and transcriptional profiles, appeared up-regulated during infections by *Pseudomonas syringae* pv. *tomato* and *Phytophthora capsici*. CRISPR-induced mutations in exons 2 and 3 of *SlDRM6* resulted in a resistant phenotype, with reduced pathogen growth and/or reduced severity of symptoms, without significant reduction of plant size in greenhouse conditions. Similarly, Nekrasov *et al.* (2017) produced a powdery-mildew resistant tomato variety, called ‘Tomelo’, by editing *SlMlo1*, one of sixteen genes (*SlMlo1-16*) involved in susceptibility to *Oidium neolycopersici*. In addition to its biological relevance, this work is also of extreme interest since the authors adopted a whole genome Illumina resequencing approach to ascertain the identity of the mutations, the presence or segregation of the T-DNA in the T_1 generation and the occurrence of off-target effects. It was therefore possible to unambiguously conclude that five T_1 plants were edited and did not carry any copy of the T-DNA, and that the technique is precise enough to avoid off-target activity, by screening 145 potential off-target sites.

CRISPR/Cas9 has helped elucidate some fundamental mechanisms of fruit development and plant physiology. Ito *et al.* (2015) induced mutations of the ripening inhibitor (*rin*) locus involved in fruit ripening, while Klap *et al.* (2017) and Ueta *et al.* (2017) used CRISPR/Cas9 to generate lines of parthenocarpic tomato by targeting, respectively, the agamous-like gene *SlAGL6* and the Aux/IAA gene *SlIAA9* involved in auxin signaling and fruit development. The characterization of parthenocarpic mutants is important both to study the molecular basis of seedless fruit production, and to obtain varieties with industrially useful features. Wang *et al.* (2017) knocked out *SlMAPK3* to investigate the mechanisms underlying drought resistance in tomato.

One novel application of CRISPR for plant genetics is the production of collections of mutant lines. Functional genetics traditionally relies on the availability of mutagenized populations and on gene silencing or knock out methods, where a single gene approach is taken to determine the function of a given locus. In mammalian cells, CRISPR/Cas9 had already been used to generate populations of mutants, and the same approach has now been extended to tomato (Jacobs *et al.* 2017). This was done by creating a pool of CRISPR plasmids, each carrying up to 3 gRNAs, and by co-transforming them into tomato cells. PCR screenings were used to identify the insertion of each T-DNA and the presence of the individual gRNAs, and the introduction of mutations at target and off-target sites was assessed by sequencing. The authors targeted the leucine-rich repeats (LRR) subfamily XII genes involved in pathogen interactions in one transformation event, and 18 putative transporter genes in a second transformation. In both cases, the technique proved useful to efficiently generate a pool of mutations which could be inherited to produce mutant lines.

2.2 CRISPR/Cas9 editing in potato (*Solanum tuberosum* L.)

Relatively fewer studies are available for CRISPR-mediated gene editing in potato, despite its good transformation efficiency and the availability of a high quality genome (The Potato Genome Sequencing Consortium, 2011). The first report of CRISPR/Cas9 editing in potato came from Wang *et al.* (2015): here it was first demonstrated that the native U6 RNA promoter from potato could successfully drive gRNA expression in a transient assay in *N. benthamiana* leaves to target the *PDS* gene, and subsequently a construct was built to target the *StIAA2* gene (involved in shoot morphogenesis) in potato. Characterization of the mutants proved that gene editing was efficient and that no obvious off-target effects could be identified. Later, Butler *et al.* (2015) reported efficient CRISPR gene editing in both diploid (the self-incompatible breeding line X914-10) and tetraploid ('Désirée') potato varieties. Two different types of vectors were used, a traditional T-DNA expression vector and a modified geminivirus T-DNA expression vector, to target the acetolactate synthase *StALS1* gene, whose mutations can confer herbicide resistance. The use of geminivirus-based vectors was further expanded by Butler *et al.* (2016), in an analogous fashion to what was reported for tomato by Čermák *et al.* (2015). In this later paper, the authors successfully used a BeYDV-based vector to achieve gene targeting. Among plant viruses, tomato rattle virus (TRV, a RNA virus) had been used in some instances for plant transformation (especially transient transformation), but showed some limitations, mostly due to its limited carrying capacity. Baltes *et al.* (2014) implemented a system to increase template copy number in the

plant nucleus; two T-DNAs are co-transformed in the cell, one carrying the viral *cis*-acting elements and the repair template, and the other carrying Rep/RepA. Butler *et al.* (2016) first demonstrated that BeYDV-based vectors could induce strong heterologous protein expression in potato by expressing β-glucuronidase (GUS) and evaluating tissue staining. Then, they stably transformed potato to express viral Rep/RepA, and used BeYDV-vectors to deliver either a TALEN- or a CRISPR/Cas9-based cassette, together with a repair template, to target the *StALS1* and introduce the mutations responsible for herbicide resistance. The strategy has proven successful and further substantiates the advantages of geminivirus-based factors for gene targeting (while they do not impact the efficiency of NHEJ). However, it poses the obvious problem of the stable integration of a viral protein into the plant genome and, for this reason, co-delivery of Rep/RepA and gene editing machinery together with viral *cis*-acting elements will probably be preferable.

Compared to tomato, sexual reproduction of potato is more problematic. *S. tuberosum* is usually propagated vegetatively, and this would pose a problem for the outcrossing of the T-DNA carrying the CRISPR/Cas9 machinery once the mutations are fixed at the target locus, because it would mean altering the genetic background and thus the agronomical properties of potato varieties (which are highly heterozygous and can be tetraploid). Andersson *et al.* (2017) addressed the issue by developing a system to transiently express CRISPR/Cas9 machinery in potato protoplasts through PEG-mediated transfection and to regenerate edited but T-DNA free plants. The gene chosen for targeting is the granule-bound starch synthase *GBSS* gene; mutations of *GBSS* are known to disrupt amylose synthesis, yielding potato tubers whose starch is composed mainly of amylopectin (the so-called "waxy phenotype"). Three gRNAs were directed at exons 8 and 9 of GBSS. The plant material was the tetraploid variety 'Kuras', and the method was successful at inducing mutations at all four alleles. A low rate of vector DNA integration (10%) was detected.

2.3 Final remarks and perspectives

CRISPR/Cas9 has been established as a fast, precise and reliable technology to perform gene editing in plants, and especially in the Solanaceae family, for which efficient regeneration protocols are available. Induction of NHEJ at target sites is expected to become a routine method to knock out genes in a clean, precise way. Gene targeting, whose efficiency in plant has always been quite low, is likely to benefit greatly from geminivirus-based vectors. Meanwhile,

other applications are being developed to further increase gene editing outcomes and to avoid integration of T-DNA.

In this thesis, CRISPR/Cas9 was used to introduce mutations in tomato endogenous genes and to address virus resistance. The first application of CRISPR/Cas9 to another member of the Solanaceae family, eggplant *(Solanum melongena* L.), is also reported, together with the optimization of a regeneration protocol for this species. These objectives have been reached by using vectors based on Golden Braid strategy described in the following paragraph.

3 - The GoldenBraid cloning standard

Synthetic biology, a discipline which has been thriving since the beginning of the millennium, aims at creating novel features in living organisms. This can be done following a bottom-up strategy, that is by assembling new living forms from their basic components, or following a top-down strategy, which means integrating new parts (genetic circuits) into existing organisms (Benner and Sismour, 2005; Cameron *et al.*, 2014). The approach adopted by synthetic biology derives from the awareness that, however elaborate, cells and organisms are organized as a hierarchical combination of functional modules. This view arises from the extensive amount of data produced by high-throughput molecular biology tools (genomics, transcriptomics, proteomics and metabolomics). The modular organization of cellular systems led to the notion that they could, in fact, be treated similarly to traditional engineering systems (electrical or mechanical) and that it was therefore possible to use novel combinations of existing modules to achieve new functions in a given organism in a predictable way, or even to devise entirely new elements once the requirements of the system became known. Synthetic biology represents a valuable and effective tool both for synthesis and for analysis. It is the basis for metabolic engineering, molecular farming and genome editing and modification, with profound implications for the production of bioactive compounds of pharmaceutical interest, of biofuels and of functional and sustainable crops. At the same time, synthetic biology also represents a fascinating and powerful approach to the unveiling of the minimal requirements of life, its governing principles and the mechanisms underlying cellular function and evolution (Benner and Sismour, 2005). As in all other fields of technology, the scale and relevance of practical applications depend at the same time on theoretical advances and on the availability of adequate technological means, with one aspect often driving the progress of the other.

When considering living cells as engineering and thus programmable entities, it is important to consider what are the characteristics of such systems, and what is required of their components to function properly. A key principle of engineering is standardization. Biological systems are far more complex than any man-made device, harboring more components and displaying dazzling emergent properties like self-duplication, repair, reproduction and optimization of energy consumption. For this reason, one of the greatest challenges of synthetic biology is to identify and create standardized, reusable, exchangeable elements that allow the efficient construction of genetic circuits, whose output can be predicted based on input and genetic background (Liu and Stewart,

2015). Of course, this predictability is hard to achieve because we do not have sufficient knowledge of all the elements involved in the behavior of cells; so much so, that what can be implemented in one *E. coli* strain is often not directly transferable to another strain of the same species (Arkin, 2008).

In the respect of assembling genetic parts, various systems have been developed to assemble DNA elements into transcriptional units (TUs) or multigenic structures. These systems combine the principles of efficiency, versatility and, of course, standardization: standardization implies adopting standard rules for assembly, which are valid independently of the identity of the genetic parts, and allows automation and sharing of materials between laboratories (Sarrion-Perdigones *et al.*, 2011). To be effective, standardization should rely on a simple and restricted set of rules; in an assembly standard, idempotency represents the maximum degree of simplicity and is achieved when any new composite part can be assembled by applying the same rules used to obtain its parent components.

The CRISPR gene editing system itself is a striking example of the virtues of simplicity, standardization and reusability in making a system efficient, economically sustainable and suitable for a wide range of applications. The tremendous success of CRISPR over other gene editing systems like ZFNs and TALENs depends on the possibility to use a standardized molecular machinery for many different experimental scopes and for many different targets, while retaining very high efficiency and specificity thanks to the only element which needs to be customized, a 20 bp sequence.

3.1 GoldenBraid synthax

GoldenBraid is a modular cloning standard and DNA assembly framework for plant synthetic biology (Sarrion-Perdigones *et al.*, 2013; Sarrion-Perdigones *et al.*, 2011; Vazquez-Vilar *et al.*, 2017). GoldenBraid (hereafter referred to as GB) is especially efficient because it makes use of a minimalist design, a limited set of vectors and only two type IIs restriction enzymes to perform multipartite or binary assemblies, with rules very close to complete idempotency. In addition, it envisages the possibility to indefinitely re-use each part in new assemblies, with the only constraint of the maximum vector size supported by the bacterial host. GB presents itself as an adaptation to synthetic biology of the Golden Gate cloning system (Engler *et al.*, 2008; Engler *et al.*, 2009). Like Golden Gate, GB is based on the use of type IIs restriction enzymes: these are characterized by the ability to digest DNA at a fixed distance a few nucleotides away from their recognition site, leaving 4 bp overhangs. The activity of the enzyme is independent of the sequence which is

cleaved, which makes type IIs restriction enzymes ideal for seamless cloning strategies: provided that they have compatible overhangs, two or more DNA fragments can be digested with the same enzyme and ligated together in a one-pot reaction, without leaving any sign of the recognition site. This relies of course on the careful positioning of the recognition sites and on the design of the appropriate 4 bp overhangs. GB uses two types of destination vectors, named alpha (α) and omega (ω), in which the restriction sites for two type IIs restriction enzymes (*Bsa*I and *Bsmb*I) are placed in opposing orientations (Fig. 9). This means that *Bsa*I can be used to clone DNA fragments into α vectors, and also to excise inserts from ω vectors; conversely, *Bsmb*I is used to excise fragments from α vectors and to insert them into ω vectors. When shuffling from α to ω vectors and viceversa, only binary assemblies are possible. GoldenBraid comprises four types of destination vectors, called pDGB α1, pDGB α2, pDGB ω1 and pDGB ω2, which can all act as recipients of binary assemblies. In a binary assembly, DNA fragments from α1 and α2 vectors can be combined in an ω-type vector using *Bsmb*I, while DNA parts contained in ω1 and ω2 vectors can be assembled into an α-type vector using *Bsa*I. Being designed for plant biology, GB vectors are available in two sets of pGreenII- and pCAMBIA-based vectors, suitable for transformation into *Agrobacterium tumefaciens* and subsequent plant transformation. In addition to the four vectors described above, a further set of four analogous vectors is available to clone DNA parts in reverse orientation (Sarrion-Perdigones *et al.*, 2013): these are named pDGB α1R, α2R, ω1R and ω2R. α and ω vectors contain different antibiotic resistance genes (kanamycin resistance for α vectors and spectinomycin resistance for ω vectors) which allow counterselection (Fig. 9).

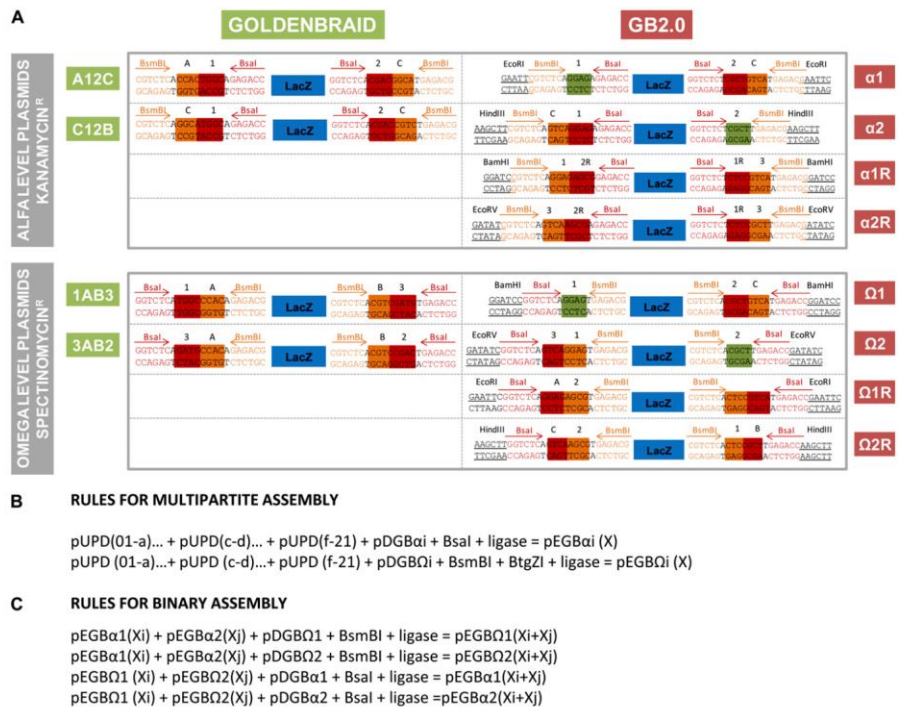


Figure 9. Arrangement of restriction sites in GB and GB 2.0 vectors and synthax rules for multipartite and bipartite assemblies. (a) Alpha and omega vectors are distinguished by the arrangement of *Bsa*I and *Bsmb*I restriction sites; in alpha vectors, *Bsa*I restriction sites are internal, which allow to remove the LacZ insert and to introduce the custom insert, while *Bsmb*I restriction sites are external and are used to excise the custom insert so that it can be cloned into an omega vector. The opposite arrangement is found in omega vectors. (b) Rules for multipartite assemblies, which allow to join various parts contained in pUPD vectors into level 1 alpha vectors using *Bsa*I, or in omega vectors using *Bsmb*I and *Btg*ZI. (c) Rules for binary assembly: once DNA parts are assembled in alpha and omega vectors, only binary assemblies are possible; the inserts of two alpha vectors can be combined into an omega vector, and viceversa the inserts from two omega vectors can be assembled into an alpha vector. Sarrion-Perdigones *et al.* (2013).

The iterative nature of GB and the complete removal of the original restriction sites allow one to assemble genetic parts into elements of increasing complexity (TUs, multigenic cassettes and regulatory elements) by indefinitely shifting in a loop from α vectors to ω vectors and back. By definition, any DNA fragment can enter the GB cloning system. In practice DNA parts, which constitute the basic elements of the standard, are divided into different categories based on their relative position and roles. GB 2.0 (Sarrion-Perdigones *et al.*, 2013) comprises 11 types of standard DNA parts (Fig. 10), broadly divided into 5' non transcribed region, transcribed region and 3' non transcribed region. Each part is specified by a particular 4 bp overhang, determining its position in the

assembly. The assembly rules behind GB, the GB grammar, effectively resemble the grammar of a spoken language: the basic units (*GB parts*, like words) are combined to form more complex structures (*GB superparts*, analogous to sentences). A GB part is, for example, a promoter, a coding sequence or a terminator, or any DNA sequence that is individually domesticated. Superparts are groups of DNA parts which have been assembled in an α- or ω-type vector and which, thanks to the seamless nature of GB cloning, behave like an indivisible unit. A typical example of a superpart is a transcriptional unit made of promoter, CDS and terminator. Superparts, just like sentences, can then be combined into constructs of higher complexity: in our case, they would typically be multigenic constructs harboring more than one TU (Fig. 11).

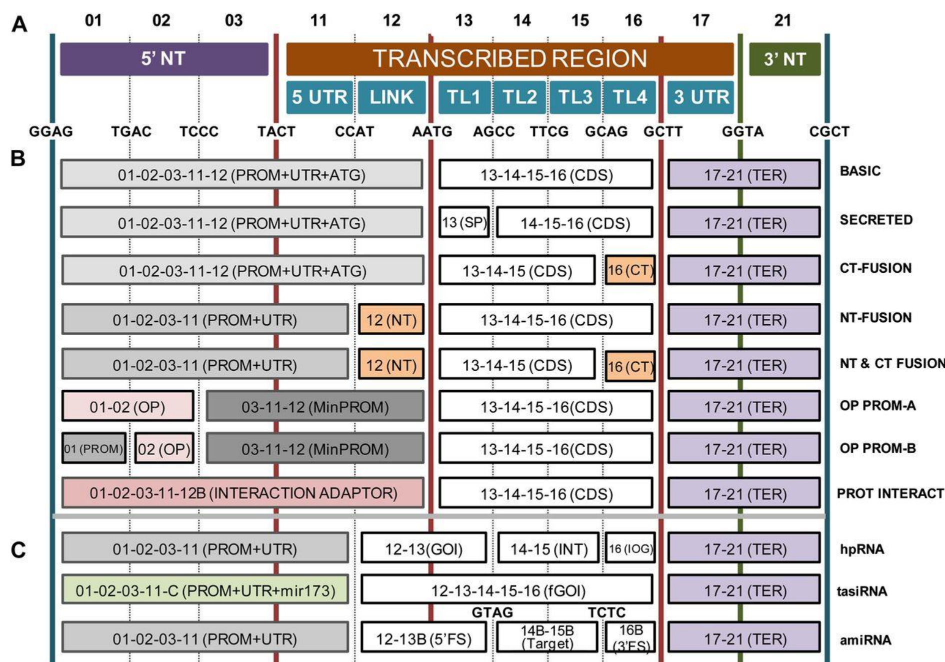


Figure 10. Eleven types of GB parts are specified by 4 nt overhangs, which determine their relative position in an assembly. (a) A schematic overview of a TU structure, with three main kinds of elements: 5' non transcribed region, transcribed region and 3' non-transcribed region. (b) Frequently used structures for the protein-coding TUs: the promoter is usually fused with the 5' UTR region, while the CDS constitutes another part, and the 3' UTR is fused with the terminator. (c) Frequently used structures for RNA silencing. Sarrion-Perdigones *et al.* (2013).

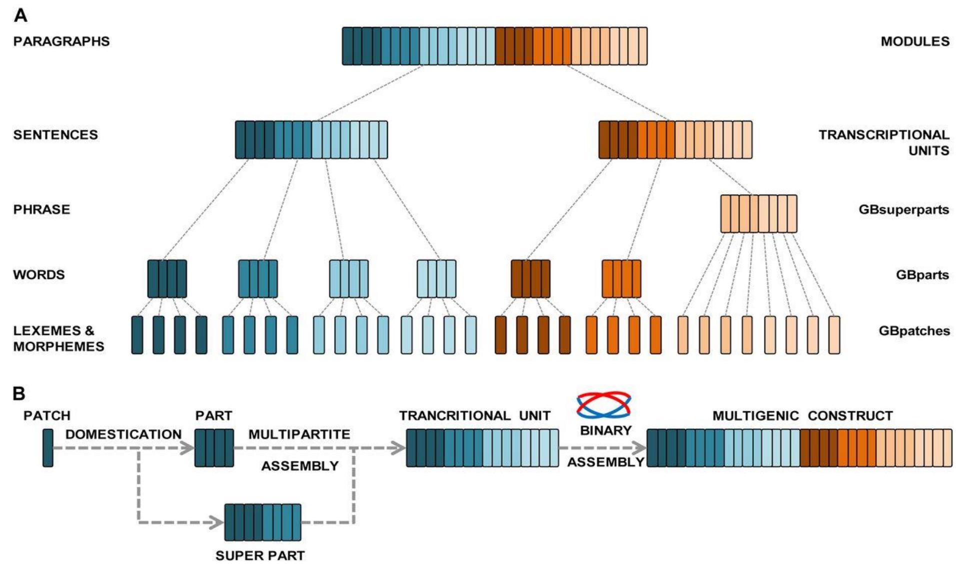


Figure 11. The rules governing the GB grammar resemble those of any spoken language. (a) Elementary parts (GB patches, analogous to lexemes and morphemes) are first assembled into GB parts (analogous to words), and then combined into superparts, transcriptional units and modules, following a limited, predetermined set of rules. (b) Flow chart of the GB assembly steps. Sarrion-Perdigones *et al.* (2013).

Interestingly, at the same time when the first description of the GB standard was published (Sarrion-Perdigones *et al.*, 2011), another cloning system called MoClo (modular cloning) was proposed by Weber *et al.* (2011). The two systems share some common principles and technical solutions, with some interesting peculiarities: while one of the most interesting features of MoClo is the possibility to perform multipartite assemblies even at higher levels, GB retains the advantage of indefinite reusability of all constructs.

3.2 GoldenBraid domestication

Every DNA fragment obtained through PCR amplification or through direct synthesis can enter the GB system through an adaptation step called domestication, which implies the addition of recognition sites for the type IIs restriction enzymes and of appropriate 4 bp overhangs, based on the type of assembly that is to be performed. A universal domesticator vector called pUPD (or pUPD2 in the latest version of the GB standard) is used to implement this step. DNA fragments are usually cloned into pUPD vectors using *Bsmb*I, and are excised thanks to *Bsa*I. One important step to consider before domestication is to scan for internal *Bsa*I or *Bsmb*I recognition sites in the sequence that is to be domesticated, which would interfere with the following cloning steps. If such

sites are present, they can be eliminated through PCR-mediated mutagenesis using primers with single mismatches to disrupt the recognition site.

The pUPD vector is referred to as the level 0 of GB cloning, where DNA parts are domesticated and enter the system. Level 1 is represented by α vectors: multiple DNA parts contained in different pUPD vectors can be inserted into an α vector through a multipartite assembly, provided their 4 bp overhangs, which specify their compatibility and relative position. Level 2 is instead represented by ω vectors: from this step on, only binary assemblies are possible. From ω vectors, it is then possible to go back to α vectors and to cycle indefinitely between the two. Actually, thanks to the presence of recognition sites for a third type IIs restriction enzyme, BtgZI, it is also possible to domesticate parts into pUPD vectors and then to perform level 1 assemblies in ω rather than α vectors (Fig.12).

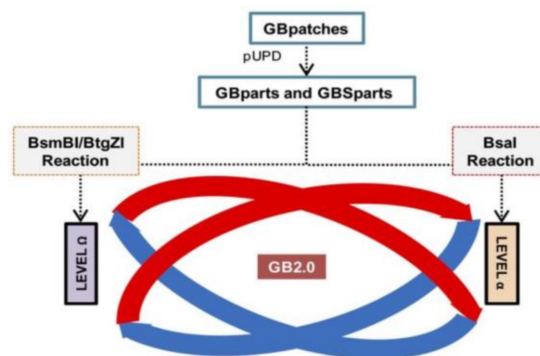


Figure 12. The domestication of DNA parts into GB. Individual sequences indicated as "GB patches" are domesticated and assembled into level 0 pUPD vector. From these, through multipartite assemblies, they can be cloned and combined into level 1 alpha or omega vectors. From here on, they can be further combined into higher level assemblies through binary assemblies. Sarrion-Perdigones *et al*. (2013).

3.3 GoldenBraid Multipartite Assemblies

Once they have been domesticated, multiple GB parts can be combined in level 1 vectors through multipartite assemblies (Fig. 13). In the multipartite assembly of a transcriptional unit, for example, the promoter, CDS and terminator contained in pUPD vectors can be inserted in level 1 alpha vectors using *Bsa*I, or in omega vectors using *Btg*ZI. Again, the compatibility and the relative positions of the GB parts are determined by their 4 nucleotide overhangs resulting from digestion of the vectors with type IIs restriction enzymes.

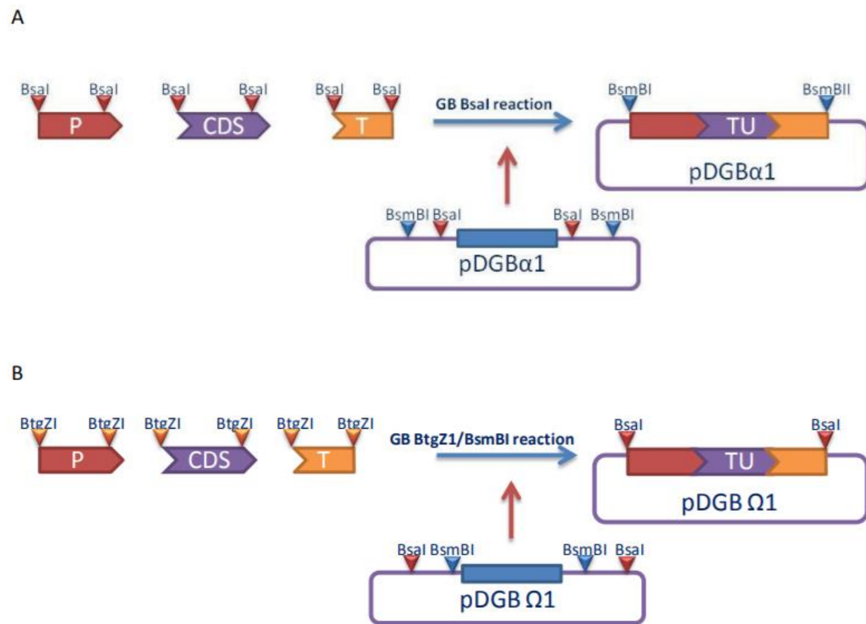


Figure 13. GB Level 1 Multipartite Assemblies. (A) The multipartite assembly of a promoter, a CDS and a terminator in an alpha vector, using *Bsa*I; (B) the multipartite assembly of a promoter, a CDS and a terminator in an omega vector, using *Btg*ZI. GB 2.0 Users Manual (http://gbcloning.upv.es/).

3.4 GoldenBraid Binary Assemblies

Transcriptional units (or in general GB superparts) assembled in level 1 vectors can be further combined into higher complexity structures through binary assemblies. For example, a first TU contained in an omega1 vector and a second TU contained in an omega2 vector can be cloned into an alpha1 destination vector through digestion with *Bsa*I and ligation (Fig. 14). The same can be done combining GB parts from an alpha1 and an alpha2 vector into an omega destination vector, using instead *Bsmb*I. In binary assemblies, only parts from the same type of vector can be combined, so GB parts from omega1 and omega2 vectors can be cloned into any alpha vector, and conversely GB parts from alpha1 and alpha2 vectors can be cloned into any omega vector. In the assembly, the overhangs left by the restriction enzymes are such that the GB part from the alpha1/omega1 vector is always placed at the left border, and the GB part from the alpha2/omega2 vector is placed at the right border of the insert. The indefinite reusability of GB constructs implies that the products of these binary assemblies can, in turn, be subject to yet another binary assembly reaction to be combined in alpha or omega vectors in a seamless fashion,

cycling between the two types of destination vectors and their respective restriction enzymes.

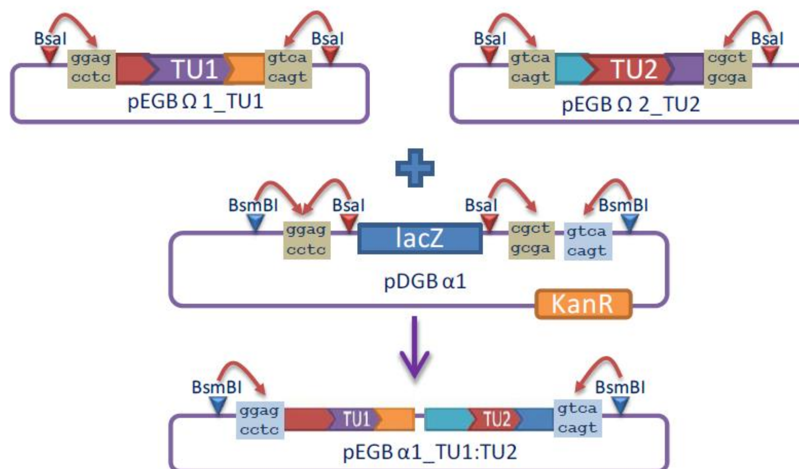


Figure 14. GB Level >1 Binary Assemblies. From level 1 alpha or omega vectors, TUs can be combined into more complex structure through binary assemblies. These new GB superparts can subsequently enter a new cloning reaction. GB 2.0 Users Manual (http://gbcloning.upv.es/).

3.5 The GoldenBraid CRISPR/Cas9 toolkit

Vazquez-Vilar *et al.* (2016) described the addition to the GB toolbox of a set of parts for the assembly of CRISPR constructs. These include different transcriptional units of the active *Cas9* gene, as well as various TUs encoding an inactivated version of Cas9 (dead Cas9, dCas9) which can act as a transcriptional modulator if associated to activator or repressor domains; they also include RNA PolIII promoters (U6 and U3) for the expression of gRNAs and scaffold RNAs.

A typical CRISPR construct contains three TUs: one constituted by the RNA PolIII promoter, the custom gRNA and the scaffold RNA; the second encoding the Cas9 endonuclease, typically under the control of the CaMV 35S promoter; and the third encoding a selection marker, usually the *NptII* gene, with the nopaline synthase promoter and terminator (Fig. 15). All DNA elements involved in the assembly are available as GB parts, with the exception of the 20 nt gRNA which, being specific for each individual construct, must be acquired by the user as a synthetic oligonucleotide.

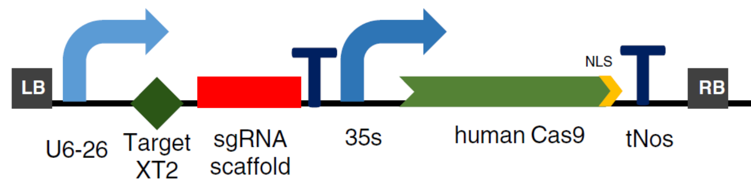


Figure 15. A typical CRISPR/Cas transcriptional unit. The gRNA (Target XT2) is placed under the control of a U6-26 PolIII promoter, and is followed by the conserved sgRNA scaffold; the Cas9 coding sequence is instead expressed with a 35S constitutive promoter, and uses the tNos terminator. In addition to these TUs, a selection marker is often added, usually the *NptII* gene for kanamycin resistance, to select individuals which have integrated the transgene. Vazquez-Vilar *et al.* (2016).

In addition, a polycistronic strategy was added to the GB CRISPR toolbox, based on the system described by Xie *et al.* (2015), which was shown to be successful in rice. Here, multiple gRNAs are stacked in a single transcript and then freed as individual gRNAs after processing of the transcript in the nucleus. Each gRNA is preceded by a tRNA: tRNAs are recognized and cleaved by RNase P and RNase Z, a system common to all Eukaryotes. These RNases recognize the tRNA structure regardless of adjacent sequences, making this a widely adaptable system. In addition, since tRNA is one of the most abundant cellular components, this processing system is thought to be constitutively active and to process a large number of substrates; tRNA genes also contain internal promoter elements which recruit RNA PolIII, meaning that their integration into CRISPR constructs might enhance transcription. The incorporation of this system into the GB cloning standard required the addition of a new assembly level, which was defined as level -1. The tRNAs and the scaffolds are carried by level -1 pVD1 vectors; when assembling tRNAs, gRNAs and scaffolds, these are first assembled into a pUPD2 vector (level 0). Up to 3 gRNAs can be assembled into a polycistronic transcript: each individual gRNA is assembled into a pUPD2 vector, and those are then assembled into a α vector (level 1) together with the PolIII promoter (typically the U6-26 Arabidopsis promoter; Fig. 16).

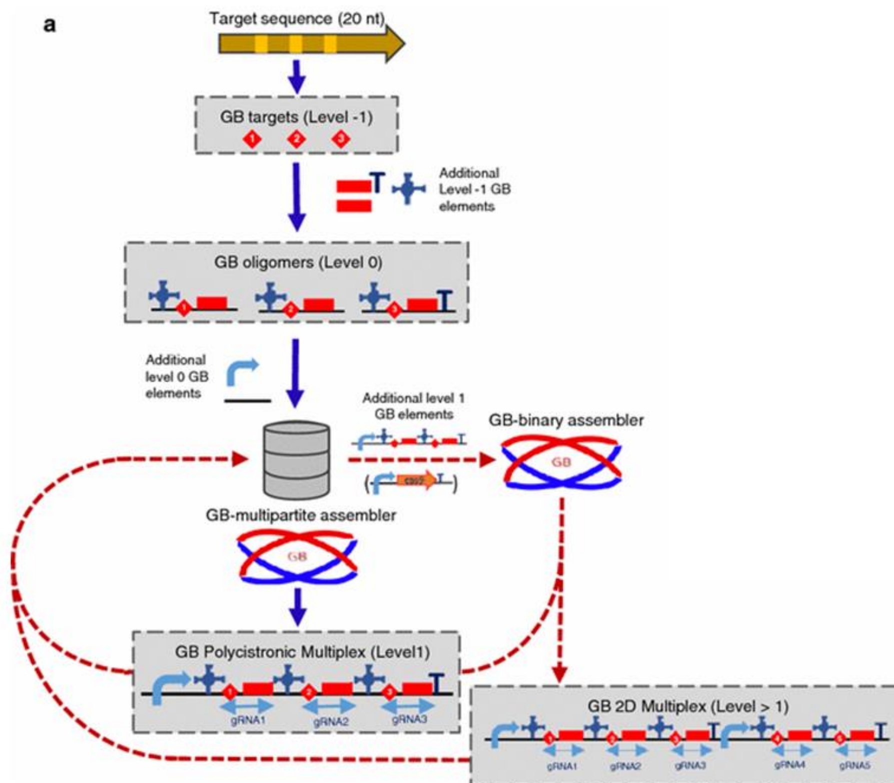


Figure 16. Assembly of multiplexing polycistronic CRISPR/Cas constructs. Individual 20 nt target sequences are integrated into level 0 pUPD vectors, together with a tRNA and a scaffold RNA, which are contained in level -1 pVD1 vectors (they are level -1 GB elements). Up to 3 gRNAs can be assembled in one polycistronic construct; the three pUPD vectors carrying the tRNA-gRNA-scaffold are then combined in an alpha vector with a single U6-26 promoter (bottom right figure, 2D multiplex). 2- or 3-gRNA TUs can be assembled. Vazquez-Vilar *et al.* (2016).

3.6 Software tools for GoldenBraid cloning

A number of software tools assisting in the design and *in silico* assembly of GB constructs have been developed and can be found at https://gbcloning.upv.es/ (Vazquez-Vilar *et al.*, 2017). Along with tools to design domestication protocols and multipartite or binary assemblies, the site acts as a catalog for GB parts. It is possible to browse the GB collection and find a description and a datasheet with standardized information for each DNA element present in the collection, from basic parts to complex multigenic constructs. Datasheets collect details about vectors, functions, categories, enzymes to be used for cloning and basic parts to make up superparts. Information is also available regarding the experiments in which these elements have been tested, enhancing

standardization and information sharing with other users and laboratories. Protocols and tutorials are also given to users to facilitate the use of GB.

Examples of setting up and applications of GB constructs can be found in Sarrion-Perdigones *et al.* (2011), Sarrion-Perdigones *et al.* (2013), Vazquez-Vilar *et al.* (2016) and Vazquez-Vilar *et al.* (2017). They include assembly of multigenic constructs for plant stable and transient transformation, the assembly of gene editing constructs based on Cas9 and Cpf1 endonucleases, tools for protein-protein interaction analysis and for gene silencing.

Bibliography

Ahloowalia, B. S. and Maluszynski, M. (2001) ‘Induced mutations – A new paradigm in plant breeding’, Euphytica, 118, pp. 167–173.

Ainley, W. M. *et al.* (2013) ‘Trait stacking via targeted genome editing’, Plant Biotechnology Journal, 11, pp. 1126–1134. doi: 10.1111/pbi.12107.

Ali, Z. *et al.* (2016) ‘CRISPR / Cas9-Mediated Immunity to Geminiviruses : Differential Interference and Evasion’, Scientific Reports. Nature Publishing Group, (May), pp. 1–11. doi: 10.1038/srep26912.

Andersson, M. *et al.* (2017) ‘Efficient targeted multiallelic mutagenesis in tetraploid potato (Solanum tuberosum) by transient CRISPR-Cas9 expression in protoplasts’, Plant Cell Reports, 36, pp. 117–128. doi: 10.1007/s00299-016-2062-3.

Arkin, A. (2008) ‘Setting the standard in synthetic biology’, Nature Biotechnology, 26(7), pp. 771–774. doi: 10.1038/nbt0708-771.

Baltes, N. J. *et al.* (2014) ‘DNA Replicons for Plant Genome Engineering’, The Plant Cell, 26(1), pp. 151–163. doi: 10.1105/tpc.113.119792.

Barrangou, R. *et al.* (2007) ‘CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes’, Science, 315(March), pp. 1709–1713.

Belhaj, K. *et al.* (2013) ‘Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system’, Plant Methods. Plant Methods, 9(1), p. 39. doi: 10.1186/1746-4811-9-39.

Benner, S. A. and Sismour, A. M. (2005) ‘Synthetic biology’, Nature Reviews Genetics, 6(7), pp. 533–543. doi: 10.1038/nrg1637.

Bortesi, L. *et al.* (2016) ‘Patterns of CRISPR/Cas9 activity in plants, animals and microbes’, Plant Biotechnology Journal, 14(12), pp. 2203–2216. doi: 10.1111/pbi.12634.

Brooks, C. *et al.* (2014) ‘Efficient Gene Editing in Tomato in the First Generation Using the Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-Associated9 System’, Plant Physiology, 166(3), pp. 1292–1297. doi: 10.1104/pp.114.247577.

Brouns, S. J. J. *et al.* (2008) ‘Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes Stan’, Science, 321, pp. 960–965.

Butler, N. M. *et al.* (2015) ‘Generation and Inheritance of Targeted Mutations in Potato (Solanum tuberosum L .) Using the CRISPR / Cas System’, PLoS ONE, pp. 1–12. doi: 10.1371/journal.pone.0144591.

Butler, N. M. *et al.* (2016) ‘Geminivirus-Mediated Genome Editing in Potato (Solanum tuberosum L .) Using Sequence-Specific Nucleases’, Frontiers in Plant Science, 7(July), pp. 1–13. doi: 10.3389/fpls.2016.01045.

Cameron, D. E., Bashor, C. J. and Collins, J. J. (2014) ‘A brief history of synthetic biology’, Nature Reviews Microbiology. Nature Publishing Group, 12(5), pp. 381–390. doi: 10.1038/nrmicro3239.

Ceasar, S. A. *et al.* (2016) ‘Insert, remove or replace : A highly advanced genome editing system using CRISPR / Cas9’, Biochimica et Biophysica Acta. Elsevier B.V., 1863, pp. 2333–2344. doi: 10.1016/j.bbamcr.2016.06.009.

Čermák, T. *et al.* (2015) ‘High-frequency, precise modification of the tomato genome’, Genome Biology, 16(1), p. 232. doi: 10.1186/s13059-015-0796-9.

Chandrasegaran, S. and Carroll, D. (2016) ‘Origins of Programmable Nucleases for Genome Engineering’, J Mol Biol, (428), pp. 963–989.

The Potato Genome Sequencing Consortium (2011) ‘Genome sequence and analysis of the tuber crop potato’, Nature, 475, pp. 189–195. doi: 10.1038/nature10158.

Curtin, S. J. *et al*. (2011) ‘Targeted Mutagenesis of Duplicated Genes in Soybean with Zinc-Finger Nucleases’, Plant Physiology, 156(June), pp. 466–473. doi: 10.1104/pp.111.172981.

Custers, R. (2017) ‘The regulatory status of gene-edited agricultural products in the EU and beyond’, Emerging Topics in Life Sciences, DOI: 10.10. doi: 10.1042/ETLS20170019.

Didovyk, A. *et al.* (2016) ‘Transcriptional regulation with CRISPR-Cas9 : principles , advances , and applications’, Current Opinion in Biotechnology. Elsevier Ltd, 40, pp. 177–184. doi: 10.1016/j.copbio.2016.06.003.

Doench, J. G. *et al.* (2014) ‘Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation’, Nature Biotechnology. Nature Publishing Group, 32(12), pp. 1262–1267. doi: 10.1038/nbt.3026.

Doudna, J. A. and Charpentier, E. (2014) ‘The new frontier of genome engineering with CRISPR-Cas9’, Science, 346(6213), pp. 1258096–1258096. doi: 10.1126/science.1258096.

Engler, C., Kandzia, R. and Marillonnet, S. (2008) ‘A One Pot , One Step , Precision Cloning Method with High Throughput Capability’, pp. 1–9. Available at: http://rsif.royalsocietypublishing.org/cgi/doi/10.1098/rsif.2008.0505.focus.

Engler, C. *et al.* (2009) ‘Golden gate shuffling: A one-pot DNA shuffling method based on type ils restriction enzymes’, PLoS ONE, 4(5). doi: 10.1371/journal.pone.0005553.

Espinoza, C. *et al.* (2013) ‘Cisgenesis and Intragenesis : New tools For Improving Crops’, Biol Res, 46, pp. 323–331.

EU (2001) ‘Directive 2001/18/EC of the European Parliament and of the Council of 12 March 2001 on the deliberate release into the environment of genetically modified organisms and repealing Council Directive 90/220/EEC’, Official Journal of the European Communities, L 106(1), pp. 1–38. doi: 2004R0726 - v.7 of 05.06.2013.

Gao, J. *et al*. (2015) ‘CRISPR / Cas9 ‑ mediated targeted mutagenesis in Nicotiana tabacum’, Plant Molecular Biology, 87, pp. 99–110. doi: 10.1007/s11103-014-0263-0.

Gelvin, S. B. (2003) ‘Agrobacterium -Mediated Plant Transformation : the Biology behind the “ Gene-Jockeying ” Tool’, Microbiology and Molecular Biology Reviews, 67(1), pp. 16–37. doi: 10.1128/MMBR.67.1.16.

Hanin, M. and Paszkowski, J. (2003) ‘Plant genome modification by homologous recombination’, Current Opinion in Plant Biology, 6, pp. 157–162. doi: 10.1016/S1369-5266(03)00016-5.

Hermans, P. W. M. *et al.* (1991) ‘Insertion Element IS987 from Mycobacterium bovis BCG Is Located in a Hot-Spot Integration Region for Insertion Elements in Mycobacterium tuberculosis Complex Strains’, Infect Immun, 59(8), pp. 2695–2705.

Hsu, P. D. *et al.* (2013) ‘DNA targeting specificity of RNA-guided Cas9 nucleases’, Nature Biotechnology. Nature Publishing Group, 31(9), pp. 827–832. doi: 10.1038/nbt.2647.

Huang, S. *et al.* (2016) ‘A proposed regulatory framework for genome-edited crops’, Nature Genetics. Nature Publishing Group, 48(2), pp. 109–111. doi: 10.1038/ng.3484.

Hunter, P. (2014) ‘“ Genetically Modified Lite ” placates public but not activists’, EMBO reports, 15(2), pp. 138–141.

Ishino, Y. *et al.* (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, pp. 5429–5433.

Ito, Y. *et al*. (2015) 'CRISPR/Cas9-mediated mutagenesis of the RIN locus that regulates tomato fruit ripening', Biochemical and Biophysical Research Communications. Elsevier Ltd, 467(1), pp. 76–82. doi: 10.1016/j.bbrc.2015.09.117.

Jacobs, T. B. *et al*. (2015) 'Targeted genome modifications in soybean with CRISPR/Cas9', BMC Biotechnology, pp. 1–10. doi: 10.1186/s12896-015-0131-2.

Jacobs, T. B. *et al.* (2017) 'Generation of a collection of mutant tomato lines using pooled CRISPR libraries', Plant Physiology, 174(4), pp. 2023–2037. doi: 10.1104/pp.17.00489.

Jinek, M. *et al*. (2012) 'A Programmable Dual-RNA – Guided', Science, 337, pp. 816–822.

Jinek, M. et al. (2014) 'Structures of Cas9 Endonucleases Reveal DNA-Mediated Conformational Activation', Science, Vol. 343 (6176), 1247997

Joung, J. K. and Sander, J. D. (2012) 'TALENs: a widely applicable technology for targeted genome editing', Nature Reviews Molecular Cell Biology. Nature Publishing Group, 14(1), pp. 49–55. doi: 10.1038/nrm3486.

Kim, H. and Kim, J. (2014) 'A guide to genome engineering with programmable nucleases', Nature Reviews Genetics. Nature Publishing Group, 15(5), pp. 321–334. doi: 10.1038/nrg3686.

Klap, C. *et al.* (2017) 'Tomato facultative parthenocarpy results from Sl AGAMOUS-LIKE 6 loss of function', Plant Biotechnology Journal, 15, pp. 634–647. doi: 10.1111/pbi.12662.

Kleinstiver, B. P. *et al.* (2016) 'Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells', Nature Biotechnology, 34, p. 869. Available at: http://dx.doi.org/10.1038/nbt.3620.

Klug, A. (2010) 'The Discovery of Zinc Fingers and Their Applications in Gene Regulation and Genome Manipulation', Annual Review of Biochemistry, 79, pp. 213–231. doi: 10.1146/annurev-biochem-010909-095056.

Li, Z. *et al.* (2015) 'Cas9-Guide RNA Directed Genome Editing in Soybean', Plant Physiology, 169(2), pp. 960–970. doi: 10.1104/pp.15.00783.

Liang, G. *et al.* (2016) 'Selection of highly efficient sgRNAs for CRISPR/Cas9-based plant genome editing', Scientific Reports. Nature Publishing Group, 6(1), p. 21451. doi: 10.1038/srep21451.

Lillestøl, R. K. *et al.* (2006) 'A putative viral defence mechanism in archaeal cells', Archaea, 2, pp. 59–72.

Liu, W. and Stewart, C. N. (2015) 'Plant synthetic biology', Trends in Plant Science. Elsevier Ltd, 20(5), pp. 309–317. doi: 10.1016/j.tplants.2015.02.004.

Makarova, K. S. *et al.* (2011) 'Evolution and classification of the CRISPR–Cas systems', Nature Reviews Microbiology. Nature Publishing Group, 9(6), pp. 467–477. doi: 10.1038/nrmicro2577.

Makarova, K. S. *et al.* (2015) 'An updated evolutionary classification of CRISPR–Cas systems', Nat Rev Microbiol, 13(11), pp. 722–736. doi: 10.1038/nrmicro3569.An.

Marraffini, L. A. (2015) 'CRISPR-Cas immunity in prokaryotes', Nature, 526, pp. 55–61. doi: 10.1038/nature15386.

Marraffini, L. A. and Sontheimer, E. J. (2008) 'CRISPR Interference Limits Horizontal Targeting DNA', Science, 322, pp. 1843–1846.

Miao, J. *et al.* (2013) 'Targeted mutagenesis in rice using CRISPR-Cas system', Cell Research, 23(10), pp. 1233–1236. doi: 10.1038/cr.2013.123.

Mikami, M., Toki, S. and Endo, M. (2015) ‘Comparison of CRISPR/Cas9 expression constructs for efficient targeted mutagenesis in rice’, Plant Molecular Biology. Springer Netherlands, 88(6), pp. 561–572. doi: 10.1007/s11103-015-0342-x.

Mojica, F. J. M., Juez, G. and Rodriguez-Valera, F. (1993) ‘Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites’, Molecular Microbiology. Blackwell Publishing Ltd, 9(3), pp. 613–621. doi: 10.1111/j.1365-2958.1993.tb01721.x.

Mojica, F. J. M. *et al.* (2000) ‘Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria’, Mol Microbiol, 36, pp. 244–246.

Mojica, F. J. M. *et al.* (2005) ‘Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements’, Journal of Molecular Evolution, 60(2), pp. 174–182. doi: 10.1007/s00239-004-0046-3.

Mojica, F. J. M. *et al.* (2009) ‘Short motif sequences determine the targets of the prokaryotic CRISPR defence system’, Microbiology, 155, pp. 733–740. doi: 10.1099/mic.0.023960-0.

Mojica, F. J. M. and Rodriguez-Valera, F. (2016) ‘The discovery of CRISPR in archaea and bacteria’, The FEBS Journal, 283, pp. 3162–3169. doi: 10.1111/febs.13766.

Nakata, A., Amemura, M. and Makino, K. (1989) ‘Unusual Nucleotide Arrangement with Repeated Sequences in the Escherichia coli K-12 Chromosome’, J Bacteriol, 171(6), pp. 3553–3556.

Nekrasov, V. *et al.* (2017) ‘Rapid generation of a transgene- free powdery mildew resistant tomato by genome deletion’, Scientific Reports. Springer US, (February), pp. 1–6. doi: 10.1038/s41598-017-00578-x.

Pan, C. *et al.* (2016) ‘CRISPR/Cas9-mediated efficient and heritable targeted mutagenesis in tomato plants in the first and later generations’, Scientific Reports. Nature Publishing Group, 6(1), p. 24765. doi: 10.1038/srep24765.

Peer, R. *et al*. (2015) ‘Targeted mutagenesis using zinc-finger nucleases in perennial fruit trees’, Planta, 241(4), pp. 941–951. doi: 10.1007/s00425-014-2224-x.

Pennisi, E. (2016). The plant engineer. Science, Vol. 353:6305, 1220-1224, DOI: 10.1126/science.353.6305.1220

Petolino, J. F. (2015) ‘Genome editing in plants via designed zinc finger nucleases’, In Vitro Cell.Dev.Biol.-Plant, 51, pp. 1–8. doi: 10.1007/s11627-015-9663-3.

Sander, JD and Joung, JK (2014) 'CRISPR-Cas systems for editing, regulating and targeting genomes' Nature Biotechnology, 32(4), pp. 347-355.

Sarrion-Perdigones, A. *et al.* (2011) ‘GoldenBraid: An iterative cloning system for standardized assembly of reusable genetic modules’, PLoS ONE, 6(7), pp. 1–11. doi: 10.1371/journal.pone.0021622.

Sarrion-Perdigones, A. *et al.* (2013) ‘GoldenBraid 2.0: A Comprehensive DNA Assembly Framework for Plant Synthetic Biology’, Plant Physiology, 162(3), pp. 1618–1631. doi: 10.1104/pp.113.217661.

Schouten, H. J., Krens, F. A. and Jacobsen, E. (2006) ‘Cisgenic plants are similar to traditionally International regulations for genetically modified organisms should be altered to exempt cisgenesis’, EMBO reports, 7(8), pp. 750–753.

Shan-e-Ali Zaidi, S., Mahfouz, M. M. and Mansoor, S. (2017) ‘CRISPR-Cpf1: A New Tool for Plant Genome Editing’, Trends in Plant Science, 22(7), pp. 550–553. doi: 10.1016/j.tplants.2017.05.001.

Shukla, V. K. *et al.* (2009) 'Precise genome modification in the crop species Zea mays using zinc-finger nucleases', Nature. Nature Publishing Group, 459(7245), pp. 437–441. doi: 10.1038/nature07992.

Sorek, R., Lawrence, C. M. and Wiedenheft, B. (2013) 'CRISPR-Mediated Adaptive Immune Systems in Bacteria and Archaea', Annual Review of Biochemistry, 82(1), pp. 237–266. doi: 10.1146/annurev-biochem-072911-172315.

Sun, N. and Zhao, H. (2013) 'Transcription Activator-Like Effector Nucleases (TALENs): A Highly Efficient and Versatile Tool for Genome Editing', Biotechnology and Bioengineering, 110(7), pp. 1811–1821. doi: 10.1002/bit.24890.

Sun, Y. *et al.* (2016) 'Engineering Herbicide-Resistant Rice Plants through CRISPR/Cas9-Mediated Homologous Recombination of Acetolactate Synthase', Molecular Plant, 9(4), pp. 628–631. doi: 10.1016/j.molp.2016.01.001.

Svitashev, S. *et al.* (2015) 'Targeted Mutagenesis, Precise Gene Editing, and Site-Specific Gene Insertion in Maize Using Cas9 and Guide RNA', Plant Physiology, 169(2), pp. 931–945. doi: 10.1104/pp.15.00793.

Svitashev, S. *et al.* (2016) 'Genome editing in maize directed by CRISPR–Cas9 ribonucleoprotein complexes', Nature Communications. Nature Publishing Group, 7, p. 13274. doi: 10.1038/ncomms13274.

de Toledo Thomazella, D. P. *et al.* (2016) 'CRISPR-Cas9 mediated mutagenesis of a DMR6 ortholog in tomato confers broad-spectrum disease resistance', bioRxiv, (July), p. 64824. doi: 10.1101/064824.

Ueta, R. *et al.* (2017) 'Rapid breeding of parthenocarpic tomato plants using CRISPR / Cas9', Scientific Reports. Springer US, 7(507), pp. 1–8. doi: 10.1038/s41598-017-00501-4.

Urnov, F. D. *et al.* (2010) 'Genome editing with engineered zinc finger nucleases', Nature Reviews Genetics. Nature Publishing Group, 11(9), pp. 636–646. doi: 10.1038/nrg2842.

Vazquez-Vilar, M., *et al.* (2016) 'A modular toolbox for gRNA–Cas9 genome engineering in plants based on the GoldenBraid standard', Plant Methods. BioMed Central, 12(1), p. 10. doi: 10.1186/s13007-016-0101-2.

Vazquez-Vilar, M. *et al.* (2017) 'GB3.0: a platform for plant bio-design that connects functional DNA elements with associated biological data', Nucleic acids research, 45(4), pp. 2196–2209. doi: 10.1093/nar/gkw1326.

Wang, L. *et al.* (2017) 'Reduced Drought Tolerance by CRISPR/Cas9-Mediated SlMAPK3 Mutagenesis in Tomato Plants', Journal of Agricultural and Food Chemistry, 65, pp. 8674–8682. doi: 10.1021/acs.jafc.7b02745.

Wang, S. *et al.* (2015) 'Efficient targeted mutagenesis in potato by the CRISPR / Cas9 system', Plant Cell Reports. Springer Berlin Heidelberg, 34(9), pp. 1473–1476. doi: 10.1007/s00299-015-1816-7.

Watanabe, T. *et al.* (2012) 'hemimetabolous insect using zinc-finger and', Nature Communications. Nature Publishing Group, 3, pp. 1017–1018. doi: 10.1038/ncomms2020.

Weber, E. *et al*. (2011) 'A Modular Cloning System for Standardized Assembly of Multigene Constructs', PLoS ONE, 6(2), pp. 1–10. Available at: https://doi.org/10.1371/journal.pone.0016765.

Weeks, D. P., Spalding, M. H. and Yang, B. (2016) 'Use of designer nucleases for targeted gene and genome editing in plants', Plant Biotechnology Journal, 14, pp. 483–495. doi: 10.1111/pbi.12448.

Westra, E. R., Buckling, A. and Fineran, P. C. (2014) ‘CRISPR-Cas systems: beyond adaptive immunity’, Nature Reviews Microbiology. Nature Publishing Group, 12(5), pp. 317–326. doi: 10.1038/nrmicro3241.

Wolt, J. D., Wang, K. and Yang, B. (2016) ‘The Regulatory Status of Genome-edited Crops’, Plant Biotechnology Journal, 4(October 2015), pp. 510–518. doi: 10.1111/pbi.12444.

Wright, D. A. *et al.* (2005) ‘High-frequency homologous recombination in plants mediated by zinc-finger nucleases’, The Plant Journal, 44, pp. 693–705. doi: 10.1111/j.1365-313X.2005.02551.x.

Xie, K., Minkenberg, B. and Yang, Y. (2015) ‘Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system’, Proceedings of the National Academy of Sciences, 112(11), pp. 3570–3575. doi: 10.1073/pnas.1420294112.

Zaidi, S. S. *et al.* (2016) ‘Engineering Plant Immunity : Using CRISPR / Cas9 to Generate Virus Resistance’, Frontiers in Plant Science, 7(November), pp. 1–10. doi: 10.3389/fpls.2016.01673.

Zetsche, B. *et al.* (2015) ‘Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System’, Cell. Elsevier Inc., 163(3), pp. 759–771. doi: 10.1016/j.cell.2015.09.038.

Zhang, B. *et al.* (2016) ‘Exploiting the CRISPR / Cas9 System for Targeted Genome Mutagenesis in Petunia’, Scientific Reports. Nature Publishing Group, 6(February), pp. 1–8. doi: 10.1038/srep20315.

Objectives

The main objective of the present work was to set up an efficient GoldenBraid-based CRISPR editing system first in tomato, as a model species, and then to apply it to eggplant. Results have been organized in the three following Chapters.

1) **Chapter I: 'Generation of a tomato chlorophyll retaining *gf1* mutant using the CRISPR/Cas9 GoldenBraid toolkit'.** This chapter includes results about the setting up of an efficient GoldenBraid-based CRISPR system to edit the single endogenous green flesh (*gf1*) gene in tomato and the evaluation of the resulting editing pattern.

2) **Chapter II: 'Engineering of tomato resistance to tomato yellow leaf curl virus (TYLCV) through a double CRISPR/Cas9 approach'.** The system set up in the first chapter was used to provide tomato with a CRISPR-based resistance against TYLCV and to assess its effects through both transient and stable transformation.

3) **Chapter III: 'A CRISPR/Cas9 editing protocol for the polyphenol oxidase (*ppo*) gene family in eggplant'.** Gene editing was reported for the first time in eggplant, by developing a CRISPR editing and *in vitro* regeneration protocol, to target the polyphenol oxidase (*ppo)* gene family.

Chapter I

Generation of a tomato chlorophyll retaining *gf1* mutant using the CRISPR/Cas9 GoldenBraid toolkit

1 - Introduction

Chlorophyll metabolism has been the subject of intensive investigation efforts over the past thirty years, and is now known into considerable detail (Tanaka and Tanaka, 2006; Masuda and Fujita, 2008; Barry, 2009; Hörtensteiner and Kräutler, 2011; Hörtensteiner, 2013; Christ and Hörtensteiner, 2014; Kuai *et al.*, 2017). Chlorophyll synthesis was the first to be elucidated, while its catabolism long remained less well understood. Chlorophyll degradation occurs mainly upon conversion of chloroplasts into gerontoplasts during leaf senescence, and into chromoplasts in the course of fleshy fruit ripening (Hörtensteiner, 2006); however, a certain rate of chlorophyll degradation and turnover is constitutive and occurs also in response to various stresses and during seed maturation. During senescence, chlorophyll degradation acts as part of a developmentally regulated process that recycles nutrients from senescent organs to parts of the plant undergoing active growth. In particular, chlorophyll degradation appears to be a prerequisite for the mobilization of the proteins contained in light harvesting complexes (LHCs), which represent the second most abundant nitrogen pool in senescent leaves, after Rubisco. Chlorophyll degradation mutants, often named staygreens, offer insight into its catabolism and regulation, and may also represent interesting visual and functional phenotypes.

1.1 Chlorophyll catabolism

Very recently, a comprehensive model for a chlorophyll degradation pathway common to higher plants, known as the pheophorbide *a* oxidase (PAO)/phyllobilin pathway, was unveiled (Kuai *et al.*, 2017). The term phyllobilins collectively refers to a class of over 40 metabolites which result from the breakdown of chlorophyll *a* (chl *a*) and share the chlorophyll *a*-derived linear tetrapyrrole, while displaying different side chain modifications and specific stereoisomeric properties. All phyllobilins are non fluorescent compounds, and are further divided into: type I phyllobilins (1-formyl, 19-oxobilins), and type II phyllobilins (1,19-dioxobilins). Different classes of phyllobilins (overall more than 40 molecules) have been identified in different species of Angiosperms, but have not yet been characterized in other plant species, raising the question of how recently did these mechanisms evolve. In Angiosperms, however, there seems to be evidence that the same pathway is active in all instances of chlorophyll degradation, from senescence to stress response and fruit ripening (Pruzinska *et al.*, 2007; Hauenstein *et al.*, 2016). The PAO/phyllobilin pathway is composed of two main phases: the first occurs at the thylakoid membranes of the chloroplast, while during the second phase chlorophyll catabolites are transported to the vacuole.

The conversion of chlorophyll *b* (chl *b*) into chlorophyll a (chl *a*) is the first step of the pathway and occurs through two reduction steps catalyzed by the enzymes chlorophyll *b* reductase (CBR) and chlorophyll *a* reductase (CAR), with the intermediate formation of 7-hydroxymethyl chlorophyll *a* (Fig. 17). This step is considered necessary as chl *a* and its products (but not their correspondent *b* pigments) are the substrates for two downstream chlorophyll catabolic enzymes, namely Mg-dechelatase and pheophorbide *a* oxidase (PAO). Two modifications need to be performed on chl *a* in order to obtain pheophorbide *a*, the substrate for PAO: removal of the Mg atom at the centre of the chlorophyll ring, and removal of the phytol chain (Fig. 18). The order in which these two steps occur was the subject of debate, but the identification of a pheophytinase (PPH) which specifically hydrolizes pheophytin but is unable to use chlorophyll as a substrate sustained the idea that Mg removal must occur first. The mechanism by which the Mg ion is lost from the chlorophyll ring remained unclear for many years, and different models were proposed, including non-enzymatic loss due to acidification of plastidial pH, the presence of chelators and specific enzymatic activities performed by Mg-dechelatases. It was only very recently (Shimoda *et al.*, 2016) that a candidate for Mg-dechelatase was identified as the staygreen (SGR) protein, thus answering at the same time to both the mystery of Mg-dechelatase identity and to that of the elusive SGR function. Finally, pheophorbide *a* is converted to primary fluorescent chlorophyll catabolite (pFCC) and then to red chlorophyll catabolite (RCC) by PAO, a Rieske-type monoxigenase located in the thylakoid membrane, and the red chlorophyll catabolite reductase (RCCR), located in the stroma (Kuai *et al.*, 2017). RCCR acts in a highly sterospecific manner depending on the species, resulting in the formation of either *p*FCC or *epi-p*FCC. Mutants for *pao* or *rccr* are known as *accelerated cell death* (*acd1* and *acd2*) in Arabidopsis, because of cellular toxicity resulting from accumulation of pheophorbide *a* and RCC, which are highly phototoxic (Hörtensteiner, 2009). During the second phase of chlorophyll degradation, its catabolites are further modified and transported from the chloroplast to the vacuole, where the acidic environment drives their tautomerization to non-fluorescent phyllobilins.

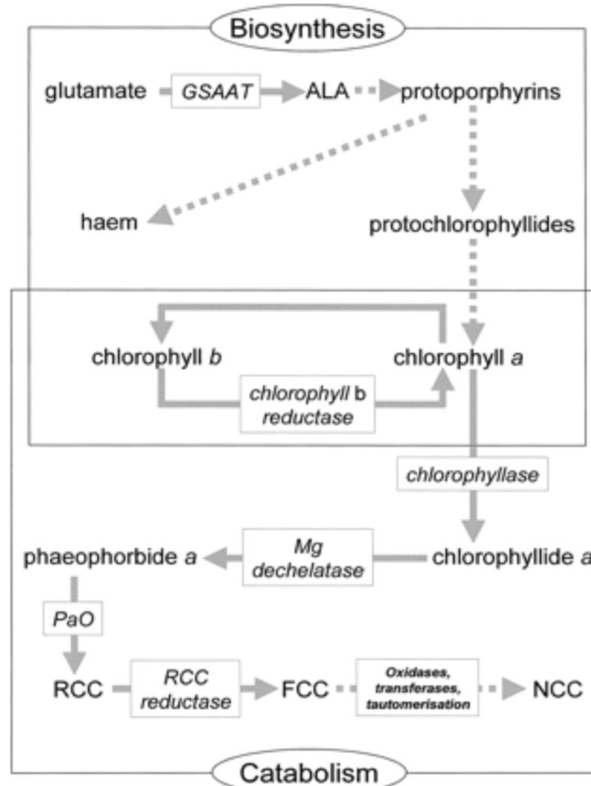


Figure 17. General chlorophyll biosynthesis and degradation pathway. GSAAT is the glutamic semialdheyde aminotransferase; ALA, 2-aminolaevulinic acid; PAO is pheophorbide a oxidase; RCC is the red chlorophyll catabolite, and FCC and NCC are the fluorescent and non-fluorescent chlorophyll catabolites. The first steps of chlorophyll catabolism occur in the chloroplast, while the final degradation of phyllobilins takes place in the vacuole. Thomas *et al.*, 2002.

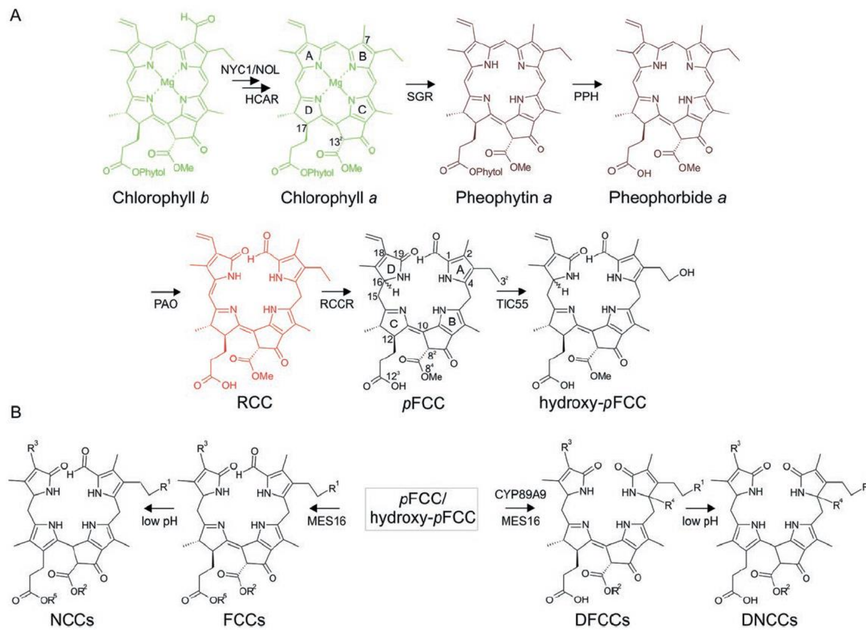


Figure 18. The PAO/phyllobilin pathway. The first part of the pathway is catalyzed by a series of enzymes in the chloroplast, while the second part occurs in the vacuole and is non-enzymatic. Kuai *et al*., 2017.

1.2 Staygreen proteins in plants

sgr was identified as the I *locus* corresponding to the green cotyledon pea phenotype used by Mendel in his study of the laws of genetic inheritance (Armstead *et al.*, 2007). SGR is a small protein with a highly conserved N-terminal region of about 150 aminoacids, which targets the protein to the chloroplast, another highly conserved central core and a cysteine-rich motif in the otherwise variable C-terminal region (Hörtensteiner, 2009). However, since the protein has no annotated functional domains, prediction of its role in the cell had always been difficult (Barry, 2009). SGR appears to be highly phylogenetically conserved across plant species and photosynthetic algae. A phylogenetic analysis, performed by Rong *et al.* (2013), showed two distinct clades in higher plants, separating SGR from SGR-like proteins in both Mono- and Dicotyledons (Fig. 19). In addition, some species (like *Arabidopsis thaliana*, *Glycine max*, *Vitis vinifera*, *Populus trichocarpa* and *Zea mays*) possess two *sgr* genes, while others (like *Solanum lycopersicum*, *Capsicum annuum*, *Oryza sativa* and *Sorghum bicolor*) have only one. Also, not all species possess *sgr-like* genes: notably, in tomato a *sgr-like* homolog can be identified on chromosome 4 by aligning the *sgr-like* sequence from

Arabidopsis, but so far no functional confirmation has been obtained for its function; interestingly, SGR-like proteins have also been found in non-photosynthetic organisms, like *Clostridium* and *Bacillus* spp..

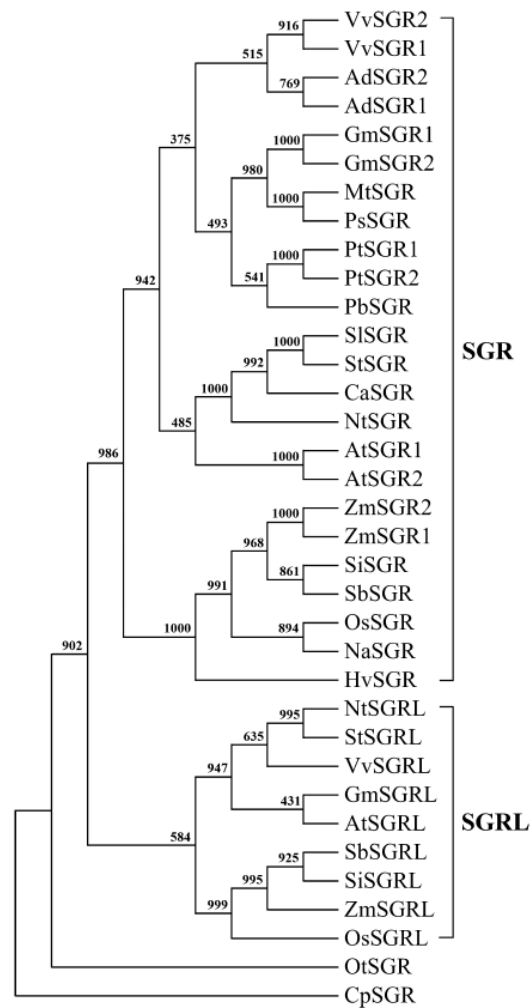


Figure 19. Phylogenetic tree of SGR and SGR-like proteins from Mono- and Dicotyledons. Two clades are clearly visible and SGR-like proteins are considerably fewer than SGR. Vv = *Vitis vinifera*, Ad = *Actinidia deliciosa*, Gm = *Glycine max*, Mt = *Medicago truncatula*, Ps = *Pisum sativum*, Pt = *Populus trichocarpa*, Pb = *Pyrus* x *bretschneideri*, Sl = *Solanum lycoperscium*, St = *Solanum tuberosum*, Ca = *Capsicum annuum*, Nt = *Nicotiana tabacum*, At = *Arabidopsis thaliana*, Zm = *Zea mays*, Si = *Setaria italica*, Sb = *Sorghum bicolor*, Os = *Oryza sativa*, Na = *Neosinocalamus affinis*, Hv = *Hordeum vulgare*, Ot = *Ostreococcus tauri*, Cp = *Clostridium perfringens*. Rong *et al.*, 2013.

