Sara Maria Francisco da Costa

Licenciada em Biologia

Towards modification of *Medicago truncatula* epigenome: genome editing with engineered nucleases

Dissertação para obtenção do Grau de Mestre em Genética Molecular e Biomedicina

Orientador: Pedro Fevereiro, Professor Doutor, Faculdade de Ciências da Universidade de Lisboa/ Instituto de Tecnologia Química e Biológica

Co-orientador: Catarina Pimentel, Doutora, Instituto de Tecnologia Química e Biológica

Júri:

Presidente: Professora Doutora Ilda Maria Barros dos Santos Gomes Sanches

Orientador: Professor Doutor Manuel Pedro Salema Fevereiro

Arguente: Doutora Ana Sofia Roldão Lopes Amaral Duque

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Setembro de 2015

Faculdade de Ciências e Tecnologia
Universidade Nova de Lisboa

Sara Maria Francisco da Costa

Licencidada em Biologia

Towards modification of *Medicago truncatula* epigenome: genome editing with engineered nucleases

Dissertação para obtenção do Grau de Mestre em Genética Molecular e Biomedicina

Orientador: Pedro Fevereiro, Professor Doutor, Faculdade de Ciências da Universidade de Lisboa/ Instituto de Tecnologia Química e Biológica

Co-orientador: Catarina Pimentel, Doutora, Instituto de Tecnologia Química e Biológica

Júri:

Presidente: Professora Doutora Ilda Maria Barros dos Santos Gomes Sanches

Orientador: Professor Doutor Manuel Pedro Salema Fevereiro

Arguente: Doutora Ana Sofia Roldão Lopes Amaral Duque

Setembro de 2015

Towards modification of *Medicago truncatula* epigenome: genome editing with engineered nucleases

Acknowledgements/ Agradecimentos

Para mim esta tese foi o despertar para uma nova realidade que me fez sair da zona de conforto. Gostaria de conseguir transmitir a todos os que estiveram presentes ao longo deste último ano o quanto me sinto grata. A falta de talento com as palavras não me permitirá fazer os devidos agradecimentos a todos os que, pelo mais simples gesto, me fizeram chegar até aqui. Assim, fica aqui o melhor que consigo transmitir.

Ao Professor Pedro Fevereiro, pela liberdade que me deu enquanto orientador por ter ouvido todas as ideias e teorias com um espirito aberto. Agradeço todo o apoio e paciência ao longo desta tese e todos os conselhos sobre a vida na ciência, incluindo os "abre-olhos" quando ficava frustrada por não ter os resultados esperados. E, também, por me ter feito ganhar os meus dias cada vez que ouvia "Bom trabalho, Sara".

À Doutora Catarina Pimentel, pelo apoio, paciência (sobretudo nesta fase final) e amizade que sempre demonstrou. Confesso que senti receio no início da tese quando soube que ia ter uma co-orientadora. Neste momento, sinto-me grata por a ter tido ao meu lado nesta tese. Obrigada por todos os "boa miúda" e os "seu verme horrível" que me incentivaram a dar o meu melhor. Agradeço as partilhas de histórias e vivências e todos os importantes conselhos sobre o futuro na ciência.

À Professora Claudina Rodrigues-Pousada por me ter recebido no seu laboratório e me ter feito sentir como sua aluna. E, sobretudo, por ter sido um exemplo de força, capacidade de luta e determinação ao longo deste ano.

A todos os membros do BCV/PlantX e do GMS pelo apoio e acompanhamento, em especial à Catarina Amaral, cuja ajuda foi mais do que preciosa.

À Soraia Caetano, pela amizade que foi crescendo e ganhando raízes ao longo deste ano, pelos desabafos e pelas gargalhadas nas alturas mais difíceis. Obrigada por me teres deixado assaltar o teu armário de soluções, extremamente organizado, e por festejares comigo todos os meus positivos. Foste e continuarás a ser um pilar fundamental.