The characterization of SGR as the Mg-dechelatase responsible for the conversion of chl *a* to pheophytin *a* was carried out through both *in vitro* and *in*

vivo assays in which it was demonstrated that SGR was capable of converting chlorophyll *a* (but not chl *b* or chlorophyllide *a*, which is often used experimentally as a substitute for chlorophyll *a*) to pheophytin *a* without the need of any additional protein (Shimoda *et al.*, 2016). Interestingly, on the other hand, SGR-like proteins could successfully convert chlorophyllide *a* to chl *a*, but were unsuccessful in using chlorophyll *a* as a substrate; this result could point to a functional evolutionary specialization of these two protein clades. This could also be consistent with the fact that *sgr-like* genes appear to be up-regulated during greening, and chlorophyllide *a* has been proposed to take part in the formation of photosystems (Chidgey *et al.*, 2014), while *sgr* genes are induced during senescence. This still fails to explain, however, the function of *sgr-like* genes in non-photosynthetic organisms.

1.3 Staygreen phenotypes

The term staygreen (or stay-green) is used to refer to a class of phenotypes in which chlorophyll degradation is severely impaired or delayed. Such mutants have been known for a long time in different plant species (*Festuca pratensis*, *Capsicum annuum*, *Solanum lycopersicum*, *Oryza sativa*, *Zea mays*, *Pisum sativum* and Arabidopsis among others; Hörtensteiner, 2009). As mentioned above, one of the seven mutants used by Mendel in his studies on genetic inheritance did in fact correspond to a *sgr* mutant. Different classes of staygreen phenotypes, resulting from mutation of different *loci,* are known and a broad division is made between so-called "functional" and "cosmetic" staygreen phenotypes. Functional staygreens are those in which senescence and ripening are effectively halted or slowed down as a consequence of the disruption of chlorophyll degradation, while in cosmetic staygreens, despite the obvious phenotype, these processes proceed normally. In particular, staygreen phenotypes have been divided into five classes (A-E, Fig. 20) based on their physiological characteristics (Thomas and Howarth, 2000). Type A and type B staygreens are classified as functional: in type A senescence is considerably delayed, but proceeds normally once it starts; in type B senescence begins on schedule, but proceeds slowly compared to the wild type. On the other hand, types C-D-E are defined as cosmetic staygreens: in type C, chlorophyll is indefinitely retained because of failure in its degradation, but photosynthetic activity is lost and senescence proceeds normally; in type D, staygreen is associated to abrupt cell death, for example due to high temperature or drying, resulting in the sudden halt of cell metabolism; finally, in type E there is an excess of chlorophyll accumulation which does not result in increased photosynthetic activity. The different classes of staygreen phenotypes depend on the mutation of different genes involved in chlorophyll degradation. In this

respect, mutations of the expression of *sgr* resulted in type C cosmetic staygreen phenotypes.

Protein	Gene	Mutant phenotype	Function
Stay-green	*SGR* *NYE1* *SID* /	*sgr* = stay-green	Binding LHCII and catabolic enzymes, stabilising catabolic complex
Chlorophyll b reductase	*NYC*	*nyc* (rice and *Arabidopsis*) = stay-green	Ferredoxin/NADPH-dependent two-step conversion of chlorophyll *b* to chlorophyll *a*
	NOL	*nol* (rice, but not *Arabidopsis*) = stay-green	
	HCAR	*hcar* = cell death, not stay-green	
'Mg dechelatase'	Identity not yet resolved	?	Removal of Mg from the macrocycle (not known whether reaction is enzymic or chemical)
Phaeophytinase	*PPH* *CRN1* *NCY3*	*pph* = stay-green	Dephytylation of phaeophytin
Phaeophorbide *a* oxygenase	*PAO* *ACD1* *LLS1*	*acd1* = cell death, not stay-green	Ferredoxin-dependent oxidative opening of macrocycle to form RCC
RCC reductase	*RCCR* *ACD2*	*acd2* = cell death, not stay-green	Ferredoxin-dependent reduction of RCC to pFCC

Figure 20. Different types of mutation causing staygreen phenotypes. Known plastid-located components of the chlorophyll degradation pathway and consequence of their disruption for expression of the stay-green character Note that SGR had not yet been recognized as the Mg dechelatase. Thomas & Ougham (2014).

In general, the staygreen trait may have different impacts on plant productivity and growth depending on the kind of mutation and on the specific physiology of the plant. In the case of functional staygreens, where the mutation impacts photosynthetic rates and the progression of senescence, the mutants can display increased yields due to prolonged carbon export from leaves as they approach senescence. This has been shown to be important mainly for cereals, where mutants of maize and sorghum with increased retention of green leaf area exhibited better adaptation to environmental stress conditions (Thomas and Ougham, 2014). One important consequence of chlorophyll retention, independently of the functionality of the underlying mutation, is a reduction in the remobilization of nutrients (nitrogen and minerals) during plant development and leaf senescence.

Even in cosmetic staygreens, chlorophyll and light harvesting complexes, with all associated proteins, are retained in the chloroplast during senescence; this is clear by observing the presence of stacked thylakoids in the chloroplast of senescent leaves and ripe fruits. With the onset of senescence, the economy of a leaf transitions from a phase in which carbon is captured from the atmosphere, fixed and exported to non-photosynthesizing organs, to a phase in which nutrients (especially nitrogen from Rubisco, photosystems and light harvesting complexes) are recycled to organs undergoing active growth, including seeds and fruits. This means that, if the degradation of the photosynthetic machinery

is impaired, a decrease in export of nitrogen will occur, possibly impacting seed quality: this is especially true for legumes such as soybean and cowpea (Thomas and Ougham, 2014), but may also be significant for cereals, where grain protein synthesis can be reduced. On the other hand, a positive correlation was found by Zhou *et al.* (2011) between the *sgr* mutation in alfalfa and its quality, with a positive impact of staygreen genotypes both on the appearance of forage and on leaf protein content. For fleshy fruits such as tomato and pepper, however, a staygreen mutation can be commercially valuable mostly because of the introduction of a novel fruit phenotype. Indeed, one of the most dramatic changes occurring in fleshy fruit during ripening is color change, which is usually determined by the synthesis of carotenoids, anthocyanins and flavonoids, accompanied by chlorophyll breakdown: this is also reflected in the shift from chloroplasts to plastoglobules (Barry, 2009). In cosmetic staygreen mutants, both types of pigments are present, and chloroplasts coexist along with plastoglobules in ripe fruits, leading to fruit phenotypes which are known as brown, black, purple or "chocolate".

The *sgr* mutation can also affect other aspects of fruit quality, as well as agronomic traits. Since specific components of fruit aroma (deriving from aminoacid, fatty acid or carotenoid metabolism) are synthesized either in the chloroplast or in the chromoplast, the simultaneous presence of organelles sharing characteristics of both may alter the volatile profile in the fruit (Barry, 2009). Roca *et al.* (2006) analyzed carotenogenesis in *C. annuum* fruits, comparing the biosynthetic rate and the total content of carotenoids between *sgr* and wild type varieties, and found that the presence of chlorophylls seems to extend the carotenogenic process and to increase their total content; however, every variety displayed a characteristic accumulation curve. Ramirez and Tomes (1964) found that, in tomato *sgr* fruits, a steep increase in β-carotene/lycopene ratio could be observed, suggesting that this kind of mutation can also affect specific carotenoid composition. In addition, the presence of chlorophyll even at later stages of fruit development might imply a greater electron flow, which is not counterbalanced by increased photosynthetic activity and might result in the formation of reactive oxygen species (ROS). The increased carotenoid content results in greater antioxidant power and allows ROS detoxification. This was also observed by Luo *et al.* (2013), who showed that tomato SGR interacts with phytoene synthase (PSY1) intervening in the regulation of the synthesis of lycopene and β-carotene in the fruit, and that *sgr* mutants have increased lycopene and β-carotene accumulation, corresponding to reduced H_2O_2 capacity.

Finally, staygreen mutations could improve shelf life and resistance to both biotic and abiotic stresses (Rivero *et al.*, 2007). Solid evidence of an association between staygreen phenotypes and drought resistance was found in sorghum (Kassahun *et al.*, 2010; Jordan *et al.*, 2012), while in rice this trait is associated to submergence tolerance (Fukao *et al.*, 2012). However, even cosmetic *sgr* mutations were found to play a positive role in limiting disease symptoms upon *Pseudomonas syringae* pv *tomato* infection in an Arabidopsis model (Mecey *et al.*, 2011). The mutants did not show a reduction in the number of pathogen cells, but symptoms were greatly reduced; in fleshy fruits, this constitutes an advantage for shelf life, marketability and quality in general.

1.4 Tomato *sgr* mutants

Type C staygreen mutants related to mutations in the *sgr locus* were described in many plant species. Important work was done starting in the 1970s in the forage crop *Festuca pratensis* (Vicentini *et al.*, 1995; Hauck *et al.*, 1997; Moore *et al.*, 2005); other species in which the *sgr* trait was well characterized include rice (Park *et al.*, 2007; Yoo *et al.*, 2007; Rong *et al.*, 2013), tomato (Barry *et al.*, 2008; Hu *et al.*, 2011; Luo *et al.*, 2013) and pepper (Efrati *et al.*, 2005; Roca and Mínguez-Mosquera, 2006; Borovsky and Paran, 2008). Until the elucidation of its role as Mg-dechelatase, however, various characteristics of the protein were unveiled in all these organisms, but its precise function remained elusive. It was observed that *sgr* mutants retained chlorophyll in senescent leaves and ripening fruits, that thylakoids failed to be degraded in these tissues and that the SGR protein could bind a number of other proteins involved in photosynthesis and chlorophyll degradation (Cheung *et al.*, 1993; Akhtar *et al.*, 1999; Grassl *et al.*, 2012). *sgr* tomato mutants were described by Kerr as early as 1956, and are characterized both by chlorophyll retention in senescent leaves and by distinctive fruit color (Kerr, 1956). In *S. lycopersicum*, the green flesh (*gf1)/sgr locus* maps to a 2.55 kb region on the long arm of chromosome 8. The first mutation described for this *locus* was a single A → T substitution responsible for changing an invariant arginine residue at position 143 into a serine, thus affecting the central conserved core of the protein. Barry and Pandey (2009) analyzed a number of tomato heirloom varieties which displayed the staygreen phenotype and discovered more mutant alleles at the *gf1 locus*. These include additional single base substitutions, a single base insertion and deletions of different sizes (2 or 1,163 bp). Also, these mutations affect different parts of the protein (as represented in Fig. 21) but they all result in the same phenotype. So far, a total of 5 mutant alleles were described; these are reported in Table 1.

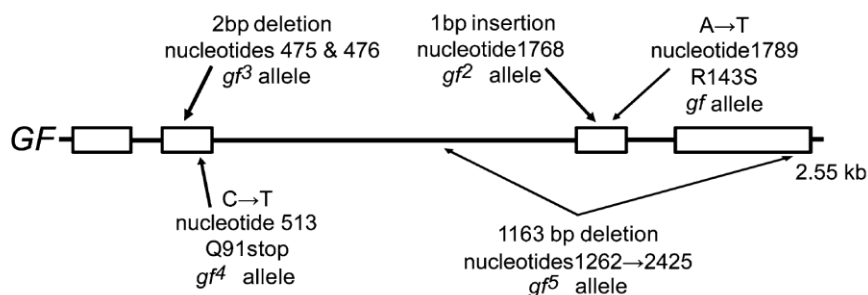


Figure 21. Gene structure and mutations of *gf1*. Barry and Pandey (2009).

Allele	Genotype	Position
gf	A → T	Nucleotide 1789 (3rd exon)
gf^2	A insertion	Nucleotide 1768 (3rd exon)
gf^3	2 bp deletion	Nucleotides 475-476 (2nd exon)
gf^4	C → T	Nucleotide 513 (2nd exon)
gf^5	1,163 bp deletion	Nucleotides 1,262-2,425 (3rd-4th exon)

Table 1. *gf* alleles Barry and Pandey (2009).

1.5 Experimental goals

Given the relevance and positive features of the *sgr* trait for overall fruit quality, we chose the tomato *gf1/sgr locus* as the first target to establish a GoldenBraid-based gene editing system in tomato. In addition to its biological relevance, it was also noteworthy that the *sgr* phenotype could be easily detected in both leaves and fruits, and that all known mutated alleles, regardless of the exact position of the mutation or its form, resulted in the same phenotype. We took advantage of the great regenerative potential of tomato *in vitro* tissue culture to obtain stable transformants expressing a CRISPR cassette targeting *gf1*. Once transformed plantlets were obtained, we first assessed whether gene editing had been successful by amplifying and sequencing the target *locus* through a Sanger approach. We also investigated potential off-target effects with an Illumina amplicon sequencing approach, by deep sequencing of putative off-target sites. A SGR phenotype was evident in both leaves and fruits. Edited, stable T_1

individuals were recovered, and the segregation of the transgene could be assessed in part of the progeny.

2 - Materials and methods

2.1 Vector design and construction

A single guide RNA was designed to target the coding sequence of the *gf1/sgr* gene (Solyc08g080090), using the design tools on Benchling (https://benchling.com). The choice of the gRNA was made based on a series of requirements:

1. that the gRNA would start with a G, a requirement associated with the use of the U6-26 Arabidopsis promoter
2. that it would be located in a structurally important region of the cds
3. that it had a satisfactory on-target activity and, likely, no off-target effects

On-target and off-target scores available in Benchling when searching for gRNAs in a given sequence are calculated based on the algorithms by Doench *et al.* (2016). These take into consideration the strength of the bond between gRNA and target, gRNA composition and the likelihood of off-target activity elsewhere in the genome. The reference genome for off-target evaluation is the *Solanum lycopersicum* str. Heinz 1706. Figure 22 shows the output of the Benchling gRNA design tool.

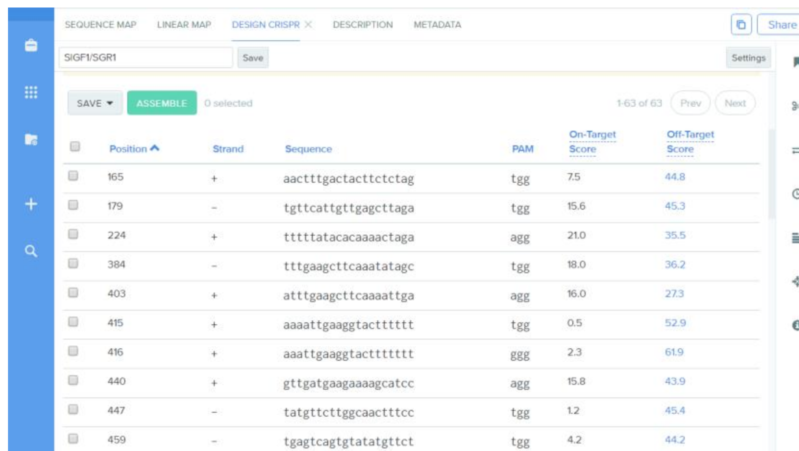


Figure 22. Benchling analysis of putative gRNAs. Guide RNA sequences are indicated, together with their position in the given sequence, and their on-target and off-target scores.

The selected sequence was domesticated for cloning into GoldenBraid pDGB3 alpha1 vectors by adding adapters for the insertion of the gRNA between the

U6-26 promoter and the scaffold RNA sequences in the assembly. This can be done using the GB CRISPR domesticator available at https://gbcloning.upv.es/ → Tools → GB CRISPR Tools, which, given the target sequence, provides the sequences of two complementary oligonucleotides comprising the gRNA sequence and the two overhangs which will allow their cloning (Fig. 23). Each strand of the gRNA-adapter sequence was synthesized as a oligonucleotide.

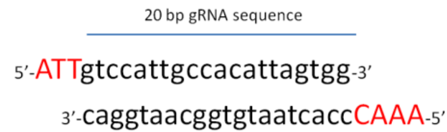


Figure 23. Design of two complementary oligonucleotides with appropriate overhangs for GB cloning into a level 1 CRISPR construct. In black is the 20 bp gRNA sequence, and in red the adapters.

Oligonucleotides were diluted to a final concentration of 1 µM; 5 µl of each were mixed and let anneal at room temperature for 30 minutes before setting up the multipartite restriction-ligation reaction. All other GB parts used in restriction-ligation reactions were diluted to a concentration of 75 ng μl^{-1}. Such reactions were performed as detailed in Supplementary Tables S1-S3 at the end of this Chapter, through three successive cloning steps which incorporated the gRNA cassette, the *hCas9* transcriptional unit (TU) and the *NptII* TU. The first assembly into a level 1 alpha vector is a multipartite assembly, while all the following steps are constituted of binary assemblies. Table 2 list the restriction-ligation conditions used in all steps of GoldenBraid cloning. *Bsa*I and *Bsmb*I enzymes, as well as BSA, were provided by ThermoScientific, while T4 Ligase was provided by Promega.

Step	Temperature	Duration	Cycles
Restriction	37°C	2 minutes	30
Ligation	16°C	5 minutes	
Hold	4°C	∞	1

Table 2. Restriction-ligation protocol for GoldenBraid cloning.

The product of each restriction-ligation reaction was transformed into *Escherichia coli* DH5α electrocompetent cells by electroporation. Cells were

resuspended in 350 µl SOC *medium* and let grow for 1 hour at 37°C with agitation, then plated on a LB-agar *medium* supplemented with 1 mM IPTG, 40 mg ml^{-1} X-Gal and the appropriate selection antibiotic (50 µg ml^{-1} kanamycin for alpha vectors, 50 µg ml^{-1} spectinomycin for omega vectors, 100 µg ml^{-1} ampicillin or 34 µg ml^{-1} chloramphenicol for pUPD and pUPD2, respectively) and incubated overnight at 37°C. White colonies were then selected and incubated overnight with agitation at 37°C in liquid LB *medium* supplemented with the appropriate antibiotic. The transformed plasmid was isolated from liquid culture using the E.Z.N.A.® Plasmid Mini Kit I (Omega bio-tek) following the manufacturer's instructions and used for subsequent restriction-ligation reactions. Plasmid identity was confirmed either by direct sequencing of the insert at the domestication step, or by restriction using two restriction enzymes per vector.

The final pDGB3 alpha2 construct Tnos:*NptII*:Pnos - U6-26:gRNA:scaffold - 35S:*hCas9*:Tnos was transformed into the rifampicin-resistant *Agrobacterium tumefaciens* LBA4404 electrocompetent cells by electroporation. Cells were then resuspended in 500 µl SOC *medium* and let grow for 2 hours at 28°C with agitation, then plated on LB-agar *medium* supplemented with 50 µg ml^{-1} rifampicin and the appropriate antibiotic as specified by vector resistance (here, 50 µg ml^{-1} kanamycin), and incubated for 3 days at 28°C. Colonies were grown in LB liquid *medium* supplemented with rifampicin and the antibiotic of choice (kanamycin), with agitation for 2 days at 28°C. The presence and identity of the plasmid were confirmed by restriction using at least two restriction enzymes.

2.2 Plant material

Tomato ‘MoneyMaker’ seeds were sterilized by washing for 30 minutes in a 2.5% sodium hypochlorite solution, then rinsed for three times in sterile water, with each rinse lasting for 5, 10 and >15 minutes, respectively. Sterile, clean seeds were transferred to a solid germination *medium* (2.5 g l^{-1} MS vitamins, 10 g l^{-1} sucrose and 10 g l^{-1} phytoagar, pH 5.8) in sterile cups and kept in the dark for 3 days before being exposed to light. Growth chamber conditions provided a 16 hours day light cycle (250 µmol photons m^{-2} s^{-1}), a relative humidity of 60-70% and 25°C. 10 days after sowing, cotyledons were used for *Agrobacterium*-mediated transformation. Cotyledon development is considered optimal for transformation when the first true leaves start to emerge. 50 seeds were used for the *gf1* CRISPR transformation.

2.3 *Agrobacterium*-mediated transformation

An *Agrobacterium* LBA4404 pre-culture, containing the CRISPR/Cas9 construct, was set up 48 hours before transformation in MGL liquid *medium* pH 7 supplemented with antibiotics (Table S4) and incubated overnight at 28°C with agitation. 24 hours before transformation, from this pre-culture a second culture was set up in TY liquid *medium* pH 5.8 supplemented with 200 µM acetosyringone with no antibiotics (Supplementary Table S4) and incubated overnight in the dark at 28°C with agitation.

Before transformation, the optical density (OD) of the bacterial culture was measured at 600 nm and the culture diluted to a final OD of 0.10-0.12 in TY *medium* with 200 µM acetosyringone. Explants of about 5 mm in length were cut from the cotyledons, dipped in the bacterial culture for 5 minutes, then blotted dry on filter paper and transferred for 48 hours to a co-culture *medium* in the dark.

2.4 Organogenesis and regeneration

The regeneration of plantlets from the transformed cotyledon explants was performed according to the method described by Qiu *et al.* (2007), with modifications. Culture *media* composition is reported in Supplementary Table S5. Briefly, after 48 h of co-culture with *A. tumefaciens*, explants were grown on a *medium* to induce the formation of *callus* and shoots (induction *medium*). Explants were moved to a fresh *medium* every 21 days, or when explants size or shoot formation required it. Shoots were then transferred to an elongation *medium* and finally moved to a rooting *medium*. Kanamycin selection was maintained at all stages of regeneration. Plants were grown *in vitro* to a size of 5-8 centimeters and then were moved to soil in the greenhouse, where they were gradually acclimated to environmental growing and humidity conditions.

2.5 DNA extraction and T7EI endonuclease assay

DNA was extracted from leaves of plantlets growing *in vitro* according to a CTAB method (Doyle, 1987) (Supplementary Box S1). Each DNA sample consisted of different leaves from the same plant. Tissues were collected from a total of 27 individuals regenerated from 13 different *calli*. Samples were named using a number (1-13) referring to the *callus* and a capital letter referring to individual plantlets within each *callus*. The quality of the extracted genomic DNA was assessed by running it on a 1% agarose gel and its quantity using the

Nanodrop spectrophotometer. DNA was diluted to a concentration of 50 ng μl^{-1} and this dilution was used for subsequent reactions.

A pair of primers (*gf1* forward and *gf1* reverse, Table 3) was designed using the Primer3 software (http://primer3.ut.ee/) to amplify a region surrounding the CRISPR target site. The resulting amplicon would be 700-900 base pairs in size and the expected Cas9 cut site would be located at 250-300 bp downstream from the 5' end of the fragment, as specified by the instructions at https://tide.deskgen.com/. The selected primers amplify a 866 bp region on *gf1*. Primer specificity was first predicted using the BLAST tool available at https://solgenomics.net/tools/blast/ and then confirmed by PCR, using the MyTaq PCR kit (Bioline). A second pair of primers was designed to amplify a 800 base pair fragment on the *hCas9* gene to assess the presence of the transgene in regenerated plantlets (*hCas9* forward and *hCas9* reverse, Table 3). The PCR reaction mix composition and cycling conditions are listed in Tables 4 and 5, respectively.

Target	**Primer sequence 5' → 3'**	**T_m**
gf1 forward	GGGTCTGGGCCAAAACTACT	59
gf1 reverse	ACAGGCACAAGCCTCTTCAC	59.4
hCas9 forward	AGGTGGCGTACCATGAAAAG	56.5
hCas9 reverse	TGTTTGCGCAACAGATCTTC	55.2

Table 3. Primers for gf1 genotyping and hCas9 amplification.

MyTaq PCR reaction set up	**Volume**
5x MyTaq reaction buffer (contains dNTPs and $MgCl_2$)	5 µl
Template	as required
Primer forward 20 µM	1 µl
Primer reverse 20 µM	1 µl
MyTaq DNA Polymerase	0.3 µl
Water	up to 25 µl

Table 4. MyTaq PCR reaction mix for genotyping and T-DNA amplification.

Step	Temperature	Time	Cycles
Initial denaturation	95°C	3 minutes	1
Denaturation	95°C	20 seconds	35
Annealing	as required	20 seconds/kb	
Extension	72°C	20 seconds	
Final extension	72°C	5 minutes	1
Hold	4°C	∞	1

Table 5. PCR cycling conditions for MyTaq polimerase reactions.

A first screening of editing efficiency was made using the T7 endonuclease I assay to detect the presence of mutations on the *gf1* alleles following the protocol supplied by New England Biolabs Inc (Tables 6 and 7). The *gf1* PCR product was first purified using the Macherey-Nagel PCR Clean-Up kit and quantified, then used to set up the following annealing reaction:

Component	19 µl annealing reaction
Purified DNA	200 ng
10X NEBuffer 2	2 µl
Nuclease-free water	up to 19 µl

Table 6. Annealing reaction mix specified by NEB for the T7 endonuclease I assay.

Step	Temperature °C	Ramp rate	Time
Initial denaturation	95		5 minutes
Annealing	95-85	-2°C/second	
	85-25	-0.1°C/second	
Hold	4°C		∞

Table 7. Denaturation and re-annealing conditions specified by NEB for the T7 endonuclease I assay.

The re-annealed PCR product was then incubated with the T7 endonuclease I enzyme for 30 minutes at 37°C (Table 8) and analyzed on a 2% agarose gel.

Component	20 µl reaction
Annealed PCR product	19 µl
T7 endonuclease I	1 µl

Table 8. Digestion mix for T7E1 endonuclease assay.

2.6 Genotyping of the T_0 generation

Subsequently, the *gf1* PCR was repeated, purified according to previously reported conditions and directly sequenced through Sanger's method for all individuals positive for the T7EI assay, 2 negative individuals and the wild type control. Each PCR product represented a population of molecules comprising the different alleles present in the plant's genome after editing. The resulting chromatograms were analyzed using the TIDE software (https://tide.nki.nl/), which performs a decomposition of the chromatogram signals around the expected cut site and calculates allele frequencies (Brinkman *et al.*, 2014). In the cases in which it was not possible to carry out an analysis with TIDE, PCR products were purified, cloned into pCR™ 2.1-TOPO™ TA Vectors (Invitrogen) and transformed into TOP10 electrocompetent *E. coli* cells (Invitrogen). The insert from the purified plasmid was sequenced using the *gf1* forward primer (Table 3) on an ABI 3130 XL capillary sequencer (DNA Sequencing Service of the IBMCP-UPV). Resulting sequences were aligned to the wild type using Benchling.

A second sampling was performed on 4 month old T_0 plants growing in the greenhouse, collecting DNA from two leaves and two fruits from each plant; samples were collected from different, distant branches on the plant. Tissues were frozen in liquid nitrogen, genomic DNA extracted according to the CTAB method and a second genotyping by amplification, sequencing and TIDE analysis was performed to assess the uniformity and stability of the mutation across different organs and developmental stages.

The ploidy level of the lines growing in the greenhouse was evaluated by flow cytometry (CyFlow® Counter, Sysmex Partec).

2.7 Evaluation of off-target effects

Possible Cas9 off-target sites for the *gf1* gRNA were identified using the CasOFFinder software (www.rgenome.net/cas-offinder/). The *S. lycopersicum* SL2.5 genome assembly was used as reference. The maximum number of mismatches was set at 5, with DNA and RNA bulge sizes set at 1. Five different off-target sites were selected based on the number and position of mismatches. The sequence of each off-target *locus* (1 kb) was retrieved using the tomato genome browser at https://solgenomics.net/ and aligned to the Viridiplantae nucleotide database on https://blast.ncbi.nlm.nih.gov/Blast.cgi to screen for annotations. The 1 kb sequence of off-target *loci* is available in the Supplementary Sequence List at the end of this Chapter.

The genotyping of the candidate off-target *loci* was performed on the 5 plant lines growing in the greenhouse using an Illumina Amplicon Sequencing Protocol (https://nematodegenetics.wordpress.com). This protocol uses a first round PCR to amplify the target sites and attach Illumina adapters (Supplementary Table S6), and a second round PCR to add Nextera indexes to the universal Illumina adapters (Supplementary Table S7). 10 µl volume PCR reactions were carried out in 384 well plates with each primer at a concentration of 1.25 µM. The first round PCR was performed according to the reaction set up described in Table 9, using the cycling conditions described in Table 10.

Kapa HiFi PCR reaction set up	Volume
2X KAPA HiFi HotStart Ready Mix	5 µl
gDNA 15 ng μl^{-1}	2 µl
Primer forward 2 µM	1.5 µl
Primer reverse 2 µM	1.5 µl

Table 9. KAPA HiFi reaction mix for the first round PCR of the Illumina Amplicon Sequencing Protocol.

Step	Temperature	Time	Cycles
Initial denaturation	95°C	3 minutes	1
Denaturation	98°C	20 seconds	20
Annealing	60-63°C	15 seconds	
Extension	72°C	15 seconds	
Final extension	72°C	1 minutes	1
Hold	4°C	∞	1

Table 10. KAPA HiFi cycling protocol for the first round PCR of the Illumina Amplicon Sequencing Protocol.

PCR products were run on a 1.5% agarose gel and purified using Ampure beads (Beckman-Coulter) at a 0.8:1 v/v concentration with respect to the sample, following the manufacturer's instruction. Purified samples were quantified and diluted to a concentration of 0.5 ng μl^{-1} to set up the second round amplification. This second amplification was performed according to the reaction set up described in Table 11, using the cycling conditions described in Table 12.

Kapa HiFi PCR reaction set up	Volume
2X KAPA HiFi HotStart Ready Mix	5 µl
First round purified PCR product 0.5 ng μl^{-1}	2 µl
Primer forward 2 µM	1.5 µl
Primer reverse 2 µM	1.5 µl

Table 11. KAPA HiFi reaction mix for the second round PCR of the Illumina Amplicon Sequencing Protocol.

Step	Temperature	Time	Cycles
Initial denaturation	95°C	3 minutes	1
Denaturation	98°C	20 seconds	10
Annealing	60°C	15 seconds	
Extension	72°C	15 seconds	
Final extension	72°C	1 minutes	1
Hold	4°C	∞	1

Table 12. KAPA HiFi cycling protocol for the second round PCR of the Illumina Amplicon Sequencing Protocol.

The products of the second round amplification were run on a 1.5% agarose gel and purified with Ampure beads 0.8:1 v/v. Finally, indexed PCR products were pooled for sequencing.

Sequencing was performed with an Illumina HiSeq2500 sequencer (Illumina Inc., San Diego, CA, USA) and 150 bp paired-end sequences were generated. Raw reads were analyzed with Scythe for filtering out contaminant substrings (https://github.com/vsbuffalo/scythe) and with Sickle to remove reads with poor quality ends (Q < 30) (https://github.com/najoshi/sickle). Reads were demultiplexed using a custom bash script targeting the dual index sequence combination (*e.g.*: TAAGGCGACTCTCTAT) in the header of each read. In this sequence, the first part is specific for each *locus* (*e.g.* off-target 1: TAAGGCGA), while the second part is specific for every individual (*e.g.* individual 1: CTCTCTAT). Indexes are listed in Supplementary Table S7.

Each putative target sequence (20 nucleotides) was used to survey cleaned fastq sequences deriving from amplified off-target sites and for recovering potentially edited sequences. Fastq sequences showing mutations with respect to the wild type target sequence, likely to be off-targets carrying mutations, were converted to fasta using a custom bash script. The resulting fasta sequences were multi-aligned in Clustalx (http://www.clustal.org/clustal2/) and alignments were manually inspected.

2.8 Genotyping of the T_1 generation

Seeds were collected from T_0 fruits and 50 seeds per line were sowed. Finally, 26 plants were kept in the greenhouse: 2 wt, 9 from line 2B, 3 from line 4C and 12 from line 12A.

A PCR was performed on the *Cas9* gene to assess whether or not the transgene had been segregated, using a new pair of primers (Table 13). The genotyping of the *gf1 locus* was performed following the same protocol used for the T_0 generation.

Target	Primer sequence 5' → 3'	T_m
hCas9 forward	GAAAGTTACCGTGAAACAGC	53.2
hCas9 reverse	CACGATTTCTCCTGTTTCTC	52.1

Table 13. Primer combination amplifying a 1.55 kb region in the *hCas9* gene.

2.9 Phenotypic evaluation

The T_0 regenerants growing in the greenhouse were inspected for visible phenotypes relating to the *gf1* mutation. In particular, senescent leaves and ripening fruits were considered indicators of chlorophyll retaining.

According to the procedure described in Barry *et al.* (2008), three leaves were collected from each of the five *gf1* plants and the wild type growing in the greenhouse, and incubated for two weeks in the dark at room temperature, floating on water, in order to observe changes in chlorophyll content and general appearance.

3 - Results

3.1 Vector design and construction

It was not possible to design a guide RNA specifically targeting a structurally important region of the SGR protein, as no annotated functional domains have been identified *in silico*. Also, spontaneous mutations occuring in different regions of the gene have an equal impact on protein function and phenotype (Barry and Pandey, 2009). Protein loss of function is less likely to occur if mutations are directed at the 3' end of the CDS, because this could generate a truncated but still functional version of the protein. For this reason, taking into consideration the position, score and sequence requirements (the first G at the 5' end due to the promoter choice) for the gRNA, the following sequence was selected in the third exon: 5' - GTCCATTGCCACATTAGTGG - 3'. The first cloning step, in which the gRNA was incorporated into its TU in the level 1 pDGB alpha1 vector, was validated by sequencing. The following steps were validated through digestion of vectors with two restriction enzymes (exemplified in Fig. 24).

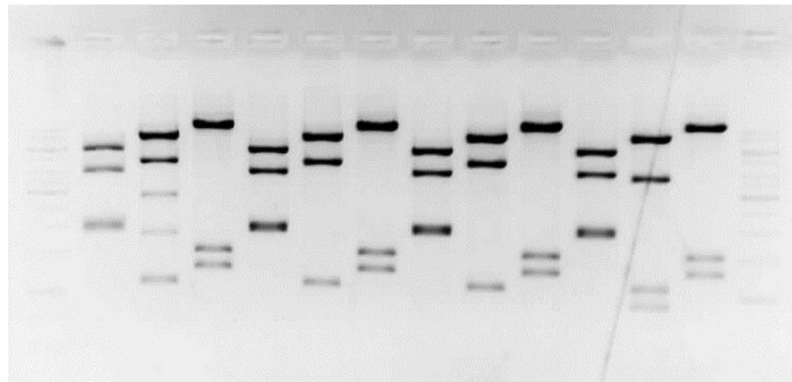


Figure 24. Example of enzymatic restriction of GB vectors. The use of multiple restriction enzymes and of highly standardized DNA parts ensures accuracy in evaluating assembly of vectors by their restriction profiles.

3.2 Plant regeneration

Following cotyledon transformation with the *gf1* CRISPR/Cas9 construct, 27 plants were regenerated and rooted *in vitro* from 13 distinct *calli* (Table 14). The introduction of the *gf1* CRISPR/Cas9 construct did not have any negative effect on *in vitro* culture.

<table>
<tr><th>Callus</th><th>Plantlet</th></tr>
<tr><td>1</td><td>A</td></tr>
<tr><td rowspan="3">2</td><td>A</td></tr>
<tr><td>B</td></tr>
<tr><td>C</td></tr>
<tr><td rowspan="3">3</td><td>A</td></tr>
<tr><td>B</td></tr>
<tr><td>C</td></tr>
<tr><td rowspan="3">4</td><td>A</td></tr>
<tr><td>B</td></tr>
<tr><td>C</td></tr>
<tr><td>5</td><td>A</td></tr>
<tr><td>6</td><td>A</td></tr>
<tr><td rowspan="3">7</td><td>A</td></tr>
<tr><td>B</td></tr>
<tr><td>C</td></tr>
<tr><td>8</td><td>A</td></tr>
<tr><td rowspan="4">9</td><td>A</td></tr>
<tr><td>B</td></tr>
<tr><td>C</td></tr>
<tr><td>D</td></tr>
<tr><td rowspan="2">10</td><td>A</td></tr>
<tr><td>B</td></tr>
<tr><td>11</td><td>A</td></tr>
<tr><td rowspan="3">12</td><td>A</td></tr>
<tr><td>B</td></tr>
<tr><td>C</td></tr>
<tr><td>13</td><td>A</td></tr>
</table>

Table 14. Identity and classification of regenerated plantlets.

3.3 Genetic characterization of the T_0 generation

The T7EI essay was performed on the 27 regenerated plantlets and showed the presence of polymorphisms in 17 of them (2B-C, 3A-B-C, 4A-B-C, 6A, 8A, 10A-B, 11A, 12A-B-C, 13A). In this assay, the purified PCR products are

denatured and let re-anneal gradually; if different alleles are present, heteroduplexes form, with one or more mismatches. Re-annealed DNA was digested with the T7EI endonuclease, which recognizes and cleaves such mismatches. Therefore, the wild type DNA has only one undigested band, while three bands are expected in edited individuals, one being made of undigested, homoduplex DNA (from either allele) and the other two resulting from the digestion of heteroduplexes. The wt and undigested bands were of the expected size (866 bp), while the two additional bands had sizes which added up to the heavier band. Individuals 12A-B-C showed a more complex situation, with a total of five bands: apart from the 866 bp band, there was a second, slightly smaller band, and three digested bands were present (Fig. 25).

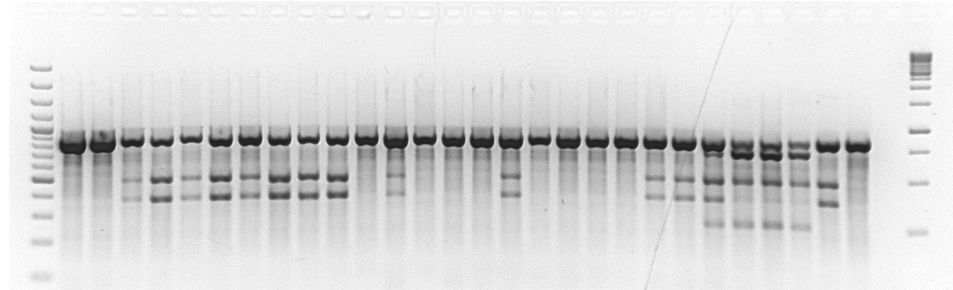


Figure 25. T7E1 endonuclease assay. 17 of 27 plants show polymorphism at the target site. Lanes, left to right, represent: 1A, 2A-B-C, 3A-B-C, 4A-B-C, 5A, 6A, 7A-B-C, 8A, 9A-B-C-D, 10A-B, 11A, 12A-B-C, 13A, wt, negative control without DNA. The outer left and right lanes are occupied by the molecular marker.

The integration of the T-DNA was verified in all T7EI-positive individuals by amplifying a fragment of the *hCas9* gene, and all proved to be positive. The undigested PCR product of the *gf1 locus* was directly sequenced for all individuals which appeared edited from the T7EI assay, together with the correspondent wt DNA and three samples which proved negative in the T7EI assay (2A, 9A and 9B). TIDE analysis of chromatograms confirmed the sensitivity of the restriction analysis, revealing a series of mutations at the *gf1 locus* in individuals which appeared to be edited, while samples 2A did not show any mutation, and 9A showed very low frequencies of mutation (two mutations with a frequency of about 2.5% each). Table 15 summarizes the results of the genotyping of the *gf1 locus*. The R^2 value indicates the percentage of variance explained by the model. Overall efficiency is calculated as the difference between R^2 and the amount of wild type allele, so it also takes into consideration noise in the data and percentages do not necessarily add up to 100 (in fact, they add up to the R^2 value). Values in the remaining columns indicate the percentage of each particular allele.

Overall, 63% of regenerated plantlets appeared edited (17 edited plants of 27 regenerated plants). Editing efficiency within each plant ranges from 21 to 98.6% as explained by the model (R^2). Among edited plants, three (2A, 6A and 11A, representing 17.6% of edited individuals) were heterozygous at the target *locus*, meaning that they retained a wild type allele; 82.4% (14 plants) were biallelic, with two different edited alleles. No homozygous edited events have been found. The most frequent mutation was a single nucleotide insertion, found in 12 individuals; T was the inserted nucleotide in 67% of cases. No insertions of more than one nucleotide were found. *Calli* 7 and 9 did not yield any transformed plant, or yielded plants with very low transformation efficiencies (like 9A) pointing to a lack of expression of the editing machinery, while in *callus* 2 only one individual did not appear mutated. Also, all plantlets regenerated from *callus* 4 had four distinct alleles, suggesting that all individuals arisen from this *callus* were tetraploid; however, single alleles were different, which indicates that gene editing occurred after polyploidization.

Sample	Overall efficiency	R^2	WT	+1 Insertion	Inserted base	-123 deletion	-20 deletion	-12 deletion	-6 deletion	-4 deletion	-3 deletion	-1 deletion
2A	1	0.99	98.3									
2B	57.5	0.94	36.5	53.4	G/T							
2C	97	0.97		49	A				47.9			
3A	98	0.98		65.6	T						32.2	
3B	96.2	0.96		47.2	G				48.9			
3C	98.6	0.99		65.1	T						33	
4A	94.8	0.95		22.2	T		25.5				22.9	23
4B	94.2	0.94		21.7	T			23.1			25.2	24.2
4C	79.2	0.79		17.3	T		21				19.4	19.3
6A	47.6	0.97	49.1	46.1	A							
8A	98	0.98		72.7	A							22.3
9A	8.1	0.97	89.2						2.7			2.4
9B	2.8	0.99	96									
10A	96.4	0.96		45.9	T							46.3
10B	97.5	0.97		46.9	T							49.6
11A	21	0.95	74.3							11.3		
13A	96.9	0.97							49.8			47

Table 15. Genotyping of T_0 individuals by PCR amplification of the target *locus*, direct sequencing and TIDE chromatogram decomposition. In green are individuals chosen for growth in the greenhouse as representatives of specific genotypes.

Samples 12A-B-C could not be analyzed with TIDE, since no significant alleles could be recognized. TIDE is usually able to identify two or more edited alleles whose frequency is statistically significant (Fig. 26). Samples analyzed from *callus* 12, however, yielded atypical results characterized by the recognition of many non-significant alleles (Fig. 27), which did not indicate neither a specific editing pattern, nor the retention of the wild type allele, rather indicating a technical issue.

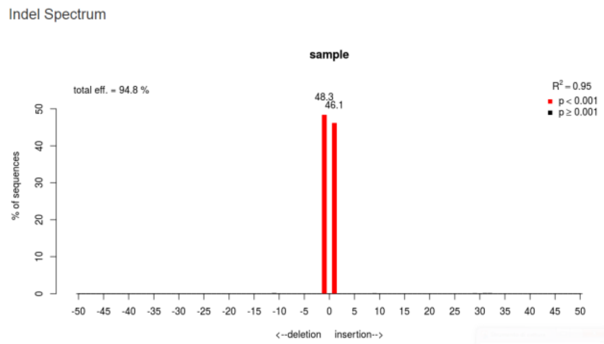


Figure 26. Typical output of a TIDE analysis. Two alleles are clearly recognizable, efficiency is high and background noise is low. Here one allele is represented by a 1 nucleotide insertion (46.1%) and the second allele is represented by a 1 nucleotide deletion (48.3%).

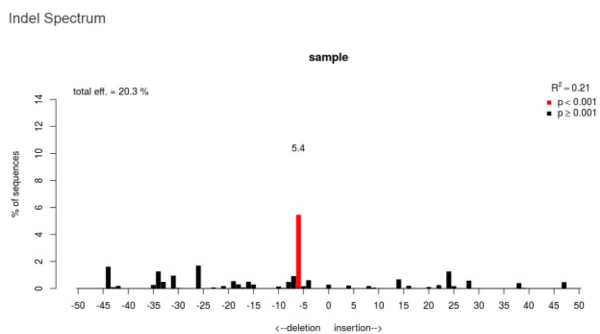


Figure 27. TIDE output for samples 12A-B-C. A single mutated allele is recognized as statistically significant, but has a very low frequency (5.4%). Total efficiency is recognized as very low, background noise is very high and many non significant mutations are detected.

Therefore, the corresponding PCR products were cloned and sequenced. This revealed the presence of two alternative alleles, one characterized by a 4 bp deletion (27.8%) and the other carrying a larger deletion of 123 bp (72.2%). This polymorphism was consistent with the bands observed during amplification and the T7EI assay. The reasons why TIDE could not detect these mutants lie in the fact that it has a maximum 50 bp window for analysis of indels, for which the 123 bp deletion was too large, and also because in these plants Cas9 did not cut at the expected site 3 bp upstream of the PAM. In the 4 bp deletion allele, the cut site was located at 1 bp upstream of the PAM, while in the 123 bp deletion allele the breakpoint was 5 bp upstream of the PAM.

Figure 28 summarizes the types of indels observed, and reports their frequencies as number of plants carrying each allele. Since the ploidy level differed between individuals, with a consistent number of tetraploids, editing events are not reported by means of frequencies of alleles in the total gene pool, but as occurrence of individuals with a particular mutation.

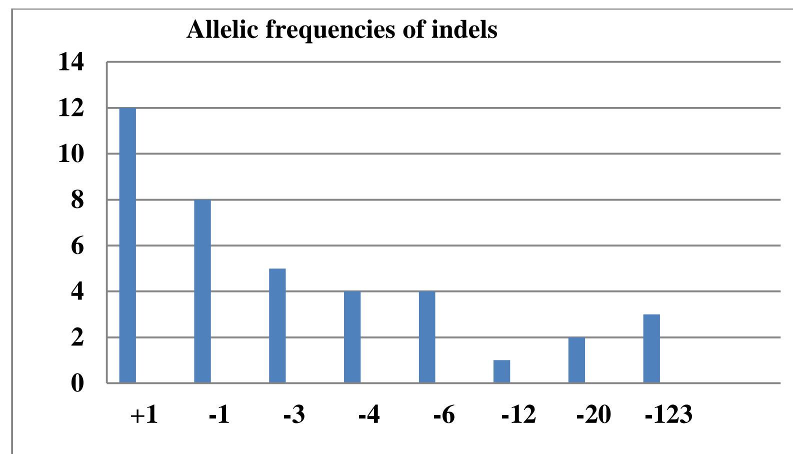


Figure 28. Allelic frequencies of indels (number of plants carrying each allele). The insertion of a single nucleotide is markedly the most abundant mutation.

Based on this data five edited plants, along with a wild type individual, were selected to be moved to the greenhouse from *in vitro* culture. The selected five distinct genotypes were: 2B as an example of a heterozygous plant, 3C, 4C as a likely tetraploid (since it possesses 4 alleles in similar proportions), 10B and 12A, for its particular editing pattern. In general, we preferred to select deletions which were not multiples of three, because this would mean removing one or more aminoacids without altering the reading frame and without necessarily disrupting protein function. The plants displayed a normal growth pattern in comparison to the wild type, except for plant 10B which was stunted. A second genotyping was performed when fruits had reached full maturity, by sampling two leaves and two fruits for each plant; fruits and leaves were collected from distant branches. The objective of this second genotyping was to assess whether CRISPR-induced mutations were stable and whether chimerism was observable between different tissues of the same individual. CRISPR-induced mutations were, indeed, stable during plant development and, in all instances in which no wild type allele had been detected during the first genotyping, the proportions of mutated alleles did not vary across different tissues (Table 16). On the contrary, in individual 2B, which was heterozygous, Cas9 continued to act during plant growth, resulting in a decrease of the wild type allele, in favor of editing. Interestingly, only one type of mutation was

retrieved in all tissues, albeit with different frequencies, thus resulting in an enrichment of a single edited allele.

Evaluation of the ploidy level through flow cytometry of the plants in the greenhouse established that 3 (3C, 4C and 10B) out of 5 edited plants were tetraploid, while 2B, 12A and the wild type were diploid. This was directly observable in the number of seeds produced by fruits of the T_0 generation. While 2B and 12A produced large amounts of mature seed, tetraploid plants failed to yield comparable amounts. Notably, seeds from 4C were viable, while those from 3C and 10B did not germinate and failed to generate progeny.

Sample	Overall efficiency	R^2	WT	+1 Insertion	Inserted base	-20 deletion	-3 deletion	-1 deletion
2B	57.5	0.94	36.5	53.4	G/T			
2B leaf 1	75.2	0.96	21.8	72.7	G/T			
2B leaf 2	70.9	0.96	25.3	67	G/T			
2B fruit 1	93.4	0.96	2.8	89.3	G/T			
2B fruit 2	96.3	0.96	0	94.1	G/T			
3C	98.6	0.99		65.1	T		33	
3C leaf 1	95.4	0.95		62.2	T		31	
3C leaf 2	95.8	0.96		63.3	T		31	
3C fruit 1	96.2	0.96		63.1	T		31.2	
3C fruit 2	94.8	0.95		60.9	T		30.4	
4C	79.2	0.79		17.3	T	21	19.4	19.3
4C leaf 1	91.6	0.92		20.8	T	26.2	22.4	22.1
4C leaf 2	91.5	0.92		20.9	T	26.6	21.6	22.4
4C fruit 1	91.6	0.92		20.9	T	26.2	22.2	22.2
4C fruit 2	92.8	0.93		21.5	T	26.5	22.6	22.1
10B	96.4	0.96		45.9	T			46.3
10B leaf 1	94.8	0.95		46.1	T			48.3
10B leaf 2	95.5	0.96		46.8	T			48.6
10B fruit 1	95	0.95		46.2	T			48.1
10B fruit 2	95.2	0.95		46.5	T			48.2

Table 16. Second genotyping of different tissues of the edited plants. In grey the results of the first genotyping, compared to the distinct tissues of the adult plant. Efficiencies and frequencies are stable for all individuals except 2B, in which the proportion of the wild type allele decreases in favor of the single base insertion.

3.4 Analysis of off-target effects

Five putative off-target sequences were selected among those identified by CasOFFinder, as representatives of different kinds and numbers of mismatches. Their features are presented in Table 17 and the sequences are available at the end of this Chapter in the Supplementary Sequence List. Two of them were chosen because they are part of protein-coding sequences and their mutations might have undesired impacts on plant metabolism. In general, no sequences are present in the tomato genome with less than 4 mismatches with respect to the *gf1* gRNA sequence, without considering bulges. When letting bulges into the analysis, the minimum number of mismatches is 2. Deep sequencing (with an average sequencing depth of 20,000 reads per *locus*) proved that indeed no such effects occurred, confirming existing data on the specificity of CRISPR/Cas9 *in planta* (Table 18). By randomly selecting 300 sequences from edited plants and 30 wt sequences and aligning them, it was possible to identify a number of SNPs at these *loci*, but no mutations related to a possible CRISPR editing activity.

Off-target	Chromosome	Position	Sequence	Annotation
1	6	7491701	agtCATTGCgAtATTAGTGGGGG	non coding
2	2	30647910	GcaaATTGCtcCATTAGTGGTGG	non coding
3	2	34542804	GTtggTgGaCACATTAGTGGTGG	cannabidiolic acid synthase (accession XM_004233099)
4	6	38237927	GgCCgTTGCCACATATAcTGGTGG	probable plastidic glucose transporter 3 (accession XM_010324389)
5	1	40900162	tTCCAcTtCC-CATTAGTGGTGG	non coding

Table 17. Characteristics of putative off-target sites for the *gf1* gRNA. Mismatches are in lowercase and bulges are represented by dashes.

	Sample	N. of sequences	N. WT target sequence	% WT	Normalized % edited
OFF TARGET 1	WT	27830	26954	96,85%	0
	2B	13578	13059	96,18%	0.69
	3C	19762	19147	96,89%	0
	4C	27090	26212	96,76%	0.09
	10B	54895	53257	97,02%	0
	12A	27017	26185	96,92%	0
OFF TARGET 2	WT	27176	26286	96,73%	0
	2B	29374	16091	96,57%	0.17
	3C	15320	14852	96,95%	0
	4C	14941	14428	96,57%	0.17
	10B	22078	21379	96,83%	0
	12A	14493	14020	96,74%	0
OFF TARGET 3	WT	33293	32344	97,15%	0
	2B	33875	15963	97,04%	0.11
	3C	13688	13305	97,20%	0
	4C	14565	14122	96,96%	0.20
	10B	21609	20948	96,94%	0.22
	12A	25735	25023	97,23%	0
OFF TARGET 4	WT	5853	5550	94,82%	0
	2B	49903	35932	96,37%	0
	3C	21256	20538	96,62%	0
	4C	26196	25185	96,14%	0
	10B	40111	38654	96,37%	0
	12A	23802	22489	94,48%	0.36
OFF TARGET 5	WT	12606	12255	97,22%	0
	2B	34547	14847	96,96%	0.27
	3C	16092	15617	97,05%	0.17
	4C	16053	15584	97,08%	0.14
	10B	8904	8447	94,87%	2.42
	12A	11453	11155	97,40%	0

Table 18. Illumina amplicon sequencing results for 5 putative *gf1* off-target *loci*. The number of sequence reads carrying the wild type sequence are comparable in all *gf1* edited individuals to the wild type, showing that no off-target activity occurred.

3.5 Genetic evaluation of the T_1 generation

Amplification of the *hCas9* gene in the T_1 generation revealed different patterns of inheritance for the 3 different lines which were analyzed. In 8 out of 12 individuals analyzed from the 12A line no amplification of *hCas9* was detectable and it is therefore reasonable to deduce that the transgene had been segregated. On the other hand, 4 individuals from line 12A and all but 1 individual from line 2B had the expected 1.55 kb band at the *hCas9 locus* (Fig. 29). Also, *hCas9* was detected in all 4 plants from line 4C. However, segregation will be analyzed also in the following generations, together with the stable inheritance of the edited alleles.

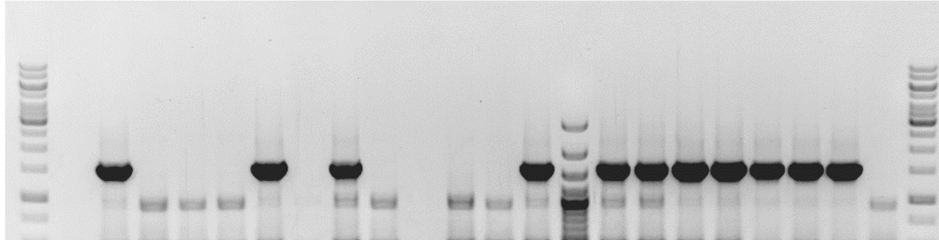


Figure 29. Amplification of the *hCas9 locus* in the T_1 progeny of the 12A and 2B lines. The outer left and right lanes are occupied by the molecular marker; the first lane to the left is a negative control, followed by 12 individuals from the 12A line and 8 individuals from the 2B line.

The *gf1 locus* was genotyped following the same procedure used in the T_0 generation. Mutations were stable and were inherited in a Mendelian fashion. Since a 119 bp polymorphism was present in line 12A, it is also possible to distinguish heterozygotes from homozygotes by amplifying the target *locus* and running the PCR product on an agarose gel (Fig. 30).

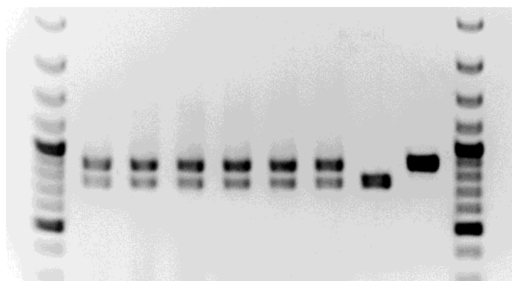


Figure 30. Amplification of the gf1 *locus* in the T_1 progeny of the 12A line. Two alleles are clearly visible in heterozygotes (left), while homozygotes have either a 743 bp band, or a 862 bp band. The outer left and right lanes are occupied by the molecular marker

.

3.6 Phenotype

The inspection of the T_0 generation plants growing in the greenhouse revealed a normal growth habit, except for plant 10B which appeared significantly stunted, a feature that could relate to its ploidy level. Nonetheless, other tetraployd individuals (3C and 4C) had normal size. Observation of older branches showed non-yellowing of leaves and green tissues in general, with the concomitant appearance of darker spots in areas between leaf veins (Fig. 31a). Full ripe fruits had normal size, with a darker pigmentation, consistent with "chocolate-like" phenotypes, due to the retention of chlorophyll along with the synthesis of ripening-associated pigments (Fig. 31b-c-d-e).

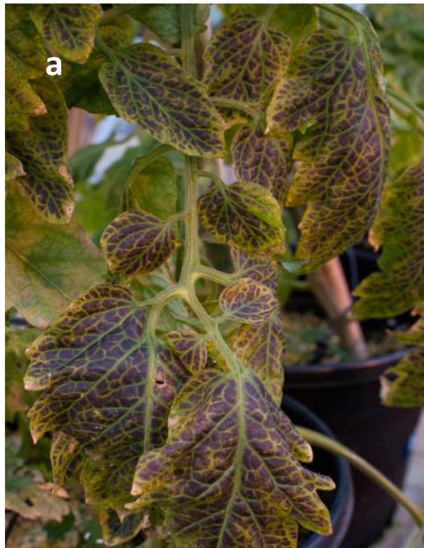


Figure 31a-b-c-d-e. (a) Non-yellowing senescent leaves; (b) *gf1* ripe fruits (background) compared to a wt 'MoneyMaker' fruit (foreground); (c,d,e) ripe *gf1* fruits.

Incubation of *gf1* and wt leaves in the dark for two weeks at room temperature revealed that chlorophyll was retained even in the absence of light, confirming the results reported by Barry *et al.* (2008). Another important feature to be noted is that, in addition to pigmentation, *gf1* leaves retained turgor and were not subject to the attack of saprophytic organisms, while wt leaves appeared wilted and showed growth of saprophytic fungi (Fig. 32a-b-c-d). No difference in treatment was applied to *gf1* and wt leaves prior to incubation, and none of them was rinsed.

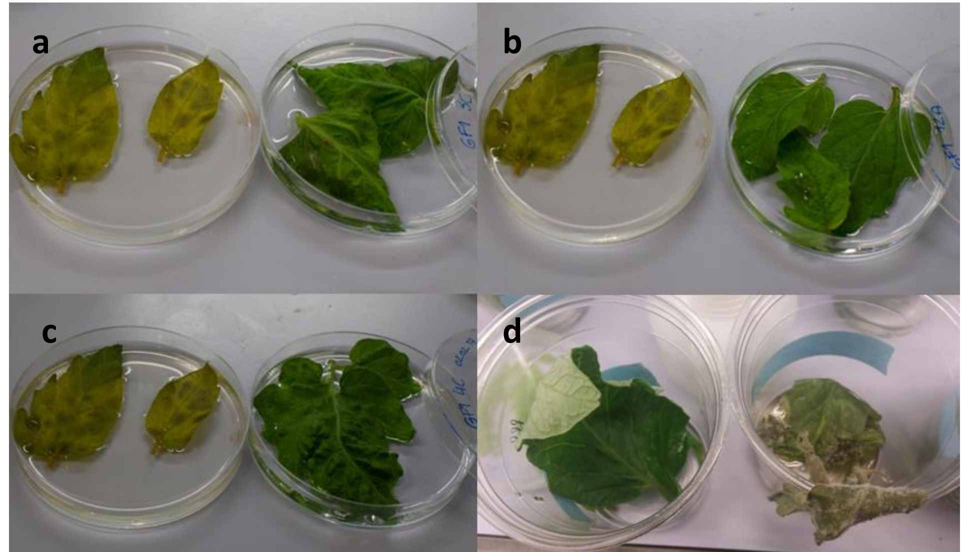


Figure 32. Comparison between wt and *gf1* leaves. Leaves were incubated for two weeks in dark, floating on water, at room temperature. (a) wt vs. 3C leaves; (b) wt vs. 12A leaves; (c) wt vs. 4C leaves; (d) 2B vs. wt leaves.

4 - Discussion

4.1 Transformation and regeneration efficiency

Regeneration efficiencies are generally good in tomato, for which numerous *in vitro* applications are known, including transformation, micropropagation, haploid regeneration and mutation selection (Gerszberg *et al.*, 2015). Among the most important factors affecting regeneration efficiency, genotype, explant source and age, and *medium* composition can be considered (Moghaieb *et al.*, 1999; Hu and Phillips, 2001; Gupta and Van Eck, 2016). In our experiments, 10 days after sowing, cotyledons of 'MoneyMaker' tomatoes were used for *Agrobacterium*-mediated transformation. 'MoneyMaker' is a variety widely used for experimental work (Chaudhry *et al.*, 2010) and tomato cotyledons are among the tissues with the highest regenerative potential, together with hypocotyls, as opposed to leaf tissue (Mamidala and Nanna, 2011). Optimal age and size for cotyledonary explants are 10 days after sowing (when the first true leaves start to emerge) and 5 mm x 5 mm, respectively, as reported in Chaudhry *et al.* (2010).
Regarding *medium* composition, the use of phytohormones and antibiotics (used both for selection and as anti-bacterial agents) was shown to play an important role in inducing *callus* formation and in the organogenesis process (Gerszberg *et al.*, 2015). The most widely used basal *medium* is the Murashige & Skoog *medium*, supplemented with auxins (indoleacetic, naphtaleneacetic and indolebutyric acid) and cytrokinins (zeatin, benzylaminopurine and thidiazuron). Different combinations of these growth regulators, often determined empirically for each variety and type of explant, are used to achieve either *callus*, shoot or root formation. Both somatic embryogenesis and organogenesis have been shown to occur during *in vitro* regeneration of tomato using the same *media*. Our regeneration *medium*, contained 2 mg l^{-1} trans-zeatin and 0.1 mg l^{-1} IAA and led to direct shoot formation. This confirms the results described in Khuong *et al.* (2013) which report that the use of 1 mg l^{-1} trans-zeatin and 0.1 mg l^{-1} indoleacetic acid (IAA) induced very little *callus* formation and favored instead direct shoot regeneration. The antibiotics used to eliminate *Agrobacterium* cells from tomato tissue culture also have an effect on the number of regenerated shoots. Mamidala and Nanna (2009) reported that some of the most widely used antibiotics, like cefotaxime and carbenicillin, while being highly effective against *Agrobacterium*, have a negative effect on shoot regeneration, especially at concentrations greater than 400 mg l^{-1}; on the

other hand, timentin (a combination of ticarcillin and clavulanic acid) has the least effect on shoot formation and is equally effective as a bacteriostatic. In our *media* we used 500 mg l^{-1} carbenicillin, and we might therefore consider improving this aspect of regeneration to reduce negative effects associated to the use of antibiotics and further improve shoot formation. Finally, the environmental conditions, especially light conditions and temperature, have important effects on development and growth; Bhatia and Ashwath (2005) reported that the most appropriate light cycle is a 16:8 light:dark day.

Many authors refer to transformation efficiency as the percentage of explants which originate a transformed plantlet. In this sense, many transformation protocols have regeneration efficiencies around 10% (McCormick *et al.*, 1986; Hu and Phillips, 2001; Yasmeen, 2009), and our values are coherent with these estimates. According to this metric, Sun *et al.* (2006) reported higher efficiencies, exceeding 40%, for the 'MicroTom' variety. However, in our experiment, the mean number of plantlets we recovered for each explant (*i.e. callus*) was 2.08, meaning that some *calli* were able to produce multiple shoots, and this increases transformation efficiency to 18% if we consider the total number of recovered plantlets instead of the number of shoot-forming *calli*. The number of recovered individual plants (27) in our study is consistent with those reported by Brooks *et al.* (2014) (29 plants) and Ito *et al.* (2015) (16 and 17 plants from two transformation events); Pan *et al.* (2016) obtained 4 tomato edited lines, with 7-25 individuals per line. While we cannot compare transformation efficiencies between these protocols and ours, we can point out that the number of plantlets we recovered is comparable to the ones reported in these publications, meaning that we ensured a comparable depth of analysis of editing profiles in our regenerants.

4.2 Editing efficiency

Overall, 63% of regenerated plantlets appeared edited (17 edited plants of 27 regenerated plants). A certain uniformity regarding editing efficiency can be found within each *callus*: *callus* 7 originated 3 shoots, all of which are not edited, while all plantlets originating from *callus* 3 were edited. The plantlets originating from a single *callus* cannot be regarded as clones, but rather as separate editing events, since they often have at least partially different mutations. Their genetic uniformity, or lack thereof, is probably to be attributed to the moment at which editing occurs: in tissues in which editing occurs earlier during *callus* formation, all daughter cells are bound to inherit the same

mutations and their genotypes are more homogenous, while if it occurs later, it leads to a chimeric *callus* with different editing patterns.
Our data show that 10 out of 17 edited individuals have editing efficiencies greater than 90%, while editing efficiency within each plant ranged from 21 to 98.6% as explained by the model (R^2). These values are also consistent with previous findings: Pan *et al.* (2016) report editing efficiencies in tomato ranging from 47 to 85%, Klap *et al.* (2017) report editing efficiencies of 19-37% and Ueta *et al.* (2017) show editing efficiencies of 30-99%. Our results prove CRISPR to be an efficient editing tool in tomato and GoldenBraid constructs to be an effective delivery method for plant transformation.

4.3 Editing patterns

Previous observations about the types of mutations introduced in plants by gene editing, showed that small indels are predominant and that the breakpoint introduced by Cas9 is usually at 3 nucleotides upstream of the PAM (that is, between nucleotides 17 and 18 of the 20 nt gRNA) (Bortesi *et al.*, 2016; Pan *et al.*, 2016; Andersson *et al.*, 2017). However, Brooks *et al.* (2014), Ito *et al.* (2015), Eid *et al.* (2016) and Nekrasov *et al.* (2017), among others, reported that it is also possible for Cas9 to cut in other positions (even downstream of the PAM). Gao *et al.* (2014) targeted the ABC transporter *NtPDR6* of *Nicotiana tabacum* and found deletions spanning across the PAM region, involving both the 3'- and 5'-adjacent regions (Fig. 33). In plant 12A we found that the cut site was either at -1 or at -5 nucleotides from the PAM, and not at -3; in all plantlets originating from other *calli*, the cut site was found at the expected position.

NtPDR6 gRNA1 9.5% (9/95) mutated PAM

```
CATGTGAGTGTAGATGCAGAAGCTAG-AGTTGGTAGTAGAGCTCTACCTACAATA WT
CATGTGAGTGTAGATGCAGAAGCTAGAAGTTGGTAGTAGAGCTCTACCTACAATA +1
CATGTGAGTGTAGATGCAGAAGCTAG-AGTTGGTGGTAGAGCTCTACCTACAATA -1/+1
CATGTGAGTGTAGATGCAGAGGCTAG-AGTTGGTAGTAGAGCTCTACCTACAATA -1/+1
CATGTGAGTGTAGATGCAGAAGCTAG-AGATGGTAGTAGAGCTCTACCTACAATA -1/+1
CATGTGAGTGTAGATGCAGAAGCTA--AGTTGGTAGTAGAGCTCTACCTACAATA -1
CATGTGAGTGTAGATGCAGAAGCT---AGTTGGTAGTAGAGCTCTACCTACAATA -2
CATGTGAGTGTAGATGCAGAAGCTAG---TTGGTAGTAGAGCTCTACCTACAATA -2
CATGTGAGTGTAGATGCAGAA------AGTTGGTAGTAGAGCTCTACCTACAATA -5
CATGTGAGTGTAGATGCAGAAGCTA---------------------CCTACAATA -20
```

Figure 33. Different breakpoints in CRISPR-edited alleles of the *NtPDR6* gene. The cut site is usually located 3 bp upstream of the PAM, but in some cases deletions can start either closer to the PAM (*e.g.* at -1, see line 8) or deletions can span the entire PAM region and adjacent bases (line 10). Gao *et al.* (2014).

In plants, insertion of one nucleotide or deletion of 1-10 nucleotides are the most frequent mutations; larger deletions of 20-150 nucleotides are rarer but do occur. Insertions of more than one nucleotide are the less frequent kind of mutation (Bortesi *et al.*, 2016). In our *gf1* edited plants, a single base insertion is the most common type of mutation (44.4% of edited plants), with T being the predominant inserted base (67% of cases). The second most common mutation is a 1 nt deletion (29.6% of edited plants); deletions of 3 to 6 nucleotides are also common, each being reported in 14.8 to 18.5% of edited individuals. Larger mutations (of 12, 20 and 123 nucleotides) are rarer, with efficiencies of 3.7, 7.4 and 11.1%, respectively. In their review, Bortesi *et al.* (2016) also reported that in plants single base insertions (usually A or T) are more common than in animals. Ma *et al.* (2015) described editing efficiencies and patterns in Arabidopsis and rice and, in the screening of 245 mutated sites from both species, they reported that the most frequent mutation was the insertion of a T (24.1%), followed by insertion of an A (19.6%). Zhang *et al.* (2014), too, tested CRISPR/Cas9 efficiency in rice by targeting 10 genes with different functions and found that in the T_0 the most common mutation was the insertion of 1 nucleotide (accounting for 53.9% of edited alleles) and A (44.8%) and T (43.4%) were the bases most frequently integrated at the cut site. This editing pattern seems to depend on the technology which is being used: Wang *et al.* (2014) edited wheat (*Triticum aestivum*) with TALENs for three homoeoalleles encoding mildew-resistance *locus* (MLO) proteins, and in their screening of mutations they did not find any allele carrying a single base insertion. Thus, the specific editing pattern seems to depend both on the type of organism (plants, as opposed to animals and fungi) and on the type of endonuclease (Cas9 *vs.* TALENs).