À Carolina Gomes, por ter sido a melhor pessoa com quem podia partilhar esta nova etapa. Dizem que, quem ultrapassa o mestrado lado a lado, cria amizade para o resto da vida. Acredito profundamente que assim seja.

À Olívia Costa pelo apoio incondicional, por acreditar em mim mais do que eu própria e por ter estado ao meu lado sempre que precisava de ajuda. Obrigada pela amizade e carinho que demonstraste desde o dia em que me conheceste e por seres uma excelente ouvinte.

E porque não só de ciência foi feita esta tese, quero agradecer à Rita Caré e ao Zé Salvado por me terem apresentado ao seu maravilhoso mundo dos rabiscos e também à Diana Branco, à Sofia Duque, à Soraia Caetano, à Carolina Gomes e ao Filipe Oliveira por me terem convencido a experimentar o surf.

Ao Gonçalo, por me ter ouvido atentamente enquanto relatava pormenorizadamente as minhas novas descobertas e sobretudo pelo apoio incansável em todos os momentos menos bons. Obrigada por teres estado ao meu lado em Salamanca, pelas palavras de ânimo e boa disposição ao longo do penoso processo que foi a escrita da tese e por teres sempre acreditado em mim.

À minha mãe, por vibrar com os meus sucessos e por me animar nas alturas mais difíceis. Obrigada por todos os "Força filha, tu consegues" que se tornaram essenciais e muitas vezes me deram o alento necessário para continuar.

Ao meu pai, pelo apoio incondicional, por me ter incutido, desde pequena, a curiosidade científica, por todos os dias me ter ouvido atentamente sobre os sucessos e insucessos desta tese e por me fazer sempre ver o lado positivo de todas as novas situações. Nunca te conseguirei agradecer o suficiente por me teres ensinado a **nunca desistir** e seres um verdadeiro exemplo de lutador. A ti, Pai, devo ter chegado até aqui.

Resumo

A seca é um dos factores que mais afecta a produtividade agrícola. O aumento da população mundial aliado às alterações climáticas levou à diminuição das fontes de água, sendo um dos maiores desafios da agricultura moderna aumentar a produção alimentar humana e animal.

A metilação *de novo* do DNA é um processo regulado por pequenos RNAs de interferência (siRNAs) que estão envolvidos na resposta das plantas ao stress abiótico. Na leguminosa modelo *Medicago truncatula*, há evidências que sugerem uma ligação entre as vias de siRNAs e resposta ao deficit hídrico.

Para compreender o papel da metilação do DNA em condições de seca, foram estabelecidas metodologias bioinformáticas e moleculares que permitiram o desenho de sistemas como os 'Clustered regularly interspaced short palindromic repeats' (CRISPR)/Cas9 e a construção de TALENS ('transcription activator-like effector nucleases'), para introduzir lesões dirigidas em dois genes que codificam enzimas envolvidas na via de metilação do DNA mediada por RNA - Dicer-like 3 (MtDCL3) e RNA-Dependent RNA polymerase (MtRDR2).

A actividade dos TALENs foi avaliada através de um ensaio em leveduras utilizando duas estratégias diferentes: electroforese em gel de poliacrilamida (PAGE) e análise da conformação de polimorfismos em cadeia simples (SSCP). Neste ensaio, a transformação tripla das células de levedura revelou-se uma alternativa rápida e eficaz relativamente às estratégias que envolvem *mating*. A técnica de PAGE pode ser uma ferramenta valiosa para testar a eficiência dos TALENs *in vivo* desde que a sua actividade seja aumentada. A estratégia baseada na análise por SSCP provou ser ineficaz devido ao elevado número de falsos positivos.

Futuramente, os TALENs e o sistema CRISPR/Cas9, permitirão a introdução de mutações nos genes da DCL3 e RDR2 permitindo o estudo da relação entre a resistência ao stress e regulação epigenética mediada por siRNAs em *M. truncatula.*

Palavras-chave

Medicago truncatula, metilação do DNA, siRNAs, TALEN, sistema CRISPR/Cas9

Abstract

Periodic drought is the primary limitation of plant growth and crop yield. The rise of water demand caused by the increase in world population and climate change, leads to one of the biggest challenges of modern agriculture: to increase food and feed production.

De novo DNA methylation is a process regulated by small interfering RNA (siRNAs), which play a role in plant response and adaptation to abiotic stress. In the particular case of water deficit, growing evidences suggest a link between the siRNA pathways and drought response in the model legume *Medicago truncatula*.

As a first step to understand the role of DNA methylation under water stress, we have set up several bioinformatics and molecular methodologies allowing the design of Clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 systems and the assembly of TALENs (transcription activator-like effector nucleases), to target both dicer-like 3 (MtDCL3) and RNA-Dependent RNA polymerase (MtRDR2), enzymes of the RNA-directed DNA methylation pathway.

TALENs efficiency was evaluated prior to plant transformation by a yeast-based assay using two different strategies to test TALENs activity: Polyacrylamide gel electrophoresis (PAGE) and Single strand conformation polymorphisms (SSCP). In this assay, yeast cells triple transformation emerged as good and rapid alternative to laborious yeast mating strategies. PAGE analysis might be a valuable tool to test TALENs efficacy *in vivo* if we could increase TALENs activity. SSCP-based approach proved to be ineffective due to the generation of several false positives.

TALENs and CRISPR/Cas9 system constructed and designed in this work will in the future certainly enable the successful disruption of DCL3 and RDR2 genes and shed the light on the relationship between plant stress resistance and epigenetic regulation mediated by siRNAs in *M.truncatula*.

Keywords

Medicago truncatula, DNA methylation, siRNAs, TALEN, CRISPR/Cas9 system

General Contents

Figure Contents

Table Contents

List of abreviations

AAD – Acidic activation domain;

AGO – Argonaute;

CRISPR – Clustered regularly interspaced short palindromic repeats;

DCL – Dicer-like;

DNA – Deoxyribonucleic acid;

DSB – Double strand break;

dsRNA – Double-strand RNA;

hc-siRNAs – heterochromatic siRNAs;

HEN – Hua enhancer;

HMA – Heteroduplex mobility assay;

HR – Homologous recombination;

MTase – Methylase;

NHEJ – Non-homologous end joining;

NLS – Nuclear localization signals;

PAGE – Polyacrylamide gel electrophoresis;

PAM – Protospacer adjacent motif;

PCR – Polymerase chain reaction;

Pol – Polymerase;

PTGS – Post-transcriptional gene silencing;

RdDM – RNA directed DNA methylation;

RDR – RNA-dependent RNA polymerase;

RFLP – Restriction fragment length polymorphism;

RNA – Ribonucleic acid;

SAPTA – scoring algorithm for predicting TALE(N) activity;

sgRNA – Single-guided RNA;

siRNAs – Small interfering RNAs;

SSCP – single-strand conformation polymorphism;

ssRNA – Single-strand RNA;

TALE – Transcription activator-like effector;

TALEN – Transcription activator-like effector nuclease;

TE – Transposable elements;

TGS – Transcriptional gene silencing;

ZFN – Zinc-finger nuclease;

RVD - repeat variable di-residues.

1 Introduction

1.1 Agriculture and water deficit

Crop quality and yield are tremendously affected by drastic changes in the environment, namely by extreme weather conditions, water deficit and nutrient depletion. To meet the increasing world food demand, strategies to minimize climate change impact on crop production need urgent research attention. Those strategies require the development of new approaches for agricultural biotechnology, such as targeted plant genome modification (Pennisi, 2010).

Among abiotic stresses, periodic drought is the primary limitation of plant growth and crop yield. This fact together with the rise of water demand caused by both the increase in world population and climate change, is on the basis of one of the biggest challenges of modern agriculture: to increase food and feed production, maintaining or diminishing water consumption (Zhang *et al*., 2014). Farmers maximize crop yield by synchronizing plant development with rainy season, but yield increase under drought conditions can only be achieve by breeding and biotechnological advances (Marris, 2008). Strategies to improve plant drought resistance and promote sustainable use of water are therefore urgent and should integrate conventional breeding and novel biotechnological approaches (Chaves *et al*., 2003).