4.4 Continued activity of Cas9 during plant development

By repeating the *gf1 locus* genotyping on different tissues of plants grown in the greenhouse, we were able to detect an increase in the proportion of edited alleles in the heterozygote individual 2B, which had retained one wt allele in the first genotyping. On the other hand, no changes were found in plants which no longer possessed a copy of the wt *gf1* allele (biallelic) since once the target site is edited, the gRNA recognition sequence does no longer exist, preventing Cas9 from introducing further changes in this region. This result seems to imply that Cas9 is able to continue acting during plant growth and development, which would be consistent with the fact that *hCas9* transcription is driven by a strong

constitutive promoter such as 35S. Similar observations were made by Mikami *et al.* (2015) in cultured rice *calli* transformed with CRISPR/Cas9 constructs, which were genotyped 1 and 2 months after transformation: extending the culture period led to an increase in the proportion of cells carrying edited alleles of the target gene (Fig. 34), and also in the number of edited plants which could be recovered from these *calli*.

gRNA	Cas9 and gRNA double-transformed line	Culture period after gRNA transformation									
		1 month					2 months				
		Mutation frequency in callus (%)	No. of regenerated plants (A)	No. of mutant plants (B)	Mutation frequency in plants (B/A) (%)	Mutant variations	Mutation frequency in callus (%)	No. of regenerated plants (A)	No. of mutant plants (B)	Mutation frequency in plants (B/A) (%)	Mutant variations
gDsRed-1	Cas9-#1_g1#5	9.6	26	2	7.6	−1, +1(A)	35.7	11	5	45.4	−1, + 1 (A,T)
gDsRed-3	Cas9-#3_g3#4	16.6	25	5	20.0	−1, −3	44.7	16	11	68.7	−1,−2, −3,−4

Figure 34. Increasing the duration of *callus* culture from 1 to 2 months increases the frequency of mutations at target sites and the number of mutant plants in rice. Mikami, *et al.* (2015).

Sun *et al.* (2015) performed gene editing of soybean hairy roots and found a variety of mutations in independent roots, indicating a continued activity of Cas9 during hairy root development. Repeating the genotyping of heterozygous or chimeric T_0 individuals which still carry a proportion of the wt allele does not seem to be a common practice in studies of CRISPR/Cas9 efficiency in plants. However, we think that, in addition to providing information about the efficiency and activity pattern of the system, this practice can be helpful for recovering a greater number of stably edited individuals and for selecting which T_1 progenies are useful to keep for further investigation.

4.5 Effect of *in vitro* culture on polyploidization

Induction of polyploidization can represent a major drawback for transformation and regeneration protocols, because of the aberrant phenotypes and reduced fertility of tetraploids. The evaluation of the ploidy level of our T_0 plants grown in the greenhouse revealed that 3 out of 5 were tetraploids (specifically, plants 3C, 4C and 10B). Spontaneous polyploidization in response to *in vitro* culture is known as a common type of somaclonal variation in different species (Chen *et al.* 2009; Jelenic *et al.*, 2001; Meyer *et al.* 2009). In particular, van den Bulk *et al.* (1990) reported that polyploidization occurs at high rates in regenerated shoots of tomato cv. 'MoneyMaker' deriving from leaf, cotyledon and hypocotyl explants, with frequencies above 50% for hypocotyl-derived plants. Ellul *et al.* (2003) analyzed the effect of

Agrobacterium-mediated transformation and regeneration on the ploidy level of tomato plants regenerated from culture of cotyledonary explants. Cotyledons are mixoploid tissues, with different ploidy levels (2C, 4C and 8C); in addition, during cell proliferation and differentiation from *callus*, endoreduplication and endomitosis can occur, which are able to induce the formation of polyploid individuals. The authors found that, testing two different cultivars of tomato and two regeneration protocols, a high proportion of polyploid plants, ranging from 24 to 80%, was recovered. The proportion of tetraploid individuals varied according to both genotype and regeneration method; the most efficient method was also the one that induced the highest proportion of tetraploids in cultivars which were more subject to polyploidization (Ellul *et al.,* 2003). These tetraploid plants, similarly to those we recovered in our *gf1* T_0, had reduced fertility and produced a lower number of seeds than diploids. An early screening of the ploidy level of regenerated plantlets thus constitutes an important improvement to assure that genotypes which are maintained for propagation and crossing are viable.

4.6 Analysis of off-target effects

The occurrence of off-target effects in plants was proven to be very low in a variety of different species (Gao *et al.*, 2015; Svitashev *et al.*, 2015, 2016; Pan *et al.*, 2016; Jacobs *et al.*, 2017; Nekrasov *et al.*, 2017), in contrast to data obtained in animal cell cultures which had pointed to off-target effects as one inconvenience of CRISPR/Cas9 editing (Bortesi *et al.*, 2016). In fact, in plants, the occurrence of off-target effects was estimated to represent a risk comparable to that of somaclonal variation deriving from tissue culture itself (Ma *et al.*, 2015). One of the factors known to increase off-target activity is a high expression level of gRNA and Cas9, which makes editing less specific (Pattanayak *et al.*, 2013). In the case of stable transformation through integration of a transgene, with the *hCas9* gene under the control of a constitutive promoter, it is not possible to control this parameter. The evaluation of off-target effects is usually carried out by identifying putative off-target sites across the genome and by analyzing them for mutations (Hendel *et al.*, 2015). Mutations can be identified either through sequencing (targeted sequencing of putative *loci*, or resequencing of the entire individual) or through other assays, like restriction assays. Different techniques imply a different level of sensitivity, with deep sequencing being the most informative kind of analysis. An off-target effect is established when the mutation is detectable above the background levels established for the control. In our case, no off-target effect was detectable

at any putative *locus,* in accordance with our expectations. Indeed, we did not expect any off-target effect for the *gf1* gRNA for a variety of reasons. First, the putative off-target *loci* had three or more mismatches with respect to the target sequence, and this significantly reduces the probability of unspecific recognition by the editing machinery (Hsu *et al.*, 2013; Kuscu *et al.*, 2014; Doench *et al.*, 2016). Second, in the majority of cases these mismatches are located in what is defined as the "seed region" of the target sequence, that is the 10-12 nucleotides directly at the 5' of the PAM, which are predicted to strongly affect binding and specificity. Finally, *gf1* does not belong to a conserved gene family, meaning that no highly conserved homologous sequences are found in the tomato genome. We can therefore conclude that the choice of our single *gf1* gRNA ensured both high editing efficiencies and high specificity.

4.7 Heritability and segregation in T_1 plants

When the CRISPR/Cas9 system was first applied to plants, concerns about the heritability of mutations were raised, due to the finding that in Arabidopsis many somatic mutations could be obtained, which were not found in the germ-line and were not efficiently inherited (Feng *et al.*, 2013). Heritability of edited alleles was improved in Arabidopsis through the substitution of constitutive promoters such as 35S which were used for *Cas9* expression, with germ-line, cell-division or egg-cell-specific promoters (Feng *et al.*, 2014; Bortesi *et al.*, 2016; Mao *et al.*, 2016). However, in all other edited Monocot and Dicot species high mutation efficiencies have been obtained for both the T_0 and the T_1 generations, pointing to high heritability of edited alleles (Miao *et al.*, 2013; Gao *et al.*, 2015; Li *et al.*, 2015; Svitashev *et al.*, 2015; Wang *et al.*, 2015; Pan *et al.*, 2016; Zhang *et al.*, 2016). We obtained Mendelian inheritance of mutations in the 2B and 12A *gf1* edited lines in the T_1 generation, confirming the heritability of edited alleles in tomato. We also confirmed that the T-DNA can be segregated and that Cas9-, transgene-free plants can be obtained in the course of one generation. Interestingly, the segregation frequency was markedly higher for the 12A line than for the 2B line. In the 12A progeny, in 8 individuals over a total of 12 plants it was not possible to detect the transgene, while in the 2B line, only 1 in 8 individuals appeared to have segregated the T-DNA. We can hypothesize than one of the reasons behind this different segregation rate might be the number of copies of the transgene which had been integrated in the plant genome; if, for example, the plant 12A had integrated only one copy of the T-DNA, while 2B had multiple copies, it would be easier and faster to eliminate it in the progeny.

4.8 Effect of *gf1* on phenotype and shelf life

The retention not only of chlorophyll, but also of turgor in senescent, detached *gf1* leaves, and the observed reduced susceptibility to fungi motivate the need for further investigation into this aspect of chlorophyll retention. The *sgr/gf1* mutations of tomato and pepper are classified as type C cosmetic staygreens and are not thought to be related to an increase in the shelf-life and delayed senescence of green tissues (Thomas & Ougham, 2014). In this type of mutants, despite a strongly reduced chlorophyll degradation rate in green tissues, other physiological aspects of senescence proceed normally, and photosynthetic rates are strongly reduced. However, interesting cues come from the work of Mecey *et al.* (2011), who identified staygreen mutations of the Arabidopsis *AtSGR1* gene as responsible for pathogen tolerance. *AtSGR1* mutants displayed reduced necrotic symptoms when infected both with a bacterial pathogen (*Pseudomonas syringae* pv. *lycopersici*) and with a necrotrophic fungal pathogen (*Alternaria brassicicola*). In both cases, the reduction of symptoms did not correspond to a reduced number of pathogen cells, meaning that mutants could support pathogen growth without developing necrotic symptoms. These results are promising in indicating a functional role for *sgr/gf1* mutants in disease tolerance.

In addition, chlorophyll content in tomato fruits was shown to be related to reduced oxidative stress and longer shelf life (Cocaliadis *et al.*, 2014). An increase in oxidative stress is a constitutive element of tomato fruit ripening, which is thought to be associated to metabolic changes leading to softening of fruit tissues. Tomato varieties with a short shelf life have high oxidative stress and low ROS scavenging activity in ripe fruits, while engineered tomato fruits with high levels of antioxidants have extended shelf life. Consistent with this, Roca *et al.* (2006) showed that in pepper the retention of chlorophyll positively correlated to an increase in the synthesis of carotenoids; in tomato, too, a high chloroplast content in green fruits leads to ripe tomatoes with more active chromoplasts (Cocaliadis *et al.*, 2014), which in turn accumulate more antioxidants and metabolites which positively affect organoleptic and nutritional quality. Inducing chlorophyll retention, even without maintaining photosynthetic activity, might thus have a functional, positive effect on fruit shelf life and reduce the susceptibility of leaves to necrotrophic and saprotrophic microorganisms.

5 - Conclusions

The GoldenBraid toolkit for CRISPR/Cas9 gene editing proved to be an efficient and time saving method to assemble a multi-TU construct carrying all the necessary elements of the gene editing machinery. An efficient regeneration protocol was available for tomato and editing efficiencies were very high, often close to 100%. This means that a mutation can be readily introduced in all progeny starting from the first generation. Analysis of off-target effects also showed that the system is highly specific and does not introduce off-target mutations at undesired sites in the genome. The phenotypes identified in the T_0 generation are consistent with the previously described *gf1* mutants, whose functional characteristics and applications to the improvement of fruit shelf life are to be further investigated.

Further phenotypic analyses will be carried out on the T_1 fruits and on the following generations. It is possible to say, however, that the mutation has been fixated and that, in some individuals, the T-DNA has been segregated, thus originating plants that are not transgenic. Overall, it can be concluded that we were able to successfully edit tomato 'MoneyMaker' plants to produce a chlorophyll retaining *gf1* mutant, and that targeting the *gf1 locus* in tomato provided a useful, readily recognizable visual system to assess the success of gene editing.

Bibliography

Akhtar, M. S. *et al.* (1999) 'Altered patterns of senescence and ripening in gf, a stay-green mutant of tomato (Lycopersicon esculentum Mill.)', *Journal of Experimental Botany*, 50(336), pp. 1115–1122. doi: 10.1093/jxb/50.336.1115.

Andersson, M. *et al.* (2017) 'Efficient targeted multiallelic mutagenesis in tetraploid potato (Solanum tuberosum) by transient CRISPR-Cas9 expression in protoplasts', *Plant Cell Reports*, 36, pp. 117–128. doi: 10.1007/s00299-016-2062-3.

Armstead, I. *et al.* (2007) 'Cross-species identification of Mendel's/locus', *Science*, 315(5808), p. 73. doi: 10.1126/science.1132912.

Barry, C. S. *et al.* (2008) 'Amino Acid Substitutions in Homologs of the STAY-GREEN Protein Are Responsible for the green-flesh and chlorophyll retainer Mutations of Tomato and Pepper', *Plant Physiology*, 147(1), pp. 179–187.

Barry, C. S. (2009) 'The stay-green revolution: Recent progress in deciphering the mechanisms of chlorophyll degradation in higher plants', *Plant Science*, 176(3), pp. 325–333. doi: 10.1016/j.plantsci.2008.12.013.

Barry, C. S. and Pandey, P. (2009) 'A survey of cultivated heirloom tomato varieties identifies four new mutant alleles at the green-flesh locus', *Molecular Breeding*, 24(3), pp. 269–276. doi: 10.1007/s11032-009-9289-4.

Bhatia, P. and Ashwath, N. (2005) 'Effect of duration of light: dark cycles on in vitro shoot regeneration of tomato', *Asian Journal of Plant Sciences*, pp. 255–260.

Borovsky, Y. and Paran, I. (2008) 'Chlorophyll breakdown during pepper fruit ripening in the chlorophyll retainer mutation is impaired at the homolog of the senescence-inducible stay-green gene', *Theoretical and Applied Genetics*, 117(2), pp. 235–240. doi: 10.1007/s00122-008-0768-5.

Bortesi, L. *et al.* (2016) 'Patterns of CRISPR/Cas9 activity in plants, animals and microbes', *Plant Biotechnology Journal*, 14(12), pp. 2203–2216. doi: 10.1111/pbi.12634.

Brinkman, E. K. *et al.* (2014) 'Easy quantitative assessment of genome editing by sequence trace decomposition', *Nucleic Acids Research*, (doi: 10.1093/nar/gku936). doi: 10.1093/nar/gku936.

Brooks, C. *et al.* (2014) 'Effi cient Gene Editing in Tomato in the First Generation Using the Clustered Regularly Interspaced Short Palindromic Repeats / CRISPR-Associated9 System', *Plant Physiology*, 166(November), pp. 1292–1297. doi: 10.1104/pp.114.247577.

Brooks, C. *et al.* (2014) 'Efficient Gene Editing in Tomato in the First Generation Using the Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-Associated9 System', *Plant Physiology*, 166(3), pp. 1292–1297. doi: 10.1104/pp.114.247577.

van den Bulk, R. W. *et al.* (1990) 'Somaclonal variation in tomato: effect of explant source and a comparison with chemical mutagenesis', *Theoretical and Applied Genetics*, 80, pp. 817–825. doi: 10.1007/BF00224199.

Chaudhry, Z. *et al.* (2010) 'Tissue Culture Studies in Tomato (Lycopersicon Esculentum) Var. Moneymaker', *Pak J Bot*, 42(1), pp. 155–163.

Chen, W. H., Tang, C. Y. and Kao, Y. L. (2009) 'Ploidy doubling by in vitro culture of excised protocorms or protocorm-like bodies in Phalaenopsis species', *Plant Cell, Tissue and Organ Culture*, 98, pp. 229–238. doi: 10.1007/s11240-009-9557-3.

Cheung, a. Y., McNellis, T. and Piekos, B. (1993) 'Maintenance of Chloroplast Components during Chromoplast Differentiation in the Tomato Mutant Green Flesh.', *Plant physiology*, 101, pp. 1223–1229. doi: 10.1104/pp.101.4.1223.

Chidgey, J. W. *et al.* (2014) 'A Cyanobacterial Chlorophyll Synthase-HliD Complex Associates with the Ycf39 Protein and the YidC/Alb3 Insertase', *The Plant Cell*, 26(3), pp. 1267–1279. doi:

10.1105/tpc.114.124495.

Christ, B. and Hörtensteiner, S. (2014) 'Mechanism and Significance of Chlorophyll Breakdown', *Journal of Plant Growth Regulation*, 33(1), pp. 4–20. doi: 10.1007/s00344-013-9392-y.

Cocaliadis, M. F. *et al.* (2014) 'Increasing tomato fruit quality by enhancing fruit chloroplast function. A double-edged sword?', *Journal of Experimental Botany*, 65(16), pp. 4589–4598. doi: 10.1093/jxb/eru165.

Doench, J. G. *et al.* (2016) 'Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9', *Nature Biotechnology*. Nature Publishing Group, 34(2), pp. 184–191. doi: 10.1038/nbt.3437.

Doyle, J. J. (1987) 'A rapid DNA isolation procedure for small quantities of fresh leaf tissue', *Phytochem. Bull.*, 19, pp. 11–15. Available at: https://ci.nii.ac.jp/naid/10011287509/en/.

Efrati, A., Eyal, Y. and Paran, I. (2005) 'Molecular mapping of the *chlorophyll retainer* (*cl*) mutation in pepper (*Capsicum* spp.) and screening for candidate genes using tomato ESTs homologous to structural genes of the chlorophyll catabolism pathway', *Genome*, 48(2), pp. 347–351. doi: 10.1139/g04-119.

Eid, A., Ali, Z. and Mahfouz, M. M. (2016) 'High efficiency of targeted mutagenesis in arabidopsis via meiotic promoter-driven expression of Cas9 endonuclease', *Plant Cell Reports*. Springer Berlin Heidelberg, 35(7), pp. 1555–1558. doi: 10.1007/s00299-016-2000-4.

Ellul, P. *et al.* (2003) 'The ploidy level of transgenic plants in Agrobacterium-mediated transformation of tomato cotyledons (Lycopersicon esculentum L . Mill .) is genotype and procedure dependent', *Theoretical and Applied Genetics*, 106, pp. 231–238. doi: 10.1007/s00122-002-0928-y.

Feng, Z. *et al.* (2013) 'Efficient genome editing in plants using a CRISPR/Cas system', *Cell Research*, 23(10), pp. 1229–1232. doi: 10.1038/cr.2013.114.

Feng, Z. *et al.* (2014) 'Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis', *Proceedings of the National Academy of Sciences*, 111(12), pp. 4632–4637. doi: 10.1073/pnas.1400822111.

Fukao, T., Yeung, E. and Bailey-Serres, J. (2012) 'The Submergence Tolerance Gene SUB1A Delays Leaf Senescence under Prolonged Darkness through Hormonal Regulation in Rice', *Plant Physiology*, 160(4), pp. 1795–1807. doi: 10.1104/pp.112.207738.

Gao, J. *et al.* (2015) 'CRISPR / Cas9 ‑ mediated targeted mutagenesis in *Nicotiana tabacum*', *Plant Molecular Biology*, 87, pp. 99–110. doi: 10.1007/s11103-014-0263-0.

Gerszberg, A. *et al.* (2015) 'Tomato (Solanum lycopersicum L.) in the service of biotechnology', *Plant Cell, Tissue and Organ Culture*, 120(3), pp. 881–902. doi: 10.1007/s11240-014-0664-4.

Grassl, J. *et al.* (2012) 'Early events in plastid protein degradation in stay-green arabidopsis reveal differential regulation beyond the retention of LHCII and chlorophyll', *Journal of Proteome Research*, 11(11), pp. 5443–5452. doi: 10.1021/pr300691k.

Gupta, S. and Van Eck, J. (2016) 'Modification of plant regeneration medium decreases the time for recovery of Solanum lycopersicum cultivar M82 stable transgenic lines', *Plant Cell, Tissue and Organ Culture*. Springer Netherlands, 127(2), pp. 417–423. doi: 10.1007/s11240-016-1063-9.

Hauck, B. *et al.* (1997) 'Leaf senescence in a non-yellowing mutant of *Festuca pratensis*: implications of the stay-green mutation for photosynthesis, growth and nitrogen nutrition', *Plant, Cell and Environment*, 20(1987), pp. 1007–1018. doi: 10.1111/j.1365-3040.1997.tb00677.x.

Hauenstein, M. *et al.* (2016) 'A Role for TIC55 as a Hydroxylase of Phyllobilins, the Products of Chlorophyll Breakdown during Plant Senescence', *The Plant Cell*, 28(10), pp. 2510–2527. doi: 10.1105/tpc.16.00630.

Hendel, A. *et al.* (2015) 'Quantifying on- and off-target genome editing', *Trends in Biotechnology*. Elsevier Ltd, 33(2), pp. 132–140. doi: 10.1016/j.tibtech.2014.12.001.

Hörtensteiner, S. (2006) 'Chlorophyll Degradation During Senescence', *Annual Review of Plant Biology*, 57(1), pp. 55–77. doi: 10.1146/annurev.arplant.57.032905.105212.

Hörtensteiner, S. (2009) 'Stay-green regulates chlorophyll and chlorophyll-binding protein degradation during senescence', *Trends in Plant Science*, 14(3), pp. 155–162. doi: 10.1016/j.tplants.2009.01.002.

Hörtensteiner, S. and Kräutler, B. (2011) 'Chlorophyll breakdown in higher plants', *Biochimica et Biophysica Acta - Bioenergetics*. Elsevier B.V., 1807(8), pp. 977–988. doi: 10.1016/j.bbabio.2010.12.007.

Hörtensteiner, S. (2013) 'Update on the biochemistry of chlorophyll breakdown', *Plant Molecular Biology*, 82(6), pp. 505–517. doi: 10.1007/s11103-012-9940-z.

Hsu, P. D. *et al.* (2013) 'DNA targeting specificity of RNA-guided Cas9 nucleases', *Nature Biotechnology*. Nature Publishing Group, 31(9), pp. 827–832. doi: 10.1038/nbt.2647.

Hu, W. and Phillips, G. C. (2001) 'A combination of overgrowth-control antibiotics improves Agrobacterium tumefaciens-mediated transformation efficiency for cultivated tomato (L. esculentum)', *In Vitro Cellular & Developmental Biology-Plant*, 37(1), pp. 12–18. doi: 10.1079/IVP2000130.

Hu, Z. L. *et al.* (2011) 'Silencing of the LeSGR1 gene in tomato inhibits chlorophyll degradation and exhibits a stay-green phenotype', *Biologia Plantarum*, 55(1), pp. 27–34. doi: 10.1007/s10535-011-0004-z.

Ito, Y. *et al.* (2015) 'CRISPR/Cas9-mediated mutagenesis of the RIN locus that regulates tomato fruit ripening', *Biochemical and Biophysical Research Communications*. Elsevier Ltd, 467(1), pp. 76–82. doi: 10.1016/j.bbrc.2015.09.117.

Jacobs, T. B. *et al.* (2017) 'Generation of a collection of mutant tomato lines using pooled CRISPR libraries', *Plant Physiology*, 174(4), pp. 2023–2037. doi: 10.1104/pp.17.00489.

Jelenic, S. *et al.* (2001) 'Mixoploidy and Chimeric Structures in Somaclones of Potato (Solanum tuberosum L .) cv . Bintje', *Food technol. biotechnol.*, 39(1), pp. 13–17.

Jiang, H. *et al.* (2007) 'Molecular cloning and function analysis of the stay green gene in rice', *Plant Journal*, 52(2), pp. 197–209. doi: 10.1111/j.1365-313X.2007.03221.x.

Jordan, D. R. *et al.* (2012) 'The relationship between the stay-green trait and grain yield in elite sorghum hybrids grown in a range of environments', *Crop Science*, 52(3), pp. 1153–1161. doi: 10.2135/cropsci2011.06.0326.

Kassahun, B. *et al.* (2010) 'Stay-green expression in early generation sorghum [Sorghum bicolor (L.) Moench] QTL introgression lines', *Euphytica*, 172(3), pp. 351–362. doi: 10.1007/s10681-009-0108-0.

Kerr EA. 1956. Green flesh, gf. *Tomato Genetics Cooperative Reports* 6, 17.

Khuong, T. T. H. *et al.* (2013) 'Optimisation of tomato Micro-tom regeneration and selection on glufosinate/Basta and dependency of gene silencing on transgene copy number', *Plant Cell Reports*, 32(9), pp. 1441–1454. doi: 10.1007/s00299-013-1456-8.

Klap, C. *et al.* (2017) 'Tomato facultative parthenocarpy results from Sl AGAMOUS-LIKE 6 loss of function', *Plant Biotechnology Journal*, 15, pp. 634–647. doi: 10.1111/pbi.12662.

Kuai, B., Chen, J. and Hörtensteiner, S. (2017) 'The biochemistry and molecular biology of chlorophyll breakdown', *Journal of Experimental Botany*, (December). doi: 10.1093/jxb/erx322.

Kuscu, C. *et al.* (2014) 'Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease', *Nature Biotechnology*. Nature Publishing Group, 32(7), pp. 677–683. doi: 10.1038/nbt.2916.

Li, Z. *et al.* (2015) 'Cas9-Guide RNA Directed Genome Editing in Soybean', *Plant Physiology*, 169(2), pp. 960–970. doi: 10.1104/pp.15.00783.

Luo, Z. *et al.* (2013) 'A STAY-GREEN protein SlSGR1 regulates lycopene and β-carotene accumulation by interacting directly with SlPSY1 during ripening processes in tomato', *New Phytologist*, 198(2), pp. 442–452. doi: 10.1111/nph.12175.

Ma, X. *et al.* (2015) 'A Robust CRISPR/Cas9 System for Convenient, High-Efficiency Multiplex Genome Editing in Monocot and Dicot Plants', *Molecular Plant*. Elsevier Ltd, 8(8), pp. 1274–1284. doi: 10.1016/j.molp.2015.04.007.

Mamidala, P. and Nanna, R. S. (2009) 'Influence of antibiotics on regeneration efficiency in tomato', *Plant Omics Journal*, 2(4), pp. 135–140.

Mamidala, P. and Nanna, R. S. (2011) 'Effect of genotype, explant source and medium on in vitro regeneration of tomato', *International Journal of Genetics and Molecular Biology*, 3(3), pp. 45–50.

Mao, Y. *et al.* (2016) 'Development of germ-line-specific CRISPR-Cas9 systems to improve the production of heritable gene modifications in Arabidopsis', *Plant Biotechnology Journal*, 14(2), pp. 519–532. doi: 10.1111/pbi.12468.

Masuda, T. and Fujita, Y. (2008) 'Regulation and evolution of chlorophyll metabolism', *Photochemical & Photobiological Sciences*, 7(10), p. 1131. doi: 10.1039/b807210h.

McCormick, S. *et al.* (1986) 'Leaf disc transformation of cultivated tomato (L. esculentum) using Agrobacterium tumefaciens', *Plant Cell Reports*, 5(2), pp. 81–84. doi: 10.1007/BF00269239.

Mecey, C. *et al.* (2011) 'A Critical Role of STAYGREEN/Mendel's I Locus in Controlling Disease Symptom Development during *Pseudomonas syringae* pv *tomato* Infection of Arabidopsis', *Plant Physiology*, 157(4), pp. 1965–1974. doi: 10.1104/pp.111.181826.

Meyer, E. M., Touchell, D. H. and Ranney, T. G. (2009) 'In Vitro Shoot Regeneration and Polyploid Induction from Leaves of Hypericum Species', *HortScience*, 44(7), pp. 1957–1961.

Miao, J. *et al.* (2013) 'Targeted mutagenesis in rice using CRISPR-Cas system', *Cell Research*, 23(10), pp. 1233–1236. doi: 10.1038/cr.2013.123.

Mikami, M., Toki, S. and Endo, M. (2015) 'Parameters affecting frequency of CRISPR/Cas9 mediated targeted mutagenesis in rice', *Plant Cell Reports*. Springer Berlin Heidelberg, 34(10), pp. 1807–1815. doi: 10.1007/s00299-015-1826-5.

Moghaieb, R. E. A., Saneoka, H. and Fujita, K. (1999) 'Plant regeneration from hypocotyl and cotyledon explant of tomato (Lycopersicon esculentum Mill.)', *Soil Science and Plant Nutrition*, 45(3), pp. 639–646. doi: 10.1080/00380768.1999.10415827.

Moore, B. J. *et al.* (2005) 'Molecular tagging of a senescence gene by introgression mapping of a stay-green mutation from Festuca pratensis', *New Phytologist*, 165(3), pp. 801–806. doi: 10.1111/j.1469-8137.2004.01269.x.

Nekrasov, V. *et al.* (2017) 'Rapid generation of a transgene- free powdery mildew resistant tomato by genome deletion', *Scientific Reports*. Springer US, (February), pp. 1–6. doi: 10.1038/s41598-017-00578-x.

Pan, C. *et al.* (2016) 'CRISPR/Cas9-mediated efficient and heritable targeted mutagenesis in tomato plants in the first and later generations', *Scientific Reports*. Nature Publishing Group, 6(1), p. 24765. doi: 10.1038/srep24765.

Park, S.-Y. *et al.* (2007) 'The Senescence-Induced Staygreen Protein Regulates Chlorophyll Degradation', *the Plant Cell Online*, 19(5), pp. 1649–1664. doi: 10.1105/tpc.106.044891.

Pattanayak, V. *et al.* (2013) 'High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity', *Nature Biotechnology*. Nature Publishing Group, 31(9), pp. 839–843. doi: 10.1038/nbt.2673.

Pruzinska, A. *et al.* (2007) 'In Vivo Participation of Red Chlorophyll Catabolite Reductase in Chlorophyll Breakdown', *the Plant Cell Online*, 19(1), pp. 369–387. doi: 10.1105/tpc.106.044404.

Qiu, D. *et al.* (2007) ‘Improved protocol for Agrobacterium mediated transformation of tomato and production of transgenic plants containing carotenoid biosynthetic gene CsZCD’, *Scientia Horticolturae*, 112, pp. 172–175. doi: 10.1016/j.scienta.2006.12.015.

Ramirez, D. A. and Tomes, M. L. (1964) ‘Relationship between Chlorophyll and Carotenoid Biosynthesis in Dirty-Red (Green-Flesh) Mutant in Tomato’, *Botanical Gazette*. The University of Chicago Press, 125(3), pp. 221–226. doi: 10.1086/336275.

Rivero, R. M. *et al.* (2007) ‘Delayed leaf senescence induces extreme drought tolerance in a flowering plant’, *Proceedings of the National Academy of Sciences*, 104(49), pp. 19631–19636. doi: 10.1073/pnas.0709453104.

Roca, M. *et al.* (2006) ‘Stay-green phenotype slows the carotenogenic process in Capsicum annuum (L.) fruits’, *Journal of Agricultural and Food Chemistry*, 54(23), pp. 8782–8787. doi: 10.1021/jf062007r.

Roca, M. and Mínguez-Mosquera, M. I. (2006) ‘Chlorophyll catabolism pathway in fruits of Capsicum annuum (L.): Stay-green versus red fruits’, *Journal of Agricultural and Food Chemistry*, 54(11), pp. 4035–4040. doi: 10.1021/jf060213t.

Rong, H. *et al.* (2013) ‘The Stay-Green Rice like (SGRL) gene regulates chlorophyll degradation in rice’, *Journal of Plant Physiology*. Elsevier GmbH., 170(15), pp. 1367–1373. doi: 10.1016/j.jplph.2013.05.016.

Shimoda, Y., Ito, H. and Tanaka, A. (2016) ‘Arabidopsis *STAY-GREEN* , Mendel’s Green Cotyledon Gene, Encodes Magnesium-Dechelatase’, *The Plant Cell*, 28(9), pp. 2147–2160. doi: 10.1105/tpc.16.00428.

Sun, H. J. *et al.* (2006) ‘A highly efficient transformation protocol for Micro-Tom, a model cultivar for tomato functional genomics’, *Plant and Cell Physiology*, 47(3), pp. 426–431. doi: 10.1093/pcp/pci251.

Sun, X. *et al.* (2015) ‘Targeted mutagenesis in soybean using the CRISPR-Cas9 system’, *Scientific Reports*. Nature Publishing Group, 5, pp. 1–10. doi: 10.1038/srep10342.

Svitashev, S. *et al.* (2015) ‘Targeted Mutagenesis, Precise Gene Editing, and Site-Specific Gene Insertion in Maize Using Cas9 and Guide RNA’, *Plant Physiology*, 169(2), pp. 931–945. doi: 10.1104/pp.15.00793.

Svitashev, S. *et al.* (2016) ‘Genome editing in maize directed by CRISPR–Cas9 ribonucleoprotein complexes’, *Nature Communications*. Nature Publishing Group, 7, p. 13274. doi: 10.1038/ncomms13274.

Tanaka, A. and Tanaka, R. (2006) ‘Chlorophyll metabolism’, *Current Opinion in Plant Biology*, 9(3), pp. 248–255. doi: 10.1016/j.pbi.2006.03.011.

Thomas, H. and Howarth, C. J. (2000) ‘Five ways to stay green’, *Journal of experimental botany*, 51(February), pp. 329–337. doi: 10.1093/jexbot/51.suppl_1.329.

Thomas, H. and Ougham, H. (2014) ‘The stay-green trait’, *Journal of Experimental Botany*, 65(14), pp. 3889–3900. doi: 10.1093/jxb/eru037.

Ueta, R. *et al.* (2017) ‘Rapid breeding of parthenocarpic tomato plants using CRISPR / Cas9’, *Scientific Reports*. Springer US, 7(507), pp. 1–8. doi: 10.1038/s41598-017-00501-4.

Vicentini, F. *et al.* (1995) ‘Chlorophyll breakdown in senescent leaves identification of the biochemical lesion in a stay-green genotype of Festuca pratensis Huds.’, *New Phytologist*, 129(2), pp. 247–252. doi: 10.1111/j.1469-8137.1995.tb04294.x.

Wang, S. *et al.* (2015) ‘Efficient targeted mutagenesis in potato by the CRISPR / Cas9 system’, *Plant Cell Reports*. Springer Berlin Heidelberg, 34(9), pp. 1473–1476. doi: 10.1007/s00299-015-1816-7.

Wang, Y. *et al.* (2014) ‘Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew’, *Nature Biotechnology*, 32(9), pp. 947–951. doi:

10.1038/nbt.2969.

Yasmeen, A. (2009) ‘An improved protocol for the regeneration and transformation of tomato (cv Rio Grande)’, *Acta Physiologiae Plantarum*, 31(6), pp. 1271–1277. doi: 10.1007/s11738-009-0364-6.

Yoo, S.-C. *et al.* (2007) ‘Quantitative trait loci associated with functional stay-green SNU-SG1 in rice.’, *Molecules and cells*, 24(1), pp. 83–94.

Zhang, B. *et al.* (2016) ‘Exploiting the CRISPR / Cas9 System for Targeted Genome Mutagenesis in Petunia’, *Scientific Reports*. Nature Publishing Group, 6(February), pp. 1–8. doi: 10.1038/srep20315.

Zhang, H. *et al.* (2014) ‘The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation’, *Plant Biotechnology Journal*, 12(6), pp. 797–807. doi: 10.1111/pbi.12200.

Zhou, C. *et al.* (2011) ‘From Model to Crop: Functional Analysis of a STAY-GREEN Gene in the Model Legume Medicago truncatula and Effective Use of the Gene for Alfalfa Improvement’, *Plant Physiology*, 157(3), pp. 1483–1496. doi: 10.1104/pp.111.185140.

Component	12 µl reaction	Final concentration
pDGB3 alpha1 destination vector	1 µl	6.25 ng μl^{-1}
GB1001 pU6-26 PolIII promoter	1 µl	6.25 ng μl^{-1}
gRNA 2 µM	1 µl	0.17 µM
GB0645 scaffold RNA	1 µl	6.25 ng μl^{-1}
10X T4 ligase buffer	1.2 µl	1X
10X BSA	1.2 µl	1X
T4 ligase	1 µl	3 U
*Bsa*I	1 µl	10 U
Water	3.6	N/A

Supplementary Table S1. First GoldenBraid reaction for the assembly of a CRISPR-Cas vector to target the *gf1 locus*. Assembly of the pDGB3 alpha1 vector carrying the U6 promoter, the gRNA and the scaffold RNA.

Component	12 µl reaction	Final concentration
pDGB3 omega2 destination vector	1 µl	6.25 ng μl^{-1}
pDGB3 alpha1 U6-26:gRNA: scaffold	1 µl	6.25 ng μl^{-1}
GB0639 35S:hCas9: Tnos	1 µl	6.25 ng μl^{-1}
10X T4 ligase buffer	1.2 µl	1X
10X BSA	1.2 µl	1X
T4 ligase	1 µl	3 U
*Bsmb*I	1 µl	10 U
Water	4.6	N/A

Supplementary Table S2. Second GoldenBraid reaction for the assembly of a CRISPR-Cas vector to target the *gf1 locus*. Assembly of the pDGB3 omega2 vector carrying the insert of the previously assembled pDGB3 alpha1 U6:gRNA:scaffold and the hCas9 TU with the 35S promoter and Tnos terminator.

Component	12 µl reaction	Final concentration
pDGB3 alpha2 destination vector	1 µl	6.25 ng µl^{-1}
pDGB3 omega2 U6-26:gRNA: scaffold - 35S:hCas9:Tnos	1 µl	6.25 ng µl^{-1}
GB1181 alpha1R Pnos:nptII:Tnos	1 µl	6.25 ng µl^{-1}
10X T4 ligase buffer	1.2 µl	1X
10X BSA	1.2 µl	1X
T4 ligase	1 µl	3 U
*Bsa*I	1 µl	10 U

Supplementary Table S3. Third GoldenBraid reaction for the assembly of a CRISPR-Cas vector to target the *gf1 locus*. Assembly of the final pDGB3 alpha2 vector carrying the gRNA and hCas9 TUs and the NptII selection marker with the Pnos promoter and Tnos terminator.

Compound	MGL *medium* pH 7	TY *medium* pH 5.8
Tryptone	5 g l^{-1}	5 g l^{-1}
Yeast extract	2.5 g l^{-1}	3 g l^{-1}
NaCl	5 g l^{-1}	-
Mannitol	5 g l^{-1}	-
Glutamic acid	1.02 g l^{-1}	-
K_2HPO_4	0.25 g l^{-1}	-
$MgSO_4.7H_2O$	0.1 g l^{-1}	0.5 g l^{-1} (2 mM)
Biotin	1 mg l^{-1}	-
Autoclave		
Acetosyringone	-	200 µM
Kanamycin/Spectinomycin	50 mg l^{-1}	-
Rifampicin	50 mg l^{-1}	-

Supplementary Table S4. Composition of MGL and TY *media* used for *Agrobacterium* growth.

	Co-culture ***medium***	**Induction** ***medium***	**Elongation** ***medium***	**Rooting** ***medium***
MS vitamins	5 g l^{-1}	5 g l^{-1}	5 g l^{-1}	2.5 g l^{-1}
Sucrose	30 g l^{-1}	30 g l^{-1}	30 g l^{-1}	10 g l^{-1}
Plant agar	6 g l^{-1}	6 g l^{-1}	6 g l^{-1}	6 g l^{-1}
pH 5.8, Autoclave				
IAA	0.1 mg l^{-1}	0.1 mg l^{-1}	0.1 mg l^{-1}	-
IBA	-	-	-	0.1 mg l^{-1}
ZR	2 mg l^{-1}	2 mg l^{-1}	2 mg l^{-1}	-
Acetosyringone	200 µM	-	-	-
Carbenicillin	-	500 mg l^{-1}	500 mg l^{-1}	500 mg l^{-1}
Kanamycin	-	100 mg l^{-1}	100 mg l^{-1}	30 mg l^{-1}

Supplementary Table S5. *Media* composition for co-culture, induction of *callus* and shoot formation, elongation and rooting. IAA = indoleacetic acid, IBA=indolebutyrric acid, ZR=zeatin riboside.

Locus	Primer Forward	Primer Reverse
OT 1	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**GATGTGAAGATTGTAGTTTG	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**GCACAGTAAATAGGGTGAAG
OT 2	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**CAAAAGGGCATTCGAACT	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**TATACAACGCCAGAGTTTCAG
OT 3	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**TTCTGACAACCCCTTTTTC	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**TGTTATCAGCAGCAATACCA
OT 4	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**GTATAATACTCTTCTAGTGTACCA	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**GTCATACCCAAGATGGTA
OT 5	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**CTCCGCCTTTTTTCATCAAATCC	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**TTTGGTTGGCGAATGGCA

Supplementary Table S6. Primer combinations for the first round amplification for 5 potential off-target *loci* for the *gf1* gRNA. In bold is the universal Illumina adapter, while the 3' terminal part of the primer is specific for each *locus*.

Individual	Nextera Code	Index Sequence	Oligonucleotide Sequence
1	S502	CTCTCTAT	AATGATACGGCGACCACCGAGATCTACAC**CTCTCTAT**TCGTCGGCAGCGTC
2	S503	TATCCTCT	AATGATACGGCGACCACCGAGATCTACAC**TATCCTCT**TCGTCGGCAGCGTC
3	S505	GTAAGGAG	AATGATACGGCGACCACCGAGATCTACAC**GTAAGGAG**TCGTCGGCAGCGTC
4	S506	ACTGCATA	AATGATACGGCGACCACCGAGATCTACAC**ACTGCATA**TCGTCGGCAGCGTC
5	S507	AAGGAGTA	AATGATACGGCGACCACCGAGATCTACAC**AAGGAGTA**TCGTCGGCAGCGTC
WT	S508	CTAAGCCT	AATGATACGGCGACCACCGAGATCTACAC**CTAAGCCT**TCGTCGGCAGCGTC
Locus	**Nextera Code**	**Index Sequence**	**Oligonucleotide Sequence**
OT 1	N701	TAAGGCGA	CAAGCAGAAGACGGCATACGAGAT**TCGCCTTA**GTCTCGTGGGCTCGG
OT 2	N702	CGTACTAG	CAAGCAGAAGACGGCATACGAGAT**CTAGTACG**GTCTCGTGGGCTCGG
OT3	N703	AGGCAGAA	CAAGCAGAAGACGGCATACGAGAT**TTCTGCCT**GTCTCGTGGGCTCGG
OT 4	N704	TCCTGAGC	CAAGCAGAAGACGGCATACGAGAT**GCTCAGGA**GTCTCGTGGGCTCGG
OT 5	N705	GGACTCCT	CAAGCAGAAGACGGCATACGAGAT**AGGAGTCC**GTCTCGTGGGCTCGG

Supplementary Table S7. Primers for the second round amplification for 5 potential off-target *loci* for the *gf1* gRNA. One series of Nextera indexes (S502-S508) is used to identify individuals, while the other (N701-N705) is used to identify *loci*. The combination of these two classes of primers allows to unambiguously label a particular PCR product from a particular individual.

CTAB EXTRACTION PROTOCOL FOR GENOMIC DNA

- Ground ~100 mg of frozen leaf tissue using the Qiagen TissueLyser II apparatus.
- Heat the CTAB buffer to 65°C in a water bath and supplement it with 2% 2-mercaptoethanol and 40 µg ml^{-1} RNAse.
- Add 700 µl of CTAB buffer to each sample and vortex.
- Keep the samples in a heaten water bath (65°C) for 30-45 minutes.
- Add 600 µl chloroform:isoamyl alcohol 24:1 and vortex.
- Centrifuge for 15 minutes at 13,000 rpm.
- Recover the aqueous phase and add 1 volume isopropanol. Mix by inverting the tubes.
- Keep the samples at -80°C for 20 minutes.
- Centrifuge for 10 minutes at 13,000 rpm at 4°C.
- Gently remove and discard the supernatant.
- Wash the pellet with 500 µl 80% ethanol, vortex and centrifuge for 5 minutes at 13,000 rpm.
- Discard the ethanol and let the pellet dry.
- Resuspend the pellet in 50-100 µl water by pipetting and vortexing.
- Store at -20°C.

Box S1. CTAB protocol for extraction of genomic DNA from plant leaf tissue.

Supplementary Sequence List

>Off-target 1 GF1
TGTTCATGTATGACTTCTCTATTGTAGGAGACTCCTTTGATCGTTGTTTAGACCATCTA
GCCGAGGTACTCAAAGATGTGAAGATTGTAGTTTGATACTAAATTGAGAGAAACGAC
GCTTCATGGTGAAAGAAGGTATTGTCTTGGGTCATCCAATCTTGAGAAATGGAATTG
AGGTTAAGGCCAAAGTTGAGTTTATTGAAAGGCTTACACTTCCCATCATTGTCGCTCC
AGAATGAGGGTAATCATTTCAAGTCATTGCGATATTAGTGGGGGTAAGGTTTTGGGG
CAAAGGCGTGAGAAGATACTTCACCCTATTTACTGTGCAAGCATATCAATTAATTCA
GCCCAAAAGAAAAACACAACCACCAAAAGAGAGCTTCTTTTGTGATTTTTGTGTTC
GAGATATTTTGCTCTTATTTGAGTGGGACTGTAGTCCTAGTGCACACAAATCATTTGG
TGTTAAGTTATTTGATAGAAAAAAGAGGATGTAAAACCACAGATTGTTAATTTGTGT
TGCTGTTCCAAGAATTTTATTTTGCTATTAAAGACTTCAAGGGGACAAAGAAACAAG
TCGCCAATCACTTGTCTAGATTGGAGGATGAACCTTTACTCGAGCTTGGTTAAAAGG
GTGAGATCAATAATATGTTT

>Off-target 2 GF1
AAACTAGCGACAAATAAAGGCAAGTAAACATACAGAAGCAAGAGAGAGATAGAGA
TAGAGAGAGACGTTGGGCCAAATCGGATCATGTTCGCCTAGACCAGTATTTGAGTCG
AATTTATTTTAGGTCTTAGGTCCATTTGGTTTATACGTGTCCTAGATAATCTTTCAATT
CGGAGAAAATCTGAGTCCATCTCTCCCTTATTCTCTGATTGTCCATATTCGCTCAATA
CAACATGTCATTCAAATCTTTAAGCGTGTCCTTGTTGTCAAAAATATACAACGCCAGA
GTTTCAGCACCCTTTGCCTAAATGGCTAAGGATCCACCACTAATGGAGCAATTTGCAT
CCTTGGAAGTATCGATGCCCCTCTTTGCATTTTCCCAGGGCTCGATACATATCAGACA
GGATGACCGGAGCTATAGGATAGTATTTTTTTTGTCCCTTGGTTCTTATACCTTTTGAA
CAAACTATAGGATAGGCAACGATATTCTTGTAAAAGTGAAAAAAGTTCGAATGCCCT
TTTGGTTAAAAAATAAATTGATGTATCCCATTGAAATTTTGACAATTATTTTCGCTCTT
AATACCTTTGGACTCCAAATATATATATATATATATAGGTGACACAATGAGCACAAT
CAGGCTAACTTTATTTTACCCTTTAATATATAAGTTCAGCAAAGCTACATACAGAAAA
AAAGAAA

>Off-target 3 GF1
GAGATTGTCATCAACTTTGTCCGTGATATGTTGCCATTTGTAGACCAGTTGAGTTGCA
TTTTGTTTCAATGTTCGTCCCACATTGAACACGGTTACCTTTTCTGGAATATCAACTAA
TTTTATTTTCCATGAGATAATGAGACCAAAACTAGTCCCTCCACCTCCTCTAATAGCC
CAAAACAGATCCTCACCCATTGATTCTCGATCCTGAATCTGTCCATTAGCATCAATTA
ATTTTGCATCAATGATGTTATCAGCAGCAATACCAAATTTTCTTGACATCATGCTATA
GCCTCCACCACTAATGTGTCCACCAACACCAACACTAGGGCAAGCCCCAGCAACAAA
GGCCAATTTATTGTTACTTTTTTCCGCGATTCTATAGTAAACTTCCCCTAAAGTCGCGC
CAGCTTGAACCCAAGCAGTCTTGTTTTTTATATCAATGGAGATTGATCTTAGGTTTCT
AAGATCAATCAGGAAAAAGGGGTTGTCAGAAATATATGAAAGTCCCTCATAGTCATG
TCCACCGCTTCGAATTCTAATTTGTAGGTCATGTAACTTGGAGCAATATATAGCTACC
TGGATTTGAGATTCTTTGGTAGGAGTGAAAATAATTGATGGTTTGAAGTTGGAGGTT
ATCCTTAAGTTATGATCTAAAAAGGAGTTATAAATGGTTGAATATGATGAGTTTTTAG
GAGTGTG

>Off-target 4 GF1
GGCAGGAGCTGGGGCGCAGGGGAATCTTTAAGGTAAATGGCGTAAAATACCTTCCGC
CAAGGTGCTGCCACTGAAGTCAAGGTCCAAAGACATGCTTTCTAGAGTGTCATTAAC
CACCCTGAATAAACCAAACCAAAGTAGTTTTAAAAATAAAAATAGAAAATGGTCAA
CAAGGTCAATGTTTAACCAACTAAAAAGGTAGGCTTTATGCAAGAAACCAAATTTGA
AAGTTCAGTCATACCCAAGATGGTAGCCAAACAAGAGTGATGAAATAATTGCCACCA
GTATATGTGGCAACGGCCGCTTCCATGAGGGGTTTCCAATTTCTTTCCACACACCATT
TTCGGCACGATCTAAATAGACACAAGGAAGAAGAATCAAAAGTCAATTTACACAAA
ATTATATCCTTTTATGTTATGTCCTATTGTTGTCTCTAAGCTAATGGTACACTAGAAGA
GTATTATACTATTAATTGGAAAATGGATGACATCATTCAACCAAGTTGGCAAATTTTT
ATTTCTTGGCAAAAAGATATAAAAATTGAAACTTGATGATTTCATAATTTGGAACAA
CATAGAGAAATAGATTATAGATCTAGGCATTATTTAAAATTAGAAAGAAAAAAAACC
ATTCTCATGCACACCCCTTTAACCCCAGATGAATTTGCAGCAAAATTAACATGTTCCT
TTCTGCTCGTACAA

>Off-target 5 GF1
TTGTCATGCATGCTTTATTAATCTTGGACTAACATATATATTGTATTTTTTTCTCCGCC
TTTTTTCATCAAATCCCACGTTCTATTTGCACTTTGCACTTAAATCCCAACTGAGCTTA
GAGTTTTTTTGTGCTTTTTACTTTTGATGACACTAATCTCAATAATACGTCTTTCCAGG
TCTTCACGTCTCACATGTTACTAGAACCACAGTGAAAAACATTCGGGGCAGAAAAGA
AGGTCGGCGTGTGTCTTTCCGGTGACCCCTCAGAAGTGATTTGTGTTGTTCCACTTCC
CATTAGTGGTGGGCAAAAAGCAAAACCACCACAGGGTCCCCTGCCATTCGCCAACCA
AACCCTAGCAAGCCACTCCATCAGATCAATGGATATAGGGACGCTGGGCGAGATCAC
CTCCTCATAAACGCTAGATGTTTGGATATTGGGTTTTAGTAAGAGGTAATTTGGCGTC
GTGGAAGAGTATGATACATAAAAGTTATCAATATAGCATAGAAGAAGAATATCGAG
CAATGGTTGTAGCAACTTGCGAGCTAGTTGGATCAAATAGCCGCTCACACAATTGAA
ATTTGAAGAAATCAATGAGATGGGTGTGATAATTAAGCAATCCTTTGTATCGAGTCA
AATCTCGTAATGAGAGGACTTAGCACATTCAGACTGACTATTATTTTTACAGAGAGT
AAATACTCG

Chapter II

Engineering of tomato resistance to tomato yellow leaf curl virus (TYLCV) through a double CRISPR/Cas9 approach

Introduction

Tomato, in comparison to other vegetable crops, including others belonging to the Solanaceae family, appears to be exceptionally susceptible to viral diseases. Intensive breeding for productivity might have reduced the genetic basis for virus resistance and areas of tomato monoculture, with homogeneous climate conditions and reduced biodiversity, can play an important role in the ability of vectors to move and spread diseases. Additional factors which favor the spread of viral diseases worldwide are the increasing international travel and exchange of plant material, and climate change, which allows the replication of insect vectors and pathogens in areas of the world where conditions did not use to be optimal (Hanssen *et al.*, 2010). Some viral diseases of tomato are rapidly emerging, like those caused by the Pepino mosaic virus (PepMV); other important diseases are caused by Tospoviruses and the Tomato Torrado virus (ToTV). In addition to this, tomato is susceptible to Begomoviruses, a numerous genus of ssDNA viruses belonging to the Geminiviridae family. Begomoviruses include important species, like Tomato yellow leaf curl virus (TYLCV), which presently constitutes the most devastating viral disease of tomato in tropical and subtropical regions.

Since no antiviral agents are available for plants, viral infections and epidemics can be limited either through control of the vector (which can be done through insecticides, insect-proof greenhouses and nets and agricultural practices) or by using resistant hosts, derived either from traditional breeding or from genetic engineering strategies. In general, all of these approaches have proven to be time-consuming and difficult to pursue: breeding for resistance or tolerance often requires introgression of resistance genes from wild relatives, followed by long cycles of crossing and selection. Transgenic approaches have been implemented in tomato to induce virus resistance, including RNA interference, expression of truncated viral proteins and expression of peptide aptamers that bind viral proteins (Ammara *et al.*, 2015; Chen *et al.*, 2015; Fuentes *et al.*, 2010; Galvez *et al.*, 2014; Praveen *et al.*, 2005; Sun *et al.*, 2014; Vanderschuren *et al.*, 2007). Some novel control strategies may arise in the near future, based on CRISPR/Cas9 approaches. At presents, efforts have focused mainly on model species, and are beginning to be translated to cultivated crops.

1.1 The tomato yellow leaf curl disease

The tomato yellow leaf curl disease (TYLCD) is characterized by upward curling of leaf margins, reduction of leaf surface, leaf yellowing, flower abortion and severe stunting, which result in extensive yield loss when infection occurs at early developmental stages (Moriones and Navas-Castillo, 2000) (Fig.

35). The disease is caused by a complex of viral species, which go under the name of tomato yellow leaf curl virus (TYLCV) and whose diffusion, starting from the late 1980s, is dependent on the worldwide spread of their vector, the whitefly *Bemisia tabaci* biotype B. This particular biotype of the whitefly has a wider host range than other biotypes, causing weeds and endemic species to constitute abundant reservoirs of the virus for adjacent crops. TYLCV can now be found in most tropical and subtropical areas of the world, including the Middle East (from where it originated), the Mediterranean basin, Asia, America, Australia and Eastern Africa (Varma and Malathi, 2003; Kenyon *et al.*, 2014; Mabvakure *et al.*, 2016).

Figure 35. TYLCV symptoms in tomato leaves.

TYLCV has a monopartite ssDNA genome (while Begomoviruses usually have bipartite genomes) and twin icosahedral capsids. The TYLCV genome is about 2.8 kb in size and encodes six proteins: the sense strand encodes the coat protein V1, which interacts with the vector and is conserved among Geminiviruses, and the pre-coat protein V2; the complementary strand encodes four proteins named C1-C4. These are: the replication initiator protein (Rep, C1), a transcriptional activator of the coat protein (TrAP, C2), a replication enhancer (REn, C3) and C4, which is implied in intercellular movement and whose open reading frame (ORF) is entirely contained in the Rep ORF (Duffy and Holmes, 2008). An intergenic region (IR) of about 200 nucleotides, which includes the plus-strand origin of replication, is also present. Geminiviruses use rolling circle replication to produce a dsDNA intermediate, which is used as a template for transcription and replication of the ssDNA genome (Hanley-Bowdoin *et al.*, 1999). First, a dsDNA copy of the viral genome is produced using as template the single strand "plus" which is, by definition, the one found in both the ssDNA and the dsDNA forms. The "plus" strand of this dsDNA molecule, known as the replicative form, serves, again, as a template to generate multiple copies of the ssDNA viral genome, which are in turn packed into new viral particles.

1.2 TYLCV resistance *loci* in tomato

For TYLCV, six resistance *loci* have been identified and mapped in tomato, named *Ty1-6*; all of them are the result of introgressions from wild relatives of tomato: *Ty1*, *Ty3*, *Ty4* and *Ty6* derive from *Solanum chilense*, *Ty2* from *S. habrochaites* and *Ty5* from *S. peruvianum* (Butterbach *et al.*, 2014). These *loci* do not strictly induce resistance (as in hypersensitive response models), but rather tolerance: virus replication is detectable in all *Ty* genotypes, but with a reduction of up to 90% with respect to sensitive genotypes. The pyramiding of these *loci* has been shown to increase tolerance (Vidavski *et al.*, 2008). *Ty1* and *Ty3* are allelic forms of the same *locus* and encode a RNA-dependent RNA polymerase which is thought to limit viral replication by inducing methylation of the V1/CP promoter (Verlaan *et al.*, 2013; Butterbach *et al.*, 2014). *Ty2* is a dominant resistance gene, which was mapped to chromosome 11, and whose tolerance mechanism is still unknown; it confers resistance to monopartite Begomoviruses, but some of these, including the tomato yellow leaf curl Sardinia virus (TYLCSV), have escaped its tolerance mechanism (Ji *et al*., 2009; Yang *et al.*, 2014; Tabein *et al.*, 2017). Similarly, *Ty4* was mapped on chromosome 3, but its precise function is still unclear (Ji *et al.*, 2009). The *Ty5* and *Ty6 loci* were mapped to the same region on chromosome 4 (Anbinder *et al.*, 2009; Hutton *et al.*, 2012; Scott *et al.*, 2015). While the *Ty6 locus* remains largely uncharacterized functionally, Lapidot *et al. (2015)* precisely mapped *Ty5* on chromosome 4 in the tomato TYLCV-resistant line TY172 and elucidated its function, identifying this *locus* as a tomato homolog of the messenger RNA surveillance factor *Pelo*. In a model proposed by Becker *et al.* (2012) for ribosome turnover in Archaea and Eukaryotes, Pelo is identified as part of a protein complex with the ABC-type ATPase 1, implicated in the dissociation of ribosomal subunits in the final stage of protein synthesis, a crucial step to make them available for a new round of translation. Since viruses encode very few proteins, they hijack host components as part of their replication mechanisms: for this reason, tolerance to viral pathogens is often associated to recessive mutations of endogenous host components involved in viral functions (Robaglia and Caranta, 2006). Eukaryotic translation initiation factors (eIF) are known to have an important role in recognizing and binding viral RNA; their mutations can confer tolerance to RNA viruses to plants and animals, with surprisingly little effect on host development and protein synthesis efficiency (Duprat *et al.*, 2002; Moury *et al.*, 2014; Robaglia & Caranta, 2006; Sanfaçon, 2015; Sato *et al.*, 2005; Wang *et al.*, 2013). However, since eIF are part of a diverse gene family, the role of each *locus* and the mechanism by which it interacts with viral components are yet to be described (Patrick and Browning, 2012; Bastet, *et al.,* 2017). *Pelo* was recognized to be

implicated in virus resistance in *Drosophila* (Wu *et al.*, 2014). In tomato, Lapidot *et al.* (2015) identified two transversions at the *Pelo locus* in TYLCV-resistant line TY172: one is the T^{1961}-to-A^{1962} transversion in the proximal promoter region of *Pelo*, the other is the T^{47}-to-G transversion in the first exon of the gene, which determines a $Valine^{16}$-to-Glycine substitution in TY172. Evaluation of the transcript levels of *Ty5* in both resistant and susceptible genotypes did not reveal a significant difference in transcription, meaning that the transversion in the promoter region does not affect gene expression. A demonstration of the function of *Ty5* came from both its overexpression in TYLCV-resistant genotypes, which resulted in increased severity of symptoms and viral titer, and from its silencing by RNA interference in wild type plants, which on the other hand resulted in a tolerant phenotype. The identification of *Pelo* mutants with tolerant phenotypes thus identifies the recycling phase of protein synthesis as a sensible step for viral replication. Of course, it is necessary to assess the impact of such mutations on host protein synthesis: *pelo* tomato plants analyzed by Lapidot *et al.* (2015) had reduced fruit size, and a small but not significant reduction of fruit number, but an otherwise normal development.

1.3 CRISPR/Cas approaches to virus resistance

Two main strategies have been described to pursue virus resistance through gene editing: (i) to edit endogenous plant genes, whose mutations result in reduced susceptibility to the pathogen, and (ii) to transform the host with a CRISPR/Cas9 construct directed at viral sequences, which acts - like it happens in Bacteria and Archaea - as a targeted immune defense (Zaidi *et al.*, 2016; Gal-On *et al.*, 2017). So far, regarding the first approach, CRISPR/Cas9 editing efforts have focused on the knock-out of translation initiation factors encoding genes to confer resistance to RNA viruses, especially Potyviruses. This approach was successfully employed by Chandrasekaran *et al.* (2016), who introduced broad RNA virus resistance in cucumber (*Cucumis sativus*) by knocking out the translation initiation factor *eIF4E*, and by Pyott *et al.* (2016) and Gomez *et al.* (2017), who addressed the same targets in Arabidopsis and cassava, respectively. For the second approach, which involves the use of gRNAs directed at viruses, the majority of the work was carried out in *N. benthamiana* through transient expression assays, and in tomato (Fig. 36). Ali *et al.* (2015) were the first to report the feasibility of this approach against Begomoviruses by targeting the replicase (Rep), the coat protein (CP) and the intergenic region (IR). gRNAs were introduced through transient transformation in leaves of *N. benthamiana* constitutively overexpressing Cas9, which were challenged with viral infectious clones 7 days after the inoculation of the

CRISPR editing tools. All targets were successful in reducing the viral titer and the infection symptoms, though the targeting of the IR yielded better results than the targeting of Rep or CP; this is thought to be due to the fact that in the second case some amount of the viral proteins could still be produced, thus attenuating but not eliminating viral replication. This first study assessed the versatility and specificity of CRISPR/Cas9: on the one hand, more than one viral species was targeted at a time, directing a single gRNA to a conserved, shared region; on the other hand, a single species was target with more than one gRNA. This approach confirmed that CRISPR/Cas9 is highly selective and able to address multiple targets at the same time.

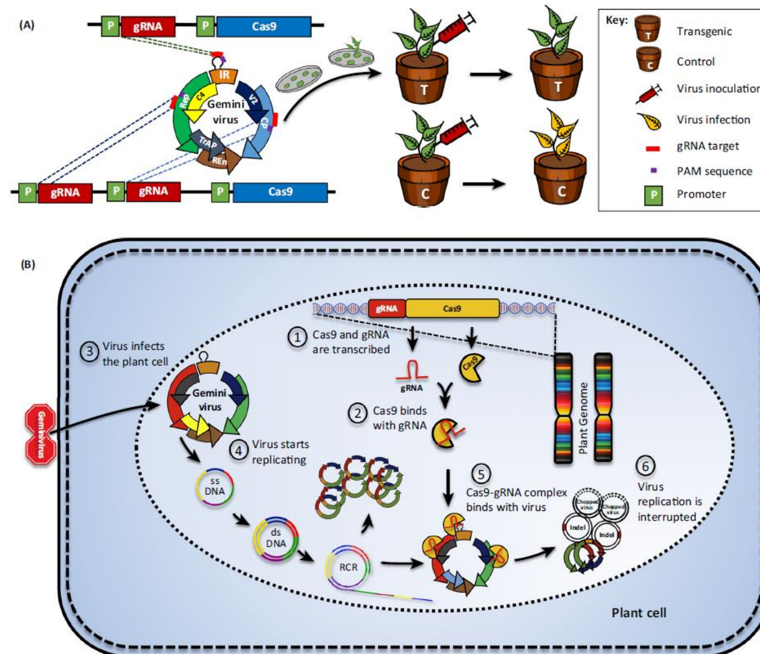


Figure 36. CRISPR/Cas9-mediated virus resistance in plants. (A) Transgenic plants expressing CRISPR/Cas9 systems directed against viruses do not develop disease symptoms when infected with the corresponding Geminivirus. (B) When the virus infects the plant, gRNAs, together with Cas9, recognize target sequences and lead to the disruption of the viral genome, which prevents it from successfully replicating. Zaidi *et al.* (2016).

One common concern when developing virus resistant genotypes is that the mutation rate of viruses might allow them to evade resistance mechanisms, restoring susceptibility. Geminiviruses are known to have high evolution rates (Duffy and Holmes, 2008), comparable to those of RNA viruses, and introducing a specific immune defense in plants might entail greater selective pressure, together with an increased tendency to recombination, which can

occur during mixed infections, and especially as a consequence of CRISPR-triggered double strand breaks (DSBs). Ali *et al.* (2016) expanded their previous work by investigating the ability of CRISPR/Cas9 to interfere with replication when targeting coding regions or the IR, and its ability to generate viral variants which could potentially evade CRISPR immunity. The system was tested both in a bipartite Geminivirus (MeMV) and in monopartite Geminiviruses (the cotton leaf curl Kokhran virus CLCuKoV and its cotton leaf curl Multan betasatellite CLCuMβ, and a severe and a mild strain of TYLCV). Both types of targets (IR and coding sequences) were effective in causing interference, but only the editing of coding sequences caused the appearance of variants arising from NHEJ repair. The infection of *N. benthamiana* plants with the sap of other infected plants expressing the CRISPR/Cas9 machinery revealed that some viral mutants were capable of infection and, though less effective, could be considered escapees generated by the NHEJ repairing mechanisms. The failure to detect NHEJ-induced variants in individuals were the IR was targeted indicates that these mutants are not capable of replication. Most recently, the same group (Tashkandi *et al.*, 2017) also obtained stable *N. benthamiana* and tomato transformants with CRISPR-induced TYLCV tolerance over multiple generations and confirmed that, consistently with observations from transient expression experiments, genetic variants capable of evasion of the immune CRISPR system could be obtained over a short span of time.

1.4 Experimental goals

In this chapter, we reported the obtaining of tomato tolerance to TYLCV through both editing approaches. The *Ty5 locus* of tomato was chosen as a target for gene editing to evaluate its potential to introduce resistance against Begomoviruses, in accordance with the results reported by Lapidot *et al.* (2015), and stable tomato transformants were obtained to this effect. To pursue the second strategy, in which the virus itself was targeted by CRISPR/Cas9, five species of TYLCV (Sardinia, Israel, Mild, New Delhi and China) were selected and, for each species, three targets were designed on the Rep and CP genes. For each species, a multiplexing editing construct was assembled using the GoldenBraid toolbox, with the three gRNAs assembled as a polycistronic transcript. The efficiency of this strategy was first evaluated for two of the TYLCV species (Sardinia and Israel) by transient expression in *N. benthamiana* leaves. To this end, the 2IR-GFP *N. benthamiana* line was used, which makes use of a GFP-based reporter system, whose expression correlates with viral replication rates. The same constructs were then stably transformed into two local Spanish tomato varieties.

2 - Materials and Methods

2.1 Vector design and construction

Tomato tolerance to TYLCV was pursued through both editing approaches: (i) by editing a plant endogenous gene and (ii) by transforming the host with a CRISPR/Cas9 construct directed at viral sequences.

2.1.1 Editing of a plant endogenous gene

A single guide RNA targeting the coding sequence of the T*y5/pelo* gene (Solyc04g009810) was designed using the design tools on Benchling (www.benchling.com), in analogy with the design of a guide RNA described for the *gf1 locus* in the Materials and Methods section of the previous Chapter. The *Ty5/pelo* gRNA was domesticated for cloning into GoldenBraid pDGB3 alpha1 vectors (Supplementary Tables S8-S10 at the end of this Chapter) and the cloning was performed as described in Chapter I. The *Ty5* CRISPR/Cas9 construct was used to transform *A. tumefaciens* LBA4404 electrocompentent cells for stable transformation of tomato.

2.1.2 Transformation of the host with a CRISPR/Cas9 construct directed at viral sequences

Genes for the replicase (Rep) and coat protein (CP) of five different species of TYLCV (Israel, Mild, Sardinia, New Delhi and China) were aligned to look for structural motifs and conserved regions. For each species, two gRNAs were designed to be directed against conserved motifs of the Rep gene, and one gRNA was designed to be directed against the CP. Each gRNA was domesticated to be inserted into a pUPD2 vector (Supplementary Table S11). Subsequently, for each species, pUPD2 vectors carrying tRNA:gRNA:scaffold constructs were combined in a multipartite assembly into a level 1 alpha vector, together with a U6-26 promoter, which controls the expression of the 3 gRNA cassette (Supplementary Table S12). To each of these constructs, the Cas9 TU and the nptII selection marker TU were further added (Supplementary Tables S13-S14).

The two constructs targeting TYLCV species Israel and Sardinia were separately transformed in *A. tumafaciens* C5801 electrocompetent cells for transient expression assays. These two constructs were chosen among the five

assembled, because of the availability of a *N. benthamiana* GFP-based reporter system for virus multiplication, specific for these 2 species (Israel and Sardinia).

Moreover, the same constructs (targeting TYLCV species Israel and Sardinia, hereafter TYLCV and TYLCSV, respectively) were used to transform *A. tumefaciens* LBA4404 electrocompentent cells to perform stable transformations of tomato, following the *A. tumefaciens* transformation protocol described in Chapter I.

2.2 Plant material

2.2.1 Stable transformation

The 'MoneyMaker' variety was used for the *ty5/pelo* transformation, while 'Muchamiel' and 'Pera' varieties were used for the TYLCV and TYLCSV transformations.

50 seeds per transformation event were sterilized by washing for 30 minutes in a 2.5% sodium hypochlorite solution, then rinsed for three times in sterile water, with each rinse lasting for 5, 10 and >15 minutes, respectively. Sterile, clean seeds were transferred to a solid germination *medium* (2.5 g l^{-1} MS vitamins, 10 g l^{-1} sucrose and 10 g l^{-1} phytoagar, pH 5.8) in sterile cups and kept in the dark for 3 days at 25°C before being exposed to light. 10 days after sowing, cotyledons were used for *Agrobacterium*-mediated transformation. Cotyledon development was considered optimal for transformation when the first true leaves start to emerge.

2.2.2 Transient transformation

Nicotiana benthamiana 2IR-GFP seeds (Morilla *et al.*, 2006), for the TYLCV- and TYLCSV-CRISPR agroinfiltrations, were sowed and grown in soil at 22°C in short day conditions (8:16 hours light:dark photoperiod).

2.3 *Agrobacterium*-mediated transient transformation

Transient expression assays in *N. benthamiana* leaves were conducted for CRISPR/Cas9 constructs targeting TYLCV and TYLCSV. *A. tumefaciens* C5801 cultures carrying the CRISPR construct (TYLCV or TYLCSV) and a *A. tumefaciens* C5801 culture carrying a control construct (an empty pGreen vector) culture were set up in LB liquid *medium* pH 7 supplemented with

antibiotics (50 mg l^{-1} kanamycin and 50 mg l^{-1} rifampicin) and incubated overnight with agitation. The optical density was measured at 600 nm with a spectrophotometer and adjusted to a value of 0.5; the culture was supplemented with 200 µM acetosyringone and incubated for at least 3 hours in the dark. Three leaves were infiltrated for each plant. Each selected leaf was infiltrated on the left side of the midrib with an *A. tumefaciens* C5801 culture carrying the CRISPR construct (TYLCV or TYLCSV) plus its respective infectious clone, and on the right side of the midrib with an *A. tumefaciens* C5801 culture carrying a control construct (an empty pGreen vector) plus the infectious clone, as illustrated in Figure 37, according to the protocol described in Elmer *et al.* (1988). Plants were infiltrated at 3 weeks after germination. To establish the optimal infiltration time of the CRISPR construct with regard to virus infection, different time points were chosen, as represented in Figure 38. This was done to allow the CRISPR/Cas9 construct to be expressed, or alternatively to allow the virus to replicate.

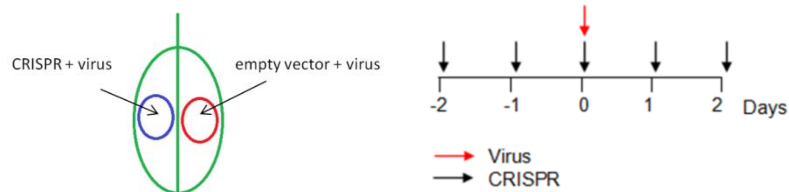


Figure 37. Inoculation pattern of *N. benthamiana* leaves with the TYLCV infectious clone and either the control empty pGreen vector or the CRISPR/Cas9 construct.

Figure 38. Inoculation planning of the CRISPR/Cas9 construct infiltration with respect to infiltration of the infectious clone (at day 0).

Four days after infiltration, leaves were collected and observed under UV light to assess differences in GFP levels for each leaf. Once we determined which time point had the greatest visual effect on virus replication, the experiment was repeated using 5 2IR-GFP *N. benthamiana* plants for each construct; here, the control was infiltrated with a CRISPR/Cas9 multiplexing vector directed at two tomato targets (XT1 and XT2 xylosyltransferase genes), which would not affect *N. benthamiana* response to the virus. Again, 3 leaves were infiltrated for each sample, following the same pattern reported in Figure 37. At 4 days post infection (dpi) leaves were analyzed under UV light. DNA was extracted from the samples according to the previously described CTAB method (see Chapter I). The total DNA extracted from plant tissues 4 dpi comprised two sources of viral DNA: what remained of the infectious clone infiltrated using the *A. tumefaciens* culture, and new copies of the virus produced via rolling circle

replication. To establish the effect of the CRISPR/Cas9 system to reduce viral replication, a selective amplification of viral DNA resulting from replication in plant cells was performed. 50 ng of total DNA extracted from leaf samples were incubated overnight with *Dpn*I (New England Biolabs), a methylation-sensitive restriction enzyme. *Dpn*I selectively binds and cuts bacterial DNA containing the G^{m6}ATC sequence, without affecting eukaryotic DNA (Barnes *et al.*, 2014). This allowed to preserve viral copies obtained in plant cells, which did not carry bacterial DNA marks. Restriction reactions were performed in a 20 µl volume according to the manufacturer's instruction. 2 µl of such reactions, corresponding to 5 ng of DNA, were used for amplification. DNA samples from leaves of the same plant were pooled for analysis. Viral DNA was amplified with species-specific primers, while the ITS *N. benthamiana* gene was amplified as reference (Table 19). PCR products were run on a 2% agarose gel to evaluate relative amounts of viral DNA.

Target	Primer sequence 5' → 3'	T_m
ITS forward	ATAACCGCATCAGGTCTCCA	64.8
ITS reverse	CCGAAGTTACGGATCCATTT	63
TYLCSV forward	GAGAAACATATGGTGCCG	59.6
TYLCSV reverse	TCATTCAGTTCGAGGG	55.8
TYLCV forward	GGAGCAGTGATGAGTTCCCCTG	69.5
TYLCV reverse	GGGGAACCACATCTCCATGTGC	72

Table 19. Primers for species-specific TYLCSV and TYLCV amplification and for amplification of *N. benthamiana* ITS reference gene.

2.4 *Agrobacterium*-mediated stable transformation

The *Ty5/pelo* construct was transformed into 'MoneyMaker' tomato explants according to the protocol described in Chapter I.
'Muchamiel' and 'Pera' tomato varieties were stably transformed with constructs harboring gRNAs directed against TYLCV and TYLCSV according to the same protocol.

2.5 Organogenesis and regeneration of stable transformed plants

The regeneration of plantlets from the transformed cotyledon explants was performed according to the method described by Qiu *et al.* (2007), with modifications. Briefly, after 48 h of co-culture with *A. tumefaciens*, explants were grown on a *medium* to induce the formation of *callus* and shoots (induction *medium*). Explants were moved to a fresh *medium* every 21 days, or when explants size or shoot formation required it. Shoots were then transferred to an elongation *medium,* and finally moved to a rooting *medium*. Regeneration *media* are reported in Chapter I, Supplementary Material. Kanamycin selection was maintained at all stages of regeneration. Plants were grown *in vitro* to a size of 5-8 centimeters and then moved to soil in the greenhouse, where they were gradually acclimated to environmental growing and humidity conditions.

2.6 Genotyping of the T_0 generation

DNA was extracted from the leaves of plantlets growing *in vitro* according to the CTAB protocol described in Chapter I, Supplementary Material. Each DNA sample consisted of different leaves from the same plant. The plants were genotyped according to the protocol described in Chapter I.
For *Ty5/pelo* CRISPR transformants in which the insertion of the transgene was confirmed, the *Ty5/pelo locus* was amplified. A first screening with the T7E1 endonuclease assay was carried out, followed by Sanger sequencing of PCR products and analysis with the TIDE software, also as described in Chapter I.

In the case of plants carrying the construct against TYLCV and TYLCSV, the screening was directed only at confirming the presence of the transgene by PCR amplification of the *hCas9* gene. Primer sequences are reported in Table 20.

Target	Primer sequence 5' → 3'	T_m
ty5 forward	TCCAATGTGGTCGAGATGAA	54.8
ty5 reverse	GGTGGAAGTGGTCGAAGTGT	58.8
hCas9 forward	AGGTGGCGTACCATGAAAAG	56.5
hCas9 reverse	TGTTTGCGCAACAGATCTTC	55.2

Table 20. Primers used for genotyping and T-DNA amplification.

2.7 Evaluation of off-target effects

Possible off-target sites for the *Ty5/pelo* gRNA were identified using the CasOFFinder software (www.rgenome.net/cas-offinder/). The *S. lycopersicum* SL2.5 genome assembly was used as reference. The maximum number of mismatches was set at 5, with DNA and RNA bulge sizes set at 1. Five different off-target sites were selected based on the number and position of mismatches. The sequence of each off-target *locus* (1 kb) was retrieved using the tomato genome browser at solgenomics.net and aligned to the Viridiplantae nucleotide database on https://blast.ncbi.nlm.nih.gov/Blast.cgi to screen for annotations.

The genotyping of the candidate off-target *loci* was performed on the 5 *Ty5/pelo* 'MoneyMaker' plant lines growing in the greenhouse according to the Illumina Amplicon Sequencing Protocol described in Chapter I. Supplementary Tables S15 and S16 list the primers used for each sample for the first and second amplification rounds.

3 - Results

3.1 *Ty5/pelo* CRISPR/Cas9 editing

3.1.1 Target identification

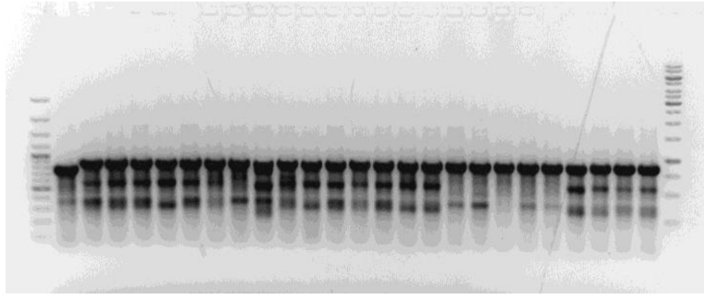
A guide RNA was designed, targeting the fourth exon of *Ty5/pelo*, a 7740 bp gene containing 16 exons, which encodes a 378 aminoacid protein. The protein contains 3 eukaryotic release factor (eRF1) domains, and the selected gRNA targets the first domain, which spans residues 1-129. The gRNA sequence was: 5' - GTTCTGCCTTGCGTATTCGC - 3'. The final destination vector was transformed into *A. tumefaciens* for stable tomato transformation.

3.1.2 *S*table transformation

Stable transformation of 'MoneyMaker' cotyledons with the T*y5/pelo* CRISPR construct allowed to regenerate 32 individuals from 17 different *calli*. Samples were named using a number (1-22) referring to the *callus* they were generated from, and a capital letter referring to individual plantlets within each *callus*. For some *calli*, shoots had been obtained, but they failed to root *in vitro* and died.

3.1.3 Genetic and phenotypic characterization of the T_0 generation

Analysis of the edited *locus* could be performed only for the *Ty5/pelo* line, since CRISPR activity the *TYLCSV* and *TYLCV* lines could have been evaluated only by infecting the plants with the corresponding viruses. Of 32 *Ty5/pelo* plants analyzed with the T7EI assay, 30 (94.12%) proved to be edited (Fig. 39). The integration of T-DNA was verified in all T7EI-positive individuals by amplifying a fragment of the *hCas9* gene, and all proved positive. The undigested PCR product of the T*y5/pelo locus* was directly sequenced for all individuals, together with the correspondent wt DNA. The results of the TIDE analysis are summarized in Table 21.



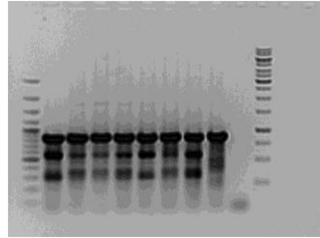


Figure 39. T7EI endonuclease assay. 30 out of 32 plants show polymorphism at the target site. Lanes, left to right and top to bottom, represent: 2A-B, 3A-B-C-D-E, 4B-C, 5A-B-D, 6A-B-C-D, 7B, 8B-C-D-E, 10A, 13A, 15A, 16B, 17A, 18A-B, 19A, 20A, 21A, 22A, wt, negative control. The outer left and right lanes are occupied by the molecular marker.

TY5	Overall efficiency	R^2	WT	+1	Ins. base	+2	-29	-22	-17	-16	-12	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1
2A	13.6	0.14																			
2B	96.4	0.96		46	T															50.2	
3A	96.7	0.97		30.1	T									4.2					53.9		8.4
3B	95.3	0.96		43.3	T												2.2		6.6	42.2	
3C	94.5	0.96	1.6	45.1	T								11.0						37.1		
3D	97.8	0.99	1.4	73.4	T				0.8							1.2	1.0	3.0	4.3		13.2
3E	99.3	0.99		89.9	T														8.9		
4B	99	1		98.8	T																
4C	39.9	0.61	21.4	6.5	T														13.8		
5A	96.7	0.97		52.5	T														42.8		
5B	96.5	0.97		21.3	T		25.6														49.7
5D	82.9	0.83		16.4	T		23.3														40.1
6A	94.8	0.97	2	64.2	T	1.6												1.7	23.9	3.4	
6B	94.4	0.96	1.5	26.4	T												56.7				9.8
6C	96.3	0.98	1.4	19.8	T									1.7			6.7	2.3	1.7	12.8	51.2
6D	89.4	0.95	5.8	20.5	T											52.5			13.4	1.8	
7B	0.7	0.99	98.6																		
8B	99.2	0.99		97.5	T																
8C	99.4	0.99		98.6	T																
8D	99.4	0.99		98.7	T																
8E	99.4	0.99		98.5	T																
10A	96.8	0.97		46.6	T						50.2										
13A	98.9	0.99		66.5	T					4										26.2	2.2
15A	94.1	0.94		42.4	T												50.9				
16B	96.8	0.97		2.6	T							9.2						53.3			31.7
17A	79.9	0.94	14.2	31.8	T			19.1							26.4				2.3		
18A	96.3	0.96		45.8	T													50.4			
18B	96.6	0.97		47	T													49.6			
19A	28.4	0.99	70.7											14					14.3		
20A	64.1	0.96	31.5	27.6	T									26.4				4.4	4.2		
21A	97.9	0.98		73.7	T															24.2	
22A	95.5	0.98	2.9	23.3	T													64.9	6.2		

Table 21. Genotyping of T_0 individuals by PCR amplification of the target *locus*, direct sequencing and TIDE chromatogram decomposition. In red are individuals chosen for growth in the greenhouse as representatives of specific genotypes.

Over 94% of individuals (30/32) were edited. At the single plant level, editing efficiencies vary between 28% (19A) and over 99% (*callus* 8). The average editing efficiency is 90.42%. The allelic frequencies of each type of indel are reported in Figure 40. Among edited plants, 12.5% were heterozygous, 15.6% homozygous and 71.9% biallelic. 90.6% of edited plants have a +1 insertion at the cut site 3 nucleotides upstream of the PAM, and the inserted nucleotide is always a T.

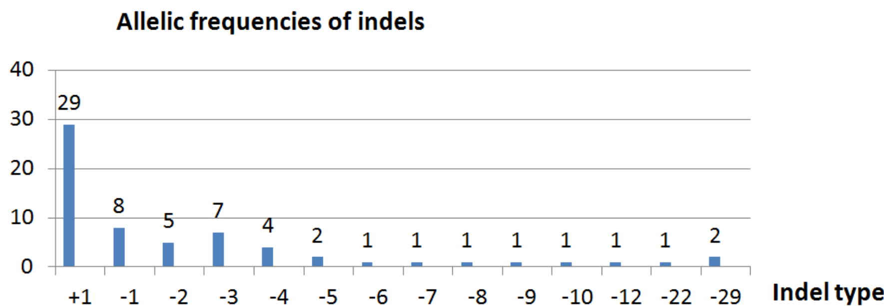


Figure 40. Allelic frequencies of indels, reported as number of plants carrying each allele.

Some plants (4B and 8B-C-D-E) were homozygous for the 1 nucleotide insertion of a T, and editing efficiencies are reported to be virtually of 100%. Five lines, highlighted in red in Table 21, were selected to be moved to the greenhouse based on their different genotypes. Deletions of three or six nucleotides were avoided, as they could result in the loss of one or two aminoacids, without possibly compromising protein function. Plants 2B and 18A had biallelic mutations, with small indels (-2 and -4, respectively, and both have the +T insertion); 10B was also biallelic, including a deletion of 12 nucleotides which resulted in the loss of four aminoacids; 5B was chimeric, with 3 alleles, one of which is a larger deletion of 29 nucleotides; finally, 8B was homozygous for the +T insertion.

Ty5/pelo plants displayed a normal phenotype during *in vitro* culture and in the first phases of growth in the greenhouse. However, after reaching a height of about 80-100 cm, they rapidly suffered yellowing and wilting, starting from the lower, older branches, resulting in plant death; flowers were formed, but failed to yield fruits. In particular, the 8B individual, which had been chosen for its homozygous genotype, was severely stunted, reaching a maximum size of about 40 cm; the other individuals were comparable in size to the wild type.

3.1.4 Analysis of off-target effects

Again, off-target activity could be assessed only for the *Ty5/pelo* CRISPR/Cas9 construct, as the TYLCSV and TYLCV constructs targeted conserved viral sequences which are not expected to have close homologs in plant genomes. Five putative off-target sequences for the *Ty5/pelo* gRNA were selected among those identified by CasOFFinder, as representatives of different kinds and numbers of mismatches. Their characteristics are presented in Table 22 and sequences can be found in the Supplementary Sequence List at the end of this Chapter. Off-targets 1, 4 and 5 corresponded to non-coding regions. Off-target 2 was located in the coding sequence of the receptor-like kinase HSL1, while off-target 3 corresponded to a coding but uncharacterized *locus*. Like for *gf1*, no putative off-target sequences had less than 4 mismatches with respect to the target sequence if bulges were not taken into consideration; with bulges, the minimum number of mismatches was 3. Candidate sequences in the same *locus* as the target were not taken into account, as their editing wouldn't result, functionally, in an off-target effect. All candidate off-target regions had at least one mismatch in the seed region of the gRNA. In this instance, too, deep sequencing (with an average sequencing depth of 15,000 reads per *locus*) proved that no relevant off-target effects were detectable for all off-target *loci* (Table 23), adding to the amount of evidence in favor of the specificity of CRISPR/Cas9 in tomato, and in plants in general. Some variation was detected for the off-target 5 *locus* in all individuals, ranging from 2.64 to 6.34% of the analyzed reads; inspection of alignments of 300 randomly selected sequences revealed that variation was not due to editing, but to the presence of an unspecific PCR product. From alignments of the off-target 4 *locus* it also emerged that a number of SNPs existed in the putative PAM region, altering the NGG pattern, which would make it impossible for the *locus* to be recognized as an off-target.

Off-target	Chromosome	Position	Sequence	Annotation
1	11	17526313	GTTtgGCCTTGCGaATTgGCCGG	non coding
2	8	51855291	tTTCTGaCcgGCtTATTCGCCGG	Receptor-like kinase HSL1 (accession XM_004245095)
3	8	42437246	tTTCTG-CTTGtGTtTTCGCAGG	uncharacterized *locus* (accession XM_004245269)
4	4	31849325	GcTCTaCCTT-CGTtTTCGCTGG	non coding
5	5	18588194	cTTCT-CCcTGCGaATTCGCCGG	non coding

Table 22. Characteristics of putative off-target sites for the *Ty5/pelo* gRNA. Mismatches are in lowercase and bulges are represented by dashes.

	Sample	N. of sequences	N. WT target sequence	% WT	Normalized % edited
OFF TARGET 1	WT	10815	10426	96,40%	0
	2B	19715	18996	96,35%	0.05
	5B	19639	18919	96,33%	0.07
	8B	15705	15149	96,46%	0
	10A	896	868	96,88%	0
	18A	14510	13976	96,32%	0.08
OFF TARGET 2	WT	9622	9344	97,11%	0
	2B	28423	27541	96,90%	0.22
	5B	13608	13176	96,83%	0.29
	8B	12964	12557	96,86%	0.26
	10A	10221	9893	96,79%	0.33
	18A	21890	21230	96,98%	0.13
OFF TARGET 3	WT	8802	8596	97,66%	0
	2B	12832	12502	97,43%	0.26
	5B	13641	13311	97,58%	0.08
	8B	7002	6810	97,26%	0.41
	10A	7158	6938	96,93%	0.75
	18A	8186	7962	97,26%	0.41
OFF TARGET 4	WT	30463	29665	97,38%	0
	2B	21131	20547	97,24%	0.14
	5B	24733	24050	97,24%	0.14
	8B	13352	13004	97,39%	0
	10A	3329	3219	96,70%	0.70
	18A	17568	17138	97,55%	0
OFF TARGET 5	WT	21061	20319	96,48%	0
	2B	16475	14983	90,94%	5.74
	5B	14881	13447	90,36%	6.34
	8B	12487	11641	93,22%	3.38
	10A	17781	16118	90,65%	6.04
	18A	24028	22570	93,93%	2.64

Table 23. Quantification of wild type off-target sequences by filtering of Illumina reads. The WT 20 bp off-target sequence was used as query. No off-target effects are detectable in transformed plants.

For off-targets 1 and 3, it is worth noting that the variability was very low (with a maximum of 0.75% variation in the off-target region) but that all variability due to the presence of indels was concentrated in the off-target region near the PAM, implying that it could be due to improper CRISPR-mediated editing, albeit with an extremely low efficiency. On the other hand, SNPs were evenly distributed across sequences.

3.2.1 Target identification

The sequences of the Rep and CP genes of five species of TYLCV (Sardinia, Israel, Mild, New Delhi and China) were aligned and functional domains and conserved regions were identified (Fig. 41-42). Three guide RNAs for each TYLCV species, two directed at the Rep gene and one directed at the CP gene, were identified. Their sequences are reported in Table 24.

gRNA	5' → 3' Sequence	Target
Sardinia Rep1	CGTCAAGTCCTACATCGACA	RCR II
Sardinia Rep2	AGATCGACGGAAGATCTGCA	RCR III
Sardinia CP	TGTGTTAGTGATGTAACTAG	Non conserved region
Israel Rep1	ATACCAGGTCGAAGAACCGT	HLHV
Israel Rep2	GATGGCAGATCAGCTAGAGG	RCR III
Israel CP	CATGGGCCTTCACATCCACG	Conserved region
Mild Rep1	ATGTGCTTATCCAATTTGAA	2 conserved regions upstream of HLHV
Mild Rep2	CGACCTGGTATCCCCAAGCA	2 conserved regions downstream of HLHV
Mild CP	CATGGGCCTTCACATCCACG	Conserved region
New Delhi Rep1	CCGCCAAGTCGTTTTAGAAT	300 nt downstream of RCR III
New Delhi Rep2	GATGGACGATCTGCTCGTGG	RCR III
New Delhi CP	AGAAGTCCCGTCGTGCCAAG	Conserved region
China Rep1	GTAGAGAATTACACGAAGAT	Conserved region upstream of HLHV
China Rep2	GATGGAAGATCGGCTAGAGG	RCR III
China CP	GCGTCACCAGAAGACAAATG	Conserved region

Table 24. gRNA sequences for TYLCV targets on the five virus species. For each species, 2 targets were selected on the Rep gene, and one on the CP.

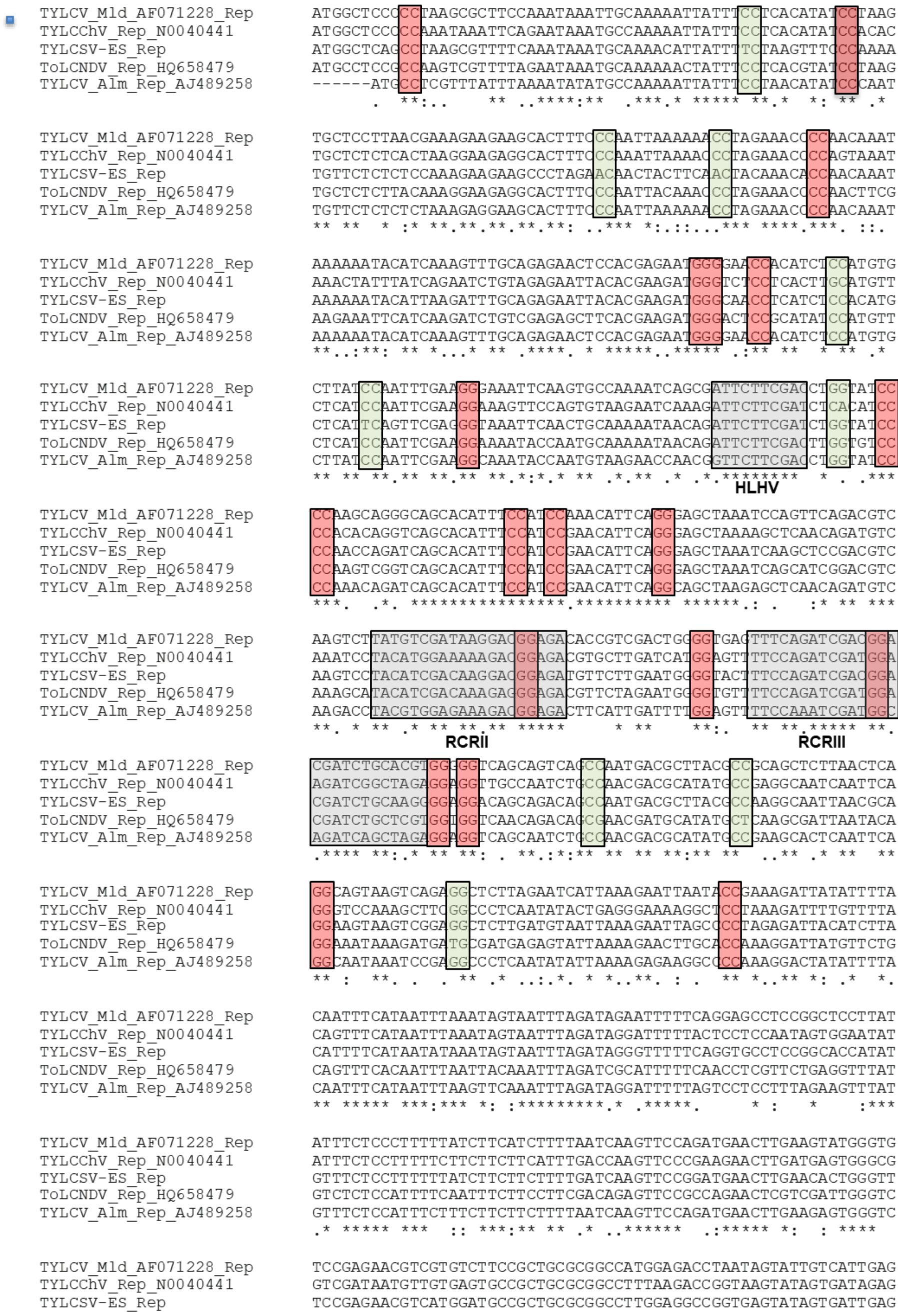


Figure 41. Alignment of Rep genes from the species Mild, China, Sardinia, New Delhi and Israel of TYLCV. Domains and conserved regions are highlighted.

GG/CC conserved in 5/5

GG/CC conserved in 4/5

Functional domains

```
CLUSTAL O(1.2.1) MULTIPLE SEQUENCE ALIGNMENT

TYLCSV-ES_CP              ATGCCGAAGCGAACAGGCGATATACTAATTTCAACGCCCGTCTCGAAGGTTCGTCGAAAA
TOLCNDV_CP_HQ658479       ATGGCGAAGCGACCAGCAGATATCATCATTTCAACGCCCGCATCGAAAGTACGCCGTCGT
TYLCV_MLD_AF071228_CP     ATGTCGAAGCGACCAGGCGATATAATCATTTCCACGCCCGCCTCGAAGGTTCGCCGAAGG
TYLCCHV_CP_N0040441       ATGGCGAAACGTCCCGCCGATATTGTCATTTCCACTCCCGCCTCGAAGGTGCGTCGCCGG
TYLCV_ALM_CP_AJ489258     ATGTCGAAGCGACCAGGCGATATAATCATTTCCACGCCCGTCTCGAAGGTTCGCCGAAGG
                          *** ****.**:.*.* .*****  *.*****.** ****  .*****.** ** **  ..

TYLCSV-ES_CP              CTGAACTTCGACAGCCCGTATACCAGCCGTGCTGCTGCCCCACTGTCCAAGGCATCA--
TOLCNDV_CP_HQ658479       CTCAACTTCGACAGCCCCTATGGAGCTCGTGCAGTTGTCCCCATTGCCCGCGTCACAAAA
TYLCV_MLD_AF071228_CP     CTGAACTTCGACAGCCCATACAGCAGCCGTGCTGCTGTCCCCATTGTCCAAGGCACAAAC
TYLCCHV_CP_N0040441       CTGAACTTCGACAGCCCCTATACGGGTCGTGTTGCTGCCCCACTGTCCGCGTCACCAG-
TYLCV_ALM_CP_AJ489258     CTGAACTTCGACAGCCCATACAGCAGCCGTGCTGCTGTCCCCATTGTCCAAGGCACAAAC
                          ** ************** **  .  .  **** :* ** ***** ** **..* ** .*

TYLCSV-ES_CP              -AGCGTCGATCATGGACTTACAGGCCCATGTATCGAAAGCCGCGGATGTACAGAATGTAC
TOLCNDV_CP_HQ658479       GC---AAAGGCCTGGACAAACAGGCCGATGAACAGAAAACCCAGAATGTACAGAATGTAT
TYLCV_MLD_AF071228_CP     AAGCGACGATCATGGACGTACAGGCCCATGTACCGGAAGCCCAGAATATACAGAATGTAT
TYLCCHV_CP_N0040441       --AAGACAAATGTGGTCAAACAGGCCCATGTACCGGAAGCCCATGATGTACCGGATGTAC
TYLCV_ALM_CP_AJ489258     AAGCGACGATCATGGACGTACAGGCCCATGTACCGAAAGCCCAGAATATACAGAATGTAT
                               :...    ***:* :******* ***:* .*.**.** . .**.***.*.*****

TYLCSV-ES_CP              AGAAGCCCTGATGTCCCCTTTGGTTGTGAAGGTCCTTGTAAAGTCCAGTCGTATGAGCAG
TOLCNDV_CP_HQ658479       AGAAGTCCCGTCGTGCCAAGGGGATGTGAAGGCCCTTGCAAGGTGCAGTCCTTCGAATCC
TYLCV_MLD_AF071228_CP     CGAAGCACTGATGTTCCCCGTGGATGTGAAGGCCCATGTAAAGTACAGTCTTATGAGCAA
TYLCCHV_CP_N0040441       AGAAGCCCTGATGTTCCAAGGGGTTGTGAAGGCCCATGTAAAGTTCAATCATGTGAGGCC
TYLCV_ALM_CP_AJ489258     CGAAGCCCTGATGTTCCCCGTGGATGTGAAGGCCCATGTAAAGTCCAGTCTTATGAGCAA
                          .****  .* *: ** **.    **:********* **:** **.** **.** *  **. .

TYLCSV-ES_CP              CGTGACGACGTCAAGCATACCGGTGTTGTTCGTTGTGTTAGTGATGTAACTAGGGGTTCT
TOLCNDV_CP_HQ658479       AGGCACGATGTCTCTCATATTGGCAAAGTCATGTGCGTTAGTGATGTTACCCGAGGAACT
TYLCV_MLD_AF071228_CP     CGGGATGATATTAAGCATACTGGTATTGTTCGTTGTGTTAGTGATGTTACTCGTGGATCT
TYLCCHV_CP_N0040441       CGTCATGATGTTTCACATACTGGTAAGGTTATTTGTGTATCGGATGTTACTCGTGGTAAC
TYLCV_ALM_CP_AJ489258     CGGGATGATATTAAGCACACTGGTATTGTTCGTTGTGTTAGTGATGTTACTCGTGGATCT
                          .*  * ** .* :. ** *  ** .: **  .  ** **::  *****:** .* **::.

TYLCSV-ES_CP              GGTATTACACATAGAGTAGGTAAACGGTTTTGTATTAAGTCAATCTATATTTTAGGGAAG
TOLCNDV_CP_HQ658479       GGACTCACACATCGCGTAGGAAAGCGATTTTGTGTGAAATCTGTGTATGTGCTGGGGAAG
TYLCV_MLD_AF071228_CP     GGAATTACTCACAGAGTGGGTAAGAGGTTCTGTGTTAAATCGATATATTTTTTAGGTAAA
TYLCCHV_CP_N0040441       GGTATTACTCACCGTGTTGGCAAGAGGTTCTGTGTTAAATCCATTTACGTCATTGGGAAA
TYLCV_ALM_CP_AJ489258     GGAATTACTCACAGAGTGGGTAAGAGGTTCTGTGTTAAATCGATATATTTTTTAGGTAAA
                          **:.* **:** .* ** ** **..*.** ***.* **.** .* **  *  * ** **.

TYLCSV-ES_CP              ATTTGGATGGATGAAAATATAAAAAAACAAAATCATACTAACCAGGTCATGTTCTTTTTA
TOLCNDV_CP_HQ658479       ATATGGATGGATGAAAACATCAAGACGAAAAACCATACTAACAGTGTTATGTTCTTTTTG
TYLCV_MLD_AF071228_CP     GTCTGGATGGATGAAAATATCAAGAAGCAGAACCACACTAATCAGGTCATGTTCTTCTTG
TYLCCHV_CP_N0040441       ATATGGATGGACGAAAATATTAAATTTAAGAATCATACAAATACAGTAATGTTTTTTCTT
TYLCV_ALM_CP_AJ489258     GTCTGGATGGATGAAAATATCAAGAAGCAGAATCACACTAATCAGGTCATGTTCTTTTTG
                          .* ******** ***** ** **.:  .*.** **:**  .  ** ***** **  *

TYLCSV-ES_CP              GTACGAGACCGAAGGCCGTATGGAACTAGTCCTATGGATTTTGGTCAAGTTTTTAACATG
TOLCNDV_CP_HQ658479       GTCCGTGACCGTCGTCCTACAGGAT---CCCCCAAGATTTCGGGGAAGTTTTTAACATG
TYLCV_MLD_AF071228_CP     GTCCGTGATAGAAGGCCCTATGGAAACAGCCCAATGGATTTTGGACAGGTTTTTAATATG
TYLCCHV_CP_N0040441       GTTAGAGATCGACGTCCAAGTGG---TACTCCTATGGATTTCCAGCAAGTCTTTAATTGT
TYLCV_ALM_CP_AJ489258     GTCCGTGATAGAAGGCCCTATGGAAGCAGCCCAATGGATTTTGGACAGGTTTTTAATATG
                          ** .*:** .*:.* ** : :**        ** .:.*****  .  *.** ***** :

TYLCSV-ES_CP              TTTGATAATGAACCTAGTACGGCTACTGTGAAGAACGATTTAAGGGATAGGTACCAAGTA
TOLCNDV_CP_HQ658479       TTTGACAATGAACCGAGCACAGCAACGGTGAAGAACATGCATCGTGATCGTTATCAAGTG
TYLCV_MLD_AF071228_CP     TTCGATAATGAGCCCAGTACCGCAACAGTGAAGAATGATTTGCGGGATAGGTTTCAAGTG
TYLCCHV_CP_N0040441       TATGACAATGAACCTAGCACTGCTACTGTAAAAAACGATTTAAGGGATCGTTTTCAAGTT
TYLCV_ALM_CP_AJ489258     TTCGATAATGAGCCCAGTACCGCAACCGTGAAGAATGATTTGCGTGATAGGTTTCAAGTG
                          *: ** *****.** ** ** **:** ** **.**.** .:  : .* ***.* *: *****

TYLCSV-ES_CP              ATGAGGAAGTTCCATGCCACGGTGGTAGGTGGTCCGTCAGGGATGAAGGAGCAGTGTCTG
TOLCNDV_CP_HQ658479       TTAAGGAAGTGGCATGCGACTGTGACGGGAGGAACATATGCATCTAGGGAGCAAGCATTA
TYLCV_MLD_AF071228_CP     ATGAGGAAATTTCATGCTACAGTTATTGGTGGGCCCTCTGGAATGAAGGAACAGGCATTA
TYLCCHV_CP_N0040441       CGTCGTAAATTTTATTCAACGGTTACGGGCGGACAGTATGCTTGTAAGGAGCAGGCGTTG
TYLCV_ALM_CP_AJ489258     ATGAGAAAATTTCATGCAACAGTTATTGGTGGGCCCTCTGGAATGAAGGAACAGGCATTA
                             .* **.*    ** * ** ** .  ** **  .. *.:*  :  *.***.**.    *.
```

Figure 42. **Alignment of CP genes from the species Sardinia, New Delhi, Mild, China and Israel of TYLCV.** Conserved regions are highlighted.

Guide RNA sequences are species-specific, with the exception of the gRNA targeting the CP of the Israel and Mild species (which are closely related). The gRNA for the CP of TYLCV Sardinia and the Rep2 gRNA for TYLCV New Delhi are the only ones not to target a conserved region; they were chosen because of the impossibility of finding a gRNA with an appropriate score in a conserved region.

3.2.2 Transient expression assays

Transient expression assays were performed to evaluate the ability of the CRISPR/Cas9 constructs directed at conserved regions of TYLCV Sardinia and TYLCV Israel to reduce viral replication in plants. The 2IR-GFP system of *N. benthamiana* (Morilla *et al.*, 2006) is a reporter system based on the green fluorescent protein GFP to detect and localize viral infections in whole plant tissues. 2IR-GFP plants have been stably transformed with a transgene containing a GFP expression cassette, containing the 35S promoter, the GFP coding sequence and the Tnos terminator, flanked by two direct repeats of the virus intergenic region (IR), which contains the origin of replication and the viral transcriptional regulators (Fig. 43). Two types of 2IR-GFP plants were developed by the group of Eduardo Bejarano, one carrying the IR of TYLCV Israel, and the other carrying the IR of TYLCSV. The presence of the 2IR-GFP transgene does not alter the infection and replication pattern of the virus and the plant develops disease symptoms, but the recognition of the IR by the virus causes the formation of an episomal circular DNA molecule resulting from transcription and circularization of the 2IR-GFP cassette. The viral Rep protein is necessary to transcribe and produce this episomal DNA, and its resulting accumulation and GFP expression are directly related to the viral titer.

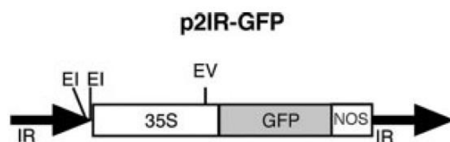


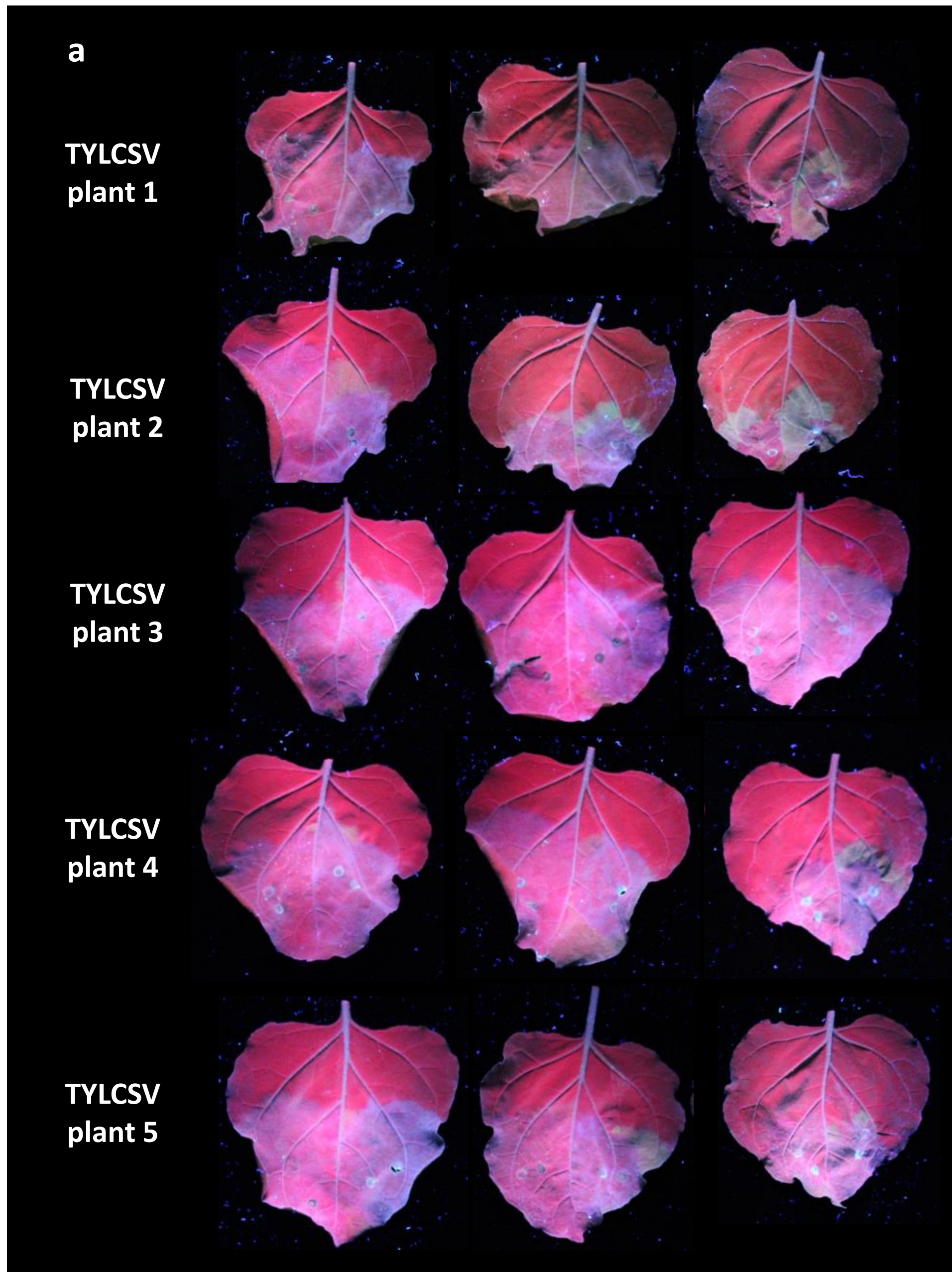
Figure 43. The 2IR-GFP cassette. EI and EV represent restriction sites for *Eco*RI and *Eco*RV. Morilla *et al.* (2006).

A reduction in the GFP levels was appreciated by visual inspection of 2IR-GFP *N. benthamiana* leaves under UV light at 4 dpi, in regions infiltrated with the CRISPR/Cas9 construct; these observations clearly pointed to co-inoculation of the CRISPR construct and the infectious clone of the virus as the optimal time point to induce a change in observable GFP levels (Table 25).

	TYLCSV + CRISPR TYLCSV	TYLCV + CRISPR TYLCV
Day -2	0	62.5
Day -1	0	50
Day 0	100	87.5
Day +1	0	25
Day +2	14.3	25

Table 25. Percentage of samples in which GFP levels were lower in treated leaves than in the control, by visual inspection.

Based on this information, the experiment was repeated focusing on co-infiltrating infectious clones and CRISPR constructs. Again, in order to standardize results, each leaf was infiltrated on the left side of the midrib with the viral infectious clone plus its corresponding CRISPR/Cas9 construct, and on the right side of the midrib with the viral infectious clone and the control vector. This allows to evaluate fluorescent signals in the context of the same leaf, excluding differences related to leaf age or metabolism. Leaves were collected at 4 dpi and observed under UV light (Fig. 44a-b).


a
TYLCSV
plant 1
TYLCSV
plant 2
TYLCSV
plant 3
TYLCSV
plant 4
TYLCSV
plant 5

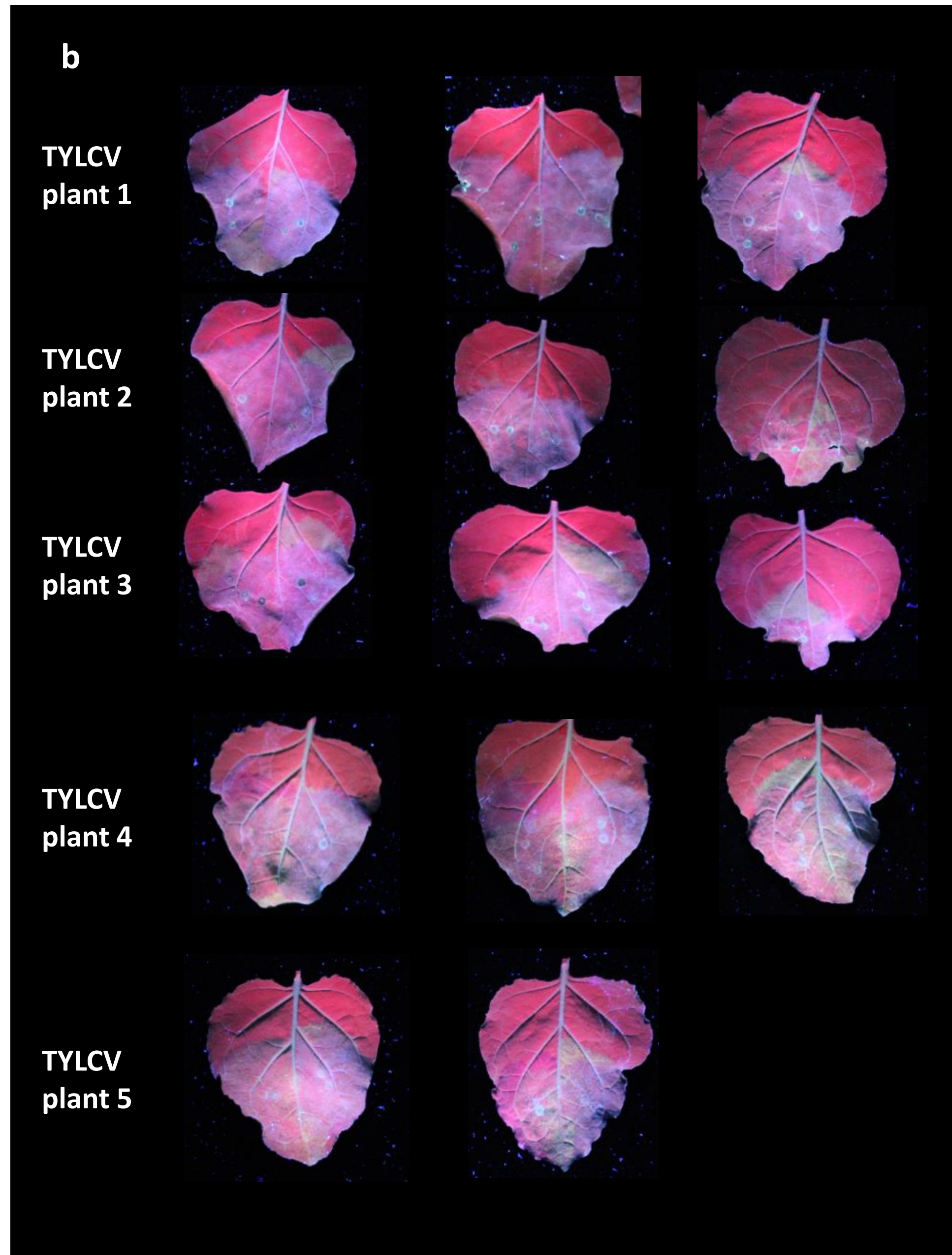


Figure 44a-b. Infiltrated TYLCSV (a) and TYLCV (b) 2IR-GFP leaves. On the left side of the midrib, for each leaf the correspondent CRISPR construct was co-infiltrated with the infectious clone of the virus, while on the right side of the midrib only the virus is present, with a control construct. Each row represents 3 leaves (replicates) from the same plant, except TYLCV plant 5, for which one leaf was damaged.

The extraction and amplification of viral DNA from the two sides of each leaf allowed to show a difference in TYLCSV and TYLCV accumulation between tissues which were or were not infected with the CRISPR/Cas9 construct. Amplification of the viral DNA was performed in the pooled genetic material from the 3 inoculated leaves for each plant, so that quantification represents a mean of the viral charge in the 3 tissues (Fig. 45a-b).

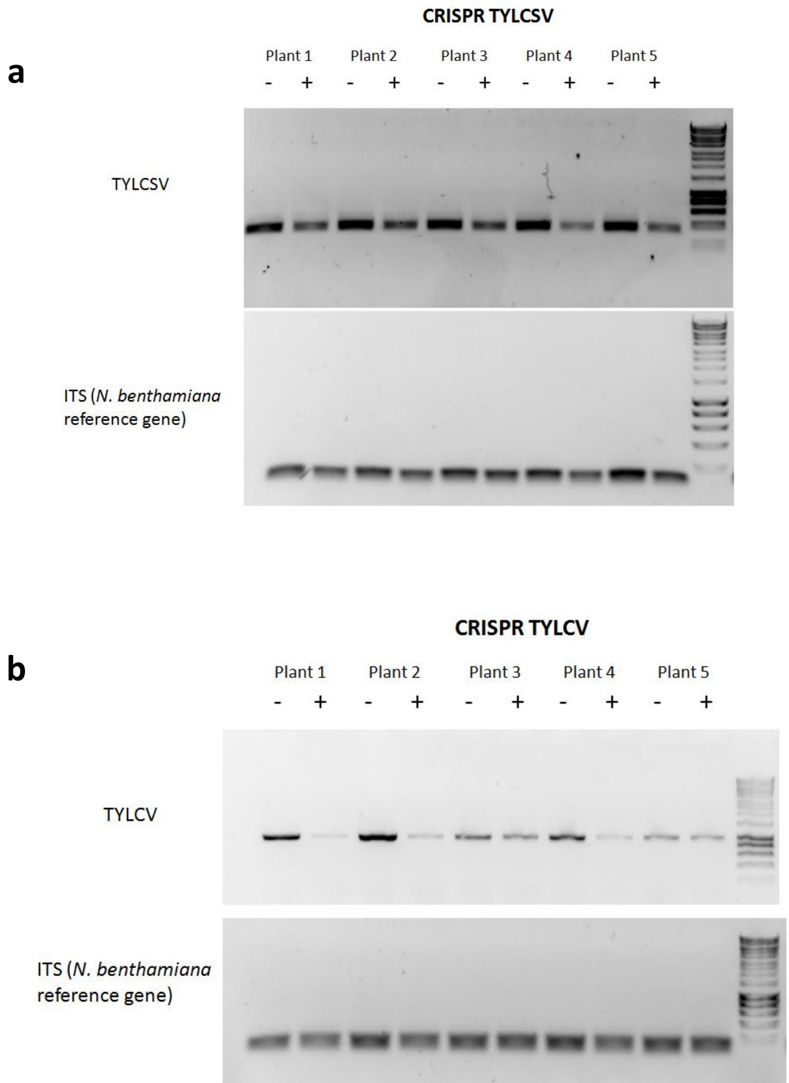


Figure 45a-b. Quantification of viral DNA. TYLCSV (a) and TYLCV (b). For each plant, the lane "-" represents tissues infected with the virus, without the CRISPR/Cas9 construct, while the lane "+" represents tissue infected with the virus and the correspondent CRISPR/Cas9 construct. The ITS *N. benthamiana* gene was used as reference. The outer right lane is occupied by the molecular marker.

The reduction is evident in all analyzed individuals and is much more apparent from DNA amplification than it is from observation of GFP levels in leaves, although some differences can readily be identified even at visual inspection. These results clearly point to a reduction in viral replication in the presence of the CRISPR/Cas9 construct, confirming previous findings (Ali *et al.*, 2015, 2016; Tashkandi *et al.*, 2017) on the ability to transpose the Prokaryotic CRISPR-mediated defense system to plants, and especially to the model species *N. benthamiana*. In addition to confirming the applicability of CRISPR/Cas9 in containing infections from DNA viruses, our results are also one of the first proofs of the functionality of the 3 gRNA multiplexing GoldenBraid CRISPR system.

3.2.3 *TYLCSV* and *TYLCV* genotyping of stable transformants

All the regenerated 'Muchamiel' and 'Pera' plants carried the transgene, as assessed by amplifying the *hCas9* gene (Fig. 46). Since the target of the CRISPR construct was not endogenous, it was not possible to further assess the efficacy of the gene editing tools. Unfortunately, although the plants were able to grow and produce fruits in the greenhouse, these were seedless and it was not possible to obtain a T_1 generation in which to assess the stability of the transgene and to challenge with the corresponding TYLCV Israel and TYLCSV to evaluate tolerance.

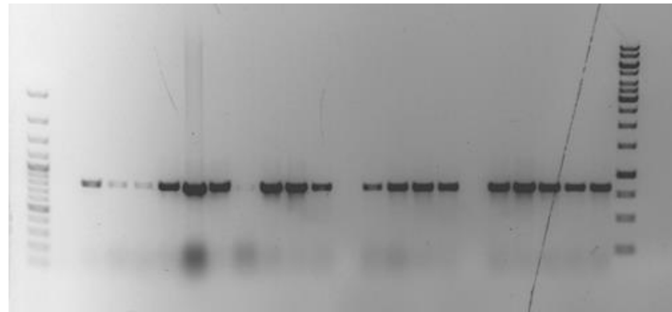


Figure 46. Amplification of the *Cas9 locus* in 'Muchamiel' and 'Pera' TYLCV plants. Left to right: negative control, positive control, 9 'Muchamiel' *TYLCV Israel* plants, 4 'Muchamiel' *TYLCV Sardinia* plants and 5 'Pera' *TYLCV Israel* plants. The outer left and right lanes are occupied by the molecular marker.

We planned to perform tolerance assays to TYLCV and TYLCSV on the T_1 generation, and therefore we were not able to assess the phenotype of our mutants in the T_0. These experiments would have been conducted in collaboration with the group of Eduardo Bejarano at the IHSM-UMA-CSIC in Malaga. Stable transformation of tomato with our virus-targeting CRISPR constructs will be repeated to obtain viable seeds.

4 - Discussion

In this work, we used both recently emerged approaches for CRISPR-mediated virus resistance. The first approach aims at editing a *locus* in the host genome, to alter host-pathogen interactions in a way which interferes with pathogen growth and replication: we did this by targeting the *Ty5/pelo locus* in tomato, which was shown to be involved in Begomovirus resistance (Lapidot *et al.*, 2015). The second approach mimics the function of CRISPR in Bacteria and Archaea, and uses gRNAs directed at viral sequences to interfere with virus multiplication: this approach was tested against two species of TYLCV. The two approaches will be discussed separately in the following paragraphs.

4.1 *Ty5/pelo* CRISPR/Cas9 editing

4.1.1 Efficiency and patterns of mutation in *Ty5/pelo* lines

The proportion of edited plants in the *Ty5/pelo* lines is extremely high, with 30 of 32 recovered plants which were effectively mutated. Also, the average editing efficiency is over 90%, with some individuals showing a mutation rate of nearly 100%. The overall editing landscape is more complex than the one described for *gf1* in the previous Chapter, with more chimerism and individuals with low-frequency different mutations, possibly indicating that in these individuals mutations occurred at different moments during the development of the plant and not before shoot formation. Other individuals, like those originating from *callus* 8, are instead highly homogeneous, with only one type of mutation.

In all edited plants, the most frequent mutation (present in 29 of 30 edited plants) was the insertion of a T at the break site 3 nucleotides upstream of the PAM. The very high occurrence of this genotype at first seemed to point to an artifact, but this hypothesis was readily discarded: the wt sequence does not have such an insertion, and aligns perfectly to the annotated *Ty5/pelo* sequence; the few non-edited individuals (like 7B) were also consistent with the wt sequence. Finally, plant 19A had a relatively low editing efficiency (28%), but both its edited alleles were small deletions. So, though highly recurring, this particular mutation was not ubiquitous and might rather be indicative of a preferential repair mechanism. As discussed in the previous Chapter, the insertion of a T or A at the break site is particularly common in CRISPR/Cas9 edited plants. In this case, its high recurrence might also be due to microhomology, since the two nucleotides in positions 16 and 17 in the gRNA

are Ts: this might increase the probability of this particular kind of repair mechanism. Apart from this single base insertion, the most common mutations are deletions of 1 to 4 nucleotides; deletions of 6 to 22 bases are present with only one occurrence for each genotype, while a 29 nucleotide deletion is present in two plants deriving from *callus* 5.

The *Ty5/pelo* gRNA seems to be more efficient than the *gf1* gRNA (described in Chapter I) in inducing mutations in a high number of shoots. However, the score calculated by Benchling when designing this gRNA was not particularly high (it was 42, so actually below the indicated optimal value of 50 for on-target activity) and this gRNA was chosen because it represented the best compromise between position and score. Some favorable characteristics were described for gRNA activity, especially based on mammalian (human and murine) cell models (Doench *et al.*, 2014, 2016). Among these are the identity of the variable base of the PAM (with cytosine being strongly preferred) and the identity of the nucleotide in position 20 at the gRNA 3' (with a strong bias towards guanine). Oddly, our *Ty5/pelo* gRNA does not possess either of these features. It is possible that models designed on mammalian cells and sequence contexts may not be readily transferred to plant cells, but also that other factors, involving the thermodynamic characteristics of the gRNA-target recognition, might play a role in determining editing efficiency.

Another factor emerging from the genotyping of *Ty5/pelo* edited plants regards the sensitivity and reliability of the T7EI assay and of Sanger sequencing in assessing mutations. Some plants (4B and 8B-C-D-E) were homozygous for the insertion of a T, and editing efficiencies as determined by Sanger sequencing and TIDE analysis were reported to be virtually of 100%. These same samples (with the exception of 8C) were positive for the T7EI assay, which shouldn't in theory be able to distinguish any heteroduplex, since the alleles are uniformly edited. It is possible to think that, however scarce, some unedited copies of the *locus* can still be present in the pool, and that the sensitivity of the T7EI assay might be able to detect them. On the other hand, Sanger sequencing of the PCR product and TIDE chromatogram decomposition, as opposed to cloning and sequencing, would not allow to detect substitutions, because the analysis is based on detection of indels. This is not regarded as a problem, however, since substitutions have not been reported among CRISPR-induced mutations in plants (Bortesi *et al.*, 2016). The most informative analysis would undoubtedly be next generation sequencing, allowing a greater sequencing depth and the analysis of all types of variants.

4.1.2 *Ty5* phenotype and alternative strategies for editing

The Pelo protein is part of a complex that mediates the dissociation of ribosomal subunits at the end of protein synthesis, a necessary step for the recycling of ribosomes and the start of a new translation cycle. Its mutation, beside affecting plant protein synthesis, heavily affects virus replication, since it depends entirely on the host's molecular machinery. A few naturally occuring mutations of this *locus* are known in plants and animals, and have been associated, among other phenotypes, to virus resistance. In the tomato TYLCV-tolerant line TY172, Lapidot *et al*. (2015) characterized a transversion in the *Ty5/pelo locus*, resulting in a Valine to Glycine substitution: both aminoacids are alifatic, and the effect of such a mutation is therefore predicted to be less severe for plant growth than an indel or a change in the open reading frame. These authors also performed gene silencing through RNA interference and did not observe abnormal development or reduced growth; RNAi-mediated gene silencing, however, is often reported to only reduce gene expression, and in this case caused a 9-fold reduction in gene expression. In both cases (Val→Gly substitution and the use of RNAi-mediated gene silencing) the plant might still be able to produce a sufficient amount of the protein to sustain ribosome function. On the other hand, the shifting of the open reading frame caused by CRISPR/Cas9 is extremely more disruptive, and, in our case, the insertion of the T between nucleotides 17 and 18 of the gRNA causes the introduction of a premature stop codon after aminoacid 91. Stop codons are introduced around the same position also with other small indels. The mutations induced by CRISPR/Cas9 in our *Ty5/pelo* plants are represented in Figure 47.

The negative phenotype observed in all transformants, which were yellowing and wilting, was likely to be attributed to the editing of the *Ty5/pelo locus* leading to a lethal phenotype because of a severe disruption of protein synthesis due to inhibition of ribosome recycling. Interestingly, a rice *pelo* mutant was recently reported to be resistant to multiple races of *Xanthomonas oryzae* pv. *oryzae*, and also to have reduced growth, but not yellowing (Zhang *et al.*, 2017). Genotyping of these mutants also identified a single base substitution (T→A) leading to the mutation of a phenylalanine to an isoleucine at position 186 in the protein sequence; these are, again, both non-polar aminoacids. Therefore, it seems that the mutations that can be tolerated at this *locus*, albeit with a reduction in plant growth (like in rice) or in number and size of fruits (like in tomato), are those which do not entirely disrupt protein function.

Figure 47. Alignment of the initial regions of the wt Ty5/Pelo protein and of the mutated versions of the protein produced by gene editing. The first line represents the wild type, while the second represents the most abundant mutation in the gene. Values on the left indicate the size of the DNA deletion.

It is possible that the selection of heterozygotes in the T_0 generation might help recover plants less severely affected by the mutation. Initially, individuals 4C and 19A, which retained a considerable proportion of the wt allele, were not considered to be grown in the greenhouse and to generate progeny, because their genotype wouldn't have been helpful to fix the mutation. Even the maintenance of heterozygosity at the target *locus*, however, is not a practical solution, because it would imply selecting heterozygotes at every generation. Alternative gene editing strategies would be less straightforward than CRISPR-mediated knock-out, and could imply the substitution of the *Ty5/pelo* wild type allele with an edited one carrying a suitable substitution. A different route for gene editing might be represented by the technique described by Shimatani *et al.* (2017), who used a dCas9-fused activation-induced cytidine deaminase (Target-AID) to introduce single base substitutions in rice and tomato. Cytidine deaminases turn a cytosine base into a uracil by deamination, thus introducing a G-U mismatch, which the cell will repair by changing the U to a T, converting the original G-C base pair to an A-T base pair. This approach is simpler, in principle, than gene targeting, and the authors reported that the efficiency of base substitutions ranged from 26.2 to 53.8% in some plant lines. Zong *et al.* (2017) also reported the use of dCas-fused cytidine deaminases to induce base

substitutions in rice, wheat and maize with efficiencies ranging from 2 to 8%, and Li *et al.* (2017) reported the same strategy in rice. By selecting an appropriate codon in the *Ty5/pelo locus*, it would be possible to induce a targeted aminoacidic substitution resembling the one described in the TY172 line using this gene editing approach.

4.1.3 Off-target activity

The analysis of five putative off-target *loci* by Illumina amplicon sequencing confirmed that the technique is highly specific. However, in contrast to what was observed for *gf1* (Chapter I), it was possible to see that, albeit occurring at very low frequencies, the majority of indels are concentrated in the PAM region of the putative off-target site. This points to some extent of unspecific editing activity, and was observed in particular for off-targets 1 and 3: both have 3 mismatches with respect to the target sequence, including one 3 nucleotides upstream of the PAM. In all cases indels occurred at frequencies lower than 1%, meaning also that they can be easily segregated in subsequent generations.

High specificity, which is extremely beneficial when targeting endogenous plant genes, might on the contrary constitute a setback when programming virus-resistance strategies against rapidly evolving microorganisms such as viruses, which could be able to evade plant defenses after a small number of generations.

4.2 Transformation of the host with a CRISPR/Cas9 construct directed at viral sequences

4.2.1 Choice of targets for the tomato yellow leaf curl virus

In our second approach to CRISPR-mediated virus resistance, we built CRISPR constructs carrying 3 gRNAs each, directed at the replicase and coat protein of TYLCV, and we chose a polycistronic approach in which all gRNAs are controlled by the same promoter. Placing multiple gRNAs under the control of a single U6 promoter as a polycistronic transcript, which is then processed to release individual gRNAs, is in principle the most efficient among the various multiplexing strategies which have been proposed. Many authors refer to multiplexing as the use of multiple gRNAs at the same time, with each gRNA being part of an independent transcriptional unit (Xing *et al.*, 2014; Lowder *et al.*, 2015; Zhang *et al.*, 2016); the assembly of these multiplexing constructs is made efficient through cloning systems such as GoldenGate and GoldenBraid. The approach described by Xie *et al.* (2015), which was incorporated into the GoldenBraid standard, assembles gRNAs in a polycistronic construct, separated

by tRNAs, reduces the size of the CRISPR/Cas9 cassette and ensures simultaneous expression; tRNAs have also been shown to enhance transcription by PolIII promoters.

Despite targeting conserved regions, because of genetic variation between the various TYLCV species it was not possible to design common gRNAs to target multiple species at the same time (Fig. 41-42), so a different construct had to be assembled for each species. Three gRNAs were chosen to target each TYLCV species, with the aim to reduce the probability of the virus being able to evade plant defenses. Target sequences were chosen among conserved functional domains, whose mutations are more likely to compromise function. Though efficient, this strategy might pose some issues, which have been already reported by Ali *et al.* (2016) and Tashkandi *et al.* (2017): especially that NHEJ-mediated repair might induce the formation of new genetic variants of these *loci* and that at least some of them might be functional, allowing the virus to evade CRISPR defenses. Also, in the case of mixed infections, which are very common in open field conditions, the induction of breaks might favor recombination between different TYLCV species. However, it is also possible to envisage that three breakpoints in the viral genome, if they occur simultaneously, would disrupt it in a way that significantly hampers repair.

Finally, it is to be considered that one possible downside of this approach to virus resistance is that it is not possible to segregate the transgene because it would imply losing the immune defense of the plant; in this respect, these tomatoes are analogous to conventional transgenics, with the difficulties this implies for regulation.

4.2.2 Transient expression of TYLCV and TYLCSV CRISPR constructs in *N. benthamiana*

N. benthamiana is widely used as a model species for functional studies in plant genetics, thanks to its amenability to agroinfiltration and virus-induced gene silencing (VIGS). In addition, it also appears to be susceptible to the vast majority of plant viruses, including the major viruses infecting tomato, and to a number of other bacterial and fungal pathogens, which make it an outstanding model for the study of plant-pathogen interactions (Goodin *et al.*, 2016). The 2IR-GFP system is a convenient method to track and assess virus replication in a visual way. In our experiments, we agroinfiltrated the same leaf with our CRISPR construct and the infectious clone of the virus on one side of the midrib, while on the other side of the midrib we infiltrated a control vector and the infectious clone. This allowed us to standardize every observation to the same leaf tissue, while the midrib prevented overlapping of the two conditions.

The first evaluation of the activity of our CRISPR constructs allowed us to understand whether it was appropriate to allow the CRISPR construct to be expressed before the plant was challenged with the virus. The greatest effect was observable when the virus and the CRISPR construct were co-infiltrated; this indicates that the CRISPR construct can be expressed efficiently at a rate that is comparable to virus replication. On the other hand, when the CRISPR construct was infiltrated 1 or 2 days after the virus, it could not contain it as efficiently. The second infiltration experiment confirmed the efficacy of co-infiltration and a difference is observable between portions of the leaves expressing the CRISPR construct targeting the virus, and the ones expressing a mock CRISPR construct directed at tomato targets. The subsequent extraction and amplification of viral DNA allowed to analyze it in a semi-quantitative way on gel: here, the difference between CRISPR-infiltrated tissues and controls appeared more pronounced than by observation of GFP fluorescence in leaves. This indicates that the 2IR-GFP reporter system is useful in providing a preliminary indication of CRISPR activity, but then a more accurate, quantitative analysis is necessary to assess its effect. This preliminary evaluation of the activity of the CRISPR constructs was done to assess their efficiency before proceeding with stable transformation of tomato.

4.2.3 TYLCV/TYLCSV stable transformation

Two Spanish local varieties, ‘Muchamiel’ and ‘Pera’, were chosen for stable transformation with the aim of introducing virus resistance in susceptible, traditional cultivars. Each variety was separately transformed with two constructs, one against TYLCV and the other against TYLCSV, which had been tested through transient expression. The transformation was successful in obtaining regenerated plants which had integrated the transgene, especially for the ‘Muchamiel’ variety. Some difficulty was found in the regeneration of the ‘Pera’ variety, because *calli* suffered sudden necrotic damages. Regenerated plants produced normal fruits, which unfortunately did not produce seeds, and it was therefore impossible to obtain a T_1 generation which would have been used for virus resistance *in vivo* assays. The failure to produce seeds is most likely to be attributed to failure of autofertilization, while no factors specifically regarding the transgene can be thought to be implicated in this phenotype.

5 - Conclusions

The possibility to apply a gene editing approach to engineering virus resistance in plants undoubtedly represents an attractive perspective but, as demonstrated by our results, also poses some challenges. Two strategies were tested in tomato, one directed at mutating endogenous tomato targets to confer resistance against Begomoviruses, the other aiming at providing the plant with a genetic, specific defense system against TYLCV, based on the CRISPR/Cas9 system. We were successful in regenerating transformed and edited plants, with high editing efficiencies, but we failed to produce a progeny from the T_0 in both of our approaches. This prevented us from conducting infection assays using TYLCV and TYLCSV, which would have provided fundamental information on plant-virus interactions in our genotypes.

Nonetheless, the present results allowed us to reach meaningful conclusions. Regarding the targeting of *Ty5/pelo*, we established that a complete disruption of the gene is lethal for the plant and has serious detrimental effects on plant physiology and development, including severe yellowing and wilting. From recent literature, it appears that only substitutions can be sustained by the plant and confer virus tolerance, which indicates that a better editing strategy might involve targeted base editing. Promising results were obtained through transient expression of the CRISPR/Cas9 costructs targeting TYLCV and TYLCSV in 2IR-GFP *N. benthamiana* plants, indicating that the system can effectively cut viral DNA, reducing its replication rate. The present results confirm the functionality of GoldenBraid multiplexing CRISPR vectors, based on the expression of a cassette of gRNAs as a polycistronic transcript, and on their subsequent processing and release. This feature further enhances the resemblance between this synthetic strategy and the natural prokaryotic CRISPR/Cas9 system. Stable transformation of tomato with these constructs represents a valuable strategy that needs to be further investigated.

Bibliography

Ali, Z. *et al.* (2015) ‘CRISPR/Cas9-mediated viral interference in plants’, *Genome Biology*, 16(1), p. 238. doi: 10.1186/s13059-015-0799-6.

Ali, Z. *et al.* (2016) ‘CRISPR / Cas9-Mediated Immunity to Geminiviruses : Differential Interference and Evasion’, *Scientific Reports*. Nature Publishing Group, (May), pp. 1–11. doi: 10.1038/srep26912.

Ammara, U. *et al.* (2015) ‘RNA interference-based resistance in transgenic tomato plants against Tomato yellow leaf curl virus -Oman (TYLCV-OM) and its associated betasatellite’, *Virology Journal*, 12(38). doi: 10.1186/s12985-015-0263-y.

Anbinder, I. *et al.* (2009) ‘Molecular dissection of Tomato leaf curl virus resistance in tomato line TY172 derived from Solanum peruvianum’, *Theoretical and Applied Genetics*, 119, pp. 519–530. doi: 10.1007/s00122-009-1060-z.

Barnes, H. E. *et al.* (2014) ‘Selective Microbial Genomic DNA Isolation Using Restriction Endonucleases’, *PLoS ONE*, 9(10). doi: 10.1371/journal.pone.0109061.

Bastet, A., Robaglia, C. and Gallois, J.-L. (2017) ‘eIF4E Resistance : Natural Variation Should Guide Gene Editing’, *Trends in Plant Science*. Elsevier Ltd, 22(5), pp. 411–419. doi: 10.1016/j.tplants.2017.01.008.

Becker, T. *et al.* (2012) ‘Structural basis of highly conserved ribosome recycling in eukaryotes and archaea’, *Nature*. Nature Publishing Group, 482(7386), pp. 501–506. doi: 10.1038/nature10829.

Bortesi, L. *et al.* (2016) ‘Patterns of CRISPR/Cas9 activity in plants, animals and microbes’, *Plant Biotechnology Journal*, 14(12), pp. 2203–2216. doi: 10.1111/pbi.12634.

Butterbach, P. *et al.* (2014) ‘Tomato yellow leaf curl virus resistance by Ty-1 involves increased cytosine methylation of viral genomes and is compromised by cucumber mosaic virus infection’, *PNAS*, 111(35). doi: 10.1073/pnas.1400894111.