1.2 *Medicago truncatula* as a model plant

Legumes are vital for sustainable agriculture, mainly because they carry out symbiotic nitrogen fixation. Food legumes of global importance include bean (*Phaesolus vulgaris*), soybean (*Glycine max*) and pea (*Pisum sativum*). Alfafa (*Medicago sativa*) is the most important forage legume species and one of the most valuable crops in the USA, yet it is not ideal for genetics or genomics research because of its large, tetraploid genome and out-crossing nature (Graham & Vance, 2003). Its close relative M*edicago truncatula*, however, is self-fertile, has a relatively small diploid genome (2n=16) and is easy to transform, which makes it useful for both genetics and genomics studies (Trinh et al., 1998; Chabaud et al., 2003; Araújo et al., 2004). For these reasons, *M. truncatula* is considered a model plant within the legume family (Cook, 1999; Young *et al*., 2005). *M. truncatula* is also regarded as a drought-adapted species, as it occurs naturally in the arid and semi-arid environments of the Mediterranean (Nunes *et al*., 2008) and, as such it is concurrently an ideal experimental model to study the molecular and biochemical mechanisms conferring drought resistance in plants (Zhang *et al*., 2014).

In response to short-term drought, plants limit water loss by undergoing physiological changes, involving stomatal closure and leaf rolling (Brodribb & Holbrook, 2003; Lawlor & Tezara, 2009). Long-term adaptation require developmental changes, such as the expansion of root system to maximize soil water capture (Benjamin & Nielsen, 2006) and requires gene expression

reprogramming and metabolism modification (Yamaguchi-Shinozaki & Shinozaki, 2006; Talamè *et al*., 2007).

1.3 Epigenetic control of gene expression

Epigenetic modifications are mitotically and/or meiotically inheritable yet reversible changes in the DNA that affect gene expression. They do not involve changes in DNA sequence and are intimately associated with plant development (Steimer *et al*., 2004). Epigenetic modifications can be propagated across generations and genes housing those modifications are the so-called epialleles (Paszkowski & Grossniklaus, 2011). Plants differing in just one epiallele can be phenotypically different. Therefore, the presence of epialleles in various loci in the genome greatly contributes to phenotypic variability in plants.

Many epigenetic changes require the recognition of particular DNA or RNA sequences (Figure 1.1). At the DNA level, recognition of those sequences within promoter regions can lead to transcriptional gene silencing (TGS), which is associated with DNA methylation and/or chromatin modifications. At the RNA level, epigenetic modifications induce post-transcriptional gene silencing (PTGS), either by promoting sequence specific RNA degradation or by inhibiting translation (Steimer *et al*., 2004).

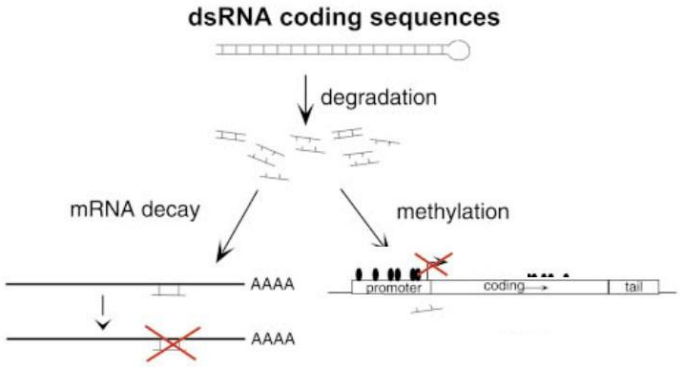

Figure 1.1 Models for PTGS and TGS. In PTGS (left side of the figure), dsRNAs are cleaved into small RNAs that pair with complementary mRNAs and mediate their degradation (indicated by the red cross). In TGS (right side of the figure), dsRNAs are cleaved into small RNAs that direct the methylation of corresponding DNA sequences (blackovals). This promoter methylation inhibits transcription, as indicated by the red cross. Adapted from Sijen *et al*., 2001.