Chandrasekaran, J. *et al.* (2016) ‘Development of broad virus resistance in non-transgenic cucumber using CRISPR / Cas9 technology’, *Molecular Plant Pathology*, 17(7), pp. 1140–1153.

Chen, H. *et al.* (2015) ‘Resistance to viral yellow leaf curl in tomato through RNAi targeting two Begomovirus species strains’, *J. Plant Biochem. Biotechnol.*, (DOI 10.1007/s13562-015-0325-7). doi: 10.1007/s13562-015-0325-7.

Doench, J. G. *et al.* (2014) ‘Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation’, *Nature Biotechnology*. Nature Publishing Group, 32(12), pp. 1262–1267. doi: 10.1038/nbt.3026.

Doench, J. G. *et al.* (2016) ‘Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9’, *Nature Biotechnology*. Nature Publishing Group, 34(2), pp. 184–191. doi: 10.1038/nbt.3437.

Duffy, S. and Holmes, E. C. (2008) ‘Phylogenetic Evidence for Rapid Rates of Molecular Evolution in the Single-Stranded DNA Begomovirus Tomato Yellow Leaf Curl Virus □ †’, *Journal of Virology*, 82(2), pp. 957–965. doi: 10.1128/JVI.01929-07.

Duprat, A. *et al.* (2002) ‘The Arabidopsis eukaryotic initiation factor (iso) 4E is dispensable for plant growth but required for susceptibility to potyviruses’, *The Plant Journal*, 32, pp. 927–934.

Elmer, J. S. *et al.* (1988) ‘Agrobacterium-mediated inoculation of plants with tomato golden mosaic virus DNAs’, *Plant Molecular Biology*, 10(3), pp. 225–234. doi: 10.1007/BF00027399.

Fuentes, A. *et al.* (2010) ‘Field Trial and Molecular Characterization of RNAi-Transgenic Tomato Plants That Exhibit Resistance to Tomato Yellow Leaf Curl Geminivirus’, *MPMI*, (http://dx.doi.org/10.1094/MPMI-08-15-0181-R).

Gal-On, A., Fuchs, M. and Gray, S. (2017) 'Generation of novel resistance genes using mutation and targeted gene editing', *Current Opinion in Virology*. Elsevier B.V., 26, pp. 98–103. doi: 10.1016/j.coviro.2017.07.024.

Galvez, L. C. *et al.* (2014) 'Plant Science Engineered plant virus resistance', *Plant Science*. Elsevier Ireland Ltd, 228, pp. 11–25. doi: 10.1016/j.plantsci.2014.07.006.

Gomez, M. A. *et al.* (2017) 'Simultaneous CRISPR/Cas9-mediated editing of cassava eIF4E isoforms nCBP-1 and nCBP-2 confers elevated resistance to cassava brown streak disease', *bioRxiv*, (doi: https://doi.org/10.1101/209874).

Goodin, M. *et al.* (2016) 'Nicotiana benthamiana : Its history and future as a model for plant - pathogen', *MPMI*, 21(8). doi: 10.1094/MPMI.

Hanley-Bowdoin, L. *et al.* (1999) 'Geminiviruses : Models for Plant DNA Replication , Transcription , and Cell Cycle Regulation', *Critical Reviews in Plant Sciences*, 18(1), pp. 71–106.

Hanssen, I. M., Lapidot, M. and Thomma, B. P. H. J. (2010) 'Emerging Viral Diseases of Tomato Crops', *MPMI*, 23(5), pp. 539–548.

Hutton, S. F., Scott, J. W. and Schuster, D. J. (2012) 'Recessive Resistance to Tomato yellow leaf curl virus from the Tomato Cultivar Tyking Is Located in the Same Region as Ty-5 on Chromosome 4', *HortScience*, 47(3), pp. 324–327.

Ji, Y. *et al.* (2009) 'Molecular Mapping of Ty-4 , a New Tomato Yellow Leaf Curl Virus Resistance Locus on Chromosome 3 of Tomato', *J. Amer. Soc. Hort. Sci.*, 134(2), pp. 281–288.

Ji, Y., Scott, J. W. and Schuster, D. J. (2009) 'Toward Fine Mapping of the Tomato Yellow Leaf Curl Virus Resistance Gene Ty-2 on Chromosome 11 of Tomato', *HortScience*, 44(3), pp. 614–618.

Kenyon, L. *et al.* (2014) 'Emergence and diversity of begomoviruses infecting solanaceous crops in East and Southeast Asia', *Virus Research*. Elsevier B.V., 186, pp. 104–113. doi: 10.1016/j.virusres.2013.12.026.

Lapidot, M. *et al.* (2015) 'A Novel Route Controlling Begomovirus Resistance by the Messenger RNA Surveillance Factor Pelota', *PLoS Genetics*, (DOI:10.1371/journal.pgen.1005538). doi: 10.1371/journal.pgen.1005538.

Li, J. *et al.* (2017) 'Generation of Targeted Point Mutations in Rice by a Modified CRISPR/Cas9 System', *Molecular Plant*, 10(3), pp. 526–529. doi: 10.1016/j.molp.2016.12.001.

Lowder, L. G. *et al.* (2015) 'A CRISPR/Cas9 Toolbox for Multiplexed Plant Genome Editing and Transcriptional Regulation', *Plant Physiology*, 169(2), pp. 971–985. doi: 10.1104/pp.15.00636.

Mabvakure, B. *et al.* (2016) 'Ongoing geographical spread of Tomato yellow leaf curl virus $', *Virology*. Elsevier, 498, pp. 257–264. doi: 10.1016/j.virol.2016.08.033.

Morilla, G. *et al.* (2006) 'A Versatile Transreplication-Based System To Identify Cellular Proteins Involved in Geminivirus Replication', *Journal of Virology*, 80(7), pp. 3624–3633. doi: 10.1128/JVI.80.7.3624.

Moriones, E. and Navas-Castillo, J. (2000) 'Tomato yellow leaf curl virus , an emerging virus complex causing epidemics worldwide', *Virus Research*, 71, pp. 123–134.

Moury, B. *et al.* (2014) 'Evolution of plant eukaryotic initiation factor 4E (eIF4E) and potyvirus genome-linked protein (VPg): A game of mirrors impacting resistance spectrum and durability', *Infection, Genetics and Evolution*. Elsevier B.V., 27, pp. 472–480. doi: 10.1016/j.meegid.2013.11.024.

Patrick, R. M. and Browning, K. S. (2012) 'The eIF4F and eIFiso4F Complexes of Plants : An Evolutionary Perspective', *Comparative and Functional Genomics*, (doi:10.1155/2012/287814). doi: 10.1155/2012/287814.

Praveen, S., Mishra, A. K. and Dasgupta, A. (2005) 'Antisense suppression of replicase gene

expression recovers tomato plants from leaf curl virus infection', *Plant Science*, 168, pp. 1011–1014. doi: 10.1016/j.plantsci.2004.12.008.

Pyott, D. E., Sheehan, E. and Molnar, A. (2016) 'Engineering of CRISPR / Cas9-mediated potyvirus resistance in transgene-free Arabidopsis plants', *Molecular Plant Pathology*, 17(8), pp. 1276–1288.

Qiu, D. *et al.* (2007) 'Improved protocol for Agrobacterium mediated transformation of tomato and production of transgenic plants containing carotenoid biosynthetic gene CsZCD', *Scientia Horticolturae*, 112, pp. 172–175. doi: 10.1016/j.scienta.2006.12.015.

Robaglia, C. and Caranta, C. (2006) 'Translation initiation factors : a weak link in plant RNA virus infection', *Trends in Plant Science*, 11(1). doi: 10.1016/j.tplants.2005.11.004.

Sanfaçon, H. (2015) 'Plant Translation Factors and Virus Resistance', *Viruses*, 7, pp. 3392–3419. doi: 10.3390/v7072778.

Sato, M. *et al.* (2005) 'Selective involvement of members of the eukaryotic initiation factor 4E family in the infection of Arabidopsis thaliana by potyviruses', *FEBS Letters*, 579, pp. 1167–1171. doi: 10.1016/j.febslet.2004.12.086.

Scott, J. W., Hutton, S. F. and Freeman, J. H. (2015) 'Fla . 8638B and Fla . 8624 Tomato Breeding Lines with Begomovirus Resistance Genes ty-5 Plus Ty-6 and Ty-6 , respectively', *HortScience*, 50(9), pp. 1405–1407.

Shimatani, Z. *et al.* (2017) 'Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion', *Nature Biotechnology*. Nature Publishing Group, 35(5), pp. 441–443. doi: 10.1038/nbt.3833.

Sun, S. *et al.* (2014) 'Transient expression of siRNA targeted against the TYLCV AV1 , AC1 and AC3 genes for high resistance in tomato', *Scientia Horticulturae*. Elsevier B.V., 179, pp. 321–327. doi: 10.1016/j.scienta.2014.10.002.

Tabein, S. *et al.* (2017) 'Pyramiding Ty-1 / Ty-3 and Ty-2 in tomato hybrids dramatically inhibits symptom expression and accumulation of tomato yellow leaf curl disease inducing viruses', *Archives of Phytopathology and Plant Protection*. Taylor & Francis, 50(5–6), pp. 213–227. doi: 10.1080/03235408.2017.1287234.

Tashkandi, M. *et al.* (2017) 'Engineering resistance against Tomato yellow leaf curl virus via the CRISPR / Cas9 system in tomato', *bioRxiv*, (doi: http://dx.doi.org/10.1101/237735).

Vanderschuren, H. *et al.* (2007) 'Engineering resistance to geminiviruses – review and perspectives', *Plant Biotechnology Journal*, 5, pp. 207–220. doi: 10.1111/j.1467-7652.2006.00217.x.

Varma, A. and Malathi, V. G. (2003) 'Emerging geminivirus problems : A serious threat to crop production', *Ann. appl. Biol.*, 142, pp. 145–164.

Verlaan, M. G. *et al.* (2013) 'The Tomato Yellow Leaf Curl Virus Resistance Genes Ty-1 and Ty-3 Are Allelic and Code for DFDGD-Class RNA – Dependent RNA Polymerases', *PLoS Genetics*, 9(3). doi: 10.1371/journal.pgen.1003399.

Vidavski, F. *et al.* (2008) 'Pyramiding of genes conferring resistance to Tomato yellow leaf curl virus from different wild tomato species', *Plant Breeding*, 127, pp. 625–631. doi: 10.1111/j.1439-0523.2008.01556.x.

Wang, X. *et al.* (2013) 'Silencing of the Host Factor eIF (iso) 4E Gene Confers Plum Pox Virus Resistance in Plum', *PLoS ONE*, 8(1). doi: 10.1371/journal.pone.0050627.

Wu, X. *et al.* (2014) 'pelo Is Required for High Efficiency Viral Replication', *PLoS Pathogens*, 10(4). doi: 10.1371/journal.ppat.1004034.

Xie, K., Minkenberg, B. and Yang, Y. (2015) 'Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system', *Proceedings of the National Academy of Sciences*, 112(11), pp. 3570–3575. doi: 10.1073/pnas.1420294112.

Xing, H. L. *et al.* (2014) 'A CRISPR/Cas9 toolkit for multiplex genome editing in plants', *BMC Plant Biology*, 14(1), pp. 1–12. doi: 10.1186/s12870-014-0327-y.

Yang, X. *et al.* (2014) 'Fine mapping of the tomato yellow leaf curl virus resistance gene Ty - 2 on chromosome 11 of tomato', *Mol Breeding*, 34, pp. 749–760. doi: 10.1007/s11032-014-0072-9.

Zaidi, S. S. *et al.* (2016) 'Engineering Plant Immunity : Using CRISPR / Cas9 to Generate Virus Resistance', *Frontiers in Plant Science*, 7(November), pp. 1–10. doi: 10.3389/fpls.2016.01673.

Zhang, X.-B. *et al.* (2017) 'A substitution mutation in *OsPELOTA* confers bacterial blight resistance by activating the salicylic acid pathway', *Journal of Integrative Plant Biology*, XXXX(Xxxx). doi: 10.1111/jipb.12613.

Zhang, Z. *et al.* (2016) 'A multiplex CRISPR/Cas9 platform for fast and efficient editing of multiple genes in Arabidopsis', *Plant Cell Reports*, 35(7), pp. 1519–1533. doi: 10.1007/s00299-015-1900-z.

Zong, Y. *et al.* (2017) 'Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion', *Nature Biotechnology*. Nature Publishing Group, 35(5), pp. 438–440. doi: 10.1038/nbt.3811.

Component	12 µl reaction	Final concentration
pDGB3 alpha1 destination vector	1 µl	6.25 ng μl^{-1}
GB1001 pU6-26 PolIII promoter	1 µl	6.25 ng μl^{-1}
gRNA 2 µM	1 µl	0.17 µM
GB0645 scaffold RNA	1 µl	6.25 ng μl^{-1}
10X T4 ligase buffer	1.2 µl	1X
10X BSA	1.2 µl	1X
T4 ligase	1 µl	3 U
*Bsa*I	1 µl	10 U
Water	3.6	N/A

Supplementary Table S8. First GoldenBraid consecutive reaction for the assembly of a CRISPR-Cas vector to target the *Ty5/pelo locus*. Assembly of the pDGB3 alpha1 vector carrying the U6 promoter, the *Ty5/pelo* gRNA and the scaffold RNA.

Component	12 µl reaction	Final concentration
pDGB3 omega2 destination vector	1 µl	6.25 ng μl^{-1}
pDGB3 alpha1 U6-26:gRNA: scaffold	1 µl	6.25 ng μl^{-1}
GB0639 35S:hCas9: Tnos	1 µl	6.25 ng μl^{-1}
10X T4 ligase buffer	1.2 µl	1X
10X BSA	1.2 µl	1X
T4 ligase	1 µl	3 U
*Bsmb*I	1 µl	10 U
Water	4.6	N/A

Supplementary Table S9. Second GoldenBraid reaction for the assembly of a CRISPR-Cas vector to target the *Ty5/pelo locus*. Assembly of the pDGB3 omega2 vector carrying the insert of the previously assembled pDGB3 alpha1 U6:gRNA:scaffold and the hCas9 TU with the 35S promoter and Tnos terminator.

Component	12 µl reaction	Final concentration
pDGB3 alpha2 destination vector	1 µl	6.25 ng μl^{-1}
pDGB3 omega2 U6-26:gRNA: scaffold - 35S:hCas9:Tnos	1 µl	6.25 ng μl^{-1}
GB1181 alpha1R Pnos:nptII:Tnos	1 µl	6.25 ng μl^{-1}
10X T4 ligase buffer	1.2 µl	1X
10X BSA	1.2 µl	1X
T4 ligase	1 µl	3 U
*Bsa*I	1 µl	10 U

Supplementary Table S10. Third GoldenBraid reaction for the assembly of a CRISPR-Cas vector to target the *Ty5/pelo locus*. Assembly of the final pDGB3 alpha1 vector carrying the gRNA and hCas9 TUs and the NptII selection marker with the Pnos promoter and Tnos terminator.

Component	12 µl reaction	Final concentration
GB0307 pUPD2 destination vector	1 µl	6.25 ng µl-1
GB1205/GB1206/Gb1207 pVD1	1 µl	6.25 ng µl-1
gRNA1/gRNA2/gRNA3 1 µM	1 µl	0.17 µM
10X T4 ligase buffer	1.2 µl	1X
10X BSA	1.2 µl	1X
T4 ligase	1 µl	3 U
BsaI	1 µl	10 U
Water	3.6	N/A

Supplementary Table S11. GoldenBraid reactions for the assembly of pUPD2 (level 0) vectors for the multiplexing assembly of gRNAs. Each of reaction allows to domesticate one gRNA for a multiplexing construct. The outcome of every reaction is a pUPD2 vector carrying a tRNA:gRNA:scaffold insert. tRNA and scaffold RNA are specific for every position of the multiplexing construct, because their external 4 bp overhangs specify their position (1, 2 or 3) relative to each other and to the promoter (which will be added in the following step). tRNAs and gRNAs are carried on pVD1 vectors: GB1205 specifies position 1 and is assembled with gRNA1, GB1206 specifies position 2 and is assembled with gRNA2, while GB1207 specifies position 3 and is assembled with gRNA3. These reactions were performed in parallel for all gRNAs and for all virus species (3 reactions per species).

Component	12 µl reaction	Final concentration
pDGB3 alpha1 destination vector	1 µl	6.25 ng μl^{-1}
GB1001 pU6-26 PolIII promoter	1 µl	6.25 ng μl^{-1}
pUPD2 gRNA1	1 µl	6.25 ng μl^{-1}
pUPD2 gRNA2	1 µl	6.25 ng μl^{-1}
pUPD2 gRNA3	1 µl	6.25 ng μl^{-1}
10X T4 ligase buffer	1.2 µl	1X
10X BSA	1.2 µl	1X
T4 ligase	1 µl	3 U
*Bsa*I	1 µl	10 U
Water	2.6	N/A

Supplementary Table S12. Level 1 GoldenBraid reaction for the assembly of a CRISPR-Cas vector with 3 gRNAs under the control of a single U6-26 promoter. Assembly of the pDGB3 alpha1 vector carrying the U6 promoter and the 3 tRNA:gRNA:scaffold inserts from previously assembled pUPD2 vectors. A separate vector was assembled for each species of TYLCV.

Component	12 µl reaction	Final concentration
pDGB3 omega2 destination vector	1 µl	6.25 ng μl^{-1}
pDGB3 alpha1 U6-26:gRNA1:gRNA2:gRNA3	1 µl	6.25 ng μl^{-1}
GB0639 35S:hCas9: Tnos	1 µl	6.25 ng μl^{-1}
10X T4 ligase buffer	1.2 µl	1X
10X BSA	1.2 µl	1X
T4 ligase	1 µl	3 U
*Bsmb*I	1 µl	10 U
Water	4.6	N/A

Supplementary Table S13. Level >1 GoldenBraid reaction for the assembly of a CRISPR-Cas vector with 3 gRNAs under the control of a single U6-26 promoter. Assembly of the pDGB3 omega2 vector carrying the insert of the previously assembled pDGB3 alpha1 U6:gRNA1-2-3 and the *hCas9* TU with the 35S promoter and Tnos terminator. A separate vector was assembled for each species of TYLCV.

Component	12 µl reaction	Final concentration
pDGB3 alpha1 destination vector	1 µl	6.25 ng µl^{-1}
pDGB3 omega2 U6-26:gRNA1:gRNA2:gRNA3 - 35S:hCas9:Tnos	1 µl	6.25 ng µl^{-1}
GB1181 alpha1R Pnos:nptII:Tnos	1 µl	6.25 ng µl^{-1}
10X T4 ligase buffer	1.2 µl	1X
10X BSA	1.2 µl	1X
T4 ligase	1 µl	3 U
*Bsa*I	1 µl	10 U
Water	4.6	N/A

Supplementary Table S14. Level >1 GoldenBraid reaction for the assembly of a CRISPR-Cas vector with 3 gRNAs under the control of a single U6-26 promoter. Assembly of the final pDGB3 alpha1 vector carrying the gRNAs and *hCas9* TUs and the *nptII* selection marker with the Pnos promoter and Tnos terminator. A separate vector was assembled for each species of TYLCV.

Locus	Primer Forward	Primer Reverse
OT 1	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**GTGAATATTTTCATTTTTGTG	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**CTTAAGACATATTCGCTATG
OT 2	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**GGAATCAATACTTTACTCACC	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**GGGTATGAAGTTTAGTAGTAAAA
OT 3	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**GTTTGTGATGCTTATTTAGC	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**AGTTGTGAATTCACCCTTAC
OT 4	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**GAAATGGCAGTTTGAGATCT	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**TTAAATGGGGAGTCGAAAAC
OT 5	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**AGCTGGAGGTCATGGGTTC	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**AAAAGTGTGCCAGGCGAGG

Supplementary Table S15. Primers for the first round amplification for 5 potential off-target *loci* for the *ty5/pelo* gRNA. In bold is the universal Illumina adapter, while the 3' terminal part of the primer is specific for each *locus*.

Individual	Nextera Code	Index Sequence	Oligonucleotide Sequence
1	S502	CTCTCTAT	AATGATACGGCGACCACCGAGATCTACAC**CTCTCTAT**TCGTCGGCAGCGTC
2	S503	TATCCTCT	AATGATACGGCGACCACCGAGATCTACAC**TATCCTCT**TCGTCGGCAGCGTC
3	S505	GTAAGGAG	AATGATACGGCGACCACCGAGATCTACAC**GTAAGGAG**TCGTCGGCAGCGTC
4	S506	ACTGCATA	AATGATACGGCGACCACCGAGATCTACAC**ACTGCATA**TCGTCGGCAGCGTC
5	S507	AAGGAGTA	AATGATACGGCGACCACCGAGATCTACAC**AAGGAGTA**TCGTCGGCAGCGTC
WT	S508	CTAAGCCT	AATGATACGGCGACCACCGAGATCTACAC**CTAAGCCT**TCGTCGGCAGCGTC
Locus	**Nextera Code**	**Index Sequence**	**Oligonucleotide Sequence**
OT 1	N706	TAGGCATG	CAAGCAGAAGACGGCATACGAGAT**CATGCCTA**GTCTCGTGGGCTCGG
OT 2	N707	CTCTCTAC	CAAGCAGAAGACGGCATACGAGAT**GTAGAGAG**GTCTCGTGGGCTCGG
OT3	N710	CGAGGCTG	CAAGCAGAAGACGGCATACGAGAT**CAGCCTCG**GTCTCGTGGGCTCGG
OT 4	N711	AAGAGGCA	CAAGCAGAAGACGGCATACGAGAT**TGCCTCTT**GTCTCGTGGGCTCGG
OT 5	N712	GTAGAGGA	CAAGCAGAAGACGGCATACGAGAT**TCCTCTAC**GTCTCGTGGGCTCGG

Supplementary Table S16. Primers for the second round of amplification for 5 potential off-target *loci* for the *ty5/pelo* gRNA. One series of Nextera indexes (S502-S508) was used to identify individuals, while the other (N706-N712) was used to identify *loci*. The combination of these two allows to unambiguously label a particular PCR product from a particular individual.

Supplementary Sequence List

>Ty5 Off-target1
GTTTTACCTGTTCGTTTTGAATGAAATTGATGTGAAGGTCCTCAACTTGTGTATGAGT
ATAATTGGCTTGGAAATGTCTGGGAAAAATGTTCTTGTGAATATTTTCATTTTTGTGC
TACTTTTCATATATGTATGCAATAACGTCCGAGAGGTCAAGTACGCCGAGTCACACA
CTGAGAGTGTCATATCTTTGAGGGTGTACTCTAGGGGTGTGACACATGGTGAGCCAT
AGCATACGATAATCGGCAAAATAAGGGTCATGCGCACTTCGATACTCGTTTGGCCTT
GCGAATTGGCCGGCATGGCCGCGAGGGCCAACAGACTACATAGCGAATATGTCTTAA
GGGACGTTCATGAGAAATTTGGCATTTTGGATGTTAGAAATCCGGATCACAAAAAAA
GATGAGGACTATAGCACATGCTAATCAGTAAAATGTTGGGCATGCGCACTTTATGTT
TGACCCAGTAAATGGGCCGACATGGCCGTAAGGGCCAAAAGACTGCATAGCCAATG
TCTTACGGATGTTCAATAGAAATTTCAACATTTTGGACGTTAGAAATCCAAATTACTA
ACAAAGATGGTGAGCTATAACCCACGAGAATTAGCAAAATGAGGGGTATGCGCTTTT
TGAGGTTTGTTTGACCTAGAAAATTTGGCGGCATGGTCATGAGAGCCAATTGGCTGA
ATCTTAAAGGTCTTAACAGATGTC

>Ty5 Off-target2
TTCAACCGGAGATTAAGCAAAGTAAGATTTTTAAGTTCACAAATAGAAGTCGGAATT
TCAACAGTTATATCTTTTTCTTGAAGTATTATACCGGTAACCTTCCCATCATCACACTC
AATCTCCGGCCAGCTACAAGGAGAAGAAGTAGAGTTCCATGAATCAAGTGCTGATGG
GTTTCCCCATTGACGTTTTATCTTCAGTAGAGTGTCACGTTCAGTAGTGGCCGGAGTT
ACAAAAATGGGTATGAAGTTTAGTAGTAAAATGAGATAAAGGGACTGTTTCCCGGCG
AATAAGCCGGTCAGAAAATGGGAAAACTCCATGAATAAAATTACGATTTGATTTTGT
AAAGTTTGGTCAAACATTGGAGGTGTGAAGAAAAGCCAAATTGATAAAAATGAATA
TATAGAGATAAAATTTTTCGATCAAAAGTTTATCATTTTTTTGATTAAAATAAGTAAA
GTGGGTGAGTAAAGTATTGATTCCATTGAGCTGCCTAACAATAAAGCTTCGAGATTT
GGGGCCCTTTTCTTTGACCATTGAAATGAAATTTCCGCACACTCTGCGTTTACTGTTC
AGCGACGCTAAGATGCAATCAGCGTTTTAAAAATGTAATTTTATTAATATTATTTTCA
TGAATATAAAATTTATTATTATCTTAAATTTGGTAAATATAAATAAAAACAGAGAAA
GTAATACTTTTT

>Ty5 Off-target 3
TTCTTAACTGGGAGTATCCCATTTAGTATTATGGTAACTAGCCATATGGTAGGAAAAT
TATAAGAATGTCGAGACAAATTAAAGAAATAAAAAGCAGAAAATATTTAGTAGGAA
GTTGTTTTGTTGTATTGTACCTTAGGAGGGCCCTCCAATGCATTAGTAGAATACAATC
TTGCACTGTCTACTATGTGACAGTATGTATGGAATGCACTTGCAAATCTCTTGTGTGA
TTTCAGTTGTGAATTCACCCTTACCGCTCTTCTACACATGATAGCTCTCCTGCGAAAA
CACAAGCAGAAATTAATTACACCATATTTACTAGAACCTACTCTTCACTTTCGATGTC
GTACCTATATCGGATTTTCCAAAAATACACTATTTTGGGAGAATCCACGTTGATATTT
TCGAAGAGTTCGAGTAACATATACTAGAACTAGCAAGAATACCTTATTCCTCTAATG
ACAGCTAAATAAGCATCACAAACAACTCCAACCAATTCTATTCTATATGGTCTTCTCT
TTCTATTTCCATCAGTGGACTCTTGTTCTTCATTCAACTGTTCCCAATAGTTCTCAGTT
ACAGATCCATCATCCTCCACTTTGTATCCAACACCCATACGGTAACGTCTTCGGTGTA
CATTTCTTGCCATGGTTATAGTTTGCACCACAAAGGGTATCCATGATAGTGTACCATC
CATTATGA

>Ty5 Off-target 4
CTGTCATTCAGACTCACTATGTGATTTGATATATGATCTGTTTCTGCTTCTTTATAAAA
AACATGTTTATATATTGTTTTAAAGTGGTATATATGAAATGGCAGTTTGAGATCTTAT
CACTAGCTATTGTTGCTTTCTCTTGTTCTTCACAGTGCAACACATATTACTAACTAAG
ATACAAATCGATTTAAATGAAGCTACGCTAATCTTGGTTGCTTCTCTGTCTTTTATTTG
TTTCTTTCTTTCTTATTTTCTCATCAATACTAGCTTAGTCCTAAGCTCTACCTTCGTTTT
CGCTGGAATGTTGTTCACATTATATATCATGTTTTCGACTCCCCATTTAAGGGTTACA
GCTTCTTGCTTTACTCTTGTTTATAGTCTTGTAGCTAATATTTCTTACTCATTCCATGTC
TTTGGGTTAGCTGTCGTTTCATTACTAATATACATCTCTACATTTATTGGTATTCTACA
TACATATAAACCTGTTACCCTAATAGTTATAAAAATGCTTTTTTCTTGTTCTTACTCAC
CATGCTAAATTTCTAAATTCCTGTCAGCCGTGTCATAATACCGTTCAAATCAATGTTT
TTTACTTCATTCATTTATTTTGTCATTGTTCGGTTTGTTTATTTTATTATGGTTTTTCTTC
CATTTAGGAGTTGTTTCCTTCAAAAAATAAAAAAAAGACTATGAGTGAGTTTGG

>Ty5 Off-target 5
TCAAATCATCCCTCTGAACTTGTCTTAAGAACCTTAACTAATGTATAAGAAACAAATG
TTAAATTAACCCGCAAAAAAAATGATGTATAAATTGTGGAGTAATGTTCCAGAACCTG
GAAATAAGTTGTAGATGAATTCTGGAAATTTAGAACAAAAAGAAAATATAAATAAA
AAAAGGTGAGGTCTGCTGGGATCAAACCCGAGACCTCACCAAAGTAAGCCTTAAATA
AATAAAAAAAAAAGTGTGCCAGGCGAGGCTTGAACCTGGAGGCTGGGGGCCGGCGAA
TTCGCAGGGAGAAGAAAAATAAATGGGGCTTGGAGGAATCGAACCCTTGACCCCTGT
ATTGCGCGGAAAAAAAGAGAAGGAAAAGAAGTGGGGGCTGGGGGGAGGGGATCGA
ACACGCGACCTTGGGGGAAGGAGGAGGAGAAAATTTATTAAGAAAATAAGTGGGAG
GCATGGGATTCGAACCCATGACCTCCAGCTACGCGCGAAAGAGTGGAAAATGGGGA
AAAAAGAGACGCGAGGCGTGGGAATCGATCCCACAACCTCAAGGCTGAAGGAGGGG
CAAAATTAATTAATAAAAGAAGTGAGGCTGTGGGGATTCGATCCCACAACCTCACTG
TTGTTCGCCCCATTTAAACAAATTAATGATAGAGGGGTTGTGGGGGTTCGAACCCAC
CACCTCACAGCCTCCTCTTCAGCG

Chapter III

A CRISPR/Cas9 editing protocol for the polyphenol oxidase (*ppo*) gene family in eggplant

1 - Introduction

Polyphenol oxidases (PPOs) belong to an ubiquitous and diverse protein family, whose members are found across plants, animals, fungi and bacteria. They are copper-containing metalloproteins which can oxidize various monophenols to *o*-diphenols (cresolase/monophenolase activity) and/or are able to oxidize such *o*-diphenols to their corresponding *o*-quinones (catecholase/diphenolase activity). These *o*-quinones can then self-polymerize or interact with other substances (such as aminoacids, proteins, quinones or phenols) to form high molecular weight dark pigments called melanins (Fig. 48) (Marusek *et al.*, 2006). Given the difficulty, in many instances, to identify the exact specificity of a protein, its substrates and its structure, PPO is actually a broad term used to refer to different types of enzymes catalyzing the conversion of different classes of phenols to *o*-quinones, especially in plants and fungi. This definition is somehow ambiguous, since it does not distinguish between tyrosinases, which have cresolase activity, and catacholases, which do not. More precisely, PPOs can be classified by their enzymatic activity into three groups: catechol oxidases (CO, o-diphenol:oxygen oxidoreductase), tyrosinases (monophenol, o-diphenol:oxygen oxidoreductase) and aureusidin synthases (2',4,4',6'-tetrahydroxychalcone 4'-O-β-D-glucoside:oxygen oxidoreductase) (Kaintz, Mauracher and Rompel, 2014). The latter category of enzymes was described only in *Antirrhinum majus* (yellow snapdragon) and *Coreopsis grandiflora*, where it participates in aurone biosynthesis. The term catechol oxidases thus refers to enzymes which only catalyze the second part of the reaction (from *o*-diphenols to *o*-quinones), while tyrosinases are enzymes which have both cresolase and catecholase activity and can catalyze the whole process.

PPOs are type-3 copper enzymes characterized by a binuclear copper centre containing two copper ions, each of them coordinated by three histidine residues. Copper allows these proteins to bind molecular oxygen in a side-on bridge mode (Jukanti, 2017); oxygen is essential for PPO activity as a primary oxidant, and copper as a prosthetic group.

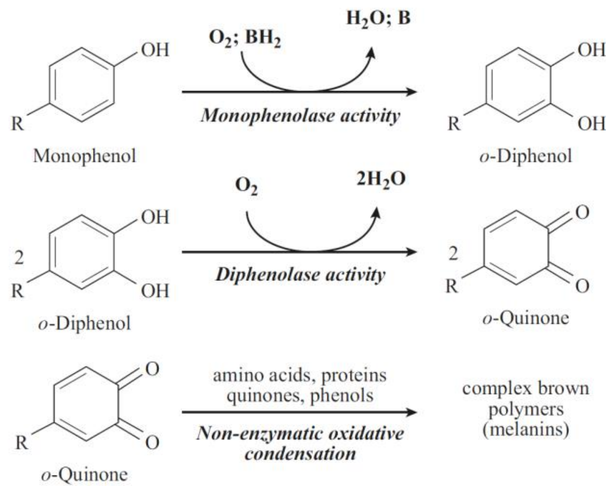


Figure 48. Molecular steps in the formation of melanins as a consequence of monophenol oxidation. Yoruk & Marshall (2003).

As mentioned above, PPOs are present not only in plants, but also in bacteria, fungi, molluscs, arthropods and vertebrates. The phylogenetic analysis of *ppo* genes across these clades indicates that, while the active site and its structure appear highly conserved, the rest of the protein is greatly variable (Fig. 49). Conservation is found mostly within groups, rather than between groups, and available information indicates that tyrosinases across different clades have distinct roles, localizations, molecular weights and secondary modifications, pointing to a great structural and functional diversity (Jaenicke and Decker, 2003). In animals, PPOs are involved in the synthesis of melanin and in cuticular hardening, but also in wound healing; in crustaceans and insects PPOs take part in defense reactions and can cause a browning of tissues analogous to that observed in plants (browning of seafood is also a concern of food industry) (Jukanti, 2017).

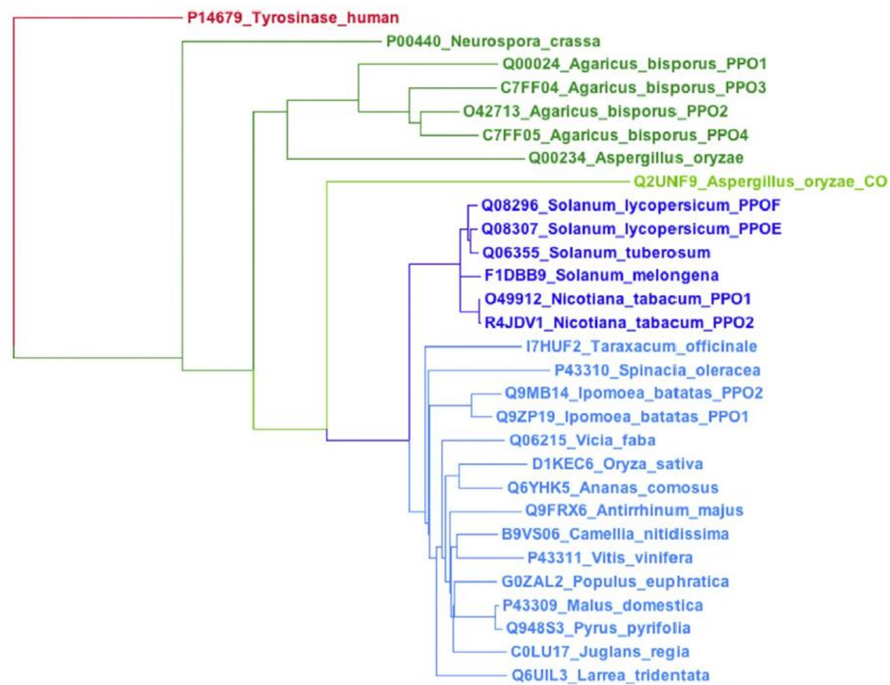


Figure 49. Phylogenetic analysis of polyphenol oxidases. The length of the sections indicates the relative distances between the sequences. Human tyrosinase is shown in red, fungal tyrosinases and catechol oxidases are shown in green, and plant tyrosinases, catechol oxidases and aureusidin synthases are shown in blue (dark blue for Solanaceae proteins). Kaintz *et al*. (2014).

1.1 PPOs in plants

PPOs have been identified in an ever-growing number of plants, from Bryophyta to Dicotyledons, but while the physical characteristics of their genes/proteins and their phylogenetic relations can be quite easily established, information on their role, substrate specificity and implication in the specialized metabolism of the cell still need to be elucidated. Tran *et al.* (2012) conducted a phylogenetic analysis of *ppo* genes across a wide range of land plants and identified some relevant features (Fig. 50). Firstly that, consistently with the colonization of land and of an oxygen-containing atmosphere, PPOs seem to have emerged along with laccases and peroxidases, prior to the divergence of the various clades of land plants; no PPOs have yet been identified in the genomes of green algae like *Chlamydomonas reinhardtii*, *Ostreococcus* spp. or *Volvox carteri*. A surprisingly high number of PPO-encoding genes (11) was found in the small-sized genome of *Selaginella moellendorffii*. In higher plants, this number is either maintained (like in *Glycine max*) or often reduced to about 6-8 genes. Notable exceptions are *Arabidopsis thaliana* and *A. lyrata*, which do

not possess any PPO-encoding gene. Plant *ppo* genes were believed to be intronless; it was later found that introns are typically present in *ppo* genes from Monocotyledons, while in Dicotyledons they are usually absent (*Aquilegia coerulea* being an exception). From phylogenetic analyses, it appears that the *ppo* family either expanded or contracted in the course of evolution, depending on the lineages. In species in which gene number is increased, the expansion is often the result of gene duplication, and *ppo* genes are found in tandem, for example on chromosome 8 of tomato, eggplant and potato (Taranto *et al.*, 2017).

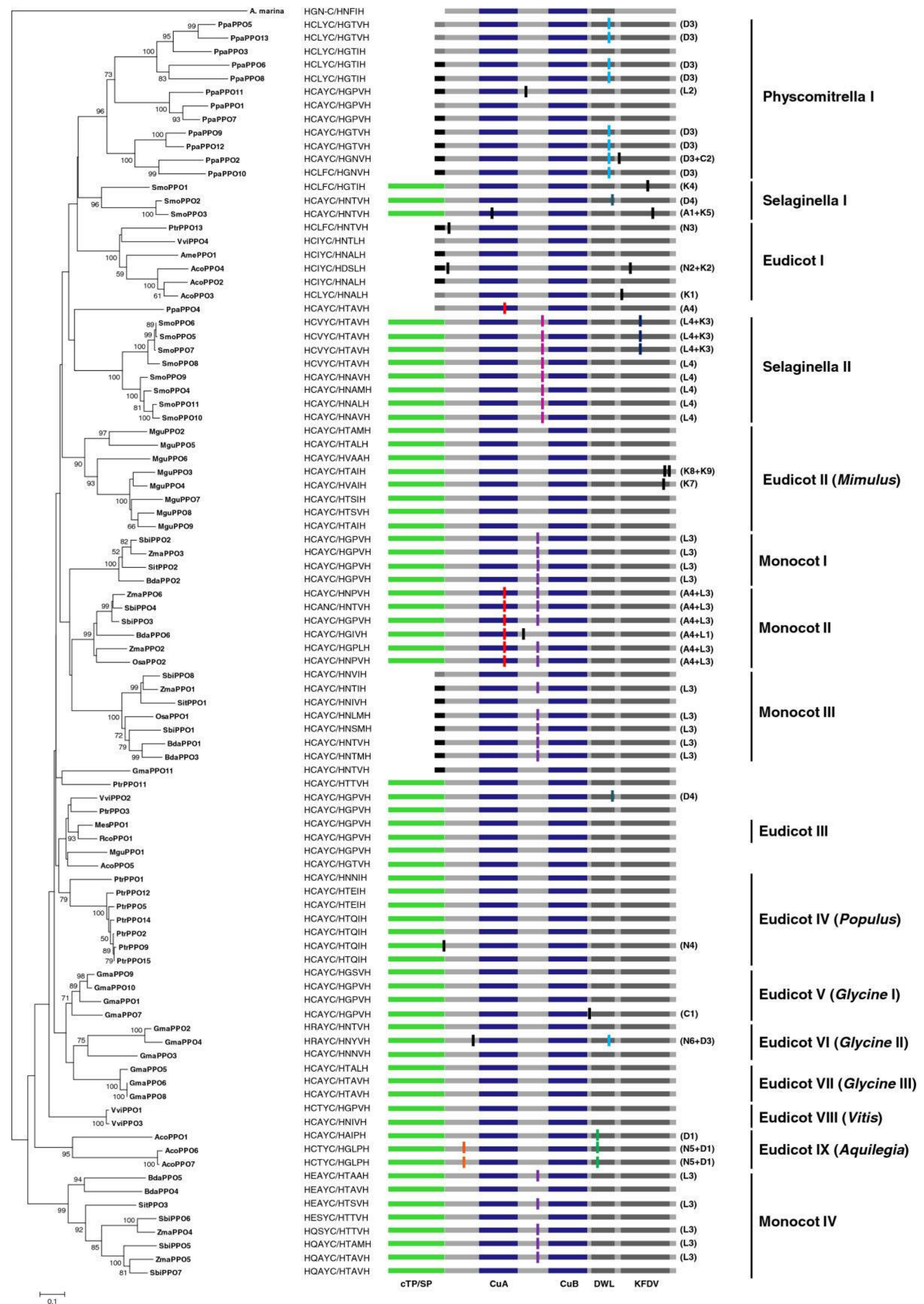


Figure 50. Phylogenetic relations and structure of plant *ppo*s. Neighbour-joining phylogenetic tree from four major land plant lineages, and visual representation of conserved regions, functional motifs, and relative intron positions. Tran *et al.*, 2012.

In higher plants, *ppo* genes are nuclear genes, but the cellular localization of PPOs is mainly in the thylakoid membrane of chloroplasts, or in vesicles in non-green plastids (Yoruk and Marshall, 2003). A distinctive feature of PPOs in many plant species is to be synthesized in the cytoplasm as precursor proteins (or pro-peptides) and to be then translocated, thanks to their N-terminal bipartite transit peptide, to the plastid, where they are located either in thylakoid membranes or in the stroma. Upon arrival at the thylakoid membrane, the pro-peptide is processed by stromal peptidases and the transit peptide is cleaved. The thylakoid membrane-bound form of the enzyme is still inactive and is a 55-70 kDa polypeptide: it needs activation through the proteolysis of the C-terminal end, which reduces its size by 15-20 kDa (Jukanti, 2017). It was suggested that in *Ipomea batata*, for which the X-ray crystal structure of the active catecholase has been determined, the C-terminal region forms a shield over the copper center and inhibits substrate binding before activation (Gerdemann *et al.*, 2002). It is interesting to note, however, that in some species (like tomato and apple) not all PPOs exhibit the hydrophobic domains necessary for membrane localization (Newman *et al.*, 1993), which may imply a functional differentiation among members of the protein family. PPOs isolated from a number of plant species are activated by SDS, which generally acts as a protein-denaturing substance. PPO structure is believed to be stabilized by disulfide bridges, which make it resistant to the denaturing effect of SDS, and this activation mechanism implies an alteration of tertiary and quaternary structure (Mari *et al.*, 1998; Yoruk and Marshall, 2003). These aspects of protein activation should be taken in consideration when studying PPO specific expression and activity, because it isn't always safe to speculate a direct relationship between gene expression and protein activity. In some cases (Lieberei *et al*., 1981; Gooding *et al.*, 2001), PPO activity was recorded without the concomitant expression of *ppo* genes, and PPOs could be stored in plastids and become active only when membranes are disrupted because of cell damage. Protein activity could also be modulated post-transcriptionally by side chain modification, and phosphorilation, glycosilation and myristoylation have been reported in PPOs in some plant species (Jukanti, 2017).

1.2 Roles of PPOs in plants

A defense role against pathogens and pests was the first to be postulated for PPOs because of their increased localized activity in response to cutting and wounding. There are various mechanisms by which PPO activity could exert a protective role against biotic stresses, including: (1) direct toxicity of *o*-quinones, (2) alkylation of cellular proteins and reduced bioavailability of

proteins in the gut of herbivores, (3) production of reactive oxygen species which trigger defense mechanisms in the plant, and (4) formation of a physical barrier by cross-linking of *o*-quinones, which creates a protective layer in the site of injury (Taranto *et al.*, 2017).

The relationship between PPO expression or activation and pathogen infections was proved in tomato by either silencing *ppo* encoding genes, resulting in increased pathogen sensibility (Thipyapong *et al*., 2004), or by over-expressing a potato *ppo* (the *ppo1* gene characterized by Hunt *et al.,* 1993) and thus reducing susceptibility to *Pseudomonas syringae* pv. *tomato* (Li and Steffens, 2002). Tomato PPO activity was also proved to be implicated in the interactions with the pathogens *Ralstonia solanacearum* and *Xanthomonas axonopodis* pv. *vesicatoria* (Kavitha and Umesha, 2008; Vanitha *et al.*, 2009). In pearl millet, wheat and chickpea, PPOs were shown to be activated in response to fungal infections by *Sclerospora graminicola* and *Fusarium* spp. (Mohammadi and Kazemi, 2002; Niranjan Raj *et al.*, 2006). In potato, PPO activity is related to resistance to the herbivore *Leptinotarsa decemlineata* (Castañera *et al.*, 1996), while in tomato it protects the plant against larvae of *Helicoverpa armigera* and *Spodoptera exigua* (Bhonwong *et al.*, 2009). In tomato and potato mutants in which *ppo* genes were silenced, plant development and growth in absence of pathogens and pests appeared normal. Interestingly, PPO activation is also reported as a consequence of abiotic stress, even in absence of cell disruption. Additional roles have been elucidated for PPOs in plant physiology, especially in tyrosine metabolism and in the biosynthesis of betalain, aurone and lignans (Jukanti, 2017). Araji *et al.* (2014) silenced the single PPO-encoding gene of walnut (*Juglans regia*), which is constitutively active in green tissues and does not seem to be activated upon stress; the inactivation resulted in the spontaneous formation of necrotic spots on leaves. Silencing of *Jrppo1* led to increased concentrations of tyramine in walnut leaves, which in turn triggered cell death; this suggests that *Jr*PPO1 could indirectly regulate cell death processes.

In plant and food research, PPOs are mainly considered for their negative impact on food quality due to discoloration and browning. This happens when, in response to cutting and wounding, the physical barriers separating the plastid-located enzymes from their substrates are destroyed. As a consequence, the aspect and the nutritional and organoleptic profiles of foods can be altered; in a few instances, like the fermentation of tea leaves, as well as in cocoa and coffee, this is a desirable process (Jukanti, 2017; Taranto *et al.*, 2017). However, in the majority of cases, high PPO activity has a negative impact on food quality, especially for processed fresh foods: pre-cut and packaged fruits

and vegetables are occupying an increasing share of the food market, and preventing them from browning might reduce waste, along with making them more appealing for the consumer. Since PPO activity is influenced by factors such as pH, temperature and oxygen, different strategies can be used to limit its effect, including the use of preservatives and of specific storage conditions. In addition, in some cases, like eggplant, breeding efforts have focused on the selection of commercial varieties with a low content of phenolics (mainly chlorogenic acid), meaning also lower levels of browning. Of course, this has a negative side effect, because it implies lowering the nutritional value of such varieties by decreasing the concentration of antioxidant phenolic compounds. A fascinating option to preserve food nutritional quality while limiting oxidative browning is to genetically manipulate and regulate the expression of *ppo* genes. This approach has been implemented in various species (among them, potato and apple) and has also helped elucidate the role of specific PPOs. The presence in the majority of species of a family of genes, often present in tandem and originated from gene duplication, also leads to suggest that a functional specialization might be present: different enzymes might be active under specific optimal conditions (temperature and pH), act upon specific substrates and have diversified roles in plant metabolism and defenses against pathogens and pests.

1.3 PPOs in Solanaceae

Six genes encoding PPOs were identified by Shetty *et al.* (2011) in *Solanum melongena*, while seven were described in tomato, and nine in potato. All *ppo* genes described by Shetty *et al.* (2011) cluster on chromosome 8, like those of potato and tomato, with the exception of potato *Stuppo9* and tomato *ppoG*, which are on chromosome 2.

Shetty *et al.* (2011) were the first to propose a distinction between clade A and clade B for eggplant PPOs, based both on protein sequence similarity and on organ-specific patterns of expression. In particular, in eggplant fruits, PPO expression is shown to be localized in the exocarp and in the pulp region surrounding the seeds; PPO activation is also induced by wounding. Like in tomato, eggplant PPO expression is generally higher in young tissues, in both vegetative and reproductive development. Taranto *et al.* (2017) extended this classification to the rest of Solanaceae PPOs (Fig. 51), and found that they separate into two clades (analogous to those reported for eggplant), whose functions seem to be diversified. The first clade includes tomato PPO A-D, potato StuPPO3/4 and eggplant class A proteins (SmePPO1-3) and includes

proteins produced by genes which are known to be expressed in roots (with the exception of *SmePPO3*); the second clade encompasses tomato PPO E and F, StuPPO1 and eggplant class B proteins (SmePPO4-6), which appear to be related to defense responses (Newman *et al.*, 1993; Thipyapong and Steffens, 1997; Thipyapong *et al.*, 2007; Newman *et al.*, 2011; Shetty *et al.* 2011; Chi *et al.*, 2014). In addition, StuPPO8 and StuPPO9 proteins form a third, distinct clade and derive from the only *ppo* genes containing introns in Solanaceae (Chi *et al.*, 2014).

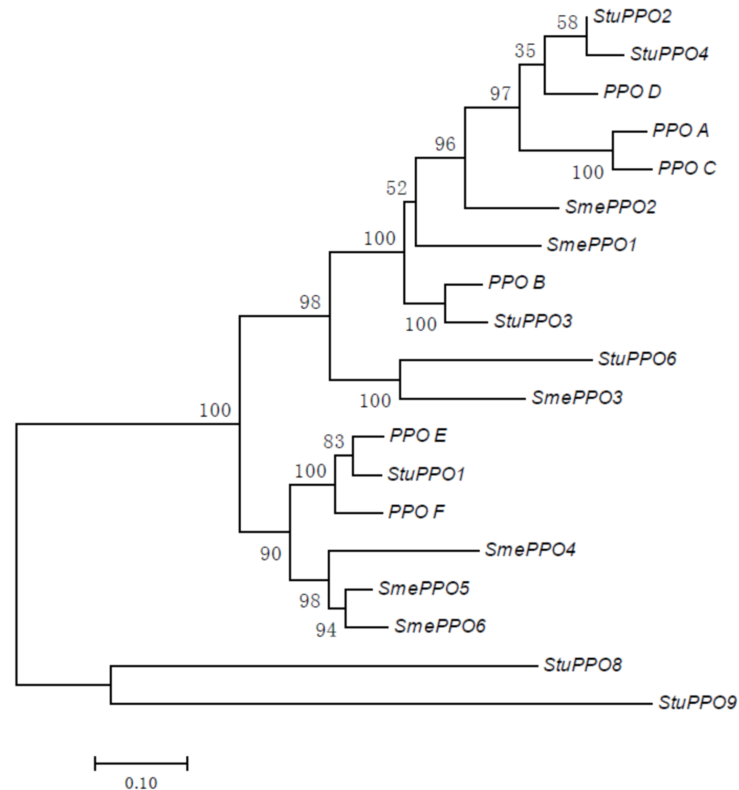


Figure 51. Phylogenetic analysis of PPOs from Solanaceae species. Two main clades are visible, one of which includes proteins which are expressed mainly in roots, the other comprising defense-related proteins. Taranto *et al.* (2017).

Chlorogenic acid is the major phenolic compound found in the flesh of eggplant berries, where it represents 70-95% of total phenolics detected by HPLC analysis (Plazas *et al.*, 2013). Plazas *et al.* (2013) also argue that, based on the screening of a wide panel of eggplant varieties and landraces, no unambiguous

relationship can be drawn between total phenolics content, chlorogenic acid content, PPO activity and browning. While there surely is a correlation between total phenolics content and browning, making it risky to breed for high phenolics content while preventing tissues from browning, more information is needed to elucidate the mechanisms behind this phenomena. The characterization of these processes will provide a margin for breeding of eggplants with enhanced antioxidant content without affecting post-harvest quality.

1.4 Genetic engineering of PPOs to reduce browning

Traditional breeding for the reduction of PPO activity and consequent lowering of oxidative browning of plant tissues, in addition to being time-consuming, often implies selecting against a high concentration of phenolic compounds, to the detriment of nutritional value and food quality. A few notable examples exist of commercially released genetically modified cultures in which *ppo*s have been silenced in order to prevent oxidative browning, extend shelf life and improve quality. One of these is the Arctic Apple® released by the Canadian company Okanagan Specialty Fruits, in which PPO encoding genes are silenced through RNA interference (Waltz, 2015a) and browning is reduced. The Arctic Apple® was cleared by the US FDA in 2015 and reached the market in 2017. A similar example is that of the Innate® potato by Simplot, where *Stuppo5* was silenced through RNAi using a truncated version of the gene; interestingly, this is a case of a cisgenic crop, in which all genetic elements derive from sexually compatible species (Waltz, 2015b). This potato was also approved by FDA in 2015 and, along with a reduction of browning, it also has lower levels of acrylamide thanks to the silencing of the asparagine synthetase-1 gene (*Asn1*) in the tuber. Notably, the USDA did not consider it was necessary to regulate a CRISPR-engineered mushroom (*Agaricus bisporus*) developed by researchers at Penn State University, in which *ppo*s had been inactivated to avoid browning (Waltz, 2016). The mushroom adds to the list of about 30 genetically modified organisms which the USDA chose not to regulate by virtue of the fact that they did not contain foreign DNA elements; specifically, it is the first example of a CRISPR-engineered food to be cleared, while the others on the list had been obtained through ZFNs and TALENs.

1.5 Experimental goals

Here, we chose to silence *ppo* genes in eggplant through a CRISPR/Cas approach. Taking advantage of the availability of a high quality sequence for the eggplant genome, we repeated a search for PPO encoding genes using the sequences identified by Shetty *et al.* (2011) as queries. We then evaluated the

expression of candidate sequences by qRT-PCR in fruits before and after cutting. Finally, we transformed eggplant cotyledons with *A. tumefaciens* carrying a CRISPR construct with gRNAs targeting various members of the *ppo* gene family, and we established an efficient protocol for *in vitro* tissue culture and shoot regeneration. We used a GoldenBraid based multiplexing construct harboring 2 guide RNAs expressed as a polycistronic transcript under the control of a single U6-26 promoter. Gene editing efficiency and off-target effects were assessed by Illumina amplicon sequencing.

2 - Materials and methods

2.1 Target identification

The six eggplant PPO protein sequences published by Shetty *et al*. (2011) were used for a Blastp search of the eggplant proteome (http://www.eggplantgenome.org/). The chosen E-value threshold was $1e^{-5}$. A phylogenetic analysis of the eggplant PPOs was conducted by comparison with the polypeptide sequences of six and nine PPOs from tomato and potato, respectively. All sequences used for tree construction are available in the Supplementary List I at the end of this Chapter. Multiple sequence alignments were carried out using the Clustal Omega online software (www.ebi.ac.uk/tools/msa/clustalo/), and the subsequent phylogenetic analysis was done using the MEGA7 software. An unrooted phylogenetic tree was generated, applying the neighbor-joining (NJ) algorithm; a confidence level was established for each node by performing a bootstrap analysis with 1,000 iterations.

2.2 Expression analysis

In order to understand which *ppo*s were activated after cutting, we performed a qPCR to assess the transcript level of the all *ppo*s in the peel and flesh of full ripe berries of the 'Slim Jim' variety, immediately and 30 minutes after cutting. RNA was extracted using an E.Z.N.A.® Plant RNA kit (OMEGA bio-tek, Norcross, USA), following the manufacturer's protocol. The single cDNA strand was synthesized from a 2 µg aliquot of RNA using a High Capacity RNA-to-cDNA kit (Applied Biosystems, Foster City, USA) as indicated by the manufacturer. Transcript abundance was quantified by running qRT-PCRs on a StepOne™ Real-Time PCR System (ThermoFisher Scientific, USA). Primer sequences, designed with Primer3 (http://bioinfo.ut.ee/primer3/), are listed in Table 26. Each 20 µl reaction was based on the Power SYBR® Green PCR Master Mix (ThermoFisher Scientific, USA) and performed according to the manufacturer's instructions. The amplification protocol comprised an initial denaturation of 95°C/5 min, followed by 40 cycles of 95°C/5s and 60°C/60s. Relative transcript abundances were calculated using the $2^{-\Delta\Delta Ct}$ method, based on the abundance of the actin transcript. Each value represented the mean of three biological replicates, which were compared using the Tukey test, implemented in the SPSS statistical software. This analysis allowed us to

identify which genes were differentially expressed in response to cutting and thus which would be appropriate candidates for gene editing.

Locus	Primer Forward	Primer Reverse
ppo1	TCTGATGGTCCTGAAGTACCC	ATGAGCAGGCTGACGGATAC
ppo2	ACAGAGTTGGCGAGAATACTGAC	TTTGGTACCAGAGTCACCGC
ppo3	AGGTCTCGGTGCTGCTAATC	TAGACGGCATAGGAGGTGGG
ppo4	TGTGGTAAATGACACAACGGGC	CTCCTCATCCACCGCATGAG
ppo5	AACTGAACTCCAAACAATGACG	CCATTCCTGGACTTGGATCG
ppo6	AGTGTTCCATACTACAAGTTCCCTC	CCAAGAGGGTCAAAAGGGTCT
ppo7	TGCGGATAAGAATTTGAATGCGGAT	GCTCTGTGAACGTGTGGCAAG
ppo8	TGCCCTTCCCAATTCTTCGG	TACGACGTGGCACCATTACC
ppo9	ATGCACCTTGTCCTCAGCTG	ACCAACCCATCGATGCACAG
ppo10	TTGGTGGTCTTCCCTTTGCC	CAACAATTGACCGGTGCTGC

Table 26. Primers used for qRT PCR of the 10 *ppo* genes identified in the eggplant genome.

2.3 Vector design and construction

The sequences of the wound-induced *ppo1*, *ppo3*, *ppo4*, *ppo5* genes were aligned with Benchling and BlastN to identify conserved regions. Blastx and Prosite were used to annotate functional domains within conserved regions. Based on this information, a gRNA was designed to target a conserved region of *ppo1* and *ppo3*, while a second gRNA was designed to target a conserved region of *ppo4* and *ppo5*. This second gRNA was found to target *ppo6* as well, due to high sequence homology between these three genes. The two selected gRNAs were domesticated to be inserted into a pUPD2 vector for the assembly of a multiplex construct, as described in Chapter II. Restriction-ligation reactions were performed as detailed in Supplementary Tables S17-S20 reported at the end of this Chapter. The final construct was then transformed into *Agrobacterium tumefaciens* LBA4404 electrocompetent cells.

Restriction-ligation reactions and *E. coli* and *A. tumefaciens* transformations were performed as reported in Chapter I, Materials and Methods.

2.4 Plant material used for transformation

Two eggplant varieties, ‘Black Beauty’ and a ‘Ecavi’ double haploid, were selected for transformation and two protocols were tested, with modifications. Seeds (100 for each variety) were sterilized by washing for 30 minutes in a 1.67% sodium hypochlorite solution (a 1:3 dilution of a commercial 5% sodium hypochlorite solution), then rinsed three times in sterile water, with each rinse lasting for 5, 10 and >15 minutes, respectively. Sterile, clean seeds were transferred to a solid germination *medium* (2.5 g l^{-1} MS vitamins, 10 g l^{-1} sucrose and 10 g l^{-1} phytoagar, pH 5.8) in sterile cups. Seeds were kept in the dark at 25°C for one week and then exposed to a 16-8 light-dark cycle at 25°C for another week. Two weeks after sowing cotyledons were used for *Agrobacterium*-mediated transformation.

2.5 *Agrobacterium*-mediated stable transformation

An *Agrobacterium* LBA4404 pre-culture was set up 48 hours before transformation in a modified MGL liquid *medium* (pH 7, Supplementary Table S21) supplemented with antibiotics and incubated overnight at 28°C with agitation. A reduction in salt concentration proved effective to reduce flocculation. 24 hours before transformation, from this pre-culture a second culture was set up in TY liquid *medium* pH 5.8 supplemented with 200 µM acetosyringone with no antibiotics and incubated overnight in the dark at 28°C with agitation (Supplementary Table S21). Before transformation, the optical density (OD) of the bacterial culture was measured at 600 nm and the culture diluted to a final OD of 0.10-0.12 in TY *medium* with 200 µM acetosyringone. Explants of about 5 mm in length were cut from the cotyledons, dipped in the bacterial culture for a minimum of 5 minutes, then blotted dry on filter paper and transferred for 48 hours to a co-culture *medium* in the dark.

2.6 Organogenesis and regeneration

A set of regeneration conditions were tested to compare their effect on regeneration efficiency and on the reduction of oxidative damage during tissue culture. To this end, two protocols were tested, with modifications.

‘Ecavi’ explants were cultured on the co-culture and induction *media* described by Arpaia *et al.* (1997) (reported in Supplementary Table S22 and Supplementary Table S23, respectively), while shoot formation was induced on

a 0.2 mg l^{-1} thidiazuron (TDZ) containing *medium* as described in Magioli *et al*. (1998) (Supplementary Table S23).

'Black Beauty' explants were co-cultured on the *media* described by Muktadir *et al.* (2016) (Supplementary Table S24), and three alternative induction *media* were tested (Supplementary Table S25): one without additives, one supplemented with ascorbic and citric acid, and finally one supplemented with polyvynilpyrrolidone (PVP). For each *medium* composition, two conditions were evaluated, one without any dark pretreatment, and the other with 3 days of dark pretreatment at the beginning of tissue culture (in addition to the 48 hours of co-culture in the dark, due to the presence of acetosyringone in the *medium*). Also, seeds were germinated for one week in the dark and for an additional week with a 16:8 light:dark cycle before transformation.

Elongation and rooting were performed on the same *media* for both varieties. Explants were moved to a fresh *medium* every 2-3 weeks. When shoots began to emerge, they were moved to an elongation *medium* and then to a rooting *medium* (Supplementary Table S26). These were not supplemented with any antioxidants, as no oxidative damage was observed from this stage onwards. Fully developed plantlets were then moved to soil and gradually acclimated to *ex vitro* conditions.

2.7 Genotyping of target *loci* and evaluation of off-target effects in 'Ecavi'

A first evaluation of editing and of the presence of the transgene was made on 10 'Ecavi' *calli*. DNA was extracted with a CTAB protocol as described in Chapter I, DNA quality was then assessed on a 0.8% agarose gel, while its concentration was measured using a spectrophotometer. The presence of the *hCas9* gene was assayed by PCR (Table 28). Both the *ppo* targets and possible off-target sites were analysed according to the Illumina Amplicon Sequencing Protocol detailed in Chapter I, Materials and Methods. In the case of *ppo4* and *ppo6*, two pair of primers were designed for each gene due to the very high similarity of the target regions and the risk of obtaining an unspecific amplicon.

Potential off-target sites for the two gRNAs directed at *ppo1-3* and *ppo4-5-6* were identified with the CasOT software, using the eggplant genome as a reference. Four off-targets were selected for each gRNA based on the number and position of mismatches and these *loci* (1 kb around the putative off-target site) were blasted against the eggplant genome and the Viridiplantae nucleotide database on https://blast.ncbi.nlm.nih.gov/Blast.cgi to determine whether they

corresponded to functional genes. Genotyping of the candidate off-target *loci* was performed using the Illumina Amplicon Sequencing Protocol reported in Table 9 of Chapter 1. Primers used in the first round PCR to amplify the target sites and attach Illumina adapters are reported in Supplementary Table S27, while Supplementary Tables S28 and S29 report primer used in the second round PCR to add Nextera indexes to the universal Illumina adapters.

Genotyping was then performed on Ecavi 11, the only plantlet we were able to obtain from *callus* culture. DNA was extracted from leaves using a CTAB protocol. DNA quality was then assessed on a 0.8% agarose gel, while its concentration was measured using a spectrophotometer. The presence of the transgene was assessed by amplifying both the *hCas9* and *nptII* genes (Table 27). Genotyping of Ecavi 11 was done by amplifying the target *loci* with the primers used for the first amplification of the Illumina Amplicon Sequencing protocol (Supplementary Table S27). Amplicons were then sequenced according to Sanger's method and analyzed with the TIDE online software as detailed in Chapter I.

2.8 Genotyping of target *loci* in 'Black Beauty' plantlets

A PCR screening was then performed on 'Black Beauty' regenerated plantlets, by extracting DNA from leaves using a CTAB protocol. Leaves were sampled when transferring plantlets from *in vitro* growth conditions to soil. DNA quality was then assessed on a 0.8% agarose gel, while its concentration was measured using a spectrophotometer. The presence of the transgene was assessed by amplifying both the *hCas9* and *nptII* genes (Table 27). Genotyping of 'Black Beauty' regenerants was made by amplifying the target *loci* with the primers used for the first amplification of the Illumina Amplicon Sequencing protocol used for genotyping of the 'Ecavi' samples (Supplementary Table S27). Amplicons were then sequenced according to Sanger's method and analyzed with the TIDE online software as detailed in Chapter I.

Target	Primer sequence 5' → 3'	T_m
hCas9 forward (Ecavi)	AGGTGGCGTACCATGAAAAG	56.5
hCas9 reverse (Ecavi)	TGTTTGCGCAACAGATCTTC	55.2
hCas forward (Black Beauty)	CTATCCTCAGGCGGCAAGAG	60
hCas9 reverse (Black Beauty)	CAGTTTTCTTGACAGCCGCC	60
nptII forward	GCCCCTGATGCTCTTCGTCC	63
nptII reverse	CCTGTCCGGTGCCCTGAATG	63

Table 27. Primer sequences for T-DNA detection.

3 - Results

3.1 PPO identification and phylogenetic analysis

In addition to the six sequences reported by Shetty *et al*. (2011), four new *loci* were recognized to encode polyphenol oxidases in the eggplant genome and were named *ppo7-10*. Newman *et al* (1993) also reported a gene called *ppoC* in tomato, which was not present in the recent annotation of the tomato genome. PPO-encoding *loci* identified in tomato, potato and eggplant are listed in Table 28 and their aminoacidic sequences are reported in the Supplementary Sequence List I at the end of this Chapter.

Taxonomic Classification	Gene	Chr.	Accession number from current annotate genome version	Accession number from previous publication
Solanum lycopersicum	*Sly*PPO_A	8	Solyc08g074680.3.1	XP_004246029.2
	*Sly*PPO_B	8	Solyc08g074683.1.1	NP_001296326.1
	*Sly*PPO_D	8	Solyc08g074682.1.1	XP_004246030
	*Sly*PPO_E	8	Solyc08g074620.3.1	NP_001318057
	*Sly*PPO_F	8	Solyc08g074630.2.1	NP_001318059
	*Sly*PPO_G	2	Solyc02g078650.2.1	Solyc02g078650.2.1
Solanum tuberosum	*Stu*PPO1	8	Sotub08g017870.1.1	PSGC003DMT400076054
	*Stu*PPO2	8	Sotub08g019390.1.1	PSGC0003DMT400048684
	*Stu*PPO3	8	Sotub08g019410.1.1	PSGC0003DMT400048681
	*Stu*PPO4	8	Sotub08g019380.1.1	PSGC0003DMT400048685
	*Stu*PPO5	8	Sotub08g019310.1.1	PSGC0003DMT400048692
	*Stu*PPO6	8	Sotub08g017890.1.1	PSGC0003DMT400076055
	*Stu*PPO7	8	Sotub08g019300.1.1	PSGC0003DMT400048703
	*Stu*PPO8	8	Sotub08g019420.1.1	PGSC0003DMT400048679
	*Stu*PPO9	2	Sotub02g021500.1.1	PGSC0003DMT400057781
Solanum melongena	*Sme*PPO1	8	SMEL_008g312510.1.01	ACT22523
	*Sme*PPO2	8	SMEL_008g312500.1.01	ADG56700
	*Sme*PPO3	8	SMEL_008g312430.1.01	ADY18409
	*Sme*PPO4	8	SMEL_008g312420.1.01	ADY18410
	*Sme*PPO5	8	SMEL_008g311990.1.01	ADY18411
	*Sme*PPO6	8	SMEL_008g312010.1.01	ADY18412
	*Sme*PPO7	8	SMEL_008g312490.1.01	not previously reported
	*Sme*PPO8	8	SMEL_008g312460.1.01	not previously reported
	*Sme*PPO9	8	SMEL_008g312520.1.01	not previously reported
	*Sme*PP10	2	SMEL_000g064350.1.01	not previously reported

Table 28. PPO protein sequences used for the generation of a phylogenetic tree representing the structure of this protein family in Solanaceae species. Sequences were retrieved from tomato, potato and eggplant annotated genomes. The current annotate genome versions are: ITAG *Solanum lycopersicum* proteins (v.2.3), ITAG *Solanum tuberosum* group Phureja DM1-3 proteins (v.1) and the SMEL V3 proteins. Table reports also the old accession numbers from previous publications*:* Newman *et al*., 1993, Chi *et al.*, 2014 and Shetty *et al.*, 2011 for tomato, potato and eggplant, respectively.

Their phylogenetic analysis revealed that, in accordance with the report by Taranto *et al.* (2017), two clusters can be distinguished among *ppo* genes in the Solanaceae, which correspond to a functional distinction between PPOs which

are preferentially expressed in roots (tomato *Sly*PPO A-D, potato *Stu*PPO2 and 4 and eggplant class A proteins, *i.e.* *Sme*PPO1-3), and PPOs whose expression is associated to damage and wounding (tomato *Sly*PPO E and F, *Stu*PPO1 and eggplant class B proteins, *i.e.* *Sme*PPO4-6) (Fig. 52). Among additional eggplant proteins identified in our analysis, *Sme*PPO7 clusters with group A PPOs, while *Sme*PPO8 and *Sme*PPO9 do not belong to either group, with *Sme*PPO9 clustering with potato *Stu*PPO8. *Sme*PPO10, instead, clusters with *Stu*PPO9 and *Sly*PPOG, that is PPOs encoded by genes residing on chromosome 2 rather than on chromosome 8 like all other members of the family.

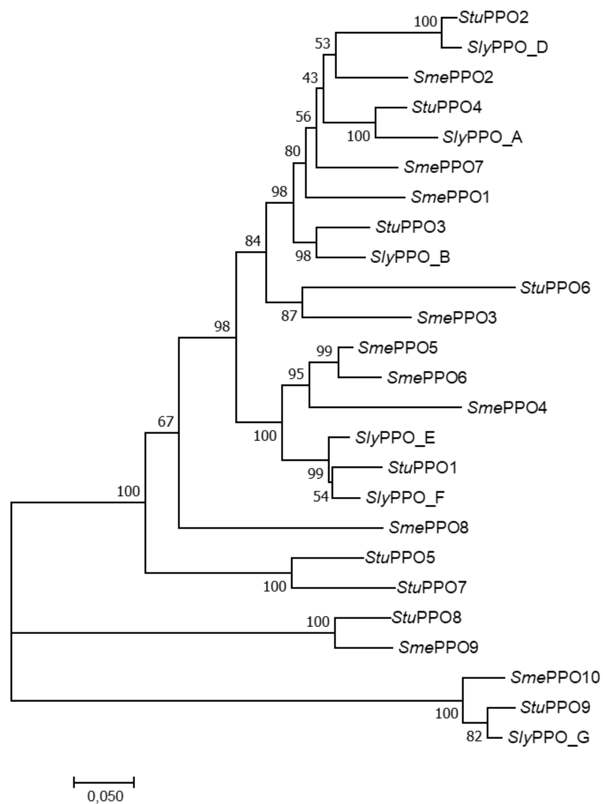


Figure 52. Phylogenetic tree of the PPOs in the Solanaceae family, including sequences from tomato, potato and eggplant. Two main clusters can be distinguished, one representing proteins which have been shown to be involved in plant defenses against herbivores, and the other comprising proteins which are expressed mainly in the roots.

*Sme*PPO10 had initially been predicted on an unanchored scaffold (12014, chromosome 0), but its high similarity with tomato and potato PPOs located on chromosome 2 let us hypothesize that it could likely be located on chromosome

2 as well. This further confirms that the structure of the gene family is conserved in Solanaceae species, and has possible implications for divergent functions.

The structural characteristics of PPO encoding genes and of PPO proteins are listed in Table 29, while the relative position of *ppo1-9* on *Solanum melongena* chromosome 8 is represented in Figure 53. Like the *ppos* from other members of the Solanaceae family, eggplant *ppos* do not possess introns. Coding sequences range in size from 1,686 to 2,466 bp, and the corresponding proteins range from 562 to 822 aminoacids; all genes except *ppo4* and *ppo3* are on the negative strand. All polypeptides present the same functional domains, namely the tyrosinase and the PPO1_DWL domain, and a protein domain of unknown function (DUF_B2219).

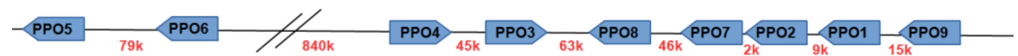


Figure 53. Relative position and organization of *ppo1-9* genes on chromosome 8 of *Solanum melongena*. All eggplant *ppo*s except *ppo10* are located on chromosome 8. In red are the distances in kbp between *loci*.

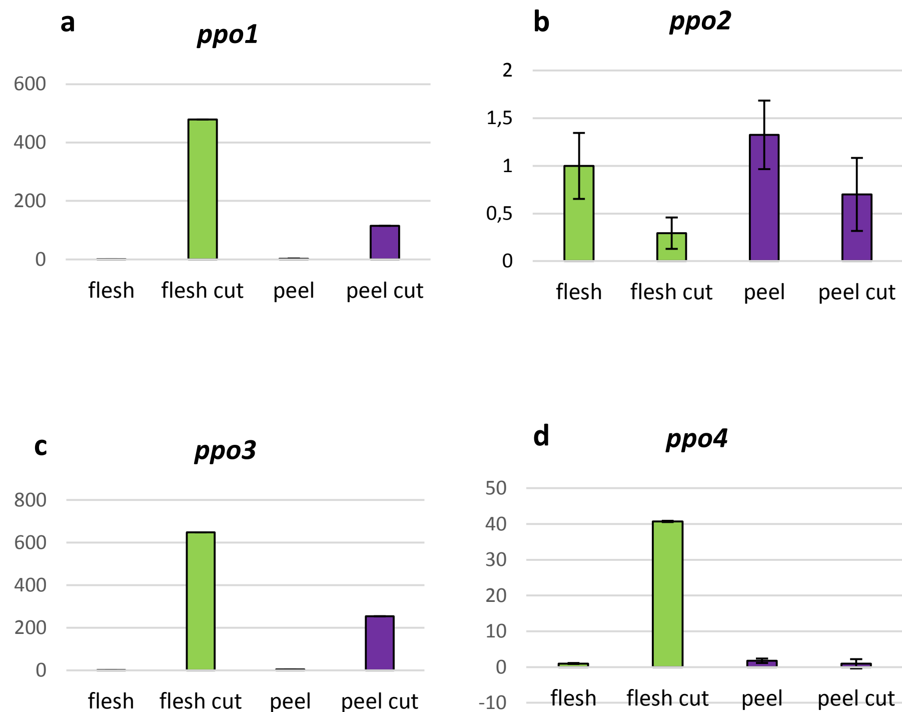
Locus	Gene name	Chr	Chromosome location	ORF length (bp)	Strand	Size (aa)	Protein Domains	Pfam domains
SMEL_008g312510.1.01	*ppo1*	8	97,412,508 : 97,414,307	1,800	-	600	PPO1_DWL - DUF_B2219 - Tyrosinase	Pfam 12142 - Pfam 1243- Pfam 00264
SMEL_008g312500.1.01	*ppo2*	8	97,401,279 : 97,403,066	1,788	-	596	PPO1_DWL - DUF_B2219 - Tyrosinase	Pfam 12142 - Pfam 1243- Pfam 00264
SMEL_008g312430.1.01	*ppo3*	8	97,284,426 : 97,286,198	1,773	+	591	PPO1_DWL - DUF_B2219 - Tyrosinase	Pfam 12142 - Pfam 1243- Pfam 00264
SMEL_008g312420.1.01	*ppo4*	8	97,238,764 : 97,240,497	1,734	+	578	PPO1_DWL - DUF_B2219 - Tyrosinase	Pfam 12142 - Pfam 1243- Pfam 00264
SMEL_008g311990.1.01	*ppo5*	8	96,314,480 : 96,316,243	1,764	-	588	PPO1_DWL - DUF_B2219 - Tyrosinase	Pfam 12142 - Pfam 1243- Pfam 00264
SMEL_008g312010.1.01	*ppo6*	8	96,395,550 : 96,397,448	1,899	-	633	PPO1_DWL - DUF_B2219 - Tyrosinase	Pfam 12142 - Pfam 1243- Pfam 00264
SMEL_008g312490.1.01	*ppo7*	8	97,397,374 : 97,399,167	1,794	-	598	PPO1_DWL - DUF_B2219 - Tyrosinase	Pfam 12142 - Pfam 1243- Pfam 00264
SMEL_008g312460.1.01	*ppo8*	8	97,349,335 : 97,351,020	1,686	-	562	PPO1_DWL - DUF_B2219 - Tyrosinase	Pfam 12142 - Pfam 1243- Pfam 00264
SMEL_008g312520.1.01	*ppo9*	8	97,429,811 : 97,432,277	2,466	-	822	PPO1_DWL - DUF_B2219 - Tyrosinase	Pfam 12142 - Pfam 1243- Pfam 00264
SMEL_000g064350.1.01	*ppo10*	2*	982,270 : 984,463	2,193	-	731	PPO1_DWL - DUF_B2219 - Tyrosinase	Pfam 12142 - Pfam 1243- Pfam 00264

Table 29. Characteristics of PPO encoding genes and of PPO proteins. *ppo1-9* cluster on chromosome 8, while *ppo10*, which was initially located on an unanchored scaffold, is probably located on chromosome 2. All PPOs share the same functional domains (PPO1_DWL and Tyrosinase, and a conserved domain of unknown function, DUF_B2219).

3.2 Expression analysis

Transcript levels were analyzed in both the peel and flesh of full-ripe eggplant berries of the 'Slim Jim' variety, before and 30 minutes after cutting (Fig. 54a-j). We showed that eggplant *ppo*s were expressed in both tissues (peel and flesh). Before cutting, eight genes were more expressed in the peel and two (*ppo5* and *ppo10*) were more expressed in the flesh. Thirty minutes after cutting, a steep increase in gene expression in the flesh was observed especially for *ppo1* (479X), *ppo3* (648X) and *ppo4* (40X). *ppo5* also appeared significantly up-regulated in the flesh after cutting, with a 2.6 fold increase in expression. Taking into account their profile patterns, *ppo1*, *ppo3*, *ppo4* and *ppo5* genes were chosen as targets for CRISPR-mediated knock out.

Expression of *ppo7*, *ppo8* and *ppo9* was localized especially in the peel, with an average 2.5 fold increase in gene expression after cutting. On the other hand, expression of *ppo2* in the fruit did not seem to be implicated in response to cutting in either tissue.



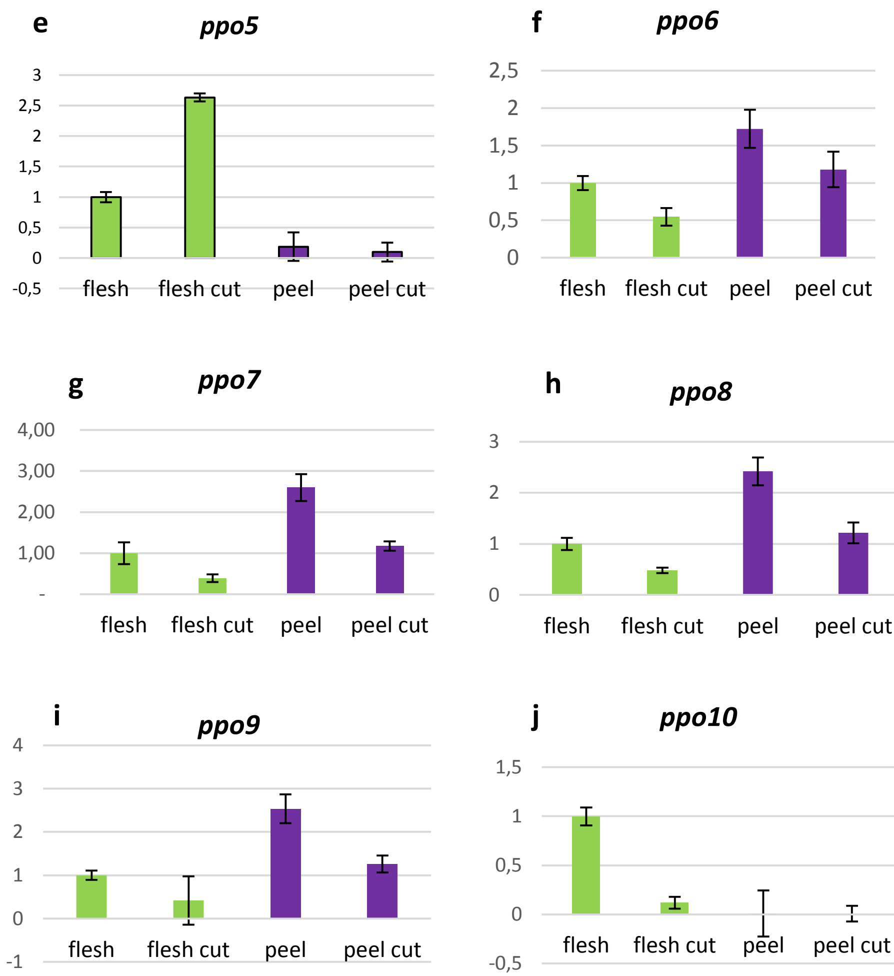


Figure 54a-j. Transcriptional levels of 10 PPO-encoding genes in the 'Slim Jim' eggplant variety. A transcriptional increase was observed in *ppo1*, *ppo3*, *ppo4* and *5* in the flesh, 30 minutes after cutting. Transcription of *ppo7*, *ppo8* and *ppo9* was instead higher in the peel, and was not activated by cutting. Expression of *ppo2* in the fruit was generally low, and did not correlate to cutting.

3.3 gRNA identification and vector design

The alignment of the *ppo1* and *ppo3* genes identified a 45 bp conserved region between positions 793 and 837 of *ppo1* and between positions 775 and 819 of *ppo3*, which corresponds to the tyrosinase domain, with a suitable PAM site. The sequence of the selected gRNA was: 5'-GTTATGGACCTTGGTTCTTT-3'. Similarly, a conserved region of 73 bp was identified between positions 1090

and 1162 of both *ppo4* and *ppo5*, also corresponding to the tyrosinase domain. The sequence of the selected gRNA for *ppo4* and *ppo5* was: 5'-TGAATGGAAAGCAATCGGA-3'. Due to high sequence similarity between *ppo4*, *ppo5* and *ppo6*, this gRNA also targets the latter gene.

3.4 Establishing an efficient regeneration protocol

The induction *medium* described by Arpaia *et al.* (1997) was efficient in generating large amounts of *callus* in the 'Ecavi' double haploid but, from this, very few compact green nodules developed and could be transferred to an elongation or rooting *medium*. For this reason, shoot differentiation was pursued by using a TDZ-supplemented *medium*, as described by Magioli *et al.* (1998). TDZ is a substituted phenylurea, used as a defoliant, which has high cytokinin-like activity in plant *in vitro* cultures, and which was shown to be active in eggplant shoot regeneration by Magioli *et al.* (1998), Swamynathan *et al.* (2010) and Muktadir *et al.* (2016), among others. TDZ acts partly through inhibition of cytokinin degradation, thus increasing cytokinin concentration and retention in plant tissues. Since the explants showed early signs of oxidative damage, the regeneration *medium* was supplemented with citric and ascorbic acids. However, regeneration efficiency remained very low, only 5 shoots were recovered for rooting, and only 1 was successfully transferred to *ex vitro* conditions.

In the second transformation experiment, performed using the 'Black Beauty' variety, three alternative *media* were tested: one without additives, one supplemented with ascorbic and citric acids, and finally one supplemented with PVP. For each *medium* composition, two conditions were evaluated, one without any dark pretreatment, and the other with 3 days of dark pretreatment at the beginning of tissue culture (in addition to the 48 hours of co-culture in the dark, due to the presence of acetosyringone in the *medium*). Also, seeds were germinated for one week in the dark and for an additional week with a 16:8 light:dark cycle before transformation. The steps of transformation and regeneration from cotyledon explants to shoots, through induction of *callus* growth, are represented in Figure 55.

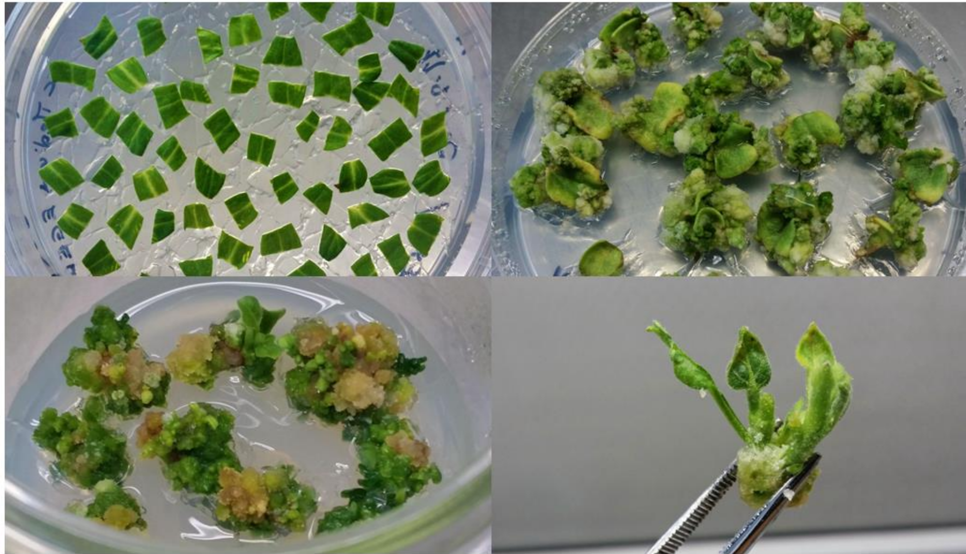


Figure 55. Regeneration of ‘Black Beauty’ explants.

Shoots apt for rooting were recovered in as short as 6 weeks, proving the protocol to be very efficient. In the basal *medium* without antioxidants, no oxidative damage was found, suggesting this might be dependent on the genotype, considering the low efficiency obtained in ‘Ecavi’. No significant differences were found between the basal and the PVP-supplemented *media* regarding the phenotype of explants. A difference was instead observed regarding the number of recovered shoots. A total of 15 shoots were rooted, 10 of which derived from the PVP-supplemented *medium*, while 4 derived from the additive-free *medium*, and only 1 from the *medium* supplemented with ascorbic and citric acids. The induction of shoot formation on PVP-supplemented *medium* appeared slower, but the total number of recovered shoots was notably higher. Instead, no differences were found between shoots which had undergone the 3-day dark pretreatment and those who had not, with an even distribution of successful events between the two conditions. In Figure 56a-b-c-d some of the regenerated plants are represented.

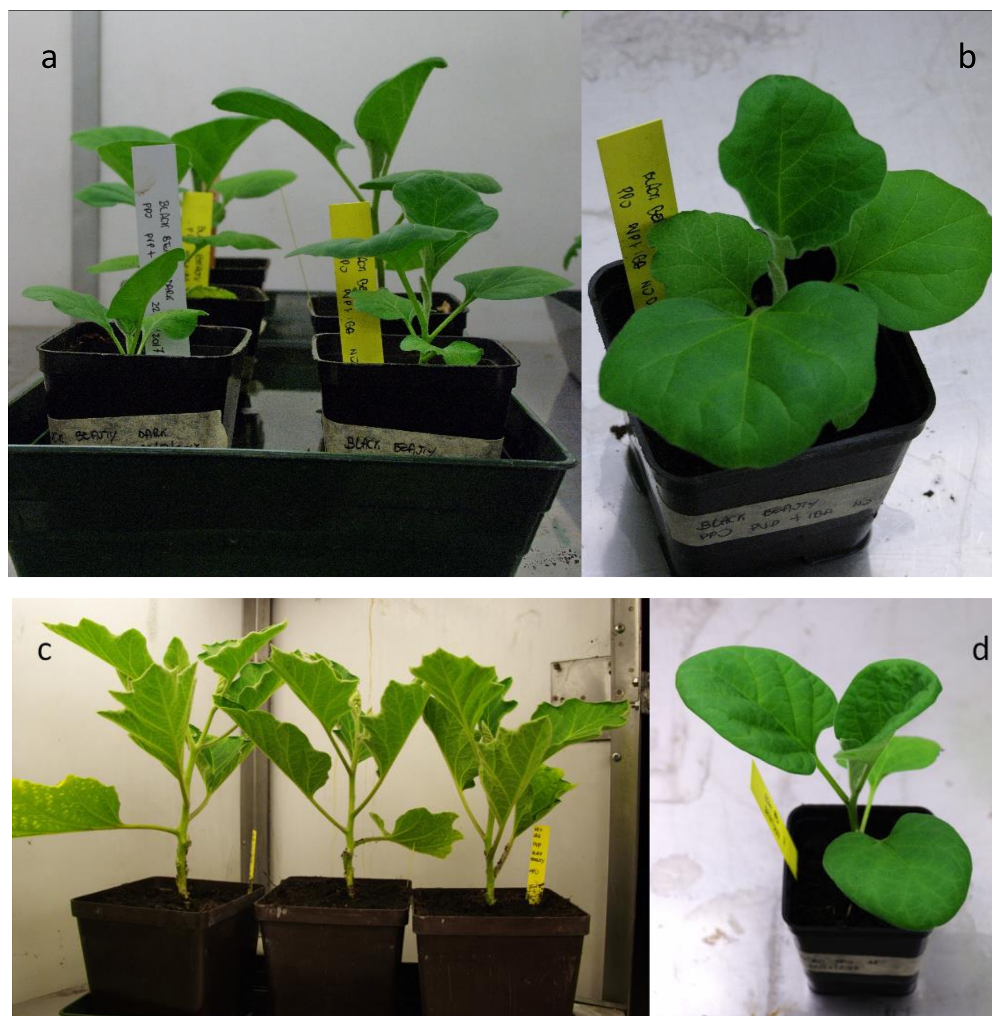


Figure 56. 'Black Beauty' and 'Ecavi' regenerated plantlets. a) and b) 'Black Beauty' regenerated plantlets; c) older 'Black Beauty' regenerants; d) the Ecavi 11 regenerated plantlet.

3.5 Genetic characterization of the T_0 generation

3.5.1 'Ecavi': genotyping of target and off-target *loci*

3.5.1.1 'Ecavi': genotyping of target *loci*

Ten 'Ecavi' *calli* were selected for an early evaluation of the outcome of gene editing of eggplant *ppo*s. The *hCas9* gene was amplified to confirm the integration of the transgene; all ten *calli* yielded amplification (Fig. 57).

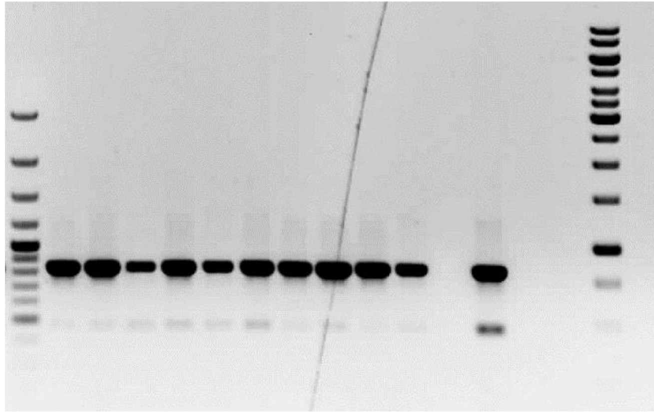


Figure 57.

***hCas9* amplification in 10 transformed 'Ecavi' *calli*.** Left to right are samples 1-10, the negative control and a positive control. The outer left and right lanes are occupied by the molecular marker.

The Illumina amplicon sequencing of the target sites and the alignment of sequence reads to gene sequences of *ppo1,3,4,5* and *6* revealed that no gene editing activity was detectable for the gRNA directed at *ppo1* and *ppo3* (editing efficiencies do not exceed 1%), while *ppo4-5-6* had been successfully edited (Table 30). Editing efficiencies for *ppo4* range between 1 and 25%, while for *ppo5* they vary between 0.5 and 61%, and for *ppo6* they vary between 0 and 33%. *Callus* 5 shows the greatest editing efficiency for all three targets. In general, each *callus* has comparable editing efficiencies for *ppo4*, *5* and *6*, with *calli* 1, 3, 5 and 7 having editing efficiencies above 10% for all targets. One hundred reads from the Illumina amplicon sequencing were randomly selected for each *locus* and each individual, and aligned to 10 randomly selected wild type sequences using ClustalX, as showed in Figure 58. These alignments were used to identify which kinds of mutations occurred in our samples, to subsequently quantify them in relation to the total of the Illumina reads. Tables 31-33 report the mutations occurring at each *locus* for *ppo4*, *ppo5* and *ppo6*, respectively.

Locus	Sample	N. of sequences	N. WT target sequence	% WT	Normalized % edited
***ppo*1**	WT	11682	11471	98,19	0
	1	29675	29154	98,24	0
	2	26249	25751	98,10	0.09
	3	21156	20738	98,02	0.17
	4	118966	116732	98,12	0.07
	5	14574	14309	98,18	0.01
	6	21781	21366	98,09	0.10
	7	13228	12963	98,00	0.2
	8	16628	16326	98,18	0.01
	9	23227	22753	97,96	0.24
	10	16408	16054	97,84	0.36
***ppo* 3**	WT	4210	4115	97,74	0
	1	10992	10761	97,90	0
	2	11538	11274	97,71	0
	3	8697	8510	97,85	0
	4	6958	6810	97,87	0
	5	12331	12057	97,78	0
	6	5069	4952	97,69	0
	7	8088	7909	97,79	0
	8	8590	8345	97,15	1
	9	13947	13641	97,81	0
	10	9754	9541	97,82	0
***ppo* 4**	WT	22748	22284	97,96	0
	1	15904	13012	81,82	16.47
	2	28697	26367	91,88	6.21
	3	14453	12136	83,97	14.28
	4	18333	16908	92,23	5.85
	5	21479	15602	72,64	25.84
	6	19990	17759	88,84	9.31
	7	17715	14647	82,68	15.6
	8	28042	26221	93,51	4.54
	9	55426	48536	87,57	10.61
	10	54040	52370	96,91	1.07
***ppo* 5**	WT	7922	7711	97,34	0
	1	9220	6903	74,87	23.08
	2	26641	24553	92,16	5.32
	3	9938	8407	84,59	13.1
	4	15678	14226	90,74	6.78
	5	18950	7167	37,82	61.15
	6	10473	8798	84,01	13.7
	7	8576	7032	82,00	15.76
	8	20144	18355	91,12	6.39
	9	37886	33806	89,23	8.33
	10	27456	26590	96,85	0.5
***ppo* 6**	WT	20747	19978	96,29	0
	1	16525	13103	79,29	17.66
	2	16300	14966	91,82	4.64
	3	9083	7796	85,83	10.86
	4	13925	12545	90,09	6.44
	5	17040	10977	64,42	33.10
	6	34554	30122	87,17	9.47
	7	13337	10774	80,78	16.11
	8	19344	18024	93,18	3.23
	9	15733	13731	87,28	9.36
	10	11199	10828	96,69	0

Table 30. Quantification of Illumina reads edited at the target *locus*. For each individual and for each *locus* the total number of reads is reported, together with the percentage of reads carrying the wild type (non edited) target sequence. The percentage of edited sequences was estimated by normalizing editing efficiency to the wild type for each target.

Figure 58. Alignment of 100 randomly selected Illumina reads on *ppo*6 from *callus* 1 against 10 randomly selected wt reads. The "CCC" sequence in the middle is the reverse complement of the PAM (highlighted in red); various polymorphisms can be seen around this region, some being only small indels, other representing larger deletions extending beyond the PAM region.

ppo4

Callus	Insertions					Deletions											
	+1	**T**	**A**	**C**	**G**	**-1**	**-2**	**-3**	**-4**	**-5**	**-6**	**-7**	**-8**	**-9**	**-10**	**-11**	**-12**
1	24,69	21,89	1,87	0,86	0,07	20,41	6,29	7,71	7,51	1,31	2,01	0,90	0,24	0,24	0,07	0,14	0,00
2	7,00	1,67	2,75	2,57	0,00	22,06	13,09	5,02	7,21	0,56	2,83	2,45	0,39	0,34	0,13	0,39	0,09
3	3,45	1,73	1,42	0,26	0,04	23,91	4,32	15,54	29,74	0,04	1,77	1,42	0,47	0,43	0,09	0,30	0,04
4	3,37	1,47	1,26	0,49	0,14	6,04	10,18	7,72	10,81	0,56	8,92	5,34	3,02	3,23	0,49	1,68	0,35
5	1,68	0,37	0,88	0,39	0,03	5,12	5,29	53,82	10,58	0,26	3,18	2,99	0,92	1,02	0,19	0,54	0,07
6	5,96	2,20	1,84	1,70	0,22	16,41	13,81	7,71	11,03	0,54	4,03	2,15	0,67	0,63	0,00	0,85	0,04
7	5,21	3,68	0,91	0,49	0,13	22,26	9,22	3,62	6,23	0,46	1,37	0,46	0,23	0,13	3,03	0,33	0,00
8	4,45	1,92	1,92	0,38	0,22	8,79	9,34	8,08	12,64	0,49	4,89	3,90	1,10	0,93	0,77	0,55	0,16
9	7,68	3,15	4,01	0,36	0,16	21,58	8,91	9,32	13,83	0,58	2,05	1,60	0,70	0,55	0,20	1,03	0,09
10	2,63	0,78	1,38	0,48	0,00	11,74	2,75	2,63	3,05	0,06	1,02	0,66	0,30	0,00	0,06	0,00	0,00

Table 31. Frequency of edited alleles per Ecavi *callus* (10 *calli* in total) for the *ppo*4 *locus*. Frequencies are calculated as percentages of non-wild type reads. T, A, C and G insertions are calculated as percentages of total edited reads.

ppo5

Callus	Insertions					Deletions											
	+1	**T**	**A**	**C**	**G**	**-1**	**-2**	**-3**	**-4**	**-5**	**-6**	**-7**	**-8**	**-9**	**-10**	**-11**	**-12**
1	2,33	0,82	1,08	0,35	0,09	46,96	18,30	2,72	6,39	2,16	1,55	1,68	0,35	0,09	0,04	0,47	0,00
2	3,93	2,54	0,91	0,34	0,14	18,67	12,74	6,08	12,74	0,43	2,78	0,91	0,24	0,14	0,34	0,00	0,00
3	37,55	1,24	35,59	0,72	0,00	4,77	5,81	3,92	6,07	1,04	2,22	4,31	0,39	0,72	0,13	0,46	0,00
4	1,24	0,21	0,41	0,41	0,21	3,93	18,39	6,82	10,06	0,34	9,09	7,23	2,89	2,07	1,03	0,90	0,00
5	0,01	0,01	0,00	0,00	0,00	0,09	0,92	0,01	1,07	0,00	47,92	45,78	0,04	0,00	0,00	0,02	0,03
6	6,21	2,27	2,57	1,19	0,18	16,00	15,11	5,55	10,57	0,60	3,22	3,22	0,24	0,18	0,24	0,42	0,00
7	4,41	2,01	1,43	0,91	0,06	7,00	13,09	3,69	5,70	0,45	12,57	12,37	0,78	0,45	0,00	0,26	0,00
8	3,19	1,79	1,06	0,34	0,00	19,34	9,78	6,43	10,45	0,61	5,31	2,91	3,41	0,84	0,11	0,67	0,17
9	5,47	1,72	3,26	0,44	0,05	12,13	22,23	4,90	12,74	0,44	3,04	4,24	1,08	0,32	0,44	1,32	0,05
10	4,16	0,81	3,35	0,00	0,00	3,12	5,67	3,58	3,93	0,00	3,24	1,97	0,81	0,00	0,23	0,46	0,23

Table 32. Frequency of edited alleles per Ecavi *callus* (10 *calli* in total) for the *ppo5 locus*. Frequencies are calculated as percentages of non-wild type reads. T, A, C and G insertions are calculated as percentages of total edited reads.

ppo6

Callus	Insertions					Deletions											
	+1	**T**	**A**	**C**	**G**	**-1**	**-2**	**-3**	**-4**	**-5**	**-6**	**-7**	**-8**	**-9**	**-10**	**-11**	**-12**
1	10,61	8,12	1,69	0,73	0,06	27,96	8,27	8,04	9,18	1,87	1,96	0,64	0,35	0,09	0,15	0,03	0,00
2	9,22	2,10	3,67	3,45	0,00	20,40	12,37	6,22	6,90	0,52	1,80	2,02	0,45	0,15	0,07	0,37	0,00
3	4,89	1,01	3,57	0,31	0,00	12,82	6,45	36,98	10,18	0,16	1,86	3,65	0,78	0,85	0,08	0,23	0,08
4	2,25	0,94	0,87	0,36	0,07	5,43	9,78	7,68	13,04	0,65	10,07	7,46	4,13	2,39	0,87	1,09	0,22
5	0,99	0,21	0,59	0,18	0,00	3,38	4,32	61,51	9,58	0,25	3,13	2,23	0,59	1,11	0,15	0,30	0,08
6	6,54	2,03	2,37	1,83	0,32	14,48	14,55	7,17	11,71	0,74	4,13	3,43	0,90	0,70	0,25	1,40	0,16
7	6,71	5,11	1,05	0,51	0,04	37,18	7,22	4,64	7,33	0,94	2,61	0,90	0,27	0,08	5,19	0,20	0,04
8	2,65	1,14	1,14	0,38	0,00	9,93	11,60	8,57	11,29	0,61	4,55	4,47	1,14	1,14	0,38	1,52	0,08
9	9,29	3,95	4,65	0,40	0,30	18,54	10,24	10,34	13,29	0,70	3,40	2,45	1,10	0,80	0,40	1,05	0,20
10	2,70	1,62	0,81	0,27	0,00	5,40	4,32	4,05	5,13	0,27	1,35	1,89	0,27	0,00	0,00	0,27	0,00

Table 33. Frequency of edited alleles per Ecavi *callus* (10 *calli* in total) for the *ppo6 locus*. Frequencies are calculated as percentages of non-wild type reads. T, A, C and G insertions are calculated as percentages of total edited reads.

To understand the reason for failure of the first gRNA in the multiplexing construct to cause gene editing of *ppo1* and *ppo3*, the gRNA sequence was revised, together with that of the multiplexing transcriptional unit. The choice of a single conserved targeting region in *ppo1* and *ppo3* led to the selection of a gRNA with a target score of 40, while optimal values are considered to be over 50 in a 0-100 scale. However, this on-target score is comparable to that of other gRNAs, which were proven to be effective (a reported in Chapter II). The pUPD2 vector in which the tRNA, gRNA and scaffold were assembled was sequenced, and the tRNA and scaffold sequences aligned perfectly to the GB1208 template sequence. However, inspection of the pVD1 GB1208 sequence by aligning it with pVD1 GB1210, its homolog GB piece for gRNA multiplexing for Monocots, revealed a 2 nucleotide deletion in the scaffold RNA sequence, which is predicted to affect tridimensional structure. We could therefore conclude that a sequence design error was present in the original plasmid used for cloning. This does not affect the release and activity of the second gRNA in the transcript, which was in fact found to be functional.

We were able to obtain only one plantlet from 'Ecavi' *calli*, called Ecavi 11. Genotyping of the target sites was performed using Sanger sequencing. Resulting chromatograms were analyzed with TIDE. This plant appears edited for all three *loci*, *ppo4, ppo5* and *ppo6* as shown in Table 34.

Ecavi 11							
Sample	**Overall efficiency**	**R^2**	**WT**	**-6 deletion**	**-4 deletion**	**-3 deletion**	**-1 deletion**
***ppo*4**	30.2	0.99	68.3	15.7	8.5		
***ppo*5**	53.3	0.95	41.5			23.4	28.8
***ppo*6**	31.3	0.98	66.5	9.1	20.3		

Table 34. Genotyping of the T_0 Ecavi 11 plantlet by PCR amplification of the target *ppo*4, *ppo*5 and *ppo*6 *loci*, direct Sanger sequencing and TIDE chromatogram decomposition.

3.5.1.2 'Ecavi': genotyping of off-target *loci*

For each gRNA, four putative off-target regions were identified (Table 35). For the gRNA targeting *ppo1* and *ppo3*, two of these sequences had two mismatches, while the others had four. As a consequence of working with a conserved gene family, many off-target sequences do, in fact, correspond to other members of the *ppo* family. Two off-target sequences (*ppo*1-3 OT1 and *ppo*4-5-6 OT2) correspond to *ppo7*, while other two (*ppo*1-3 OT2 and *ppo*4-5-6

OT3) correspond to *ppo2* and another corresponds to *ppo3* (*ppo*4-5-6 OT1). The remaining three do not map to an annotated *locus*. Sequences of 1 kb around the off-target *loci* are available in the Supplementary Sequence List II at the end of this Chapter and were used for alignments and primer design.

Off-target	Chromosome	Site	Sequence	Annotation
*ppo*1-3 OT1	SMEL3Ch08	97398340-97398363	GTTcTGGA_tCTTGGTTCTTT-CGGT	*ppo7*
*ppo*1-3 OT2	SMEL3Ch08	97402248-97402271	GTaATGGA_tCTTGGTTCTTT-TGGT	*ppo2*
*ppo*1-3 OT3	SMEL3Ch08	49334189-49334212	GTaATGGc_tCTgGGTTCTTT-TGGA	none
*ppo*1-3 OT4	SMEL3Ch00	15092:16193-16216	tTTAcGGA_atTTGGTTCTTT-TGGG	none
*ppo*4-5-6 OT1	SMEL3Ch08	97285544-97285566	TGAATGG_AAAcaAATCGGA-GGGA	*ppo3*
*ppo*4-5-6 OT2	SMEL3Ch08	97398006-97398028	cGAATGG_AAgGCAATaGGA-GGGA	*ppo7*
*ppo*4-5-6 OT3	SMEL3Ch08	97401914-97401936	cGAATGG_AAAGCgATaGGA-GGGA	*ppo2*
*ppo*4-5-6 OT4	SMEL3Ch06	88269542-88269564	TGAtTGG_AAActAATCGGA-TGGT	none

Table 35. Characteristics of putative off-target sites for the gRNAs directed at *ppo*1-3 and *ppo*4-5-6. Mismatches are in lowercase and bulges are represented by dashes.

Illumina sequencing results are reported in Table 36 for *ppo*1-3 off-targets, and in Table 37 for *ppo*4-5-6 off-targets. For some of the *loci* (*ppo*1-3 OT1 and OT3, and *ppo*4-5-6 OT1, OT2 and OT4) clearly no significant off-target effects are detectable, with the greatest editing efficiency being 1.64%. Here, variability is comparable to that of the untransformed wild type, and may result either from naturally occurring SNPs, or from sequencing errors (which are bound to occur at great sequencing depths such as the ones we obtained here). For other *loci* (*ppo*1-3 OT2 and OT4, and *ppo*4-5-6 OT3) significant variability exists in the wild type, as well as in the transformed samples: by analyzing randomly selected reads it was possible to see that more than one sequence was amplified and that thus variability is due to the presence of different amplicons, but not to editing. It is possible to conclude that no off target activity can be observed against *ppo7* concerning *ppo*1-3 OT1 and *ppo*4-5-6 OT2, as well as for *ppo3* concerning *ppo*4-5-6 OT1, which represents a good indication of the specificity of CRISPR/Cas9 in this gene family.

	Sample	N. of sequences	N. WT target sequence	% WT	Normalized % edited
PPO 1-3 OT1	WT	23986	23240	96,89%	0
	1	22076	21491	97,35%	0
	2	14645	14199	96,95%	0
	3	10764	10443	97,02%	0
	4	7587	7347	96,84%	0
	5	5057	4865	96,20%	0.71
	6	21529	20883	97,00%	0
	7	11850	11507	97,11%	0
	8	6895	6681	96,90%	0
	9	19834	19323	97,42%	0
	10	17461	16977	97,23%	0
PPO 1-3 OT2	WT	32242	15439	47,88%	-
	1	26978	15774	58,47%	-
	2	46885	24171	51,55%	-
	3	2578	1303	50,54%	-
	4	20899	11486	54,96%	-
	5	15479	7959	51,42%	-
	6	24955	12980	52,01%	-
	7	19844	12203	61,49%	-
	8	26747	13882	51,90%	-
	9	37090	19277	51,97%	-
	10	16920	9392	55,51%	-
PPO 1-3 OT3	WT	14708	14328	97,42%	0
	1	14569	14053	96,46%	0.99
	2	19598	19032	97,11%	0.32
	3	8886	8655	97,40%	0.02
	4	12997	12637	97,23%	0.20
	5	12304	11921	96,89%	0.54
	6	17340	16840	97,12%	0.31
	7	15745	15294	97,14%	0.29
	8	8936	8699	97,35%	0.07
	9	20440	19844	97,08%	0.35
	10	15311	14846	96,96%	0.47
PPO 1-3 OT4	WT	23370	15508	66,36%	-
	1	13149	11358	86,38%	-
	2	15850	10387	65,53%	-
	3	9827	7652	77,87%	-
	4	13463	11687	86,81%	-
	5	16632	12424	74,70%	-
	6	23117	16658	72,06%	-
	7	21603	18647	86,32%	-
	8	16460	11739	71,32%	-
	9	22340	16279	72,87%	-
	10	33063	25616	77,48%	-

Table 36. Quantification of Illumina reads edited at putative off-target *loci* for the PPO1-3 gRNA. For each individual and for each locus the total number of reads is reported, together with the percentage of reads carrying the wild type (non edited) off-target sequence. The percentage of edited sequences is estimated by normalizing editing efficiency to the wild type for each target.

	Sample	N. of sequences	N. WT target sequence	% WT	Normalized % edited
PPO 4-5-6 OT1	WT	17926	17377	96,94%	0
	1	28723	27874	97,04%	0
	2	26236	25428	96,92%	0.02
	3	24196	23449	96,91%	0.03
	4	20621	20030	97,13%	0
	5	18214	17497	96,06%	0.9
	6	18610	17745	95,35%	1.64
	7	15496	14836	95,74%	1.24
	8	23224	22456	96,69%	0.26
	9	33027	31975	96,81%	0.13
	10	16234	15608	96,14%	0.83
PPO 4-5-6 OT2	WT	17855	16501	92,42%	0
	1	25650	24643	96,07%	0
	2	25361	23439	92,42%	0
	3	21205	20283	95,65%	0
	4	13231	12755	96,40%	0
	5	16975	16014	94,34%	0
	6	23392	21722	92,86%	0
	7	11425	10939	95,75%	0
	8	23079	21344	92,48%	0
	9	31506	30182	95,80%	0
	10	18490	17614	95,26%	0
PPO 4-5-6 OT3	WT	27671	16913	61,12%	-
	1	33704	27755	82,35%	-
	2	35227	23550	66,85%	-
	3	30086	26883	89,35%	-
	4	20282	18634	91,87%	-
	5	28492	21881	76,80%	-
	6	34156	24689	72,28%	-
	7	17211	14172	82,34%	-
	8	8882	5965	67,16%	-
	9	21656	14386	66,43%	-
	10	28703	23793	82,89%	-
PPO 4-5-6 OT4	WT	26977	26174	97,02%	0
	1	50579	49363	97,60%	0
	2	41507	40426	97,40%	0
	3	34541	33719	97,62%	0
	4	274	266	97,08%	0
	5	39677	38664	97,45%	0
	6	13905	13483	96,97%	0.05
	7	38926	37940	97,47%	0
	8	22217	21610	97,27%	0
	9	39754	38784	97,56%	0
	10	27890	27241	97,67%	0

Table 37. Quantification of Illumina reads edited at putative off-target *loci* for the PPO4-5-6 gRNA. For each individual and for each *locus* the total number of reads is reported, together with the percentage of reads carrying the wild type (non edited) off-target sequence. The percentage of edited sequences is estimated by normalizing editing efficiency to the wild type for each target.

3.5.2 ‘Black Beauty’: genotyping of target *loci*

For ‘Black Beauty’ plantlets, we performed Sanger sequencing of the target sites and analyzed the resulting chromatograms with TIDE. The integration of the transgene was confirmed by amplification of the *hCas9 locus*. Results are reported in Tables 38, 39 and 40 for *ppo4-5-6*, respectively. No genotyping was performed on *ppo1* and *ppo3*, because of previous indications from the Illumina sequencing on ‘Ecavi’ that no editing activity was detectable for this gRNA. Unfortunately, of 15 plants, the number of edited individuals is very low: 4 for *ppo4*, 7 for *ppo5* and 2 for *ppo6*. The maximum editing efficiency is 75.3% for *ppo5*.

Because of the lack of activity of the first gRNA in the construct, a new construct was assembled to target *ppo1* and *ppo3*: here, 2 gRNAs were designed for each gene, each individually under the control of a U6-26 promoter. This construct will be used for transformation of ‘Black Beauty’ cotyledon explants.

ppo4									
Sample	**Overall efficiency**	**R^2**	**WT**	**+1 Insertion**	**Inserted base**	**-6 deletion**	**-4 deletion**	**-3 deletion**	**-1 deletion**
BB 1	3.1	0.99	95.8						
BB 2	0.6	0.98	97.3						
BB 3	66.5	0.97	30.4	48.0	G			13.6	
BB 4	38.4	0.98	59.8						32.3
BB 5	0.3	0.97	96.7						
BB 6	0.7	0.98	97.6						
BB 7	0	0.98	98.4						
BB 8	5.0	1.0	94.8						
BB 9	9.0	0.99	90.0						3.6
BB 10	6.4	0.99	93.1						
BB 11	6.9	0.99	92.5						
BB 12	5.3	0.99	94.1						2.5
BB 13	4.7	0.99	94.3						
BB 14	5.4	0.99	93.8						
BB 15	5.7	0.99	93.2						

Table 38. Genotyping of T_0 individuals by PCR amplification of the target *ppo4 locus*, direct Sanger sequencing and TIDE chromatogram decomposition.

ppo5										
Sample	**Overall efficiency**	**R^2**	**WT**	**+1 Insertion**	**Inserted base**	**-6 deletion**	**-4 deletion**	**-3 deletion**	**-2 deletion**	**-1 deletion**
BB 1	6.0	0.99	92.9							
BB 2	7.7	0.99	91.1							
BB 3	75.3	0.99	23.4						74.7	
BB 4	13.3	0.98	85.0							9.2
BB 5	5.3	0.99	93.4							
BB 6	3.2	0.99	95.8							
BB 7	6.1	0.98	92.0							
BB 8	1.3	1.0	98.3							
BB 9	13.0	0.99	85.9						4.0	5.2
BB 10	1.0	0.99	98.3							
BB 11	21.7	0.99	77.0				3.9		6.2	8.6
BB 12	16.7	0.98	81.1							10.4
BB 13	5.2	0.99	94.2							
BB 14	9.8	0.99	88.8							3.7
BB 15	15.5	0.98	82.2							9.0

Table 39. Genotyping of T_0 individuals by PCR amplification of the target *ppo5 locus*, direct Sanger sequencing and TIDE chromatogram decomposition.

ppo6										
Sample	**Overall efficiency**	**R^2**	**WT**	**+1 Insertion**	**Inserted base**	**-6 deletion**	**-4 deletion**	**-3 deletion**	**-2 deletion**	**-1 deletion**
BB 1	3.1	0.99	96.0							
BB 2	2.5	0.99	96.7							
BB 3	56.5	0.95	38.1	19.0	G			34.6		
BB 4	30.2	0.99	68.5				4.0			25.2
BB 5	1.8	0.99	96.8							
BB 6	1.5	0.99	97.4							
BB 7	3.4	0.99	95.5							
BB 8	2.6	0.99	96.2							
BB 9	3.5	0.99	95.5							
BB 10	2.9	0.99	96.2							
BB 11	2.7	0.99	96.1							
BB 12	2.9	0.99	95.6							
BB 13	2.8	0.99	96.0							
BB 14	2.5	0.98	96.0							
BB 15	1.6	0.99	97.7							

Table 40. Genotyping of T_0 individuals by PCR amplification of the target *ppo6 locus*, direct Sanger sequencing and TIDE chromatogram decomposition.

4 - Discussion

4.1 Identification of new PPO encoding *loci*

The search of the high quality eggplant genome with the previously annotated eggplant sequences encoding PPOs allowed us to identify a total of ten *ppo* genes in *S. melongena*. Extensive structural similarities can be found between these gene families in the three Solanaceae species. Ten genes have been identified to encode PPOs in eggplant, nine in potato, while six genes are annotated in tomato. Protein size is highly conserved, varying between 562 and 633 aminoacids (aa) for most proteins, with very few exceptions: potato PPO6 is only 430 aa in size, while some other proteins are considerably larger (eggplant PPO9 is 822 aa, and eggplant PPO10 is 731 aa). PPOs across the three species have the same conserved functional domains (Fig. 59):

i. the central common tyrosinase domain (Pfam00264), shared by polyphenol oxidases and hemocyanins, which binds the two copper ions in the catalytic site through two sets of three histidine residues;
ii. a conserved C-terminal domain of unknown function (DUF_B2219, Pfam 12142), characterized by a conserved KFDV motif, which is typical of Eukaryotes and is found in association with the tyrosinase domain, especially in plant polyphenol oxidase family;
iii. the polyphenol oxidase middle domain PPO1_DWL (Pfam 12143), with the conserved DWL motif and located between the previous two, found in Bacteria and Eukaryotes and associated to the tyrosinase domain.

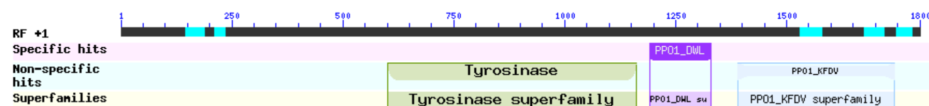


Figure 59. Annotated protein domains of eggplant PPO1. Common structure of PPO proteins, with three conserved annotated domains. All characterized PPOs from potato, eggplant and tomato have the same domains.

Interestingly, it seems that in eggplant too, similarly to what is observed in tomato and potato, all *ppos* cluster on chromosome 8, with the exception of a gene on chromosome 2 (*Smeppo10* in eggplant, *Stuppo9* in potato, and *Slyppo*_G in tomato, Table 28). In eggplant, genes located on chromosome 8 are physically divided in two main groups, with *ppo5* and *ppo6* located 840 kbp away from the other *ppo*s, which instead cluster in a more restricted region. In

particular, *ppo*s *1*, *2* and *7* are closely associated. Among previously annotated eggplant *ppo* genes (Shetty *et al.*, 2011), *loci* which are physically closely associated show a greater degree of sequence homology, which strongly suggests that they originated through replication of an ancestral *ppo* gene.

4.2 Functional specialization of PPOs in the Solanaceae family

Differential expression patterns have been identified for *ppo* genes in all three Solanaceae species (eggplant, tomato and potato). Specifically, two clusters have been characterized in phylogenetic analyses (Fig. 52), which identify two functional classes of PPOs, named A and B (Shetty *et al.*, 2011; Docimo *et al.*, 2016; Taranto *et al.*, 2017). These two classes include structurally diverse genes, whose expression has also been shown to be differentially regulated across plant tissues and in response to *stimuli* such as wounding. Shetty *et al.* (2011), in their study of eggplant PPOs, also highlighted that class A and B PPOs might be subject to differential evolutionary pressures: while analysis of substitutions in class A genes (*ppo1-2-3*) identified mainly neutral substitutions, in class B genes (*ppo4-5-6*) a positive selection appeared to be in place, which determined the occurrence of substitutions leading to the emergence of discrete functional and structural features. This is important, because it would imply a greater functional specialization among class B genes in comparison to class A.

In eggplant, the expression of all *ppo* genes was shown to be greater in young tissues and to diverge during plant development in mature and reproductive organs. Shetty *et al.* (2011) reported expression of *ppo1*, *ppo3* and *ppo4-6* in the fruit and in pre-anthesis flowers; *ppo2* exhibited a distinct transcriptional pattern, with no expression in fruits, but a predominant expression in roots and young leaves. Class A PPOs have been shown to be expressed in roots, with eggplant *ppo3* constituting an exception. In fruits, *ppo* expression was found to be concentrated in the exocarp (peel), in seeds and in areas of the berry surrounding the seeds, while expression was low in the mesocarp (flesh) (Shetty *et al.*, 2011). Also, *ppo* expression and localization was high around wounds induced either by herbivores or by knife cutting. Interestingly, in our analysis it appeared that, in response to cutting, some *ppo* genes (namely *ppo1*, *ppo3*, *ppo4* and *ppo5*) were strongly activated in the flesh of eggplant berries, with more dramatic increases in transcription than those observed in the flesh. Consistently with previous data, *ppo2* transcriptional levels were low in the fruit and transcription was not induced by cutting.

The involvement of PPOs in the response to mechanical damage, caused either by cutting or by the presence of herbivores, was extensively analyzed for a number of plant species, including tomato (Li and Steffens, 2002; Thipyapong *et al.*, 2004; Kavitha and Umesha, 2008) and potato (Castañera *et al.*, 1996). In potato, *ppo10* expression was shown to be elicited in response to fungal pathogens and pests (Chi *et al.*, 2014). The response of plants to mechanical damages involves the activation of multiple, complex pathways, whose regulation is often still unclear (Docimo *et al.*, 2016). In eggplant, oxidative browning in response to cutting was shown to be the result of the interaction of a number of metabolic responses, whose dynamics varied according to genotype (Docimo *et al.*, 2016) and which included transcriptional activation of *ppo*s as well as of genes involved in the synthesis of phenolic compounds. Docimo *et al.* (2016) showed that genotypes in which severe browning was observed in response to cutting had a transient and steep activation of class A *ppo* genes, while non-browning genotypes had a more gradual and durable activation of both classes of *ppo*s. The final browning phenotype thus seems to depend on the interaction of multiple factors, including *ppo* expression, total phenolics content and also the way in which the plant integrates environmental *stimuli* to elicit defense responses (Mishra *et al.*, 2013; Plazas *et al.*, 2013; Docimo *et al.*, 2016). We observed the activation of four *ppo* genes, belonging to both class A and class B, in the berry in response to cutting. The other members of the gene family, on the contrary, did not appear differentially expressed. *ppo* expression profiles point to a functional specialization within the gene family, which is crucial for the design of an appropriate editing strategy directed at reducing detrimental oxidative browning in fruit tissues (especially the flesh, which has the greatest impact for consumers' perception), while retaining the functions of PPOs in response to damaging and pests.

4.3 Editing a conserved gene family: gRNA design, multiplexing and genotyping

Targeting a numerous and conserved gene family, whose members have different functions and expression patterns, poses some challenges regarding both the design of the gRNAs and the evaluation of edited genotypes and off-targets.

Our initial experimental design aimed at targeting simultaneously all *ppo*s activated after cutting, and probably more involved in the browning of fruit tissues. To do this, we used a multiplexing approach, described by Xie *et al.* (2015) and Vazquez-Vilar *et al.* (2016), in which gRNAs are expressed as a

polycistronic transcript and subsequently processed thanks to the presence, before each gRNA sequence, of a tRNA which is recognized and cleaved. This kind of approach shows some advantages in comparison to the construction of vectors carrying a distinct transcriptional unit for each gRNA: (i) the construct has a smaller size; (ii) the two gRNAs are bound to be expressed at the same time, eliminating differences in gRNA expression levels for the two targets; and (iii) multiplexing is predicted to have a positive effect on gRNA expression, because tRNAs used to ensure processing and release of gRNAs can act as transcriptional enhancers for PolIII. In addition, the use of this multiplexing strategy eliminates the need to select a gRNA starting with a G: the constraint derives from the use of the U6 promoter, which requires the transcript to begin with a guanine, but in a multiplexing construct this requirement is met by the first tRNA at the beginning of the transcript. This broadens the pool of suitable gRNAs and allows to select targets with good specificity scores beginning with A, C or T, which would otherwise not be useful. Unfortunately, the choice of a multiplexing approach constituted a disadvantage in our particular case, since the gRNA directed at *ppo1* and *ppo3* was not effective, probably because of a 2 nucleotide deletion in the original pVD1 plasmid used to assemble the polycistronic construct, which affected functionality of the scaffold. Nevertheless, the approach remains effective in principle, and our transient expression data reported in Chapter II seem to confirm that GoldenBraid CRISPR multiplexing vectors are effective in targeting multiple *loci* at the same time.

The very high degree of sequence homology between members belonging to the two PPO classes (A and B) made it possible to select only two gRNAs, one targeting *ppo1* and *ppo3* (class A) and one *ppo4-5-6* (class B). Despite its potential high functionality, this approach has two main drawbacks, one functional, and the other technical. From the functional point of view, targeting conserved regions in a conserved gene family exponentially increases the likelihood of affecting off-target sites within the family (like *ppo2* and *ppo7*, in our case). On the technical side, instead, the high degree of conservation makes it difficult to design specific primers for the genotyping of edited *loci*. This difficulty had already been reported by a number of authors analyzing *ppo* transcriptional levels and activity (Chi *et al.*, 2014; Docimo *et al.*, 2016), which often led to the choice of amplifying, silencing or analyzing groups of *ppo*s together (for example amplifying all members of class B *ppo*s with the same primers). In our case, it was easier to achieve specificity for the evaluation of *ppo* transcriptional levels through qRT PCR, but it was more difficult to design discriminating primers for *ppo4* and *ppo6* for CRISPR genotyping, because of

amplicon size constraints dictated by the position of the target site. This is one of the reasons which led us to choose a deep sequencing approach to precisely estimate editing activity at each *locus*. The very high similarity in the target region between *ppo4* and *ppo6,* in fact, determined that both samples contained amplicons from the two *loci*, but they could be discriminated when aligning Illumina reads thanks to the presence of a number of conserved SNPs in the sequenced region.

4.4 Editing patterns and efficiency in *calli* of the 'Ecavi' variety

Because of the difficulties experienced in regenerating shoots from the 'Ecavi' variety, we conducted a preliminary screening of the activity of the CRISPR/Cas9 system by extracting DNA from ten randomly selected *calli*. This meant, however, that the situation that we were analyzing in these samples was necessarily more complex, from the genetic point of view, than that of a plantlet, because *calli* are composed of a mass of undifferentiated cells and are more chimeric. The editing efficiency reported through sequencing of DNA from *callus* is not expected to reflect the actual editing efficiency which could be assessed in regenerated shoots, because the cellular composition of such undifferentiated tissue is thought to be extremely more complex and chimeric. Although *hCas9* could be detected in all *calli*, it is likely that not all cells in the analyzed portion of the *callus* had integrated the transgene or were expressing it. In addition, whole plantlets are subject to greater selective pressure during their development, being exposed to a selective *medium* for a longer time than *calli*. However, the data we obtained was informative, by proving that editing was in fact occurring in our samples in a targeted way. The Illumina amplicon sequencing approach was chosen because it allowed to obtain a comprehensive view of the editing efficiency and of the types of mutations occurring in our samples, and because, thanks to its precision, it enabled us to discriminate between reads with a high degree of homology, in case primers were not specific. Sequencing allowed us to establish that only one gRNA (the one directed at *ppo4-5-6*) was active, and that the majority of *calli* exhibited some degree of editing at target *loci*. An exception is *callus* 10, in which virtually no mutated sequences were detectable for all three *loci*. In other *calli*, like *callus* 5, editing appeared to be uniformely more efficient, with all targets edit with efficiencies exceeding 25%, with values even as high as 61%. The alignment of 100 randomly selected reads with 10 wt sequences for each sample (see Fig. 58) led to identify which kinds of mutations were present, and these were quantified relative to the total of the non-wild type reads for each sample (Tables 31-32-33). As for tomato, the most recurrent mutations are small indels, in particular 1

nt deletions, which represent the highest share of mutated sequences. A number of alleles are present, all at relatively low frequencies, ranging from 1 nt to 6 nt deletions; we searched for deletions as large as 12 nt, but deletions larger than 6 nt are very rare. Single base insertions are present in all samples, but their frequency, except for some *calli* (like *callus* 1) is not as high as those observed in tomato, and is often significantly below 10% of edited sequences; thymine is still the most frequently inserted nucleotide, but it is not possible to identify a strong bias towards the insertion of this particular base. Interestingly, *callus* 5, which showed the greatest editing activity, has only two edited alleles present at high frequencies for each *locus*: for *ppo4* and *ppo6* these are 3 nt (50-60%) and 8 nt (10%) deletions, while for *ppo5* they are 6 nt (47.92%) and 7 nt (45.78%) deletions. This is important, because it means that mutations can be introduced very early in plant tissue culture and be propagated and fixed in a high number of cells, even in an undifferentiated tissue. Most deletions and insertions occur starting at the expected position 3 nt upstream of the PAM, and we quantified these particular alleles. However, alignments showed that a smaller proportion of edited sequences had indels at other positions with respect to the PAM (-2, -1 or in correspondence of the PAM). Interestingly, some *calli* (like *callus* 1 at the *ppo6 locus*) also show larger mutations (45-48 base pairs) which extend beyond the PAM site at its 3' end, effectively deleting all the gRNA complementary site and the PAM. When such larger deletions occur, repair seems to involve a greater degree of rearrangement at the cut site, including insertion of new bases, with no homology with adjacent sequences. Despite the lower efficiency due to the type of analyzed tissue and the fact that one gRNA was not active, these results still constitute a first evidence of the effectiveness of CRISPR/Cas9 mediated editing against a gene family in eggplant.

4.5 Editing specificity and off-target analysis in *calli* of the 'Ecavi' variety

The same Illumina amplicon sequencing approach was taken to analyze putative off-target *loci* in 'Ecavi' *calli*. This approach is extremely precise and informative and provides a deep insight into possible unspecific editing activity. As discussed above, one of the major risks about editing a conserved gene family is that off-targets are most likely to be other members of the same family. This is particularly problematic if, like in the case of *ppo*s, different members of the gene family have specialized functions and distinct activation patterns, meaning that editing other genes beyond the targets may have an impact on separate metabolic functions (in this case, defense). Overall, no off-target activity was detected, although among off-target sites were *ppo2*, *ppo3* and *ppo7*. All samples were further analyzed by aligning reads with ClustalX

and scanning alignments for specific mutations. Some variation was detected in three off-target *loci* (*ppo*1-3 OT2 and OT4, and *ppo*4-5-6 OT3), which was not due to unspecific editing, but to the presence of different sequences in the alignments. These additional sequences were very similar to the off-target *loci* which were being analyzed, but could not be attributed to other members of the *ppo* family, and might thus represent allelic variants of these *loci*. We can therefore conclude that editing of *ppo*s is specific and is limited to target sites, with other members of the gene family (namely *ppo2*, *ppo3* and *ppo7*) unaffected by CRISPR/Cas9 activity against *ppo4-5-6*.

4.6 Editing patterns and efficiency in plantlets of the 'Black Beauty' variety

The genotyping approach adopted for our second editing experiment was instead analogous to the one used for tomato plantlet genotyping, and consisted of sequencing target PCR products with a Sanger approach, and analyzing the resulting chromatogram to identify indels. Genotyping results are reported in Tables 38-40. The number of edited plants is low, and editing efficiencies are also significantly below those observed in tomato: while for *gf1* and *Ty5/pelo* in tomato we could recover plantlets with a proportion of edited *loci* above 90%, here the highest editing efficiency is 75.3%, and is found only in one plant for one *locus*. Of 15 'Black Beauty' transformed plants, only 4 are edited for *ppo4*, 7 for *ppo5* and 2 for *ppo6*. All plants are chimeric and retain a proportion of the wild type allele: the presence of multiple edited alleles within the same plant suggests that shoots have not originated from mutated cells, but rather that mutations have been accumulating over the course of plant development.

Plant 3 shows the greatest editing efficiency for all three *loci*: editing efficiencies are 66% for *ppo4*, 75.3% for *ppo5* and 55% for *ppo6*. Plant 4 is also mutated, although with a lower efficiency (max. 32% for *ppo4*). Plants 9 and 12 are edited with very low efficiencies (max. 10%) for *ppo4* and *ppo5*, but not for *ppo6*. In addition, plants 11, 14 and 15 have very low frequencies of edited alleles for *ppo5*. From this data, it seems that in whole regenerated plants a difference can be observed regarding editing efficiency for the different *loci* targeted by the same gRNA. This was also at least partly evident in the Illumina sequencing of target *loci*, in which proportions of edited alleles are consistently higher for *ppo5* than for *ppo4* and *ppo6*, especially for *callus* 5. The types of mutations found in 'Black Beauty' edited plantlets are mainly small deletions (1 to 4 bp, with a 1 bp deletion as the most frequent mutation) and, in plant 3, a single nucleotide insertion (a guanine). This pattern confirms data obtained from genotyping of the 'Ecavi' *calli*, where the most common mutation is a

single nucleotide deletion and in which no obvious preference for inserting a thymine is observable.

Given our previous work in tomato, reported in Chapters I and II, editing efficiencies in eggplant are markedly lower than expected, and might be traced to low levels of expression of Cas9 and gRNAs, as determined by Pan *et al.* (2016), or to the efficiency of the specific gRNAs. It is less likely that they might be due to a lack of integration of the transgene, since the plants grew on a kanamycin-containing *medium* until they were transferred to a rooting *medium* and thus withstood antibiotic selection. Moreover, almost 50% of them (especially for *ppo5*) report at least 3.7% of an edited allele, meaning that, however low, some CRISPR/Cas9 activity occurred. It is possible that repeating a genotyping analysis when plants reach fruit development may show that the proportion of edited alleles has increased, in analogy to what we observed for *gf1*, which would be beneficial for obtaining a mutated progeny for all three genes and which would offer an interesting insight into the ability of Cas9 to act in adult tissues.

Together with the 15 'Black Beauty' plants, we also analyzed the only 'Ecavi' plant (Ecavi 11) we were able to obtain from the first editing experiment. The results for its genotyping are reported in Table 34. This plant is edited for all three *loci*, with an efficiency of about 30% for *ppo4* and *ppo6*, and of 53.3% for *ppo5*, and shows a series of small deletions (from 1 to 6 nt).

Overall, genotyping data from both eggplant edited *calli* and plants points to a lower editing efficiency for this species in comparison to tomato. It also shows an editing pattern that is at least partially different in the two species, with a strong preference in eggplant for small deletions and a lower frequency of single nucleotide insertions, which were instead the hallmark of CRISPR/Cas9 gene editing in the *Ty5/pelo* mutants, and which were also recurrent in *gf1*.

4.7 Factors affecting *in vitro* regeneration efficiency of eggplant

In our experiments, one of the major goals was to obtain an efficient *in vitro* regeneration system for eggplant.

Regeneration efficiencies depend on the combination of three factors: genotype, growth regulators and the type of explant which is being employed. Unfortunately, we did not dispose of previous information about the regeneration potential of each variety, thus we tested a variety of conditions, including the use of different kinds of antioxidants (PVP, ascorbic and citric

acids), different combinations of growth regulators and a dark pretreatment at the beginning of *in vitro* culture.

Dark pretreatments, both during germination and during explant culture, are known to increase adventitious shoot formation in cotyledon, leaf and hypocotyl explants. A positive correlation between a dark pretreatment and regeneration efficiency has been found also for eggplant (Muktadir *et al.*, 2016), as well as in a number of other species (Mohamed *et al.*, 1992; Hsia and Korban, 1998; Cao *et al.*, 2002). The exact mechanism by which dark increases the regeneration potential of plant tissue cultures is unknown, but there seems to be an influence on the preservation of endogenous growth regulators and on the reduction of starch and oxidative phenolics accumulation. Darkness reduces cell wall accumulation of compounds, thus reducing cell wall thickness, and the development of vascular tissue: these factors tend to favor cell division and reallocation of growth regulators towards the regeneration sites (Trigiano and Gray, 1999).

Browning of plant tissues during *in vitro* culture, leading to necrosis, is one of the major causes for failure of tissue culture and shoot regeneration. Browning is associated to the oxidation of phenolics, whose release is prompted by the manipulation and cutting of explants. Different measures can be taken to control browning, among which the most common are frequent moving of explants and *calli* to fresh *medium*, dark treatments and addition of antioxidant or adsorbent compounds to culture media. Citric and ascorbic acids are among the *medium* additives commonly used to prevent browning and necrosis. These organic acids do not interfere with the exudation of phenolics in culture *media*, but they control browning by preventing their oxidation. Ascorbic acid alone (Abdelwahd *et al.*, 2008; Li *et al.*, 2015) or in combination with citric acid (Menin *et al.*, 2015) was reported to have a positive effect on *callus* and/or shoot formation in species as different as faba beans, banana and globe artichoke.

Polyvinylpyrrolidone (PVP), like activated charcoal, acts instead as an adsorbent and has been reported to be beneficial in plant *in vitro* culture to prevent oxidation (Menin *et al.*, 2015) and is notably used during the extraction of cellular components, also in eggplant, to bind and remove phenolics (Collonnier *et al.*, 2001; Ikeda *et al.*, 1999). Concerns about the use of PVP in culture media regard mainly its ability to bind not only phenolics, but also nutrients and phytohormones (Bhat and Chandel, 1991).

Eggplant tissues are rich in phenolic compounds, the most abundant of which is chlorogenic acid. The construct used to transform cotyledons carried gRNAs to

target *ppo* genes, and this might imply a reduction in PPO synthesis and activity in transformed tissues, which would in itself be beneficial to reduce browning. However, *calli* are undifferentiated tissues undergoing rapid cell division, and transformation efficiency might not be high enough to ensure that a sufficient number of cells actually carry the desired mutation; thus, it was advisable to consider measures to limit such oxidative damages.

In eggplant, we observed that citric and ascorbic acid do not affect *callus* proliferation, but are inhibitory of shoot formation, while retaining its efficiency in contrasting browning of tissue and necrosis. This might be consistent with the described effect of ascorbic acid as a mitotic stimulator in tobacco cells (de Pinto *et al.*, 1999).

PVP addition, instead, exerts a positive effect on shoot regeneration in eggplant. For 'Black Beauty' regeneration, the induction of shoot formation on PVP-supplemented *medium* appeared slower, but the total number of recovered shoots was notably higher (10 of a total of 15 rooted plantlets). Instead, no differences were found between shoots which had undergone the 3-day dark pretreatment and those who had not, with an even distribution of successful events between the two conditions. For the 'Ecavi' variety, failure to efficiently regenerate shoots from *calli* may have depended on the interaction of different factors, one of them being the use of ascorbic and citric acids as antioxidants. Although they were removed from the later phases of *in vitro* culture (elongation and rooting), they may have exerted a negative effect on the ability of plant tissue to form shoots. In this respect, PVP appears to be a more appropriate antioxidant.

5 - Conclusions

Members of the PPO protein family have a defense role against pathogens and pests in plants, but in food research PPOs are considered mainly for their negative impact on food quality due to oxidative browning. Different strategies, including the use of preservatives and of specific storage conditions, have been used to limit oxidative browning; in recent years, the genetic manipulation and regulation of the expression of *ppo* genes was also employed. In fact, the knock-out of polyphenol oxidases in different plant species (apple and potato especially), and more recently in mushrooms, has drawn interest to the production of food products with reduced browning. This approach can be very interesting for a species, such as eggplant, rich in phenolic compounds (especially chlorogenic acid). To date, the selection of commercial varieties with low flesh browning has resulted in the indirect selection of genotypes with low concentrations of phenolics, and thus reduced nutraceutical properties.

In this chapter, we reported for the first time a successful protocol for CRISPR/Cas9 gene editing in eggplant and for the analysis of edited alleles and the screening of off-target sites. The use of one guide RNA directed simultaneously at three members of the *ppo* gene family allowed us to effectively establish the specificity of the system, by producing targeted mutations in all three targets, without detectable off-target effects on other members of the same gene family (especially *ppo2*, *ppo3* and *ppo7*). Moreover, a regeneration protocol was successfully set up for the 'Black Beauty' genotype. Further genotypes will need to be evaluated and, especially, it will be crucial to test our system with eggplant varieties with a high content of phenolic compounds, in order to obtain varieties enriched in antioxidants and with a low level of oxidation; also, these will provide us information about the impact of phenolics on regeneration efficiency. So far, we were only able to evaluate the positive outcome of the regeneration protocol and the genotype associated to the CRISPR/Cas9 editing of *ppo4-5-6* in T_0 plants. The phenotypic effect on the browning of fruit tissues will be evaluated in fruits of the T_0 generation and confirmed in the T_1 generation.