Besides a prominent role in the regulation of gene expression during plant development (Zilberman *et al*., 2007), histone modifications and DNA methylation are also important for maintaining genomic integrity (Lauria & Rossi, 2011).

Two of the four DNA bases, adenine and cytosine, may undergo modifications, but adenine modifications have been poorly investigated (Kumar *et al*., 2013). In plant nuclear genomes, cytosine residues are methylated in CG, CHG and CHH (where H= A, T, C) contexts (Vanyushin & Ashapkin, 2011). In higher eukaryotes, cytosine-5 methylation was proposed to be a genome defence system established to protect genomes against endogenous transposable elements (TEs) and exogenous virus invasions (Zilberman *et al*., 2007). TEs, including DNA transposons and retrotransposons, are usually

heavily methylated and loss of methylation are often associated with their transcriptional activation and mobilization (Chan *et al*., 2005). DNA methylation also occurs within gene coding regions (Zhang *et al*., 2006), but its significance remains a matter of debate (Humbeck, 2013).

In plants, DNA methylation patterns result from two mechanisms: maintenance or *de novo* DNA methylation (Chan *et al*., 2005). Maintenance of methylation is the process by which pre-existing methylation patterns are preserved after DNA replication (Zhang *et al*., 2010b). The *de novo* methylation is a process in which cytosine residues are methylated, resulting in the formation of new methylation patterns. This process is also known as RNA-directed DNA methylation (RdDM) because putative sites for methylation are targeted via homology with small interfering RNAs (siRNAs) (Figure 1.2) (Matzke *et al*., 2009; Henderson & Jacobsen, 2007). RdDM is active in all tissues and cell types in vegetative and reproductive phases of plants (Kumar *et al*., 2013).

siRNAs are non coding small RNAs (typically 24 nucleotides (nt)), arising from non-coding regions of the genome (TEs and other repetitive elements),and therefore called heterochromatic siRNAs (hc-siRNAs), by means of the RNA interference pathway (Zhang *et al*., 2007). Polymerase IV (Pol IV) generates a single stranded RNA (ssRNA) transcript of the target region by using methylated DNA as template (Herr *et al*., 2005). Then, the RNA dependent RNA polymerase 2 (RDR2) interacts with Pol IV and converts ssRNA into double-stranded RNA (dsRNA) (Xie *et al*., 2004; Lu *et al*., 2006). Dicer-like 3 (DCL3) chops this dsRNA , originating 24 nt siRNAs (Xie *et al*., 2004; Henderson *et al*., 2006). siRNAs are stabilized by 3'-terminal ribose methylation mediated by the HUA enhancher 1 (HEN1) (Yu *et al*., 2005). Argonaute 4 (AGO4) or AGO 6 (Li *et al*., 2006) are then loaded with mature siRNAs and direct cytosine methylation by recruiting the *de novo* DNA Methylase (MTase) DRM2 (Matzke *et al*., 2009).

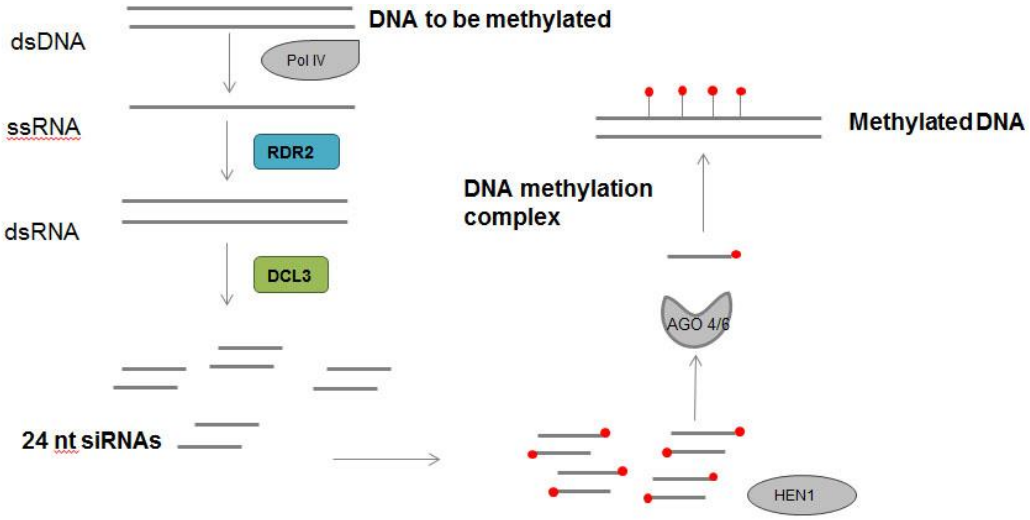

Figure 1.2 RdDM pathway. Transcription of non-coding regions (heterochromatic regions, repeat regions or transposons) by Pol IV results in the formation of ssRNA which is converted into dsRNA by the action of RDR2. This dsRNA is processed into predominantly 24-nt long siRNAs by DCL3. These siRNAs are methylated by HEN1 and associate with AGO4 or AGO6 to form a RNA-directed DNA methylation effector complex that lead to the methylation of target loci.

1.4 Epigenetic modulation of gene expression and stress response

Epigenetic changes are known to play a relevant role in plant stress response such as DNA damage, drought, high salinity and pathogen infection (Richards, 2011; Paszkowski & Grossniklaus, 2011). In fact, plants deficient in small RNA biogenesis exhibit impaired transgenerational response to stress (Boyko & Kovalchuk, 2010). DCL2, DCL3 and DCL4 proteins have some functional redundancy and are required for biogenesis of several classes of small RNAs (Boyko & Kovalchuk, 2010). In the *Arabidopsis dcl3* mutant, dsRNAs arising from the RdDM pathway can be processed by multiple DCLs, although preferentially processed by DCL3, in wild-type plants (Kasschau *et al*., 2007). The analysis of the progeny of *dcl2* and *dcl3* mutant plants revealed the absence of significant changes in DNA methylation in response to heat or UVC stress conditions as compared to the progeny of wild-type plants, possible due to the functional overlap between DCL proteins (Boyko & Kovalchuk, 2010). Recent observations in our laboratory further reinforce the relationship between epigenetic modification and stress response as it was shown that, in *M. truncatula*, DCL3 transcript levels increase under water deficit conditions (Capitão *et al*., 2011).

In the RdDM pathway, RDR2 is upstream DCL3 (Figure 1.2). RDR2 seems to have a crucial role in this pathway, since knock-out RDR2 mutants do not produce heterochromatic siRNAs (Daxinger *et al*., 2009). This finding suggests that no other RDR proteins (RDR1 or RDR6) are able to counteract the loss of RDR2, (Kasschau *et al*., 2007).

Interestingly, the analysis of small RNAs in the *Arabidopsis rdr2* mutant has shown that the small RNA population was highly enriched for other types of small RNAs (Lu *et al*., 2006), that were not however able to compensate RDR2 loss, leading to the up-regulation of many DNA TEs, retrotransposons and genes (Jia *et al*., 2009).

The fully understand (and control) of the mechanisms that rule epigenetic regulation of gene expression under a context of stress conditions, such as drought, would surely open new possibilities for sustained improvement of crops.

1.5 Genome editing using nucleases

Targeted genome engineering has an enormous impact on academia, society and industry, as this approach enables researchers to modify genomic loci of interest in a very precise manner.