We have established a successful protocol for gene editing in eggplant, adding to the list of Solanaceae species for which CRISPR/Cas9 represents an alluring option for the introduction of specific mutations.

Bibliography

Abdelwahd, R. *et al.* (2008) ‘Use of an adsorbent and antioxidants to reduce the effects of leached phenolics in in vitro plantlet regeneration of faba bean’, *African Journal of Biotechnology*, 7(8), pp. 997–1002.

Araji, S. *et al.* (2014) ‘Novel Roles for the Polyphenol Oxidase Enzyme in Secondary Metabolism and the Regulation of Cell Death in Walnut’, *Plant Physiology*, 164(3), pp. 1191–1203. doi: 10.1104/pp.113.228593.

Arpaia, S. *et al.* (1997) ‘Production of transgenic eggplant (Solanum melongena L.) resistant to Colorado Potato Beetle (Leptinotarsa decemlineata Say)’, *Theoretical and Applied Genetics*, 95(3), pp. 329–334. doi: 10.1007/s001220050567.

Bhat, S. R. and Chandel, K. P. S. (1991) ‘A novel technique to overcome browning in tissue culture’, *Plant Cell Reports*, 10, pp. 358–361.

Bhonwong, A. *et al.* (2009) ‘Defensive role of tomato polyphenol oxidases against cotton bollworm helicoverpa armigera and beet armyworm spodoptera exigua’, *Journal of Chemical Ecology*, 35(1), pp. 28–38. doi: 10.1007/s10886-008-9571-7.

Cao, X., Hammerschlag, F. A. and Douglass, L. (2002) ‘A Two-step Pretreatment Significantly Enhances Shoot Organogenesis from Leaf Explants of Highbush Blueberry’, *HortScience*, 37(5), pp. 819–821.

Castañera, P., Steffens, J. C. and Tingey, W. M. (1996) ‘Biological performance of Colorado potato beetle larvae on potato genotypes with differing levels of polyphenol oxidase’, *Journal of Chemical Ecology*, 22(1), pp. 91–101. doi: 10.1007/BF02040202.

Chi, M. *et al.* (2014) ‘Reduced polyphenol oxidase gene expression and enzymatic browning in potato (Solanum tuberosum L.) with artificial microRNAs’, *BMC Plant Biology*, 14(1), p. 62. doi: 10.1186/1471-2229-14-62.

Collonnier, C. *et al.* (2001) ‘Source of resistance against Ralstonia solanacearum in fertile somatic hybrids of eggplant (Solanum melongena L .) with Solanum aethiopicum L’, *Plant Science*, 160, pp. 301–313.

Docimo, T. *et al.* (2016) ‘Insights in the fruit flesh browning mechanisms in Solanum melongena genetic lines with opposite postcut behavior’, *Journal of Agricultural and Food Chemistry*, 64(22), pp. 4675–4685. doi: 10.1021/acs.jafc.6b00662.

Gerdemann, C. *et al.* (2002) ‘Comparative modeling of the latent form of a plant catechol oxidase using a molluskan hemocyanin structure’, *Journal of Inorganic Biochemistry*, 89(1–2), pp. 155–158. doi: 10.1016/S0162-0134(01)00399-3.

Gooding, P. S., Bird, C. and Robinson, S. P. (2001) ‘Molecular cloning and characterisation of banana fruit polyphenol oxidase’, *Planta*, 213(5), pp. 748–757. doi: 10.1007/s004250100553.

Hsia, C. and Korban, S. S. (1998) ‘Effect of Growth Regulators , Dark Treatment and Light Intensity on Shoot Organogenesis from Leaf Tissues of Evergreen Azalea’, *Journal of Horticultural Science and Biotechnology*, 73(1), pp. 53–60. doi: 10.1080/14620316.1998.11510943.

Hunt, M. D. *et al.* (1993) ‘cDNA cloning and expression of potato polyphenol oxidase’, *Plant Molecular Biology*, 21(1), pp. 59–68. doi: 10.1007/BF00039618.

Ikeda, T. *et al.* (1999) 'Growth dependence of ovaries of facultatively parthenocarpic eggplant in vitro on indole-3-acetic acid content', *Scientia Horticolturae*, 79, pp. 143–150.

Jaenicke, E. and Decker, H. (2003) 'Tyrosinases from crustaceans form hexamers', *Biochemical Journal*, 371(2), p. 515 LP-523. Available at: http://www.biochemj.org/content/371/2/515.abstract.

Jukanti, A. (2017) *Polyphenol Oxidases (PPOs) in Plants*. doi: 10.1007/978-981-10-5747-2.

Kaintz, C., Mauracher, S. G. and Rompel, A. (2014) *Type-3 copper proteins: Recent advances on polyphenol oxidases*. 1st edn, *Advances in Protein Chemistry and Structural Biology*. 1st edn. Elsevier Inc. doi: 10.1016/bs.apcsb.2014.07.001.

Kavitha, R. and Umesha, S. (2008) 'Regulation of defense-related enzymes associated with bacterial spot resistance in tomato', *Phytoparasitica*, 36(2), pp. 144–159. doi: 10.1007/BF02981327.

Li, L. and Steffens, J. C. (2002) 'Overexpression of polyphenol oxidase in transgenic tomato plants results in enhanced bacterial disease resistance', *Planta*, 215(2), pp. 239–247. doi: 10.1007/s00425-002-0750-4.

Li, Z. *et al.* (2015) 'Cas9-Guide RNA Directed Genome Editing in Soybean', *Plant Physiology*, 169(2), pp. 960–970. doi: 10.1104/pp.15.00783.

Lieberei, R., Biehl, B. and Voigt, J. (1981) 'Serological studies on phenolase from spinach leaves', *Phytochemistry*, 20(9), pp. 2109–2116. doi: https://doi.org/10.1016/0031-9422(81)80095-7.

Magioli, C. *et al.* (1998) 'Efficient shoot organogenesis of eggplant (Solanum melongena L.) induced by thidiazuron', *Plant Cell Reports*, 17(8), pp. 661–663. doi: 10.1007/s002990050461.

Mari, S. *et al.* (1998) 'Unfolding and refolding of active apple polyphenol oxidase', *Phytochemistry*, 49(5), pp. 1213–1217. doi: 10.1016/S0031-9422(98)00142-3.

Marusek, C. M. *et al.* (2006) 'Comparative analysis of polyphenol oxidase from plant and fungal species', *Journal of Inorganic Biochemistry*, 100(1), pp. 108–123. doi: 10.1016/j.jinorgbio.2005.10.008.

Menin, B. *et al.* (2015) 'Cynara cardunculus L. var. scolymus as a system for the production of caffeoylquinic acids', *Journal of Horticultural Science and Biotechnology*, 88(5), pp. 537–542. doi: 10.1080/14620316.2013.11513003.

Mishra, B. B., Gautam, S. and Sharma, A. (2013) 'Free phenolics and polyphenol oxidase (PPO): The factors affecting post-cut browning in eggplant (Solanum melongena)', *Food Chemistry*. Elsevier Ltd, 139(1–4), pp. 105–114. doi: 10.1016/j.foodchem.2013.01.074.

Mohamed, M. F., Read, P. E. and Coyne, D. P. (1992) 'Dark Preconditioning , CPPU , and Thidiazuron Promote Shoot Organogenesis on Seedling Node Explants of Common and Faba Beans', *J. Amer. Soc. Hort. Sci.*, 117(4), pp. 668–672.

Mohammadi, M. and Kazemi, H. (2002) 'Changes in peroxidase and polyphenol oxidase activities in susceptible and resistant wheat heads inoculated with Fusarium graminearum and induced resistance', *Plant Science*, 162(4), pp. 491–498. doi: 10.1016/S0168-9452(01)00538-6.

Muktadir, M. A. *et al.* (2016) 'Regeneration efficiency based on genotype, culture condition and growth regulators of eggplant (Solanum melongena L.)', *Agriculture and Natural Resources*.

Elsevier Ltd, 50, pp. 38–42. doi: 10.1016/j.anres.2014.10.001.

Newman, S. M. *et al.* (1993) ‘Organisation of the tomato polyphenol oxidase gene family’, *Plant Molecular Biology*, 21(6), pp. 1035–1051. doi: 10.1007/BF00023601.

Newman, S. M., Tantasawat, P. and Steffens, J. C. (2011) ‘Tomato polyphenol oxidase B is spatially and temporally regulated during development and in response to ethylene’, *Molecules*, 16(1), pp. 493–517. doi: 10.3390/molecules16010493.

Niranjan Raj, S., Sarosh, B. R. and Shetty, H. S. (2006) ‘Induction and accumulation of polyphenol oxidase activities as implicated in development of resistance against pearl millet downy mildew disease’, *Functional Plant Biology*, 33(6), pp. 563–571. doi: 10.1071/FP06003.

Pan, C. *et al.* (2016) ‘CRISPR/Cas9-mediated efficient and heritable targeted mutagenesis in tomato plants in the first and later generations’, *Scientific Reports*. Nature Publishing Group, 6(1), p. 24765. doi: 10.1038/srep24765.

de Pinto, M. C., Francis, D. and De Gara, L. (1999) ‘The redox state of the ascorbate-dehydroascorbate pair as a specific sensor of cell division in tobacco BY-2 cells’, *Protoplasma*, 209(1–2), pp. 90–97. doi: 10.1007/BF01415704.

Plazas, M. *et al.* (2013) ‘Diversity and relationships in key traits for functional and apparent quality in a collection of eggplant: Fruit phenolics content, antioxidant activity, polyphenol oxidase activity, and browning’, *Journal of Agricultural and Food Chemistry*, 61(37), pp. 8871–8879. doi: 10.1021/jf402429k.

Shetty, S. M., Chandrashekar, A. and Venkatesh, Y. P. (2011) ‘Eggplant polyphenol oxidase multigene family: Cloning, phylogeny, expression analyses and immunolocalization in response to wounding’, *Phytochemistry*. Elsevier Ltd, 72(18), pp. 2275–2287. doi: 10.1016/j.phytochem.2011.08.028.

Taranto, F. *et al.* (2017) ‘Polyphenol oxidases in crops: Biochemical, physiological and genetic aspects’, *International Journal of Molecular Sciences*, 18(2). doi: 10.3390/ijms18020377.

Thipyapong, P., Hunt, M. D. and Steffens, J. C. (2004) ‘Antisense downregulation of polyphenol oxidase results in enhanced disease susceptibility’, *Planta*, 220(1), pp. 105–117. doi: 10.1007/s00425-004-1330-6.

Thipyapong, P. and Steffens, J. C. (1997) ‘Tomato Polyphenol Oxidase (Differential Response of the Polyphenol Oxidase F Promoter to Injuries and Wound Signals).’, *Plant physiology*, 115, pp. 409–418. doi: 10.1104/pp.115.2.409.

Thipyapong, P., Stout, M. J. and Attajarusit, J. (2007) ‘Functional analysis of polyphenol oxidases by antisense/sense technology’, *Molecules*, 12(8), pp. 1569–1595. doi: 10.3390/12081569.

Tran, L. T., Taylor, J. S. and Constabel, C. P. (2012) ‘The polyphenol oxidase gene family in land plants: Lineage-specific duplication and expansion’, *BMC Genomics*, 13(1). doi: 10.1186/1471-2164-13-395.

Trigiano, R. N. and Gray, D. J. (1999) *Plant Tissue Culture Concepts and Laboratory Exercises*. 2nd edn. CRC Press.

Vanitha, S. C., Niranjana, S. R. and Umesha, S. (2009) ‘Role of phenylalanine ammonia lyase and polyphenol oxidase in host resistance to bacterial wilt of Tomato’, *Journal of Phytopathology*, 157(9), pp. 552–557. doi: 10.1111/j.1439-0434.2008.01526.x.

Vazquez-Vilar, M. *et al.* (2016) ‘A modular toolbox for gRNA–Cas9 genome engineering in plants based on the GoldenBraid standard’, *Plant Methods*. BioMed Central, 12(1), p. 10. doi: 10.1186/s13007-016-0101-2.

Waltz, E. (2015a) ‘Nonbrowning GM apple cleared for market’, *Nature biotechnology*. Nature Publishing Group, 33(4), pp. 326–327. doi: 10.1038/nbt0415-326c.

Waltz, E. (2015b) ‘USDA approves next-generation GM potato’, *Nature biotechnology*, 33(1), pp. 12–13. doi: 10.1038/nbt0115-12.

Waltz, E. (2016) ‘Gene-edited CRISPR mushroom escapes US regulation’, *Nature*, 532(7599), p. 293. doi: 10.1038/nature.2016.19754.

Xie, K., Minkenberg, B. and Yang, Y. (2015) ‘Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system’, *Proceedings of the National Academy of Sciences*, 112(11), pp. 3570–3575. doi: 10.1073/pnas.1420294112.

Yoruk, R. and Marshall, M. M. R. (2003) ‘Physicochemical properties and function of plant polyphenol oxidase: a review’, *Journal of Food Biochemistry*, 27(352), pp. 361–422. doi: 10.1111/j.1745-4514.2003.tb00289.x.

Component	12 µl reaction	Final concentration
pUPD2 destination vector	1 µl	6.25 ng $µl^{-1}$
pVD1 GB1208/GB1207	1 µl	6.25 ng $µl^{-1}$
gRNA *PPO1-3* / gRNA *PPO4-5-6* 2 µM	1 µl	0.17 µM
10X T4 ligase buffer	1.2 µl	1X
10X BSA	1.2 µl	1X
T4 ligase	1 µl	3 U
*Bsmb*I	1 µl	10 U
Water	4.6	N/A

Supplementary Table S17. GoldenBraid reactions for the assembly of pUPD2 (level 0) vectors for the multiplexing assembly of gRNAs. Each of reaction allows to domesticate one gRNA for a multiplexing construct. The outcome of every reaction is a pUPD2 vector carrying a tRNA:gRNA:scaffold insert. tRNA and scaffold RNA are specific for every position of the multiplexing construct, because their external 4 bp overhangs specify their position (1 or 2) relative to each other and to the promoter (which will be added in the following step). tRNAs and gRNAs are carried on pVD1 vectors: GB1208 specifies position 1 of a 2 gRNA multiplexing construct, and is assembled with gRNA1 (targeting *PPO1-3*), while GB1207 specifies position 2 and is assembled with gRNA2 (targeting *PPO4-5-6*). These reactions were performed in parallel.

Component	12 µl reaction	Final concentration
pDGB3 alpha1 destination vector	1 µl	6.25 ng $µl^{-1}$
pUPD2 tRNA:gRNAppo1-3:scaffold	1 µl	6.25 ng $µl^{-1}$
pUPD2 tRNA:gRNAppo4-5-6:scaffold	1 µl	6.25 ng $µl^{-1}$
10X T4 ligase buffer	1.2 µl	1X
10X BSA	1.2 µl	1X
T4 ligase	1 µl	3 U
*Bsa*I	1 µl	10 U
Water	4.6	N/A

Supplementary Table S18. Level 1 GoldenBraid reaction for the assembly of a CRISPR-Cas9 vector with 2 gRNAs. Assembly of the pDGB3 alpha1 vector carrying the U6 promoter and the 2 tRNA:gRNA:scaffold inserts from previously assembled pUPD2 vectors.

Component	12 µl reaction	Final concentration
pDGB3 omega 2 destination vector	1 µl	6.25 ng μl^{-1}
pDGB3 alpha1 gRNA *ppo1-3* - gRNA*ppo4-5-6*	1 µl	6.25 ng μl^{-1}
GB0639 pDGB3 alpha2 35S:hCas9:Tnos	1 µl	6.25 ng μl^{-1}
10X T4 ligase buffer	1.2 µl	1X
10X BSA	1.2 µl	1X
T4 ligase	1 µl	3 U
*Bsmb*I	1 µl	10 U
Water	4.6	N/A

Supplementary Table S19. Level >1 GoldenBraid reaction for the assembly of a CRISPR-Cas9 vector with 2 gRNAs. Assembly of the pDGB3 omega2 vector carrying the insert of the previously assembled pDGB3 alpha1 U6:gRNA1:gRNA2:scaffold and the hCas9 TU with the 35S promoter and Tnos terminator.

Component	12 µl reaction	Final concentration
pDGB3 alpha1 destination vector	1 µl	6.25 ng μl^{-1}
pDGB3 omega2 gRNA *ppo1-3* - gRNA*ppo4-5-6* - hCas9	1 µl	6.25 ng μl^{-1}
GB1181 pDGB3 omega1R Tnos:nptII:Pnos	1 µl	6.25 ng μl^{-1}
10X T4 ligase buffer	1.2 µl	1X
10X BSA	1.2 µl	1X
T4 ligase	1 µl	3 U
*Bsa*I	1 µl	10 U
Water	4.6	N/A

Supplementary Table S20. Level >1 GoldenBraid reaction for the assembly of a CRISPR-Cas9 vector with 2 gRNAs. Assembly of the final pDGB3 alpha1 vector carrying the 2 gRNAs and hCas9 TUs, and the NptII selection marker with the Pnos promoter and Tnos terminator.

Compound	MGL *medium* pH 7	TY *medium* pH 5.8
Tryptone	5 g l^{-1}	5 g l^{-1}
Yeast extract	2.5 g l^{-1}	3 g l^{-1}
NaCl	0.1 g l^{-1}	-
Mannitol	5 g l^{-1}	-
Glutamic acid	1.15 g l^{-1}	-
KH_2PO_4	0.25 g l^{-1}	-
$MgSO_4.7H_2O$	100 g l^{-1}	0.5 g (2 mM)
Biotin	1 mg l^{-1}	-
Autoclave		
Acetosyringone	-	200 µM
Kanamycin/Spectinomycin	50 mg l^{-1}	-
Rifampicin	50 mg l^{-1}	-

Supplementary Table S21. Composition of the modified MGL and the TY media for LBA4404 culture.

Compound	Concentration
MS basal salt mixture	4.5 g l^{-1}
MES	0.5 g l^{-1}
Glucose	20 g l^{-1}
Phytoagar	10 g l^{-1}
pH 5.8, Autoclave	
Gamborg vitamins	1 ml l^{-1}
Trans-Zeatin	0.5 mg l^{-1}
BAP	0.3 mg l^{-1}
KIN	0.2 mg l^{-1}
NAA	0.1 mg l^{-1}
Acetosyringone	200 µM

Supplementary Table S22. Co-culture *medium* composition, from Arpaia *et al.* (1997).

Induction *medium*		**Shoot induction *medium***	
Compound	**Concentration**	**Compound**	**Concentration**
MS basal salt mixture	4.5 g l^{-1}	MS basal salt mixture	5 g l^{-1}
MES	0.5 g l^{-1}	MES	0.5 g l^{-1}
Glucose	20 g l^{-1}	Glucose	20 g l^{-1}
Phytoagar	10 g l^{-1}	Phytoagar	10 g l^{-1}
Ascorbic acid	5 mg l^{-1}	Ascorbic acid	5 mg l^{-1}
Citric acid	5 mg l^{-1}	MES	0.5 g l^{-1}
pH 5.8, Autoclave		pH 5.8, Autoclave	
Gamborg vitamins	1 ml l^{-1}	Gamborg vitamins	1 ml l^{-1}
Trans-Zeatin	0.5 mg l^{-1}	TDZ	0.2 µM
BAP	0.3 mg l^{-1}		
KIN	0.2 mg l^{-1}		
NAA	0.1 mg l^{-1}		
Kanamycin	30 mg l^{-1}	Kanamycin	30 mg l^{-1}
Carbenicillin	400 mg l^{-1}	Carbenicillin	400 mg l^{-1}

Supplementary Table S23. Media composition for the induction of *callus* and shoot formation, from Arpaia *et al.* (1997) and Magioli *et al.* (1988), respectively. Each *medium* was tested with a normal light cycle of 16-8 hours light-dark, at 24°C.

Compound	Concentration
MS basal salt mixture	4.5 g l^{-1}
MES buffer	0.5 g l^{-1}
Sucrose	30 g l^{-1}
Phytoagar	10 g l^{-1}
pH 5.8, Autoclave	
Gamborg vitamins	1 ml l^{-1}
Trans-Zeatin	2 mg l^{-1}
Indoleacetic Acid	0.1 mg l^{-1}
Acetosyringone	200 µM

Supplementary Table S24. Co-culture *medium* composition, from Muktadir *et al.* (2016).

Basal composition		Supplemented with ascorbic and citric acid		Supplemented with PVP40	
Compound	**Concentration**	**Compound**	**Concentration**	**Compound**	**Concentration**
MS basal salt mixture	5 g l^{-1}	MS basal salt mixture	5 g l^{-1}	MS basal salt mixture	5 g l^{-1}
Sucrose	30 g l^{-1}	Sucrose	30 g l^{-1}	Sucrose	30 g l^{-1}
Phytoagar	10 g l^{-1}	Phytoagar	10 g l^{-1}	Phytoagar	10 g l^{-1}
		Ascorbic acid	5 mg l^{-1}		
		Citric acid	5 mg l^{-1}		
pH 5.8, Autoclave		pH 5.8, Autoclave		pH 5.8, Autoclave	
Gamborg vitamins	1 ml l^{-1}	Gamborg vitamins	1 ml l^{-1}	Gamborg vitamins	1 ml l^{-1}
Trans-Zeatin	2 mg l^{-1}	Trans-Zeatin	2 mg l^{-1}	Trans-Zeatin	2 mg l^{-1}
Indoleacetic Acid	0.1 mg l^{-1}	Indoleacetic Acid	0.1 mg l^{-1}	Indoleacetic Acid	0.1 mg l^{-1}
				PVP40	200 mg l^{-1}
Kanamycin	30 mg l^{-1}	Kanamycin	30 mg l^{-1}	Kanamycin	30 mg l^{-1}
Carbenicillin	400 mg l^{-1}	Carbenicillin	400 mg l^{-1}	Carbenicillin	400 mg l^{-1}

Supplementary Table S25. Media composition for the induction of *callus* and shoot formation, from Muktadir *et al.* (2016). Each *medium* was tested with a normal light cycle of 16-8 hours light-dark, at 24°C.

Elongation *medium*		Rooting *medium*	
Compound	**Concentration**	**Compound**	**Concentration**
MS basal salt mixture	5 g l^{-1}	MS basal salt mixture	2.5 g l^{-1}
Sucrose	30 g l^{-1}	Sucrose	30 g l^{-1}
Phytoagar	10 g l^{-1}	Phytoagar	8 g l^{-1}
pH 5.8, Autoclave		pH 5.8, Autoclave	
Gamborg vitamins	1 ml l^{-1}	Gamborg vitamins	0.5 ml l^{-1}
Kanamycin	30 mg l^{-1}	Indolebutyric acid	0.2 mg l^{-1}
Carbenicillin	200 mg l^{-1}	Carbenicillin	200 mg l^{-1}

Supplementary Table S26. Composition of elongation and rooting media, based on Muktadir *et al.* (2016).

Locus	Primer Forward	Primer Reverse
ppo1	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**CCATAGATGGTACCTGTACTTCT	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**GTTGTTACTCATCAACTGTAATTG
ppo3	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**TTGTACTTCTACGAGAGGATGT	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**AGTTATTACTCATCAACTGTAGTTC
ppo4	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**CTCTTGGAACGGATCCAAAG	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**AATTCCGAGTTCAACCAATC
ppo5	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**CTCTGGGAACCGATCCAAGT	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**GATTCCGAGTTCAACCAATC
ppo6	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**TCTCGGAACGGATCCAAAG	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**AATTCCGAGTTCAACCAATC
***ppo4* new**	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**TCACACTCCAGTCCACAT	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**AATTCCGAGTTCAACCAATC
***ppo6* new**	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**TCACAATGCGGTCCACAA	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**AATTCCGAGTTCAACCAATC
***ppo1-3* OT 1**	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**TTCCATAGATGGTACTTGTACTTC	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**AGTTATTACTCATCAGCTGTAATTG
***ppo1-3* OT 2**	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**CCATAGATGGTACTTGTACTTCT	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**GTTATTACTCATCAACTGTAGTTC
***ppo1-3* OT 3**	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**ATTATTCTTGCTTTTCCAGGGG	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**GAGATTTATGTCCAAGTAATTGGG
***ppo1-3* OT 4**	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**GACTGAAAATATACGTACCC	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**GTTCTGTAGTAGAGATAGAATTCAG
***ppo4-5-6* OT 1**	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**GGAGTGCAGGGAACCATT	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**GAATTCTGAGTCCAGCCAATC
***ppo4-5-6* OT 2**	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**GGGAACCGTTGAAGTCAT	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**GAACTCCGAGTTCAACCAA
***ppo4-5-6* OT 3**	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**AACCATTGAAGTCATCCCT	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**GAGTTCAACCAATCTTTATGTG
***ppo4-5-6* OT 4**	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**TTGCGGGATAAAGAAGTTG	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**GAACTCTTATCAATCCATCACA

Supplementary Table S27. Primers for the first round amplification for *ppo1*, *ppo3*, *ppo4*, *ppo5* and *ppo6*, and for potential off-target *loci* for the *ppo1-3* and *ppo4-5-6* gRNAs. In bold is the universal Illumina adapter, while the 3' terminal part of the primer is specific for each *locus*.

Individual	Nextera Code	Index Sequence	Oligonucleotide Sequence
1	S502	CTCTCTAT	AATGATACGGCGACCACCGAGATCTACAC**CTCTCTAT**TCGTCGGCAGCGTC
2	S503	TATCCTCT	AATGATACGGCGACCACCGAGATCTACAC**TATCCTCT**TCGTCGGCAGCGTC
3	S505	GTAAGGAG	AATGATACGGCGACCACCGAGATCTACAC**GTAAGGAG**TCGTCGGCAGCGTC
4	S506	ACTGCATA	AATGATACGGCGACCACCGAGATCTACAC**ACTGCATA**TCGTCGGCAGCGTC
5	S507	AAGGAGTA	AATGATACGGCGACCACCGAGATCTACAC**AAGGAGTA**TCGTCGGCAGCGTC
6	S508	CTAAGCCT	AATGATACGGCGACCACCGAGATCTACAC**CTAAGCCT**TCGTCGGCAGCGTC
7	S510	CGTCTAAT	AATGATACGGCGACCACCGAGATCTACAC**CGTCTAAT**TCGTCGGCAGCGTC
8	S511	TCTCTCCG	AATGATACGGCGACCACCGAGATCTACAC**TCTCTCCG**TCGTCGGCAGCGTC
9	S513	TCGACTAG	AATGATACGGCGACCACCGAGATCTACAC**TCGACTAG**TCGTCGGCAGCGTC
10	S515	TTCTAGCT	AATGATACGGCGACCACCGAGATCTACAC**TTCTAGCT**TCGTCGGCAGCGTC
WT	S516	CCTAGAGT	AATGATACGGCGACCACCGAGATCTACAC**CCTAGAGT**TCGTCGGCAGCGTC

Supplementary Table S28. Primers for the second round amplification for *ppo1*, *ppo3*, *ppo4*, *ppo5* and *ppo6*, and for potential off-target *loci* for the *ppo1-3* and *ppo4-5-6* gRNAs. The S Nextera indexes (S502-S216) identify individuals.

Locus	Nextera Code	Index Sequence	Oligonucleotide Sequence
ppo1	N701	TAAGGCGA	CAAGCAGAAGACGGCATACGAGAT**TCGCCTTA**GTCTCGTGGGCTCGG
ppo3	N702	CGTACTAG	CAAGCAGAAGACGGCATACGAGAT**CTAGTACG**GTCTCGTGGGCTCGG
ppo4	N703	AGGCAGAA	CAAGCAGAAGACGGCATACGAGAT**TTCTGCCT**GTCTCGTGGGCTCGG
ppo5	N704	TCCTGAGC	CAAGCAGAAGACGGCATACGAGAT**GCTCAGGA**GTCTCGTGGGCTCGG
ppo6	N705	GGACTCCT	CAAGCAGAAGACGGCATACGAGAT**AGGAGTCC**GTCTCGTGGGCTCGG
***ppo4* new**	N706	TAGGCATG	CAAGCAGAAGACGGCATACGAGAT**CATGCCTA**GTCTCGTGGGCTCGG
***ppo6* new**	N707	CTCTCTAC	CAAGCAGAAGACGGCATACGAGAT**GTAGAGAG**GTCTCGTGGGCTCGG
***ppo1-3* OT 1**	N710	CGAGGCTG	CAAGCAGAAGACGGCATACGAGAT**CAGCCTCG**GTCTCGTGGGCTCGG
***ppo1-3* OT 2**	N711	AAGAGGCA	CAAGCAGAAGACGGCATACGAGAT**TGCCTCTT**GTCTCGTGGGCTCGG
***ppo1-3* OT 3**	N712	GTAGAGGA	CAAGCAGAAGACGGCATACGAGAT**TCCTCTAC**GTCTCGTGGGCTCGG
***ppo1-3* OT 4**	N714	GCTCATGA	CAAGCAGAAGACGGCATACGAGAT**TCATGAGC**GTCTCGTGGGCTCGG
***ppo4-5-6* OT 1**	N715	ATCTCAGG	CAAGCAGAAGACGGCATACGAGAT**CCTGAGAT**GTCTCGTGGGCTCGG
***ppo4-5-6* OT 2**	N716	ACTCGCTA	CAAGCAGAAGACGGCATACGAGAT**TAGCGAGT**GTCTCGTGGGCTCGG
***ppo4-5-6* OT 3**	N718	GGAGCTAC	CAAGCAGAAGACGGCATACGAGAT**GTAGCTCC**GTCTCGTGGGCTCGG
***ppo4-5-6* OT 4**	N719	GCGTAGTA	CAAGCAGAAGACGGCATACGAGAT**TACTACGC**GTCTCGTGGGCTCGG

Supplementary Table S29. Primers for the second round amplification for 5 potential off-target *loci* for *ppo1*, *ppo3*, *ppo4*, *ppo5* and *ppo6*, and for potential off-target *loci* for the *ppo1-3* and *ppo4-5-6* gRNAs. The N Nextera indexes (N701-N719) identify *loci*. The combination of the two series of indexes allows to unambiguously label a particular PCR product from a particular individual.

Supplementary Sequence List I

>Sotub08g017870.1.1_StuPPO1
MSSSSTTTLPLCTNKSLSSSFTTNNSSFLSKPSQLFLHGRRNQSFKVSCNVNNNVGEHEKN
LDAVDRRNVLLGLGGLYGAANLAPLASASPIPPPDLKSCGVAHVKEGVDVLYSCCPPVP
DDIDSVPYYKFPPMTKLRIRPPAHAADEEYVAKYQLATSRMRELDKDSFDPLGFKQQANI
HCAYCNGAYKVGGKELQVHFSWLFFPFHRWYLYFYERILGSLINDPTFALPYWNWDHP
KGMRIPPMFDREGSSLYDDKRNQNHRNGTIIDLGHFGKEVDTPQLQIMTNNLTLMYRQM
VTNAPCPSQFFGAAYPLGTKPSPGMGTIENIPHTPVHIWTGDTPRQKNGENMGNFYSAGL
DPIFYCHHANVDRMWDEWKLIGGKRRDLSNKDWLNSEFFFYDENRNPYRVKVRDSLDS
KKMGFSYAPMPTPWRNFKPIRKTTAGKVNTASIAPVTKVFPLAKLDRAISFSITRPASSRT
TQEKNEQEEILTFNKIAYDDTQYVRFDVFLNVDKTVNADELDKAEFAGSYTSLPHVHGN
NTNHVTSVIFKLAITELLEDNGLEDEDTIAVTLVPKVGGEGVSIESVEIKLEDC*

>Sotub08g019390.1.1_StuPPO2
MASLCNSSNTSLKTPFTSSSTSLSSTPKPSSTFHPWKTYQMFKVSCKVTNNNGDQNQNVE
TNSVDRRNVLLGLGGLYGVANAIPLAASASPTPPPDLSSCSKATINETTVVPYSCCAPKPD
DMEKVPYYKFPSMTKLRVRQPAHEANEEYIAKYNLAVSKMRDLDKTQPLNPIGFKQQA
NIHCAYCNGAYRIGGKELQVHNSWLFFPFHRWYLYFYERIVGKLIDDPTFALPYWNWDH
PKGMRFPAMYDREGTSLFDVTRDQSHRNGAVIDIGFFGNEVETTQLQLMSNNLTLMYRQ
MVTNAPCPRMFFGGPYDLGSNVELPGTIENIPHGPVHIWSGTVRGSTLPNGAISNGENMG
HFYSAGLDPVFFCHHSNVDRMWSEWKATGGKRTDITHKDWLNSEFFFYDENENPYRVK
VRDCLDTKKMGYDYKPMATPWRNFKPLTKASAGKKVNTSSIPPVSQVFPLAKLDKAISF
SINRPTSSRTQQEKNAQEEMLTFSSIRYDNRGYIRFDVFLNVDNNVNANELDKAEFAGSY
TSLPHVHRAGETNHIATVDFQLAITELLEDIGLEDEETIAVTLVPKRGGEGISIEGATISLAD
C*

>Sotub08g019410.1.1_StuPPO3
MASVCNSSSTTTTLKTPFISSNTSLSSTPKPSQLFLHGKRNQMFKVSCKVTNNNGDQNVE
TNSVDRRNVLLGLGGLYGVANAIPLAASAAPTPPPDLKTCGKATISDGPLVGYTCCPPPM
PTNFDNIPYYKFPSMTKLRIRSPAHAVDEEYIAKYNLAISRMKDLDKTEPLNPLGFKQQA
NIHCAYCNGAYVFGDKVLQVHNSWLFFPFHRWYLYFYERILGKLIDDPTFALPYWNWD
HPKGMRLPPMFDREGTSIYDERRNQQVRNGTVMDLGSFGDKVETTQLQLMSNNLTLMY
RQMVTNAPCPLLFFGAPYVLGNNVEAPGTIENIPHIPVHIWAGTVRGSTFPNGDTSYGED
MGNFYSAGLDSVFYCHHGNVDRMWNEWKAIGGKRRDLSEKDWLNSEFFFYDENKKPY
RVKVRDCLDAKKMGYDYAPMPTPWRNFKPKTKVSAGKVNTSSLPPVNEVFPLAKMDK
VISFSINRPASSRTQQEKNEQEEMLTFDNIKYDNRGYIRFDVFLNVDNNVNANELDKVEF
AGSYTSLPHVHRVGENDHTATVTFQLAITELLEDIGLEDEETIAVTLVPKKGGEGISIENVE
IKLLDC*

>Sotub08g019380.1.1_StuPPO4
MASLCNSSSTTLKTPFTSSSTSLSSTPKPSQLFLHGKRNKTFKVSCKVTNNNGDQNQNVE
TNSVDRRNVLLGLGGLYGVANAIPLAASASPTPPPDLSSCSKATINETTEVPYSCCAPTPK
DMTKVPYYKFPSMTKLRIRPPAHALDEAYIAKYNLAISRMKDLDKTQPTNPIGFKQQAN
IHCAYCNGAYVIDGKVLQVHNSWLFFPFHRWYLYFYERILGSLINDPTFALPYWNWDHP
KGMRFPPMFDIPGTALYDERRGEQIHNGTIIDLGSFGDQVQTTQLQLMTNNLTLMYRQL
VTNAPCPLMFFGGPYTLGSDVESPGTVEVIPHSPVHIWAGTRRGSILPDGKTSNGEDMGH
FYSAGLDPVFYCHHSNVDRMWKEWKAIGGKRTDIQNKDWLNSEFFFYDESGNPFKVKV

RDCLDTKKMGYDYEPMPTPWRNFKPKTKASAGKKVNTSSIPPVSQVFPLAKLDKAISFSI
NRPASSRTQQEKNAQEEMLTFNEIKYDNRDYIRFDVFLNVDNNVNANELDKAEFAGSYT
SLPHVHRVGDTNHTATATLQLAITELLEDIGLEDDETIAVTLVPKKGGISIGGVEIKLADC*

>Sotub08g019310.1.1_StuPPO5
MSKLRKRPVAQDVTKEYIAKYQLETKRMKELDKDDPRSFMQQANIHCAYCNGAYKFG
DEILQVHQSWLFFPFHRWYLFFYDRILGKLIDDPTFALPYWNWDHPKGMRLPPMFDRAN
TPLYDARREDMGNFYSAALDPVFYCHHANVDRMWKVWKGLCGKRRDIIDPDWLNSEF
FFYDENKNPYRVRVGDCLDTKKMGYDYAPAAIPWINCRPTRKGREGKVVVTNIKPANK
VFPIANLNKPISFSINRPTTSRSQKDKDEKEEVLIFKGLKYDTNKYIRFDVFLNEDEDMNT
DELDKVEFAGSYVNLPHVHAHNQKMDSDEMFQLGITELLEDIELEDDDTITVTVVPKTG
GDVISIQSVAIELLDG*

>Sotub08g017890.1.1_StuPPO6
MKRCKAKTLIPAMASSILPLCTTNIPSSFSNNTNSSFLSKPSQLFLHGRRSQSFKVSCNYSE
HDKNNLHDDAVDRRNVLLGLGGLYGAANLAPLATAAPSPPPDLKTCSTATVTPGGPAV
DYSCCPPPIPTDMSTIPYYKFPPMNKPRIRSPAHLADEEYIAKYNLAITRMKNLDKTEPLNP
LGFKQQANIHCAYCNSAYRIGGKELQVHNSWLFFPFHRWYIYFYERILGSLIDDPTFALPY
WNWDNPKGMRLPPMFDREGTSLYDERRNPQVRNRTVMDLGTFGKYKVDTTELLLINN
NLILMYRQMVTNAPCPLLFFGAPYVLGTNIEAPGTIENIPHITVHIWTGTVRDSTFPNGDTS
YGEDMGNFYSAGLDPVFYCHHGNVDRMWNEWKALGGRRTDLTQPDWLNSEFFFYDEN
KNPYLVKVQDCLDMKKMGYDYAPSPTPWLNFKPNKKYSTGILDTSLVPPARKIFPIFKLD
VNTSFSLDRPASSRTQVEKDNKEEILTFSFIKYDNREYRRFDVYLNADKNVNIEDLNQKE
YAGSYTSLPHTHKPGDDDHVSSATFQLAITEVLEENGLENEEKIMVTLVPRKGGEGLAIG
CVEIMLVDC*

>Sotub08g019300.1.1_StuPPO7
MASSCSINSMCLSLGEQSSKTLIITTPSSFFAKPSRRSQNFHVSCNNNANNGDEHDKNVAR
RNVLLGLGGLYGASNLAPLASASPIPAPNLKSCGKATKTGSTKEVGYSCCPLTPDDWNNI
PYYKFPPISKLRKSPVAQDVTEEYIAKYQLATQRMKDLDKKNPCSFMQQANIHCAYCNG
AYKFGDEVLQRILGKLIDDPTFALPYWNWDHPKGMRLPPMFHRETTSLYDARRNPHVR
NGTIIDFSSSTDEVYTDVKKTVTNNLTVMYHQMITNAACPLQFFGARYVLGNNNMNDR
GTIENSPHTPVHIWTGTVKGTVLGDGKPSFGEDMGNFYSAALDPVFYSHHANVDRMWNI
WKGLGWKKKDITDTDWLNSEFFFYDEHERPYRVRVGDCLDTKMGYDYAPADIPWINCR
PTRKGREGKVDLTNIEPANKVFPIANLNKPISFCINRPTTSRSQKDKDEKEEVLIFKGLKYD
TSIYISFDVFLNEDEDVNTNELDKVEFAGSYVNLPHVHAHNKRMDYGETFQLDIIELLEDI
GLEDDDTITVTVVPKKGGEVISIQSVAIEFLEG*

>Sotub08g019420.1.1_StuPPO8
MASNFLLTSCTTISSSPSKIFVRPKRIDNFKVNCEIKNSNTINNDDNEGKSFPGKLDRRNVL
LGLGGLYGASNLIGVTNEPFALGAPVPPPDFSTCSTASLPDGSQVPFSCCPPLPKDLTNIPT
YKLPNVSKVKIRPAAHNVTQEYITKYNTAIQKMKSLDKDDPLSFIQQANIHCAYCNSGYK
ELGFPGVPLQVHFSWLFFPFHRWYLYFFERILGSLIGDDTFALPFWNYDSQVGMQFPSLY
NDVNSSLYDPNRNQNHFPPNVIDLGFTTIDLDASDQQKINNNLTMMYRQMLTNAPCPQL
FFGNPIRGGEQPIRGMGTIENVPHNSVHRWVGNPNNKFRENMGTFYSAARDPIFYAHHA
NVDRMWTIWKTLGGNRRDFNDRDWLDSAFLFYDENRTLVKVTVQDCINNEKLGYKYE
NVPIPWKNYKPVPRKQKLKKNPKNVKPSTEIFPSTLKKTLSFSIKRPKISRTQQDKDIEEEL
LVFNNMTFDENEYIRFDVFINEDEGVKSKVLDRTEYVGSFANLPHVHAAGNNTGSSSSGT
PAVMSLAISEILEDLGLEDEEEIVVVVVPKFGGKEITIASVEIDTLACAN

>Sotub02g021500.1.1_StuPPO9
MFMNTSQTSKKMSSIPLPTTNTLSSSTTTTFSNLHSSPFFAKTSKISSIRKHNVHRNFQVSC
KTIDDNSHEHNNSPIDISKKNDSLNNMIDRRNVLLGLGGLYGASTLVGGLPFALAAPVNG
PDVTKCGAADLPPGAAPVNCCPPMSGNIIDFQLPSSTALRTRPAAHSVDSAYIEKFNRGIQ
LMKQLPDDDPRSFKQQANVHCAYCDGAYDQLGFPNSELQVHSSWIFLPFHRCYLYFFER
ILGSLINDPTFAMPFWNWDHPDGMRMPGLYTNSTSSLYDRLRDRRHQPPTMVDLDFNGT
DPNISNAQQTSQNLTNMYRQMVSLGRTPETFLGDPYRAGGVPGGAGSLENMAHGAVHV
WTGDRTQANFENMGDFYSAARDPIFYAHHSNIDRLWTVWKTLGGRRQDFTDPDFLNTS
FLFYDEKAQMVRIRVRDVLDSSKLGYVYQNVTNPWINSRPTPRVSRALSSVRRLVEAKA
ADDNSNVMNFPRPKEIFPTKLDHVIKVMVKRPNKKKRNKKEKNEKEEILIVEGLEVESDV
FVKFDVLINDEDETLVSPDNAEFVGSFVNVPHHSHGKGEKNSKRKTKLKLAITELLEDLD
AENDENVLVTFVPKNGSGAVKIGGVKIVLED*

>Solyc08g074680.3.1_PPO_A
MYSPMNLHPKLAIMASLCSNSSSTSLKTPFTSSTTCLSSTPKASQLFLHGKRNKTFKVSCK
VTNTNGNQDETNSVDRRNVLLGLGGLYGVANAIPLAASAAPTPPPDLSSCNKPKINATTE
VPYFCCAPKPDDMSKVPYYKFPSVTKLRIRPPAHALDEAYIAKYNLAISRMKDLDKTQPD
NPIGFKQQANIHCAYCNGGYSIDGKVLQVHNSWLFFPFHRWYLYFYERILGSLIDDPTFG
LPFWNWDHPKGMRFPPMFDVPGTALYDERRGDQIHNGNFIDLGSFGDQVETTQLQLMT
NNLTLMYRQLVTNSPCPLMFFGGPYTLGSTVEAAGTVENIPHSPVHIWVGTRRGSVLPDG
KISNGEDMGNFYSAGLDPLFYCHHSNVDRMWNEWKATGGKRTDIQNKDWLNSEFFFY
DENGNPFKVRVRDCLDTKKMGYDYQPTATPWRNFKPKTKASAGKVNTGSIPPESQVFPL
AKLDKAISFSINRPASSRTQQEKNAQEEVLTFNAIKYDNRDYIRFDVFLNVDNNVNANEL
DKAEFAGSYTSLPHVHRVGDPKHTATATLRLAITELLEDIGLEDEDTIAVTLVPKKGDISI
GGVEIKLADC

>Solyc08g074683.1.1_PPO_B
MASVVCNSSSSTTTTTLKTPFTSLGSTPKPSQLFLHGKRNKTFKVSCKVINNNGNQDETNS
VDRRNVLLGLGGLYGVANAIPLAASATPIPSPDLKTCGRATISDGPLVPYSCCPPPMPTNF
DTIPYYKFPSMTKLRIRTPAHAVDEEYIAKYNLAISRMRDLDKTEPLNPLGFKQQANIHCA
YCNGAYIIGGKELQVHNSWLFFPFHRWYLYFYERILGKLIDDPTFALPYWNWDHPKGMR
LPPMFDREGSSLYDERRNQQVRNGTVLDLGSFGDKVETTQLQLMSNNLTLMYRQMVTN
APCPLLFFGAPYVLGNNVEAPGTIETIPHIPVHIWAGTVRGSKFPNGDVSYGEDMGNFYS
AGLDPVFYCHHGNVDRMWNEWKAIGGKRRDISEKDWLNSEFFFYDEHKNPYRVKVRD
CLDTKKMGYDYAPMPTPWRNFKPKSKASVGKVNTSTLPPANEVFPLAKMDKTISFAINR
PASSRTQQEKNEQEEMLTFNNIRYDNRGYIRFDVFLNVDNNVNANELDKAEFAGSYTSL
PHVHRAGENDHIAKVNFQLAITELLEDIGLEDEDTIAVTLVPKKGGEGISIENVEIKLVDC

>Solyc08g074682.1.1_PPO_D
MASLCSNSSTTSLKTPFTSLGSTPKPCQLFLHGKRNKAFKVSCKVTNTNGNQDETNSVDR
RNVLLGLGGLYGVANAIPLAASAAPTPPPDLSSCSIARIDENQVVSYSCCAPKPDDMEKV
PYYKFPSMTKLRVRQPAHEADEEYIAKYNLAVSKMRDLDKTQPLNPIGFKQQANIHCAY
CNGAYRIGGKELQVHNSWLFFPFHRWYLYFYERIVGKLIDDPTFALPYWNWDHPKGMR
FPAMYDREGTFLFDETRDQSHRNGTVIDLGFFGNEVETTQLQMMSNNLTLMYRQMVTN
APCPRMFFGGPYDLGTNVELPGTIENIPHGPVHIWSGTVRGSTLPNGAISNGENMGHFYS
AGLDPVFFCHHSNVDRMWSEWKATGGKRTDITHKDWLNSEFFFYDENENPYRVKVRDC
LDTKKMGYDYKPMATPWRNFKPLTKASAGKVNTSSIPPVSQAFPLAKLDKAVSFSINRPT
SSRTPQEKNAQEEMLTFNSIRYDNRGYIRFDVFLNVDNNVNANELDKAEFAGSYTSLPHV
HRAGETNHIATVDFQLAITELLEDIGLEDEDTIAVTLVPKRGGEGISIENATISLADC

>Solyc08g074620.3.1_PPO_E
MSSSSSITTTLPLCTNKSLSSSFTTTNSSLLSKPSQLFLHGRRNQSFKVSCNANNVDKNPDA
VDRRNVLLGLGGLYGAANLAPLATAAPIPPPDLKSCGTAHVKEGVDVIYSCCPPVPDDID
SVPYYKFPSMTKLRIRPPAHAADEEYVAKYQLATSRMRELDKDPFDPLGFKQQANIHCA
YCNGAYKVGGKELQVHFSWLFFPFHRWYLYFYERILGSLINDPTFALPYWNWDHPKGM
RIPPMFDREGSSLYDEKRNQNHRNGTIIDLGHFGKEVDTPQLQIMTNNLTLMYRQMVTN
APCPSQFFGAAYPLGSEPSPGQGTIENIPHTPVHIWTGDKPRQKNGEDMGNFYSAGLDPIF
YCHHANVDRMWNEWKLIGGKRRDLTDKDWLNSEFFFYDENRNPYRVKVRDCLDSKK
MGFDYAPMPTPWRNFKPIRKSSSGKVNTASIAPVSKVFPLAKLDRAISFSITRPASSRTTQE
KNEQEEILTFNKISYDDRNYVRFDVFLNVDKTVNADELDKAEFAGSYTSLPHVHGSNTN
HVTSVTFKLAITELLEDIGLEDEDTIAVTLVPKAGGEEVSIESVEIKLEDC

>Solyc08g074630.2.1_PPO_F
MSSSTPNTLPLLSTNKSLSSPFTNNHSTFLSKPSQPFLHGRRCQSFKVSCNVGEHDKNLDA
VDRRNVLLGLGGFYGAANLAPLASAAPIPPPDLKSCGVAHIDDKGTEVSYSCCPPVPDDI
DSVPYYKFPPMTKLRIRPPAHAADEEYVAKYQLATSRMRELDKDPFDPLGFKQQANIHC
AYCNGAYKIGGKELQVHFSWLFFPFHRWYLYFYERILGSLINDPTFALPYWNWDHPKGM
RIPPMFDREGSSLYDEKRNQNHRNGTIIDLGHFGKDVETPQLQIMTNNLTLMYRQMVTN
APCPSQFFGAAYPLGSDPEPGMGTIENIPHTPVHIWTGDSPRQGHGEDMGNFYSAGLDPL
FYCHHANVDRMWNEWKLIGGKRRDLSNKDWLNSEFFFYDENRNPYRVKVRDCLDSKK
MGFDYAPMPTPWRNFKPIRRSSSGKVNTASIAPVSKVFPLAKLDRAISFSITRPASSRTTQE
KNEQEEILTFNKMAYDDTKYVRFDVFLNVDKTVNAEELDKAEFAGSYTSLPHVHGNND
NHVKDVTFTLAITELLEDIGLEDEDTIAVTLVPKVGGEGVSIESVEIKLEDC

>Solyc02g078650.2.1_PPO_G
MFMNTPQTSKVMSSILFPTTNTLSSSTTNTFSNLHSSPFFAKTSKISSIRKHNVHRNFQVSC
KTIDDNNHEHNSPVDISKKNDSSNNMIDRRNVLLGLGGLYGASTLVGGLPFVLAAPVNG
PDVTKCGAADLPPGAELVNCCPPMSGSIIDFQLPSSSTPLRTRPAAHSVDSAYIEKFNRAIQ
LMKQLPDDDPRSFKQQANIHCAYCDGAYDQLGFPNSELQVHSSWIFLPFHRCYLYFFERI
LGSLINDPTFAMPFWNWDHPDGMRMPGLYTNPTSSLYDRLRDRRHQPPTMVDLDFNGT
DPNISTAQQTSQNLTNMYRQMVSLGRTPETFLGDPYRAGGVPGGAGSLENMAHGAVHV
WTGDRTQANFENMGDFYSAARDPIFYAHHSNIDRLWTVWKTLGGRRQDFTDPDFLNTS
FLFYDEKAQMVRIRVRDVLDSSKLGYVYQNVRNQWINSRPTPRVSRALSSVRRLVEARA
ADDNNNNIMNFPRPKEIFPTKLDHVIKVMVKRPNKKKRNKKEKNEREEILIVEGLEVESD
VFVKFDVLINDEDETLISPDNAEFAGSFVNVPHHSHGKGEKNSKRKTKLKLAITELLEDL
DAENDDNVLVTFVPKNGSGAVKIGGVKIVLED*

>SMEL_008g312510.1.01 PPO1
MASVCNTSTATLKSSFIPSPNSLGSTPKPSQLFLHGKRNQAFKVSCKVTNNNGDQNQNVV
DTNSVDRRNVLLGLGGLYGVANAIPLAASATPIPAPNAPSCGTATISDGPEVPYTCCPPG
MPEDIEKIPYYKFPSATKLRIRQPAHAVDEELIAKYNLAISKMRELDTTDHFSPLAFKQQA
NIHCAYCNGAYKIGGKELQVHNSWLFFPFHRWYLYFYERILGKLIDDPTFALPYWNWDH
PKGMRLPPIFDRQGTALYDERRSTQVRNGTVMDLGSFGDKVQTTQLQLMSNNLTLMYR
QMVTNAPCPLLFFGAPYVLGNNVEAPGTVEVIPHIPVHIWVGTARGSKFPDGSTSYGEDM
GNFYSAGLDPVFYCHHSNVDRMWNEWKQIGGKRRDISQRDWLNSEFFFFDENKNPYRV
RVRDCLDTKTMGYDYAPMPTPWRNFKPKTKASSGKANTSAFPPASQVFPLAKMDKVITF
SIKRPASSRTQQEKNEKEEMLTFNNIKYDNREYVRFDVFLNVDNNVNANELDKAEFAGS
YTSLPHVHRASQTDHVATATLQLAITELLEDIGLEDEDTIAVTLVPKKGGEGISIEGVEISL
ADC*

>SMEL_008g312500.1.01 PPO2
MASLCNSSNTTLKTPFTSLGSTPKPSQLFLHGKRKQTFKVSCKVSNNNGDQNQNEVEKN
SVDRRNVLLGLGGMYGAANFAPLAASAAPTPPPDLSSCSIAKITETEEVSYSCCAPTPDDL
NKIPYYKFPSMTKLRIRQPAHAADEEYIAKYNLAISRMKHLDTTEPLNPIGFKQQANIHCA
YCNGAYKIGDKVLQVHNSWLFFPFHRWYLYFYERILGSIIDDPTFALPYWNWDHPKGMR
MPAMFDREGTALYDQVRNQSHRNGRVMDLGSFGDEVQTTELQLMSNNLTLMYRQMVT
NAPCPRMFFGAPYVLGNNVEAPGTIEVIPHGPVHVWTGTVPGTTLPNGRTSHGENMGHF
YSAGLDPVFFCHHSNVDRMWSEWKAIGGKRRDISHKDWLNSEFFFYDENGDPFRVKVR
DCLDTKKMGYDYAPMPTPWRNFKPITKASVGKVDTSSLPPVSQVFPLAKLDKAISFSINR
PASSRTQQEKNEQEEMLTFNNIKYDNRNYVRFDVFLNVDSNVNADELDKAEFAGSYTNL
PHVHRVGENTDHVATATLQLAITELLEDIGLEDEDTIAVTLVPKKGGEGISIEGVEISLADC
*

>SMEL_008g312430.1.01 PPO3
MASSFLPLCIHPSFSNTSESSFLPKPSQLFLQRRHNQRFKVSCNANKHEKDNLDVVDRRN
VLLGLGGLGAANLAPLTANAAPSPPPDFKTCGIATITADGPPVPYTCCPPPMPSNVNTIPY
YKLPSMTKVRIRQPAHTVDEEFIAKYNLAISRMKELDEKEPLNPLGFKQQANIHCAYCNG
AYKIGEKVLQVHQSWLFFPFHRWYLYFYERMLGKLIDDPTFALPYWNWDLPKGMRLPP
MFDREGSPLYDERRNPQVRNGTVMDLGSFGDQVQTTELQLMSNNLTLMYRHMVTNAS
CPLLFFGGRYVLGSTQGVQGTIEKIPHTPVHIWVGTKKDSILPNGKKSYGEDMGNFYSAA
LDPVFYCHHSNVDRMWNEWKQIGGKRRDLSQKDWLDSEFFFYDENKNPYLVKVRDCL
DTKKMGYDYAPSSTVWRNFKPNKKNTDGKVNTGSLPSATKIFPIFKLDKAISFSINRPASS
RTQQEKNEQEELLTFSYIKYDNREYIRFDVFVNVDKNVKADELDKIEYAGSYTSLPHVHK
DGDKDHIATATLQLALTELLEDIGLENEETIAVTLVPKKGGEGLSIGCVEIKLEDC*

>SMEL_008g312420.1.01 PPO4
MSSSSSTLPLCNSKSLFFSFCNSPFLPQPSKLFLQRTRSQRFKVSCNANNVGEHDKNLDAV
DRRNVLLGLGGLYGAANLAPLAANAAPIPPPELKTCGRAVVNDTTGELVKYSCCPPIPDD
IDSVPYYKFPSMTKLRIRPPAHAVDEEYIAKYQLATSQMRELDKDPFGPIGFKQQANIHCA
YCNGAYKAGGKELQVHFSWLFFPFHRWYLYFYERILGSLINDPTFALPYWNWDHPKGM
RLPPMFDVEGSSLYDAKRNQSHRNGKIIDLGFFGQETETTELQTMTNNLTLMYRQMVTN
APCPLLFFGNPYPLGTDPKPGMGTIENIPHTPVHIWTGDSPRQPNGEDMGNFYSAGLDPV
FYCHHANVDRMWNEWKAIGGKRRDLADKDWLNSEFFFYDENRNPFKVKVRDCLDSKK
MGFDYAPMPTPWRNFKPVRRTTSGKANTRSIPPASKVFPTCETRQSDFIFHRQTSFVKDS
KAEKNEQEGDTNIRQIQYDDSQYVRFDVFLNVDKTVKALELDQPEFAGSYTSLPHVHGD
KDRAPVTFKLAITELLEDNNLEDEESIVITLIPKAGGDGISIQNAVIDLVDC*

>SMEL_008g311990.1.01 PPO5
MSSSSSTTTLPLCTNKSLSSFTNSSFLAKPSQLFLHRSRSQSFKVSCNANNVGEHDKNLDA
VDRRNVLLGLGGLYGAANLAPLAANAAPIPPPDLKSCSKAHINPDKEVTYSCCPPIPQDID
SVPYYKFPPMTKLRIRPPAHAVDEEYIAKYQLATSRMRELDKDPFDPLGFKQQANIHCAY
CNGAYKVGGKELQVHFSWLFFPFHRWYLYFYERILGSLINDPTFALPYWNWDHPKGMRI
PPMFDREGSSLYDEKRNQNHRNGKIIDLGFFGTETQTTELQTMTNNLTYMYRQMVTNAP
CPLLFFGNPYPLGTDPSPGMGTIENIPHNPVHIWTGDSPRQPNGEDMGNFYSAGLDPVFY
CHHANVDRMWNEWKAIGGKRRDLADKDWLNSEFFFYDENRNPFKVKVRDCLDSKKM
GFDYAPMPTPWRNFKPIRKTTSGKANIGSIPPASKVFPIAKLDRAISFSINRPASSRTQAEK
NEQEEILTFNKVKYDDSQYVRFDVFLNVDKTVNADELDKAEFAGSYTSLPHVHGDNNT
HVTSVTFNLAITELLEDIGLEDEDTIALTLVPKQGGEGISIDNAEIVLVDC*

>SMEL_008g312010.1.01 PPO6
MSSSSATLPLCTNKSLSSFTNSSFLAKPSQLFLHRSRSQSFKVSCNANNVGEHDKNLDAID
RRNVLLGLGGLYGAANLAPLAASAAPIPSPDPKSCSKAHIKPNKEVPYSCCPPPPQDIDSV
PYYKFPPMTKLRIRPPAHAVDEEYIAKYQLATSRMRELDKDPFDPLGFKQQANIHCSYCC
GAYKVGGKVLQVHSSWLFFPFHRWFLYFYERILGSLINDPTFALPYWNWDHPKGMRIPP
MFDHEGSSLYDEKRNQNHRNGKIINLGFSCKETQTTELQTMTNNLTLMYRQMVTNAPCP
LLFFGNPYPLGTDPKPGMGTIENIPHNAVHNWTGDQPRQPNGEHMGTFYSAGLDPVFYS
HHANVDRMWNEWKAIGGKRRDLADKDWLNSEFFFXFYSAGLDPVFYCHHANVDRMW
NEWKAIGGKRRDLADKDWLNSEFFFYDENRNPFKVKVRDCLDSKKMGFDYAPMPTPW
RNFKPIRKTTSGKANIGSIPPASKVFPIAKLDRAISFSINRSASSRTQAEKNEQEEILTFNKV
QYDDSQCVRFDVFLNVDKTVNADELDKPEFAGSYTSLPHVHGDNNTHVTSVTFKLVITE
LLEDIGLEDEDTIAVTLVPKEGGEGISIENAEIVLMDC*

>SMEL_008g312490.1.01 PPO7
MASVCNSSTTTLKSPLTSSPSLRSTPKPSQLFLHGKRNQTFKVSCKVTNSNGDQNQNVVE
TNSVDRRNVLLGLGGMYGVANAIPLAASATPVPPPDLTSCRKAKITETEEVAYLCCAPKP
DDMKKVPYYKFPSATKLRLRQPAHAADEEFIAKYNLAISKMKDLDKTEPTNPIGFKQQA
NIHCAYCNGAYAIDDKVLQVHNSWLFFPFHRWYLYFYERILGKLVDDPTFALPYWNWD
HPKGMRLPSMFDRQGTDLYNERRNPQVRNGTVLDLGSFGDQIQTTQLQLMSNNLTLMY
RQLVTNAPCPLMFFGAPYVLGNNVEAPGTVEVIPHGPVHVWTGTRAGSILPDGSRSHGE
DMGNFYSAGLDPVFYCHHSNVDRMWNEWKAIGGKRRDISHKDWLNSEFFFYDENGNPF
RVKVRDCLDTKKMGYDYAPMPTPWRNFKPKTKASSGKVNTSSLPPASKVFPLAKLDKAI
SFSINRPASSRTQQEKNEQEEMLTFSEIKYDNREYIRFDVFLNADKNLNADELDKAEFAGS
YTSLPHVHRAGDTNHIATATLGLAVTELLEDIGLEDENTIAVTLVPKKGGEGISIGGVQIT
LADC*

>SMEL_008g312460.1.01 PPO8
MTSFSSKILTTPSPFFTKPCQRSQSFNVSCEKKFEANFDRRNVLLGLGGMYGASNLEPLAA
CASPIPPPDLKSCNRATISEGPEVPYSCCPRKPEDLDNIPYYKLPSMSKLRKRPAAQDVNE
EYTAKYQLATKKMKELDDDDPLGFKQQANIHCAYCNNAYKIDGKVLQVHQSWLFFPFH
RWYLYFYERILGSLINDPTFALPYWNWDHPKGMCLPPMFDVEGSSLYDERRNPHVRNG
AIIDLGSFGDEVKTAELQMRTNNLTLMYRQMITNAPCPSQFFGARYVLGTDPKGQGTIEN
IPHTPVHIWTGTVRGPNNLGNGATSYGEDMGNFYSAGLDPVFFCHHANVDRMWNIWKQ
LGGKRRDLMDNDWLNSEFFFYDENRNPYRVRVRDCLDSKKMGYVYAPMPTPWRNFKP
TKKSQVGKVSSNSINPASKVFPLAKLDRAISFSINRQRDKNEQEEVLTFKGIKYDNSKYIR
FDVFVNAEESVNADELDKIEFVGSYVSMAPLVHGHKMESDESFQVAIREVLEDNGLADD
DTITVTLVPKKGGQLLSIQSVDIEFVTG*

>SMEL_008g312520.1.01 PPO9
MASNFLITSCPIISSSPSKFSQSPSKIFVSAKCNGNFKVNCSNNDNEGKSFSEKLDRRNVLV
GLGGLYGASNFIGVANEPFALGAPVPPPDLSTCNTASLPDGSPVPFTCCPPLPKDLSNIPTY
KLPDVSSKLKIRPAAHNVTQEYITKYSIAIQKMKSLDKDDPLNFMQQANIHCAYCNSGYK
ELGFPGVPLQVHFSWLFFPFHRWYLYFFERILGSLIGDDTFALPFWNYDSKVGMQLPSLY
NDVNSPLYDPNRNQNHLPPNVVDIGFTTIDLDVSDQQKINNNLTMMYRQMITNAPCPQL
FFGNPIRGGEQPIRGMGTIENIPHNAVHRWVGNPNNKFRENMGTFYSAARDPIFYAHHAN
IDRMWTIWKTLGAKRRDFTDRDWLDSAFLFYDETRTLVKVTVQDCINNEKLGYKYENV
PIPWKNYRPVPRKQKVKKNANKVKPSTEIFPSTLKKTLSFSIKRPNISRTQQDKDTQEELL
VFNNMTFDENEYIRFDVFINEDEGVKAKVLDRIEYVGSFANLPHVHNGGQNTSSATPAIM
SLAITEILEDLGLEDEEEIVVVVVPHSGGKEITIGSVEINTLACAN*

>SMEL_000g064350.1.01 PPO10
MSSVPLHTTNTISSSTTTTFSNLHPSPFFGKTPKISSIKKHNGHRNFKVSCKSTDENNNHEH
NSPIDISKKSTNSSSNIIDRRNVLLGLGGLYGASTLVGGLPFAFAAPVNGPDVNKCGAADL
PPGAAPVNCCPPMTGNIIDFQLPSSTTIRTRPAAHSVDSAYIEKFNRGIQLMKELPDNDPRS
FKQQANVHCAYCDGAYDQLGFPNTELQVHSSWIFLPFHRCYLYFFERILGSLINDPTFAM
PFWNWDHPDGMRLPGMYANSTSALYDRLRDRRHQPPIMVDLDFNGTDPNISDSQQTSQ
NLTNMYRQMISLGRTPETFLGDPYRAGGLPGGGGSLENMAHGSVHVWTGDRTQPNFEN
MGDFYSAARDPIFYAHHSNIDRLWSVWKTLGGRRQDFTDPDFLNTSFLFYDEKAQMVRI
RVRDVLDSSKMGYVYQNVTNPWINSRPTPRVSRALSSIRRLVEARAADDNVMTFPRPKD
VFPTKLDHVIKVMVKRPNKNKRNKKEKNEKEEILVVEGLEVETDVFVKFDVLINDEDET
LTSPGNAEFAGSFVNVPHHRHGKGDKNSKRKTKLKLAITELLEDLDAENDENVLVTFVP
KNGSGAVKIGNVKIVLED*

Supplementary Sequence List II

>PPO1-3 off-target 1

TAAGCATTCTTAAAAAATATAAAAAACACATTTTTAAAGATGAATATATTTCTCATTT
TATTAACTCCCATTATTTAGAGAATTTTAAATATCCATTTATATTTTCATGAAAAGTA
GTATCACGTTAATTTTTTTTTACCTTTCTTCCGCTTCAATTGAGCAAGCAGAATAATAC
TCATAATGATTATTCAACATTTTGTGTAAGATAAAATAAGTTTATTCAAACATGCGCA
ATAGTGTTTCAACATGAAAAGTAACAGTTTTTTGAAAATTTTAACAATCCGCAAGAG
TTATCTGGACACCTCCAATGGAGATACCTTCTCCACCTTTCTTTGGTACCAGAGTCAC
CGCAATAGTGTTTTCATCTTCCAATCCAATGTCCTCCAACAGTTCAGTTACCGCGAGC
CCCAAAGTAGCAGTGGCGATATGGTTGGTATCACCAGCTCTGTGAACGTGTGGCAAG
CTTGTATAGCTCCCCGCAAACTCTGCCTTGTCAAGCTCATCCGCATTCAAATTCTTAT
CCGCATTCAGGAACACATCGAACCTTATATACTCTCTGTTATCATATTTTATTTCGCTG
AACGTTAGCATCTCCTCTTGTTCATTTTTCTCCTGTTGAGTCCTTGATGAAGCCGGCCT
ATTGATGGAAAACGAAATGGCTTTGTCCAGTTTAGCCAGTGGGAACACCTTGCTAGC
TGGCGGAAGTGAACTTGTATTCACTTTCCCTGATGATGCCTTTGTTTTTGGCTTGAAG
TTGCGCCACGGGGTTGGCATTGGGGCGTAATCATACCCCATCTTCTTTGTATCCAAAC
AGTCTCGGACTTTCACACGGAATGGGTTTCCATTTTCATCGTAGAAAAAGAACTCCG
AGTTCAACCAATCTTTATGTGAGATATCCCTTCTTTTCCCTCCTATTGCCTTCCATTCG
TTCCACATCCGGTCCACATTGCTGTGATGGCAGTAGAAAACCGGGTCCAAACCAGCT
GAGTAGAAATTACCCATGTCCTCGCCGTGTGACCTTGAACCATCAGGCAAAATCGAA
CCTGCCCTTGTACCAGTCCAAACGTGGACAGGACCGTGAGGGATGACTTCAACGGTT
CCCGGCGCTTCCACGTTATTCCCCAGAACGTAAGGCGCACCGAAGAACATAAGGGGA
CATGGTGCATTAGTTACCAATTGACGGTACATTAGAGTTAAGTTATTACTCATCAGCT
GTAATTGAGTTGTTTGGATTTGGTCACCGAAAGAACCAAGATCCAGAACGGTCCCAT
TACGGACTTGTGGATTACGCCTTTCGTTGTAAAGGTCAGTCCCTTGACGATCGAACAT
AGAAGGCAAACGCATGCCCTTAGGATGGTCCCAGTTCCAGTATGGCAAAGCGAAAGT
TGGATCATCGACCAGTTTTCCCAAGATTCTCTCATAGAAGTACAAGTACCATCTATGG
AACGGGAAGAAAAGCCATGAGTTATGAACTTGTAACACTTTGTCATCAATTGCGTAA
GCTCCGTTACAATAAGCACAATGTATATTGGCTTGTTGCTTGAAGCCAATAGGGTTTG
TAGGTTCTGTCTTATCAAGATCTTTCATTTTGCTAATAGCCAAATTGTACTTGGCGAT
AAACTCCTCATCAGCAGCATGAGCAGGCTGACGAAGGCGGAGCTTAGTCGCAGAAG
GGAACTTGTAATATGGAACTTTTTTCATATCGTCAGGCTTGGGAGCGCAACACAAGT
ATGCCACCTCCTCCGTTTCAGTAATCTTGGCTTTTCTACAAGATGTGAGATCAGGGGG
TGGTACAGGAGTAGCCGATGCCGCTAATGGTATAGCATTAGCAACACCATACATCCC
TCCTAAACCAAGAAGCACATTTCTCCTATCAACAGAATTTGTTTCAACTACATTTTGG
TTTTGGTCACCATTACTATTGGTAACCTTGCATGAGACTTTGAAAGTTTGGTTACGTTT
TCCATGGAGGAATAGTTGAGAGGGCTTTGGAGTGGAACGTAAAGAAGGGGAGGAAG
TAAGAGGAGATTTGAGAGTTGTAGTACTACTATTGCACACACTTGCCATTGCTAGCTT
CTGGATGAGAGATCCATAGGGCTATGCCTTGCCTTATATAATATGTTAGCACCAAAT
AAAATATTTCACCATTTGAAAATATTCCAGACGCTGTAGGTAAGGAAATTTTTAATA
GTATATGTCATTAAATTTTTATTCTTAGAGATCTTACTAAATAGAGAGCGAGATAATT
TTAGTCAAGATAATAGTCTATGTATAACTAATATGTACATATTCGTGTATAACCAATG
TCAAACTATGAGCTATTAATGTAGGATATACCATATGAAATAGTTTTAATGTAAATA
ATCTCAATATTACCAATACACCATTTTCAATGTTGTTAGAACCCTTACTACTCACCTA
ACACCTCTTTTCCTCTCAAAAGTTGGAGACAAAACGTGA

>PPO1-3 off-target 2

CCTTTAGAGTTGATCTTAGCATTTTTCCAGGTTGCAGTTATATAGATATTTTTTATAA
CATAATTGTAACAGATTTATTTTTGAATAACAACTAAAGGAAATCTAGCGTCAAGCT
GATTTCGACATGAAAAGTAACATGGAAAAGCAAAACAGAGTATGGAAAGAGAACAA
TCCATCAAAGTTATCTCAGTAATTCTCCCCTCAAATTAACCAGGACTTAACAATCCGC
AAGACTGATCTCTACACCTTCAATGGAGATACCTTCACCGCCTTTCTTTGGTACCAGA
GTCACCGCAATAGTGTCTTCATCTTCCAGACCAATGTCCTCCAACAGTTCAGTTATCG
CCAGCTGCAAAGTAGCAGTCGCGACATGGTCAGTATTCTCGCCAACTCTGTGAACAT
GTGGCAGATTAGTATAGCTCCCAGCGAACTCTGCCTTGTCAAGCTCGTCAGCGTTCAC
ATTACTGTCCACATTCAGGAACACGTCGAACCTTACATAGTTTCTATTATCATATTTT
ATGTTGTTGAACGTTAGCATCTCCTCTTGTTCATTTTTCTCCTGTTGAGTCCTTGATGA
AGCCGGCCTATTGATGGAAAACGAAATGGCTTTGTCCAGCTTCGCGAGTGGGAATAC
CTGGCTTACCGGGGGAAGTGAACTTGTATCAACTTTCCCTACTGATGCCTTTGTTATT
GGCTTGAAGTTGCGCCACGGTGTTGGCATGGGTGCGTAATCGTACCCCATCTTCTTCG
TGTCCAAACAATCTCGGACTTTCACACGGAATGGGTCTCCATTTTCATCGTAGAAAAA
GAACTCCGAGTTCAACCAATCTTTATGTGAGATATCCCTTCTTTTCCCTCCTATCGCTT
TCCATTCGCTCCACATCCGGTCCACGTTACTGTGGTGGCAAAAGAAAACCGGGTCCA
AACCAGCTGAGTAGAAATGACCCATGTTCTCACCGTGTGACGTTCTACCATTAGGCA
AGGTTGTACCTGGCACTGTACCAGTCCAAACGTGGACAGGACCGTGAGGGATGACTT
CAATGGTTCCTGGGGCTTCAACGTTATTCCCAAGAACGTAAGGCGCGCCAAAGAACA
TCCGAGGACATGGGGCATTAGTCACCATTTGACGGTACATTAGAGTTAAGTTATTACT
CATCAACTGTAGTTCAGTTGTTTGGACTTCATCACCAAAAGAACCAAGATCCATTACT
CTTCCATTGCGGTGACTTTGGTTACGTACTTGGTCGTAAAGGGCAGTCCCTTCACGAT
CGAACATGGCAGGCATACGCATGCCCTTTGGATGGTCCCAATTCCAATATGGCAAAG
CAAAAGTTGGATCATCAATGATGGATCCAAGGATTCTCTCGTAGAAGTACAAGTACC
ATCTATGGAAAGGGAAGAAAAGCCATGAATTATGAACTTGTAACACTTTGTCACCAA
TTTTGTAAGCACCGTTGCAATAAGCACAATGGATATTAGCTTGTTGCTTAAACCCAAT
AGGGTTTAAAGGTTCTGTTGTATCAAGATGTTTCATTCGGCTGATAGCCAAATTGTAC
TTTGCAATATACTCCTCATCAGCAGCATGAGCAGGCTGACGAATACGGAGCTTAGTC
ATAGAAGGGAACTTGTAATATGGAATTTTGTTCAAATCATCAGGAGTGGGAGCGCAA
CAACTGTATGACACCTCCTCAGTTTCGGTAATCTTGGCTATACTACAAGACGAGAGAT
CAGGGGGTGGTGTAGGAGCAGCTGATGCAGCTAATGGTGCAAAATTAGCAGCACCA
TACATCCCTCCTAAACCAAGAAGAACATTTCTTCGATCAACAGAATTTTTTTCAACTT
CGTTTTGGTTTTGGTCACCGTTATTATTGGAAACTTTGCATGAAACTTTGAACGTCTGT
TTACGTTTTCCATGGAGGAAAAGTTGAGAGGGCTTAGGAGTGGAACCTAAAGAAGTG
AAGGGAGTTTTGAGGGTTGTATTACTACTATTGCACAAACTTGCCATTGCTAGCTTTG
TATGAAGATTCATAGGCAACATCTTGCCTTTTATAATAAGTATAAATTAACACAATAA
ATTATTAATCCATTTGAATCAATATTAGAGGCGGCTAATACATCAATAATTTGGTGAC
GCGGTAGGTGGAATAGCGTTCATACATCATTAAGTTTTTATTCGTAGGAGCTTTGTAG
TAAACATAATTTGGCGTACATCTTTTAATCTTCGATAAGGAGTCATATTTATCCCTAT
ACTCTAACGAAAAGGTTATATTTACCCTTAGTTATAGTTTTTCAACATATTTGCTCTTG
TCGTCTAACTTTGAGCTCGTATATACCCTTCTCCGTTATTTACCATGGCATTTCTTTAC
ATCCACCCACCCACAAAAAATCAAAGTAAACAAACAGGTTTTCTTTAATTTATCAAA
ATCTATCTTCTCTGGAGTATTTCCAACCCTTTTTAA

>PPO1-3 off-target 3

TGAATCAATTTGTAAGAACTGCCATTGCTCTTCTTATTAACTTTATTGAGGCAATTAA
TGATATTGAATTCCAGTTTGCGAAGTGCCTTGAATCTTCATCCACTATAAGGTTAAAA
GAAAAAAAACTACCTCTGTTCCACTTAGTTTGATACTTTCATTATTAATCTGTTCCAA

AAACAATTAATAAGATTGAAGAATAAGATTGAAGAACATGAGCTGTAAATCTATCTG
TTGAGATGTATTTCACTGTCATATATATATATATTTTTTTGGGGTAAAATTCACTGTCATA
TTTGGATGTGTAGTATGAGTGTGCTCTGTATAATAGGATATGAAGTATCACAGCTTAA
TTACAAGAGTCTGTTGAATGAAATTTCCAAACACATGGAAGCCTTCAAATTTGCCATT
CATATTCTGAACTCTGTTTCAATTGCCACTTTCTGCAGTGTCTTTGAGAAATCTCTTC
GAGCTCGGTTGTTGAAGGTCTGCCCAACAGTTGTCTTCCACTCAGTTAATTTTCCTAG
GTATTTACTGCCAGAGAAGAAAAACGCAAAACTTACATTCTTTGTTTCTCTTACAGGC
AAATCCTGACACTGATTATCAATGTCTAGAAACATTTGGTTCTATGTTCTCTCAAGAT
ATTCCTATACATGCAGTAAGCAAGCTATAACCTTACCACCGTCAATTTATTCTAGGAA
GTCTCCTCATCTTTTAGACATGACTGGAATTTTCAGGGTACAACAATCAGTTTTCGAC
GAACAACTGATGGACATCTAATTACTGAGAGTAAGTTCACTGTGTCTAAGAAGAAAC
ACTCTTCTAATTTCTTACTCTTTTAGTAAGGGGATTACGCAAAACATCTCACTTTTGG
ACTACTATATTTGCAGTTGGTGGCAACCATATTGGAGCAGTTCAGAGCAGAGAATTG
TGTAGTAAGTGAACTGTCCTTCAATTTATTTCAAGGCGCTTCTGGATATACTGTGCAA
AGTGGCATGATAATATGATATTGATTCTTGGGAAGCTAAGTAGGAGAAAATGTATGG
GAAAAAGACAGATACTTAACAGTTATTATTCTTGCTTTTCCAGGGGCAATTTTTGATA
TGTACATTGGCGATGGTCCTATATGTGAGGAAACAAAAGAAGAGATTGGCAAAAAT
GTTGCAAGTATGATTAGGGGGTGTTAAAAAGATCGGTTTTATCTTTGTGCTTCCACAA
GAGCTTATTTGCTGGTAATGGCTCTGGGTTCTTTTGGAACTTGACTTCATAGTGTCCC
AATTACTTGGACATAAATCTCAAAATGAACATAGGTTTCACAAGGATAAATGATCTT
TAGGCTTATATATGTAGAATCCATTTATTTTTGACAGGGCTCTTATATTCAAAATTGA
TTAGAATACAAAATTTCTTGTGACATTGTTGGCAAGCTGCCATTACAGAGCTGCCCCT
CTGCCAACTGTTCTAAATGAAGAGAGCTTTTTCTGCATTTTAGGCTTAAGCTTCTCTC
CTATTCTGTAAAATATTGTTTCTTCTTGTCAAGTAATTTGTAGTATTATTTTGTATCTT
ATGCAACATTTTGCTATTTCAAGTTGTCAACTTTCATATATTACCCCGGTTAACATAG
TAGTTATGAGAAGATGTTTCACAAGATTGGGATGCAAATCGTAATTTAGATGAAAGT
TTGAGTGAACACTCGGTATTTCATGAAAGCCATTCCTAAGCAAAATAAAAAAATCCT
GCTAGTCAGATGGAATATAAACTAAAATGTGTAAGAATTGTAGAATTGGTTGTTGTG
GGTAGCAACATGGCTGCATATACAGGCCAACATACTTCAAATTCAGTGGAGAATCGT
GGAATTCATCCACCGCCTCCTAAAAATTGGCCCCATTTTTCACTTAATCATCTGAAGT
GTATACATTTTGAACTTTTGTAGTTGCACTTCATTGTGTTATTTTGACAATTCTCGCTG
ATATTACATAATGGGTGTAACTGTAGGGGTGTTCATGGTTCGGTTTGGATCGGTTATT
ATTTATAATCATAATTAAATCAATTTAGTTGGTTTCTAAAATTTCTAAAATCAAACCA
AGCCAAATCAAAAAAATAACCATCGATTTGGTTTATTTTTTTGGTTTTTTTTCTATTTG
AAAGTAATAAGTTTGTTAAGGGCATACAAATTTCAATAGACACATACAACTAGTACA
TGAAGAATCCACAGAAAGCTATTGTGTATTCAATTAAGCAATAGTACTATAGTTATT
GTTAAGAAATACAAAAATTATTATACACGAACAAAGTGATTTAGATAAATAATACTA
AAGATCAAATCTTTATAGAAAAAAAAAAAAACAAAGCTCAATAATTAAATAACTATTC
GGGGAGGGGTTACAAGAAAACTACATATCAAAATACTAACAAGCGATTTTCCAGATA
TAAATGGTAAAGTTAAAAATTTAAGTTAATTAACA

>PPO1-3 off-target 4

GCATTATCTGGATCTAAAGAACTAGTCAAGATATGATATATAAGTCATATCTTTGTAG
CAACTGAAATCTTTTGCATAAACAAAAAAAAAAAGAAATCTGATTCTAAGTTGTAAA
GCAAAATAGACAAAAAAGATGTGGATAAACGGAAGGATGAGAGAAAGAGAAAAAG
AATACCAATGATATAAAATTCCAATATGTAAGGTCTATGAGTAATCTCATAAAAAGC
AGTGTAATAAAGCATCAATACGCATTCATACCATACTAAAAAGAATTTTCTTATAGA
AATGGAACAAAAAATTAAGAGCTTCGAGCCAATAAAGACTGAGAAGATTGACTCAA
GAACCAATTTCATTCTGAGCTCCATTGTAGAATTCGGACCTAACCATTAAGTAAGAA

GTGATGGGAACGACGGAACTTGTGAATGNCATTAAGAGCTTCGAGCAAATAAAGAC
TGAGAAGATTGACTCAAGAACCAATTTCATTCTGAGCTCCATTGTAGAATTCGGACCT
AACCATTAAGTAAGAAGTGATGGGAACGACGGAACTTGTGAATGCAAAAGATTCTAT
TGAAAAGGAATCTTAATGATTCACTGGTCGGGATGGCGGAATGAACCAAAGATTAAT
TCATGTATTCTGAGATCTGAGAAGTCACGAGTTAACCCTACAAATGAAATAGGGATT
GAAAGAGTAAATATTCGCCCGCGAAAACGTTTTTATTGAAAAATTTAAGAGACAAT
ACAATAAAGGACAAAATAAGAATTTGGTATAATAATACAATTTTTTTTCTATTTAGA
AAACTACAAAGTTTAGTTATCTATTCAAACAAGATATACAAATTACTAATAGATTTA
ATGAAATTTCAAAGAAGCCCACGTTCAAGGTATTACTCAGTAAATACATATATATAT
ATATCTTAACTTAAGATTGACTATTCTAGCTTAATTCAAATTGCATTTTTTTGAATCCT
TTTATTTGCGAGGAGCTGGATGAGAAGAAACTCTCACGTCCGGTTCTGTAGTAGAGA
TGGAATTCAGAACCAACCATCAACTATAACCCCAAAAGAACCAGATTCCGTAAACAA
CATAGAGGAAGAATGAAGGGAATATCTTATCGAGGTAATCGTATTTCTTTCGGTAAN
TATCCGGTTCTGTAGTAGAGATAGAATTCAGAACAAACCATCAACTATAACCCCAAA
AGAACCAAATTCCGTAAACAACATAGAGGAAGAATGAAGGGAATATCTTATCGAGG
TAATCGTATTTCTTTCGGTAAATATGCTCTTCAGGCACTTGAACCTGCTTGGATTACA
TCTCGACAAATAGAAGCAGGCCGACGGGCAATGACACGAAATGCACGTCGTGGTGG
AAAAATATGGGTACGTATATTTTCAGTCAAACCAGTTACACTAAGACCCGCAGGAAC
ACGTATGGGTTCAGGAAAAGGATCCCCCGAATATTGGGTAGCTGTTGTTAAACCCGG
GCGAATACTTTATGAAATGGGTGGAGTAACAGAAAATATAGCCAGAAGGGCTATTTC
ACTAGCAGCATCTAAAATGCCTATACGAACTCAATTCATTATTTCGTAATGTAATAGA
AAAAATATAGATATAGAAAGGGCTCTTAGATATGAATCAAAACCGCATGTTTCTTTT
TTTTTTTTTGACAAAAAATATTTCTTTTTTTCTTCGCCCTTTGCATTCAAAGAACGGAT
TAAAAAAATGATTCAACCTCAGACCCATTTAAATGTAGCGGATAATAGCGGNCAGAA
AGGGCTCTTAGATATGAATCAAAACAGCATGTTTATTTTTTTTTTTTTTTGACAAAAA
TATTTCTTTTTTTCATCGCCCTTTGCATTCAAAGAATGGATTAAAAAAATGATTCAAC
CTCAGACCCATTTAAATATAGCAGATAATAGCGAGGCTCGAGAATTGATATGTATTC
GAATCATAGGAGCTAGCAATCATCGATATGCTCATATTGGTGATGTTATTGTTGCTGT
GATCAAAGAAGCAGTACTAAATATGCCCCTACAAAGATCAGAAGTAGTCAGAGCTGT
AATTGTGCATACCTGTAAAGAACTCAAACGTGACAACGAGATGATAATATGATATGA
TGACAATGCTGCAGTTGTTATTGATCAAGAAGAAAATCCAAAAGGAACTCGAATTTT
TGGTGCAATCGCTCGGGAATTGAGAGACTTAAATTTTACTAAAATAGTTTCATTAGCT
CTCGAGGTATTATAAAATGAAATTGTGATCTGTTTCAAGTAGGGTATTTGAAAGAAA
TAGATTAAATAGTAGATTGTGTCTCACGCATATATCTTTAATAATTCATATCATATTC
CAAAATAAATAAGTAAAAAACACGTTGATTATATAAAATTTGAAAACCCCAATAATT
TTAGTTCATCATGGGCAGGGACACTATTGCTGAGATAATAACTTCTATACGAAATGCT

>PPO4-5-6 off-target 1

GTATGAGAAATTTCTAACATGTCGACTAACAATCAGCTGCATGTTTTACGTTTGTAGT
AGGTTCCATGAATTGTGTCTATAAGTAGTGATGAAACCAAGCAAAAACAAAAACACT
CATCCCAGCAATGGCTTCTTCTTTTCTACCTTTGTGTATCCACCCTTCTTTCTCCAATA
CCTCCGAATCATCTTTCTTACCAAAACCCTCTCAGCTTTTCCTCCAGAGAAGGCACAA
TCAGCGTTTCAAGGTTTCATGCAACGCCAACAAGCATGAAAAAGATAACCTTGACGT
TGTTGATAGGAGAAATGTACTTTTAGGCCTAGGAGGTCTCGGTGCTGCTAATCTTGCA
CCATTAACAGCCAACGCGGCTCCCTCACCTCCCCCTGATTTCAAAACTTGTGGTATAG
CCACTATAACAGCTGATGGTCCTCCAGTACCCTATACTTGTTGCCCACCTCCTATGCC
GTCTAACGTGAACACCATTCCATATTACAAGCTCCCTTCTATGACTAAGGTCCGTATC
CGTCAGCCTGCTCATACTGTTGATGAGGAGTTTATTGCCAAGTACAATTTGGCTATAA
GCCGAATGAAGGAACTTGATGAGAAAGAACCTTTAAACCCTCTTGGATTTAAGCAAC

AAGCTAATATTCATTGTGCTTATTGTAACGGTGCTTATAAAATTGGTGAGAAAGTGTT
ACAAGTTCATCAATCATGGCTTTTCTTCCCGTTTCATAGATGGTACTTGTACTTCTACG
AGAGGATGTTGGGAAAACTCATTGATGATCCCACTTTTGCTTTGCCATATTGGAATTG
GGACCTTCCAAAAGGCATGCGCTTACCTCCCATGTTCGATCGTGAAGGGTCTCCCCTT
TATGACGAAAGACGTAATCCGCAAGTCCGCAATGGAACCGTTATGGACCTTGGTTCT
TTCGGTGACCAAGTCCAAACAACTGAACTACAGTTGATGAGTAATAACTTAACTCTA
ATGTACCGTCATATGGTAACTAATGCTTCATGCCCTCTTTTATTCTTCGGTGGCCGTTA
CGTTCTCGGGAGTACCCAGGGAGTGCAGGGAACCATTGAAAAAATCCCTCACACTCC
TGTACACATTTGGGTCGGTACAAAGAAAGATTCAATTCTACCTAATGGTAAAAAGTC
ATACGGTGAGGATATGGGTAATTTCTACTCAGCTGCTTTGGACCCGGTTTTCTATTGC
CATCACAGCAATGTGGACCGAATGTGGAATGAATGGAAACAAATCGGAGGGAAAAG
AAGGGATCTCTCGCAAAAAGATTGGCTGGACTCAGAATTCTTTTTCTATGATGAGAA
CAAAAACCCTTACCTTGTGAAAGTGCGAGACTGTTTGGACACAAAGAAAATGGGATA
TGATTACGCGCCAAGCTCCACCGTATGGCGTAACTTCAAGCCAAACAAAAAGAACAC
AGATGGGAAAGTGAATACAGGTTCACTTCCGTCAGCGACCAAGATATTCCCAATCTT
TAAGCTGGACAAAGCCATTTCATTTTCCATTAATAGGCCGGCTTCCTCGAGGACTCAG
CAGGAGAAAAATGAACAAGAGGAGTTACTAACATTCAGCTACATAAAATATGATAA
TAGAGAATATATAAGGTTCGATGTGTTCGTGAACGTGGACAAGAACGTGAAAGCAG
ATGAGCTTGACAAGATAGAGTATGCGGGGAGCTATACGAGCTTGCCACATGTTCACA
AAGATGGTGATAAAGATCATATCGCGACTGCTACTTTGCAGCTCGCGTTAACTGAAC
TGTTGGAGGACATTGGTTTGGAAAATGAGGAGACTATTGCAGTGACTTTGGTTCCAA
AGAAAGGTGGTGAAGGTCTCTCTATTGGTTGTGTGGAGATCAAGCTTGAAGATTGTT
AATTTGGGGGGAGAATTAGTGTATTAATCTGCTGAGATTGCACTTTGATTCTGTTTTT
TTTCCCCCTTTTCCATGTCACTTTTCATGTTGAAATCAGCTTGACGCTCGATTTTTTCT
GCATTTATTTTCGCAAGAATAAATTGTTGGAAAAAATGCTACAATATTAATATTATAC
TAGAATTAAATAACTTATCACAAATACTGTGTCATGGCAAGCTACTGCTATGAAAGT
GATTTTGGCCCGACTTCGGTGCAAATCGCTCTAGCCGATCACTCCCCAAGATTCCATA
GTTACTTTCAAACATGTGCACTCGTGGCTGAAAGTTTGGAGATTGCCTAAAAAATGTT
TAAGAAGAAGAAAGGAGGAAGTTATTATTATTATTATTATTCTTTTGCTTTTATTGGT
ATCTTTTATATTGTTCACAAAGCCAACTTATTTAGGATTTTGAGGGGAAGGAGTTGCA
AAAAAGAGAGTTGAATTTGAGAAACTTAAAGTGAATATCAATGGTATTAAAGAACC
ATTATTATCATATTTACAGAAAACTCCTCTTCCTTTCTTTTGTT

>PPO4-5-6 off-target2
GCTACTACTGTTAAGAATTCAAGCAACGATGGAGGGATTGAAAGACTTGGAGAAGA
AGAGAATAGAGAATTTTTATTGATTGAAAAATGAAGTCGTACAACTGAACTTCTTCA
AACTTACACCAACTAGTATCTCTTTATATAGAAAGCAAAATAACCCTTCATCCCCACT
CATTAAATCTGTGTAAGTCAGCTCCACTGTGAAAAACGAAATTCATCGGAATATGTT
AAAGAGTAAAAAATTAATCATATTGAAACAATAACATGTGAATTATTCTAGGATAAC
ATATACATTTTCTTATAAGCATTCTTAAAAAATATAAAAAACACATTTTTAAAGATGA
ATATATTTCTCATTTTATTAACTCCCATTATTTAGAGAATTTTAAATATCCATTTATAT
TTTCATGAAAAGTAGTATCACGTTAATTTTTTTTTACCTTTCTTCCGCTTCAATTGAGC
AAGCAGAATAATACTCATAATGATTATTCAACATTTTGTGTAAGATAAAATAAGTTT
ATTCAAACATGCGCAATAGTGTTTCAACATGAAAAGTAACAGTTTTTTGAAAATTTTA
ACAATCCGCAAGAGTTATCTGGACACCTCCAATGGAGATACCTTCTCCACCTTTCTTT
GGTACCAGAGTCACCGCAATAGTGTTTTCATCTTCCAATCCAATGTCCTCCAACAGTT
CAGTTACCGCGAGCCCCAAAGTAGCAGTGGCGATATGGTTGGTATCACCAGCTCTGT
GAACGTGTGGCAAGCTTGTATAGCTCCCCGCAAACTCTGCCTTGTCAAGCTCATCCGC
ATTCAAATTCTTATCCGCATTCAGGAACACATCGAACCTTATATACTCTCTGTTATCA

TATTTTATTTCGCTGAACGTTAGCATCTCCTCTTGTTCATTTTTCTCCTGTTGAGTCCTT
GATGAAGCCGGCCTATTGATGGAAAACGAAATGGCTTTGTCCAGTTTAGCCAGTGGG
AACACCTTGCTAGCTGGCGGAAGTGAACTTGTATTCACTTTCCCTGATGATGCCTTTG
TTTTTGGCTTGAAGTTGCGCCACGGGGTTGGCATTGGGGCGTAATCATACCCCATCTT
CTTTGTATCCAAACAGTCTCGGACTTTCACACGGAATGGGTTTCCATTTTCATCGTAG
AAAAAGAACTCCGAGTTCAACCAATCTTTATGTGAGATATCCCTTCTTTTCCCTCCTA
TTGCCTTCCATTCGTTCCACATCCGGTCCACATTGCTGTGATGGCAGTAGAAAACCGG
GTCCAAACCAGCTGAGTAGAAATTACCCATGTCCTCGCCGTGTGACCTTGAACCATC
AGGCAAAATCGAACCTGCCCTTGTACCAGTCCAAACGTGGACAGGACCGTGAGGGAT
GACTTCAACGGTTCCCGGCGCTTCCACGTTATTCCCCAGAACGTAAGGCGCACCGAA
GAACATAAGGGGACATGGTGCATTAGTTACCAATTGACGGTACATTAGAGTTAAGTT
ATTACTCATCAGCTGTAATTGAGTTGTTTGGATTTGGTCACCGAAAGAACCAAGATCC
AGAACGGTCCCATTACGGACTTGTGGATTACGCCTTTCGTTGTAAAGGTCAGTCCCTT
GACGATCGAACATAGAAGGCAAACGCATGCCCTTAGGATGGTCCCAGTTCCAGTATG
GCAAAGCGAAAGTTGGATCATCGACCAGTTTTCCCAAGATTCTCTCATAGAAGTACA
AGTACCATCTATGGAACGGGAAGAAAAGCCATGAGTTATGAACTTGTAACACTTTGT
CATCAATTGCGTAAGCTCCGTTACAATAAGCACAATGTATATTGGCTTGTTGCTTGAA
GCCAATAGGGTTTGTAGGTTCTGTCTTATCAAGATCTTTCATTTTGCTAATAGCCAAA
TTGTACTTGGCGATAAACTCCTCATCAGCAGCATGAGCAGGCTGACGAAGGCGGAGC
TTAGTCGCAGAAGGGAACTTGTAATATGGAACTTTTTTCATATCGTCAGGCTTGGGAG
CGCAACACAAGTATGCCACCTCCTCCGTTTCAGTAATCTTGGCTTTTCTACAAGATGT
GAGATCAGGGGGTGGTACAGGAGTAGCCGATGCCGCTAATGGTATAGCATTAGCAA
CACCATACATCCCTCCTAAACCAAGAAGCACATTTCTCCTATCAACAGAATTTGTTTC
AACTACATTTTGGTTTTGGTCACCATTACTATTGGTAACCTTGCATGAGACTTTGAAA
GTTTGGTTACGTTTTCCATGGAGGAATAGTTGAGAGGGCTTTGGAGTGGAACGTAAA
GAAGGGGAGGAAGTAAGAGGAGATTTGAGAGTTGTAGTACTACTATTGCACACACTT
GCCATTGCTAGCTTCTGGATGAGAGATCCATAGGGCTATGCCTTGCCTTATATAATAT
GTTAGCACCAAATAAAATATTTCACCATTTGAAAATATTC

>PPO4-5-6 off-target3
ATACTATCCTTGACCTTTATGTTTTCTAGGAGTTCTTTATTCTTTCAAGAATATCATTT
AATCTTTCAATATTATCTCTTAGTTTTTCAAGAAATCTTTTAAAAGATTGTGTTACATG
TATCTAGTTATTTGTGCAAATCTTTTAAATTGGTTTTAGTAGTATAAATAGAGATGTA
ATTGATGATTGTATATAGATCGAGTGGCCTTAAGTAGCCTAATATAACTTGAGCATTC
TCTAAGAAAGTTTTTTCTCTTCCTATTCTTGTTTTTAACACTATGAATAATTTGAGGTT
GAAATTCGTGTCAAGCTAGGTTTAGAGGTGTTTTTAATACTTTTTTCGATATATAAAT
GTGGTTCACCCTTTAGAGTTGATCTTAGCATTTTTCCAGGTTGCAGTTATATAGATATT
TTTTTATAACATAATTGTAACAGATTTATTTTTGAATAACAACTAAAGGAAATCTAGC
GTCAAGCTGATTTCGACATGAAAAGTAACATGGAAAAGCAAAACAGAGTATGGAAA
GAGAACAATCCATCAAAGTTATCTCAGTAATTCTCCCCTCAAATTAACCAGGACTTA
ACAATCCGCAAGACTGATCTCTACACCTTCAATGGAGATACCTTCACCGCCTTTCTTT
GGTACCAGAGTCACCGCAATAGTGTCTTCATCTTCCAGACCAATGTCCTCCAACAGTT
CAGTTATCGCCAGCTGCAAAGTAGCAGTCGCGACATGGTCAGTATTCTCGCCAACTC
TGTGAACATGTGGCAGATTAGTATAGCTCCCAGCGAACTCTGCCTTGTCAAGCTCGTC
AGCGTTCACATTACTGTCCACATTCAGGAACACGTCGAACCTTACATAGTTTCTATTA
TCATATTTTATGTTGTTGAACGTTAGCATCTCCTCTTGTTCATTTTTCTCCTGTTGAGTC
CTTGATGAAGCCGGCCTATTGATGGAAAACGAAATGGCTTTGTCCAGCTTCGCGAGT
GGGAATACCTGGCTTACCGGGGGAAGTGAACTTGTATCAACTTTCCCTACTGATGCCT
TTGTTATTGGCTTGAAGTTGCGCCACGGTGTTGGCATGGGTGCGTAATCGTACCCCAT

CTTCTTCGTGTCCAAACAATCTCGGACTTTCACACGGAATGGGTCTCCATTTTCATCG
TAGAAAAAGAACTCCGAGTTCAACCAATCTTTATGTGAGATATCCCTTCTTTTCCCTC
CTATCGCTTTCCATTCGCTCCACATCCGGTCCACGTTACTGTGGTGGCAAAAGAAAAC
CGGGTCCAAACCAGCTGAGTAGAAATGACCCATGTTCTCACCGTGTGACGTTCTACC
ATTAGGCAAGGTTGTACCTGGCACTGTACCAGTCCAAACGTGGACAGGACCGTGAGG
GATGACTTCAATGGTTCCTGGGGCTTCAACGTTATTCCCAAGAACGTAAGGCGCGCC
AAAGAACATCCGAGGACATGGGGCATTAGTCACCATTTGACGGTACATTAGAGTTAA
GTTATTACTCATCAACTGTAGTTCAGTTGTTTGGACTTCATCACCAAAAGAACCAAGA
TCCATTACTCTTCCATTGCGGTGACTTTGGTTACGTACTTGGTCGTAAAGGGCAGTCC
CTTCACGATCGAACATGGCAGGCATACGCATGCCCTTTGGATGGTCCCAATTCCAAT
ATGGCAAAGCAAAAGTTGGATCATCAATGATGGATCCAAGGATTCTCTCGTAGAAGT
ACAAGTACCATCTATGGAAAGGGAAGAAAAGCCATGAATTATGAACTTGTAACACTT
TGTCACCAATTTTGTAAGCACCGTTGCAATAAGCACAATGGATATTAGCTTGTTGCTT
AAACCCAATAGGGTTTAAAGGTTCTGTTGTATCAAGATGTTTCATTCGGCTGATAGCC
AAATTGTACTTTGCAATATACTCCTCATCAGCAGCATGAGCAGGCTGACGAATACGG
AGCTTAGTCATAGAAGGGAACTTGTAATATGGAATTTTGTTCAAATCATCAGGAGTG
GGAGCGCAACAACTGTATGACACCTCCTCAGTTTCGGTAATCTTGGCTATACTACAA
GACGAGAGATCAGGGGGTGGTGTAGGAGCAGCTGATGCAGCTAATGGTGCAAAATT
AGCAGCACCATACATCCCTCCTAAACCAAGAAGAACATTTCTTCGATCAACAGAATT
TTTTTCAACTTCGTTTTGGTTTTGGTCACCGTTATTATTGGAAACTTTGCATGAAACTT
TGAACGTCTGTTTACGTTTTCCATGGAGGAAAAGTTGAGAGGGCTTAGGAGTGGAAC
CTAAAGAAGTGAAGGGAGTTTTGAGGGTTGTATTACTACTATTGCACAAACTTGCCA
TTGCTAGCTTTGTATGAAGATTCATAGGCAACATCTTGCCTTTTATAATAAGTATAAA
TTAACACAATAAATTATTAATCCATTTGAATCAATAT

>PPO4-5-6 off-target4

CTGTCTTTAACTCATCACCATGCCCTAAAAAGCTTGCAATGATTTATATACTCATACA
TATTCATGTGTCTTGTAGGACAAATTTTGACATAACATGCTATACCCCCACCCCAACC
CAAAAACTAATTTCATGCATAGCATCACTAATTTCATGTATAGCATCACTTTTCTACT
TGAGTTTGGCCAAATGTTTCATTTCTTTTTGCTAGGAAGCAATTCATCTACTAAAGTG
ATACACAAAAGTATACTGTGGAATAGTAGTTTCATACTCCCATCGTATTAATTTACGT
TGTGACATTTGACTGGGTCTGAATTTGAAGTTTAAGGGAAAAAAAAGAAAGATTTTA
ATATTTTTTTATCCAAAGGAATTATTTAATTTGGGTATAAAATAAAAACTTTAAAATT
AATTTTTAAGGATATAGAAATGTGTCGCATGAAATTGAAAAGATCGTTATTTTACCCG
AGGTGATGACATTAGCAGAGAGTTATCATATTCATAGTACATGACTTTAAATTTACG
GATCGTTATTTTACCCGAGGTGATGACATTAGCAGAGAGTTATCATATTCATAGTACA
TGACTTTAAATTTACGGAGTCAAAAGTGTGCTTTAATATTGTGAAAGTCATATTTTTC
TTTACACACGTGCTTACTTAATTGTCTGGTCAACGGATCACCCCAGCAAATAGTACAT
GAGCTTTAAATTTCATATCAGGTCATGTTATGATCAAAAGACAAATTAGGAGGGACC
TAATAAGAACTTGACTTCAATGCTTTTACAATTTGAATGATGGTTTTTTTTAATAAA
AAAAGAGTTCATTAGTGATGCATATGAACTTTTAATGCTGAGTTAGTTAAGAATTAG
CGTGACGTTGTCAAGACAATAGGAGATGGAACGCAAATTAAAGTAGAGACATATCA
ATATAATGTAACATAACAATAACATATTCATTATAATTTCATAATTTTGTAAGTGAAA
TCTAAGAGAAATAGAATGTATGCATACCCTTATTTTGATATTTGCGGGATAAAGAAG
TTGTTTTCAATAGATTCTTAGCTCAAGGAGATTATTCTCAAATAAAGGTTTGAAAGAA
AAGCAAGATTAAAAGGTCAAGTTAAAATATTGAATATTAAAGATAATCAAAGTAAA
AGAAACATTGATTGGATTGATGGTTTGTGAGCATTGAATCTTGAAAGATGTGATTGG
AAACTAATCGGATGGTTGATAGGCCGAAAGGCTTGATTGTGATGGATTGATAAGAGT
TCTTATGAGATTCATTCATTGGATTGATGGTATATGAGCATTGAATCTTGGGAGGAGT

ATCAAGCACCGAATTGAGCAAGAGTAAGTCCAACTCGAACTCAATAACTACGTCGCC
AAATGTAGGAGAGGATTAAACCGTTAAGTCGGATGTTTTCCTATTTTATTATCCTGAA
AGAATAGGACTTGATTAATTGTTTGATCCATGATTGGTTGATTCATTCGTTATCCTAG
CAAAGTATTGAATAGTTGTGGCAACATCAATAGGATTGTTGGACAATCAGTATTGAC
TCATAAGTGATTGTTGTCGGATAAGAGAAACTCCCATATAGGTCTTGATAGTACTTTG
AATGGATTGGATTGTATATGATTGGTCTCGGTCTAAATCTTACTCTTGCTTTAGACTT
ATGACATCTAGATCGAGATCCTTATTTTCTTGATTTGAGTTAGTTGAGTTGATAGTAT
TCTACCTTGATGCATGTTTCATTCTGCCATTTTACATACATTTCACGTACTGACGTCCA
TTTGGACCTGCATCATTTTATGATGAAGTCAAGTACTAGAGGTCGTCAACAGACACTT
CGTTGCGGATCCTTTCGCATTCAGCTGATTGGTGTGACCTCCATACTTTCGGGAGATT
TTCACATTATTATCATCTAACTATTAGAGTTTAGATTTGCTTTTATTTTAATTTAAGGT
AGTCATAGAGTTGTCATTGGCACCACATAAACAGCATGGATAGAGGCTTTATAGATT
AGACAGTGATTGATGATTTGGGATGTTGTCTTTTGACTGCTAGTTTTCTTTCTTACTAG
TCTTCCGATTGGTTTTGCTAGATTGGTTGGCCTCTGGCTTAGAGTTATCTCTTGATGAT
TGAATATTCCGCAATTAGATTATGATTGAATGAATGTGATAAGATTAAGTGATTTGCT
TGGGGACTAACAATAGTCTCTGAGTATCGTCCACGTCAAGAGTACCCTCCCGGAACG
TGACACTTGTAATGCTAGAACCCAGCCTATGACTAGCGCCTGATGCTTCAGCAATAT
AGATTTTAGTCGAGACATTAATCCTCCTCAATAGCATCCTCTTCAAGATAACTAATAG
TATAGTTGATATGACTAAGTACTTCACATGCCCCACCTCCCCATCCCCTTGTAGAGTG
TAATTGAGCAGTTTAGATAACAGATTGTGTTGAATC

General Discussion and Conclusions

1. What to expect of the CRISPR era

CRISPR/Cas9 has revolutionized the field of gene editing in plants, and will likely continue to do so for years to come. Programmable nucleases like ZFNs and TALENs had already opened exciting perspectives for custom genetic modification of crops and model plant species (including gene knock out, homologous recombination, sequence replacement and regulation of gene expression), but two characteristics of CRISPR/Cas9 gene editing, combined, are predicted to mark a difference with respect to previous genetic engineering approaches. First, the simplicity and versatility of the system: ZFNs and TALENs, for their molecular characteristics, required the synthesis of a new protein for every editing target, which made experimental design complicated and expensive (Kim & Kim, 2014; Shukla *et al.*, 2009; N. Sun & Zhao, 2013; Yan *et al.*, 2013); CRISPR/Cas9, on the other hand, is exceptionally simple at the structural level, with a single endonuclease which can be directed at diverse targets by just providing it with a 20 nucleotide custom RNA. This characteristic makes it readily transferable to new species and new targets, if adequate transformation systems are available: this marks an improvement, in terms of time and costs, over other programmable nucleases. Second, with respect to 'traditional' genetically modified organisms, it is possible from the T_1 generation (sometimes also in the T_0 generation, as reported in Andersson *et al.,* 2017 and Svitashev *et al.,* 2015) to recover mutated individuals which do not carry any foreign sequence in their genome; in addition, CRISPR-induced mutations are practically undistinguishable, in form, from spontaneous mutations resulting from small insertions and deletions, making these plants analogous, in substance, to naturally occurring variants. For this reason, CRISPR-derived plants could probably have a significant advantage in terms of regulation, which in turn would have a paramount impact on the cost of developing new edited plant varieties (Puchta, 2017). At present, indeed, the great economic burden of developing genetically modified crops does not result from plant transformation and testing, but rather from an onerous registration process which causes the developing of a transgenic crop to have a cost in the order of tens of millions of dollars.

Different applications of gene editing can be achieved with CRISPR/Cas9 with varying efficiency: all rely on the induction of double strand breaks, and the specific editing outcome depends on the way these will be repaired (Ceasar *et al*., 2016). Repair by non-homologous end joining (NHEJ) is by far the most efficient system in plants, and an extensive literature exists which confirms its success in knocking out plant genes (Puchta, 2017). A less efficient, but still feasible application of CRISPR-mediated gene editing consists in gene

targeting, that is the precise insertion or replacement of a DNA sequence at a target *locus*. This repair pattern is triggered by the induction of paired double strand breaks, and by providing the cell not only with the editing machinery, but also with a repair template. Examples of this approach have been reported by Schiml *et al.* (2014) in Arabidopsis, by Čermák *et al.* (2015) in tomato, by Svitashev *et al.* (2015) in maize and by Endo *et al.* (2016) and Yu *et al.* (2017) in rice.

Of course, the first reports of gene editing in plants came from model species, such as *Arabidopsis thaliana*, *Nicotiana benthamiana*, and even tomato itself, which serves as a model for fleshy fruit development and whose abundance of genomic data and good regeneration potential make it amenable to a great number of functional studies (Gupta & Van Eck, 2016). Other crops to be readily edited with CRISPR/Cas9 were the ones, like rice, maize and soybean, which represent the greatest shares of global agriculture for food, feed and industrial applications (Li *et al.*, 2015; Miao *et al*., 2013; Sun *et al*., 2016; Svitashev *et al.*, 2015, 2016; Wang *et al.*, 2015). Having been established as a reliable gene editing technology, however, CRISPR/Cas9 holds great promises for the development of sustainable, innovative crops with improved agronomical traits, nutritional quality and favorable industrial characteristics, including crops adapted to changing environmental conditions and to specific agricultural systems (Ricroch *et al*., 2017). Moreover, if CRISPR-derived plants will not be considered as GMOs, and if regulatory authorities will not consider it necessary to regulate them, an efficient, precise and simple gene editing technology will be made affordable for smaller companies and research institutions. This, in turn, would benefit the improvement of crops which do not occupy a prominent position in terms of cultivated area or market value, but which do have a cultural, ecological, gastronomic value, or which represent important economic resources for specific areas and populations (SIGA & SIBV, 2017).

In our work, we chose tomato as a model species to tune a GoldenBraid-based gene editing protocol; to this end, we first chose an endogenous target gene, and then addressed the issue of virus resistance, both for its innovative multi-target approach and for the relevance of Geminivirus-related diseases in tomato. The choice of eggplant for CRISPR/Cas9, on the other hand, depended on the experience our group acquired working on this species over the past few years and on its importance as a horticultural crop worldwide, and constituted an opportunity to set up, for the first time, a gene editing protocol for this species. In addition, the *ppo* gene family received interest in recent years for the

development of genetically engineered crops (namely apple and potato) with reduced browning.

2. The great bottleneck: regeneration and organogenesis.

CRISPR/Cas9 gene editing proved to be highly efficient in a range of different species, and is able to introduce heritable mutations; the system is specific, accurate and fast (Bortesi *et al.*, 2016). However, successful gene editing has three fundamental requirements, all of which strongly depend on basic research.

The first requirement is represented by the availability of a high quality genomic sequence of the target organism, which is required to design a specific gRNA and to evaluate its quality; the genome sequence of the organism to edit is also necessary to evaluate potential off-target effects, especially when working with gene families. Complete genome sequences are now available for many relevant crops (including tomato, potato, eggplant and pepper, in the Solanaceae family) and genomic information is becoming more readily available thanks to high-throughput sequencing technologies and reduction of costs.

The second requirement is to possess an efficient way to deliver the CRISPR/Cas9 machinery to plant cells: this can be done either by integrating the corresponding TUs in the plant genome, or through transfection of protoplasts with a ribonucleoprotein complex of the Cas9 and the single guide RNA, which can directly edit the target locus without being integrated in the genome (Andersson *et al.*, 2017; Svitashev *et al.*, 2016); another method which does not require integration is virus-mediated transformation (Baltes *et al.*, 2014; Wang *et al.*, 2017). Transformation protocols exist for many species and are based mainly on *Agrobacterium*-mediated transformation or on the biolistic approach. In case the transgene is integrated, it is necessary to go through segregation in order to obtain an edited but transgene-free plant. This does not constitute a limitation for species like tomato and eggplant, which normally reproduce sexually and for which generation times are short. It can, instead, be a limitation for species which are routinely propagated in a vegetative way: these include species like potato, and the majority of fruit trees, which have long generation times. For these species, the availability of a gRNA/Cas9 delivering technology not involving stable integration of a transgene would decidedly improve development of edited crops. In biolistic- and *Agrobacterium*-mediated transformation it is possible, albeit with very low

efficiency, to recover mutants without an integrated transgene (Svitashev *et al.*, 2015).

The third requirement for successful gene editing is the availability of regeneration methods to obtain whole, viable individuals from undifferentiated, transformed cells and this currently represents the true bottleneck and the major obstacle to routine gene editing in plants. The optimization of regeneration protocols to induce *callus* formation and, from this, differentiated and specialized plant tissues is a highly empiric, trial-and-error field of plant science. Significant variability exists within each plant species regarding regeneration efficiency, depending on genotype, growth conditions, tissues used for transformation, age of such tissues, growth regulators, use of antibiotics and application of specific treatments (like dark incubation and use of antioxidants) (Gerszberg *et al.*, 2015; Trigiano & Gray, 1999). The interaction of these factors with plant cells are still largely unknown, and it is only recently that their molecular mechanisms started to be unveiled (Ikeuchi *et al.*, 2016). For this reason, the first question to be addressed when planning gene editing in plants is whether a suitable regeneration protocol is available.

Regarding the first two requirements, we did not encounter any difficulty with both species we worked with. Genomic resources for tomato are abundant, as it was the first species of the Solanaceae family whose genome was sequenced. A high quality genome sequence for eggplant was recently released by a consortium of Italian institutions (http://www.eggplantgenome.org/): the availability of the complete genome allowed us to expand the search of targets and the characterization of the *ppo* gene family we wanted to edit in eggplant. To deliver the CRISPR/Cas9 editing machinery to cells, we relied in both cases on *Agrobacterium*-mediated transformation. For tomato, we could rely on previous experience in transformation and regeneration and we were be able to exploit a highly efficient, established protocol for cotyledon transformation. For eggplant, on the other hand, we needed to determine which conditions were most beneficial for organogenesis. A number of protocols exist for *Agrobacterium*-mediated eggplant transformation and subsequent recovery of transformed plants, with adjustments for genotype, shoot differentiation, somatic embriogenesis and growth conditions (Arpaia *et al.*, 1997; Ikeda *et al.*, 1999; Magioli *et al.*, 1998; Muktadir *et al.*, 2016; Swamynathan *et al.*, 2010). With respect to tomato, eggplant has additional issues which need to be addressed, especially production and release of phenolic compounds in the culture *medium*, which can be oxidized and cause necrosis in plant tissues, and a greater difficulty in producing shoots. Various conditions and combinations of growth regulators proved efficient in induction of *callus* growth; shoot

formation from *calli* was instead more difficult to obtain, and depended mostly on the achievement of the right balance of cytokinins or cyotkinin-like compounds (like trans-zeatin and thidiazuron) with respect to auxins. Rooting is another critical phase of *in vitro* culture, which is strongly dependent on plant interactions with growth regulators and antibiotics, which are known to inhibit formation of roots. In the case of eggplant, we proved the beneficial effect of supplementing culture media with PVP, which (in contrast to ascorbic and citric acids) does not show any adverse effect on plant tissues and significantly increases the number of regenerated shoots. On the other hand, we did not observe a correlation between dark pretreatments and number of regenerated shoots.

3. Critical aspects of gene editing

One of the greatest concerns regarding CRISPR gene editing was that of the possible off-target activity of programmable endonucleases. Some off-target mutations had been reported especially in mammalian cell cultures (Fu *et al.*, 2013; Hsu *et al.*, 2013), but often derived from the use of mutation-prone cancer cells or from the use of poorly specific gRNAs. However, extensive characterization of activity patterns of CRISPR/Cas9 in a wide range of organisms drastically downsized this concern (Bortesi *et al.*, 2016). Illumina amplicon sequencing approaches, or whole genome resequencing approaches, have often been implemented to detect off-target mutations and assess the specificity of the CRISPR/Cas9 editing system (Baysal *et al.*, 2016; Doench *et al.*, 2016; Nekrasov *et al.*, 2017), and always resulted in a confirmation of the rarity of such effects. In this work, we used an Illumina amplicon sequencing approach to evaluate gene editing activity at 5 potential off-target sites for each of 2 gRNAs in tomato, and 4 potential off-target sites for each of 2 gRNAs in eggplant. Our results confirm the specificity of the system, especially in tomato in which the targets did not belong to a gene family. In eggplant, the context was made more complex by the fact that we were trying to differentially target members of a conserved gene family, which resulted in many potential off-targets being highly homologous genes. Some off-target mutations have been found in tomato by Jacobs *et al.* (2017) when targeting leucine-rich repeat (LRR) subfamily XII, but careful evaluation of putative off-target sites upon designing gRNAs and CRISPR constructs is generally sufficient to ensure high specificity to gene editing. In addition, off-target mutations are, by definition, less likely to occur than target gene editing, given also the great susceptibility of the CRISPR/Cas9 machinery to even a little number of mismatches between

gRNA and target sequence. This means that, should they occur, they would likely be segregated in the progeny and could easily be excluded, without representing a drawback for plant breeding.

Conversely, it must be taken into consideration that the efficiency of the CRISPR/Cas9 system often results in complete knockout of the target *locus*. While this is commonly considered a positive feature, which allows to rapidly obtain stable mutants, it can be detrimental for those phenotypes which can negatively impact plant growth. An example of this is the *Ty5/pelo* tomato mutant we obtained (described in Chapter II), which failed to generate a progeny because of severe yellowing, stunting and wilting, likely resulting from impaired protein synthesis caused by the CRISPR-induced mutation. Together with the accurate selection of the gRNA, careful selection of target genes is also essential for the outcome of CRISPR gene editing. Various reports exist of gene editing applied to pathogen resistance in plants through targeting of endogenous genes (Chandrasekaran *et al.*, 2016; de Toledo Thomazella *et al.*, 2016; Nekrasov *et al.*, 2017). When pursuing this approach, the tradeoff between pathogen tolerance or resistance and plant growth must be carefully evaluated. In general, disease resistance is controlled by single, strong *loci* with specific functions in plant-pathogen interactions. Especially in the case of viruses, which rely on the host molecular machinery for replication, translation and movement, tolerance is instead mediated by the mutation of genes involved in fundamental cell processes which support viral multiplication. Addressing these *loci*, as we did for *Ty5/pelo*, can pose serious threats to plant metabolism, especially if no redundant genes can at least partially supply for the loss of function.

The use of the CRISPR/Cas9 system as a plant immune defense tool against viruses provides a specific and clean system to limit viral infections, as opposed to expensive and inefficient agricultural practices and time-consuming plant breeding for resistance, but it also appears to be one of the most challenging practical applications of gene editing to crop improvement, so far. It is possible that specificity itself might constitute one of the main drawbacks of this approach. CRISPR/Cas9 is a microbial system, evolved by Prokaryotes to fend off viral infections, and is adaptive: although prokaryotic mutation rates are not comparable to those of viruses, and viruses therefore have an advantage to evade immune defenses, prokaryotic cells maintain a variable repository of spacers, which dynamically reflects environmental conditions and is adapted to recent encounters. On the other hand, when transforming a plant with a CRISPR array of viral targets, the targets recognized by the plant are bound to remain unchanged over a long time, while viruses are free to mutate at much greater rates. The choice of multiple targets in conserved functional domains of viral

genes such as Rep and CP is predicted to increase the durability of the system, but Ali *et al.* (2016) and Tashkandi *et al.* (2017) have already proved that TYLCV is potentially able to evade such defenses over a small number plant generations, which would require the development of plants carrying novel sets of targets every few years, very much in the same way as vaccines are developed. Our approach consisted of targeting three distinct *loci* in the replicase and coat protein of different species of TYLCV. Functional domains were chosen to counterbalance the ability of the virus to rapidly produce variants of these genes, by focusing on regions subject to greater evolutionary constraints.

4. Conclusive Remarks and Perspectives

Although promising and efficient, the applications of CRISPR/Cas9 for crop improvement are still in their infancy. Many different issues, ranging from gRNA design to plant regeneration *in vitro*, need to be evaluated and optimized according to plant species or even to particular genotypes within a species. The work which is currently being carried out in model species and, increasingly, in important grain and vegetable crops represents a valuable body of knowledge for a future wide application of gene editing to plant breeding.

The results reported in this thesis provide an important contribution to the progress of knowledge on the application of gene editing to Solanaceae species. The GoldenBraid toolkit for CRISPR/Cas9 gene editing proved to be an efficient and time saving method to assemble a multi-TU construct carrying all the necessary elements of the gene editing machinery. Tomato 'MoneyMaker' plants were successfully edited to produce chlorophyll retaining *gf1* mutants from which a T_1 progeny with fixed mutations was obtained and whose individulas, in some cases, segregated for the transgene. Editing efficiencies were very high, often close to 100%, and the system demonstrated to be highly specific, as it did not introduce off-target mutations at undesired sites in the genome.

Although a gene editing approach to engineering virus resistance in plants constitutes an attractive perspective, it presents some challenges as well. In tomato, transformed and edited plants were regenerated with high editing efficiencies, but the production of a progeny from the T_0 was unsuccessful. This hampered the possibility to assess the resistance to TYLCV and TYLCSV and did not allow to gather key information on plant-virus interactions in the genotypes under study. On the other hand, promising results could be obtained

through transient expression of the CRISPR/Cas9 constructs targeting TYLCV and TYLCSV in 2IR-GFP *N. benthamiana* plants, indicating that the system can effectively cut viral DNA and reduce virus replication rates. The reported results also confirmed the functionality of GoldenBraid multiplexing CRISPR vectors, based on the expression of a cassette of gRNAs as a polycistronic transcript, and on their subsequent processing and release. This feature further enhances the similarity between this synthetic strategy and the natural prokaryotic CRISPR/Cas9 system.

A furher goal of the present thesis was also the development of an efficient protocol for transformation and shoot regeneration in eggplant, which provides the first example of CRISPR/Cas9 gene editing in the species. Constructs were designed to specifically target members of the *ppo* gene family and reduce browning of tissues due to damaging and cutting. Our protocol will have to be tested in a larger number of genotypes, especially varieties with a high content of polyphenols.

No obvious limitations have been shown, so far, to depend strictly on the CRISPR/Cas9 gene editing technique. However, additional editing approaches and transformation methods to those routinely used are proving beneficial especially in some applications. Optimized versions of the *Sp*Cas9 with increased on-target specificity and reduced off-target effects were obtained either through rational design (Kleinstiver *et al*., 2016a; Slaymaker *et al*., 2016) or through a yeast-based screening of a library of *Sp*Cas9 variants (Casini *et al.*, 2018). After Cas9, other endonucleases started to emerge as potential tools for genome engineering: one of them is Cpf1. While the fundamental mechanism of inducing double strand breaks and triggering repair pathways is conserved, the ability of this protein to leave cohesive rather than blunt ends is predicted to benefit homologous recombination mechanisms (Hyeran Kim *et al.*, 2017; Kleinstiver *et al.*, 2016b; Shan-e-Ali Zaidi *et al.*, 2017; Zetsche *et al.*, 2015). Beside Cpf1, other novel endonucleases were identified, including C2c1 and C2c2, two other Class II CRISPR-associated proteins capable of targeting DNA and RNA, respectively. The use of different endonucleases with distinct sequence recognition patterns can significantly improve the number of recognized targets in the genome, boosting the multiplexing potential of gene editing. The use of modified viral vectors for plant transformation and production of great numbers of copies of the editing machinery and of repair templates is also being successfully used to improve gene targeting efficiencies (Butler *et al.*, 2016). Furthermore, in order to develop viral vectors carrying *Cas9* genes, alternative *Cas9* orthologs with smaller protein sizes are being

evaluated, from different bacterial sources, such as *Campylobacter jejuni* (Mitsunobu *et al.*, 2017).

Bibliography

Ali, Z. *et al*. (2016) 'CRISPR / Cas9-Mediated Immunity to Geminiviruses : Differential Interference and Evasion', Scientific Reports. Nature Publishing Group, (May), pp. 1–11. doi: 10.1038/srep26912.

Andersson, M. *et al.* (2017) 'Efficient targeted multiallelic mutagenesis in tetraploid potato (Solanum tuberosum) by transient CRISPR-Cas9 expression in protoplasts', Plant Cell Reports, 36, pp. 117–128. doi: 10.1007/s00299-016-2062-3.

Arpaia, S. *et al*. (1997) 'Production of transgenic eggplant (Solanum melongena L.) resistant to Colorado Potato Beetle (Leptinotarsa decemlineata Say)', Theoretical and Applied Genetics, 95(3), pp. 329–334. doi: 10.1007/s001220050567.

Baltes, N. J. *et al.* (2014) 'DNA Replicons for Plant Genome Engineering', The Plant Cell, 26(1), pp. 151–163. doi: 10.1105/tpc.113.119792.

Baysal, C. *et al.* (2016) 'CRISPR/Cas9 activity in the rice OsBEIIb gene does not induce off-target effects in the closely related paralog OsBEIIa', Molecular Breeding, 36(8), pp. 1–11. doi: 10.1007/s11032-016-0533-4.

Bortesi, L. *et al*. (2016) 'Patterns of CRISPR/Cas9 activity in plants, animals and microbes', Plant Biotechnology Journal, 14(12), pp. 2203–2216. doi: 10.1111/pbi.12634.

Butler, N. M. *et al.* (2016) 'Geminivirus-Mediated Genome Editing in Potato (Solanum tuberosum L.) Using Sequence-Specific Nucleases', Frontiers in Plant Science, 7(July), pp. 1–13. doi: 10.3389/fpls.2016.01045.

Casini, A. *et al.* (2018) 'A highly specific SpCas9 variant is identified by in vivo screening in yeast', Nature Biotechnology, doi:10.1038/nbt.4066

Ceasar, S. A. *et al.* (2016) 'Insert, remove or replace : A highly advanced genome editing system using CRISPR / Cas9', Biochimica et Biophysica Acta. Elsevier B.V., 1863, pp. 2333–2344. doi: 10.1016/j.bbamcr.2016.06.009.

Čermák, T. *et al.* (2015) 'High-frequency, precise modification of the tomato genome', Genome Biology, 16(1), p. 232. doi: 10.1186/s13059-015-0796-9.

Chandrasekaran, J. *et al.* (2016) 'Development of broad virus resistance in non-transgenic cucumber using CRISPR / Cas9 technology', Molecular Plant Pathology, 17(7), pp. 1140–1153.

Doench, J. G. *et al*. (2016) 'Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9', Nature Biotechnology. Nature Publishing Group, 34(2), pp. 184–191. doi: 10.1038/nbt.3437.

Endo, M., Mikami, M. and Toki, S. (2016) 'Biallelic Gene Targeting in Rice', Plant Physiology, 170(February), pp. 667–677. doi: 10.1104/pp.15.01663.

Fu, Y. *et al.* (2013) 'High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells', Nature Biotechnology. Nature Publishing Group, 31(9), pp. 822–826. doi: 10.1038/nbt.2623.

Gerszberg, A. *et al*. (2015) 'Tomato (Solanum lycopersicum L.) in the service of biotechnology', Plant Cell, Tissue and Organ Culture, 120(3), pp. 881–902. doi: 10.1007/s11240-014-0664-4.

Gupta, S. and Van Eck, J. (2016) 'Modification of plant regeneration medium decreases the time for recovery of Solanum lycopersicum cultivar M82 stable transgenic lines', Plant Cell, Tissue and Organ Culture. Springer Netherlands, 127(2), pp. 417–423. doi: 10.1007/s11240-016-1063-9.

Hsu, P. D. *et al.* (2013) 'DNA targeting specificity of RNA-guided Cas9 nucleases', Nature Biotechnology. Nature Publishing Group, 31(9), pp. 827–832. doi: 10.1038/nbt.2647.

Ikeda, T. *et al.* (1999) 'Growth dependence of ovaries of facultatively parthenocarpic eggplant in vitro on indole-3-acetic acid content', Scientia Horticolturae, 79, pp. 143–150.

Ikeuchi, M. *et al.* (2016) 'Plant regeneration: cellular origins and molecular mechanisms', Development, 143(9), pp. 1442–1451. doi: 10.1242/dev.134668.

Jacobs, T. B. *et al.* (2017) 'Generation of a collection of mutant tomato lines using pooled CRISPR libraries', Plant Physiology, 174(4), pp. 2023–2037. doi: 10.1104/pp.17.00489.

Kim, H. *et al.* (2017) 'CRISPR/Cpf1-mediated DNA-free plant genome editing', Nature Communications. Nature Publishing Group, 8, pp. 1–7. doi: 10.1038/ncomms14406.

Kim, H. and Kim, J. (2014) 'A guide to genome engineering with programmable nucleases', Nature Reviews Genetics. Nature Publishing Group, 15(5), pp. 321–334. doi: 10.1038/nrg3686.

Kleinstiver, B.P. *et al.* (2016a) 'High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects', Nature, 529, pp. 490-495.

Kleinstiver, B. P. *et al.* (2016b) 'Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells', Nature Biotechnology, 34, p. 869.

Li, Z. *et al.* (2015) 'Cas9-Guide RNA Directed Genome Editing in Soybean', Plant Physiology, 169(2), pp. 960–970. doi: 10.1104/pp.15.00783.

Magioli, C. *et al.* (1998) 'Efficient shoot organogenesis of eggplant (Solanum melongena L.) induced by thidiazuron', Plant Cell Reports, 17(8), pp. 661–663. doi: 10.1007/s002990050461.

Miao, J. *et al.* (2013) 'Targeted mutagenesis in rice using CRISPR-Cas system', Cell Research, 23(10), pp. 1233–1236. doi: 10.1038/cr.2013.123.

Mitsunobu, H. *et al.* (2017) 'Beyond Native Cas9: Manipulating Genomic Information and Function', Trends in Biotechnology, Vol.35, No. 10, pp. 983-996

Muktadir, M. A. *et al*. (2016) 'Regeneration efficiency based on genotype, culture condition and growth regulators of eggplant (Solanum melongena L.)', Agriculture and Natural Resources. Elsevier Ltd, 50, pp. 38–42. doi: 10.1016/j.anres.2014.10.001.

Nekrasov, V. *et al*. (2017) 'Rapid generation of a transgene- free powdery mildew resistant tomato by genome deletion', Scientific Reports. Springer US, (February), pp. 1–6. doi: 10.1038/s41598-017-00578-x.

Puchta, H. (2017) 'Applying CRISPR / Cas for genome engineering in plants : the best is yet to come', Current Opinion in Plant Biology. Elsevier Ltd, 36, pp. 1–8. doi: 10.1016/j.pbi.2016.11.011.

Ricroch, A., Clairand, P. and Harwood, W. (2017) 'Use of CRISPR systems in plant genome editing : toward new opportunities in agriculture', Emerging Topics in Life Sciences, 1, pp. 169–182.

Schiml, S., Fauser, F. and Puchta, H. (2014) 'The CRISPR/Cas system can be used as nuclease for in planta gene targeting and as paired nickases for directed mutagenesis in Arabidopsis resulting in heritable progeny', Plant Journal, 80(6), pp. 1139–1150. doi: 10.1111/tpj.12704.

Shan-e-Ali Zaidi, S., Mahfouz, M. M. and Mansoor, S. (2017) 'CRISPR-Cpf1: A New Tool for Plant Genome Editing', Trends in Plant Science, 22(7), pp. 550–553. doi: 10.1016/j.tplants.2017.05.001.

Shukla, V. K. *et al.* (2009) 'Precise genome modification in the crop species Zea mays using zinc-finger nucleases', Nature. 459(7245), pp. 437–441. doi: 10.1038/nature07992.

SIGA and SIBV (2017) 'Position document on genome editing techniques applied to agriculture'. Available at: http://primaigeni.it/genome-editing-techniques.pdf.

Slaymaker. I.M *et al.* (2016) 'Rationally engineered Cas9 nucleases with improved specificity', Science, 351:6268, pp. 84-88

Sun, N. and Zhao, H. (2013) ‘Transcription Activator-Like Effector Nucleases (TALENs): A Highly Efficient and Versatile Tool for Genome Editing’, Biotechnology and Bioengineering, 110(7), pp. 1811–1821. doi: 10.1002/bit.24890.

Sun, Y. *et al.* (2016) ‘Engineering Herbicide-Resistant Rice Plants through CRISPR/Cas9-Mediated Homologous Recombination of Acetolactate Synthase’, Molecular Plant, 9(4), pp. 628–631. doi: 10.1016/j.molp.2016.01.001.

Svitashev, S. *et al.* (2015) ‘Targeted Mutagenesis, Precise Gene Editing, and Site-Specific Gene Insertion in Maize Using Cas9 and Guide RNA’, Plant Physiology, 169(2), pp. 931–945. doi: 10.1104/pp.15.00793.

Svitashev, S. *et al*. (2016) ‘Genome editing in maize directed by CRISPR–Cas9 ribonucleoprotein complexes’, Nature Communications. Nature Publishing Group, 7, p. 13274. doi: 10.1038/ncomms13274.

Swamynathan, B., Nadanakunjidam, S. and Ramamourti, A. (2010) ‘In-Vitro Plantlet Regeneration Through Somatic Embryogenesis in Solanum melongena (Thengaithittu Variety)’, Academic Journal of Plant Sciences, 3(2), pp. 64–70. Available at: http://idosi.org/ajps/3(2)10/3.pdf.

Tashkandi, M. *et al.* (2017) ‘Engineering resistance against Tomato yellow leaf curl virus via the CRISPR / Cas9 system in tomato’, bioRxiv, (doi: http://dx.doi.org/10.1101/237735).

de Toledo Thomazella, D. P. *et al.* (2016) ‘CRISPR-Cas9 mediated mutagenesis of a DMR6 ortholog in tomato confers broad-spectrum disease resistance’, bioRxiv, (July), p. 64824. doi: 10.1101/064824.

Trigiano, R. N. and Gray, D. J. (1999) Plant Tissue Culture Concepts and Laboratory Exercises. 2nd edn. CRC Press.

Wang, M. *et al.* (2017) ‘Gene targeting by homology-directed repair in rice using a Geminivirus-based CRISPR/Cas9 system’, MOLECULAR PLANT. Elsevier Inc., (doi: 10.1016/j.molp.2017.03.002). doi: 10.1016/j.molp.2017.03.002.

Wang, S. *et al.* (2015) ‘Efficient targeted mutagenesis in potato by the CRISPR / Cas9 system’, Plant Cell Reports. Springer Berlin Heidelberg, 34(9), pp. 1473–1476. doi: 10.1007/s00299-015-1816-7.

Yan, W., Smith, C. and Cheng, L. (2013) ‘Expanded activity of dimer nucleases by combining ZFN and TALEN for genome’, pp. 1–6. doi: 10.1038/srep02376.

Yu, Q. *et al.* (2017) ‘CRISPR / Cas9-induced Targeted Mutagenesis and Gene Replacement to Generate Long- shelf Life Tomato Lines’, Scientific Reports. Springer US, (September), pp. 1–9. doi: 10.1038/s41598-017-12262-1.

Zetsche, B. *et al.* (2015) ‘Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System’, Cell. Elsevier Inc., 163(3), pp. 759–771. doi: 10.1016/j.cell.2015.09.038.

Acknowledgements

There are many people I need to thank for all that I learned during the past three years, which I hope laid the foundations for many more years of learning to come. I have never been one for big declarations, but I feel I can say that I enrolled at university, almost ten years ago, because I wanted to be a researcher and because I loved genetics. Put it another way, because in ten years I would have wanted to be where I am now. I must say I am quite pleased with the way things worked out.

For this, my deepest thanks go to

Prof. Sergio Lanteri, for the opportunities and the trust he gave me, for which I have always felt truly grateful and lucky.

Prof. Cinzia Comino, for being the tutor of this thesis, for her availability and passion and for going at things - literally - at full speed. There are some aspects of my PhD work, even the writing of this thesis, I mostly owe to this quality of hers.

Dr. Andrea Moglia, for being a mentor, and such a good one. It is always a pleasure to work together and learn from him.

Prof. Alberto Acquadro, for his precious time and the most necessary bioinformatic support. One of the goals for the upcoming months is to get a little less bioinformatically dumb and bother him less.

Prof. Antonio Granell and **Dr. Diego Orzaez** for the indescribably warm welcome they gave me in their group at the IBMCP-CSIC in Valencia and for the way my work evolved during my stay there: it is the very core of this thesis, and will likely represent the basis of my work for the near future.

Joan Bernabé Orts, for teaching me and a whole bunch of Italian students GoldenBraid and CRISPR. Thank you on behalf of all of us.

The technicians, who teach you to behave and some of the most fundamental things you need in your everyday work. Thanks to **Silvia Presa**, **Anna Milani** and **Danila Valentino** for their experience, their time and their patience. And thanks to Danila for sharing her office, too!

Asun Fernández-del-Carmen, for knowing everything about everything and sharing it.

All the members of the Plant Genetics and Breeding group at the University of Turin, **Professors Ezio Portis** and **Lorenzo Barchi**, and my colleagues **Elisa**

Cerruti and **Dionís Borras** for the time spent together and their contribution to my work and my education.

Prof. Eduardo R. Bejarano and his group in Málaga, especially **Pepe Cana Quijada**, for yet another warm welcome, a very interesting glimpse at plant virus research and an apartment by the sea.

Prof. Jaime Prohens, **Dr. Mariola Plazas**, **Pietro Gramazio** and **Edgar García** at the COMAV in Valencia, for their unmatched kindness, for eggplant seeds and for regeneration hints.

The students who worked with me, for the passion they shared and the real and most welcome help they offered. Striving to be an acceptable teacher undoubtedly made me better at my work and helped me define the boundaries of the things I knew. I especially want to thank **Francesco Blasio**, for coffee, beer and food, and for almost never saying no to weekends in the lab.

All the people I worked with during my stay in Valencia: thanks to the Spanish who made me love their country, their culture and their language so much, and to the Italians who never made me miss my home. They are some of the most interesting and motivated people I have met and it was a real pleasure to share work and life with them. I do believe Valencia made me different, and I like myself better now.

Truth is, I'm not sure I would have been brave enough to embark on a PhD had I not known I could count on my family. Among the many things I owe my parents is the notion of what a purposeful life is, and I know I am lucky to have been given every means to pursue one.

Torino, February 2018