A recent and powerful strategy for genome editing relies on engineered DNA endonucleases. These enzymes are able to introduce site-specific double-strand breaks (DSB) in a chosen locus, which after repair by non-homologous end joining (NHEJ) or homologous recombination (HR), generate the desired genetic modification (Figure 1.3). DNA repair through the NHEJ pathway may lead to frameshift mutations (generated by indels within the coding region of the target gene) and to the consequent expression of a truncated and/or non-functional protein (Figure 1.3 (A)). The HR-dependent DSB repair mechanism requires a homologous DNA segment as a template to copy the information across the break (Figure 1.3 (B)) (Cathomen & Joung, 2008). The choice of DSB repair

pathway inevitably impacts on the fidelity of the repair. Indeed, while HR is generally viewed as a conservative DSB repair pathway, NHEJ operates with poor fidelity and nucleotide deletions and/or insertions (indels) are frequently detected at repaired junctions (Decottignies, 2013).In plant species, DNA repair is typically effected by the NHEJ and not by HR pathway (Britt & May, 2003; Weinthal *et al*., 2010).

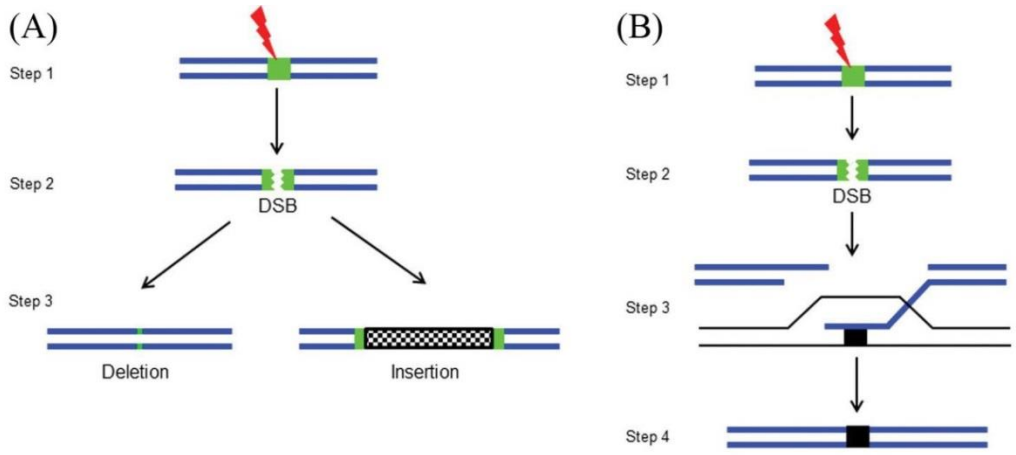

Figure 1.3 Models for NHEJ and HR mechanisms. (A) NHEJ mechanism. In step 1, DNA is damaged by an external agent, represented by the lightning symbol, and this result in a DSB (step2). In step 3, DNA is either removed by repair enzymes and religated resulting in a delection or DNA is inserted. (B) HR mechanism. In step 1 DNA is damaged by an external agent, represented by the lightning symbol, and this results in a DSB (step2). In step 3, damaged DNA binds to donor DNA at a point of homology and repair enzymes copy information from the donor DNA. Th break is repaired resulting in gene conversion in the chromossome as symbolized by the black box. Adapted from Wright *et al*., 2014.

The generation or targeted DSBs requires endonucleases with high fidelity of DNA recognition, in order to create site-specific cleavage. These enzymes must recognize a long DNA sequence with high specificity in order to avoid cytotoxic off-target DNA cleavage. In recent years, great efforts have been made to devise efficient and accessible genome editing nucleases, of which zing finger nucleases, transcription activator-like effector nucleases (TALENs), and the CRISPR/Cas9 system are good examples.

1.5.1 Zinc Finger Nucleases

Zinc finger nucleases (ZFNs) are artificial DNA nucleases constructed by fusing several zinc finger domains to the cleavage domain of the endonuclease FokI. FokI, a type IIS restriction enzyme (Pingoud *et al*., 2005), composed of two separate domains, an N-terminal DNA-binding domain and a C-terminal DNA-cleavage domain (Li *et al*., 1992). FokI exists as an inactive monomer in solution and becomes an active dimer upon binding to its target DNA. Therefore, two FokI molecules are involved in double-stranded DNA cleavage (Bitinaite *et al*., 1998). Custom-designed ZFN functions as a heterodimer that recognizes a two DNA sequences with a spacer between them (Figure 1.4) (Porteus & Carroll, 2005).

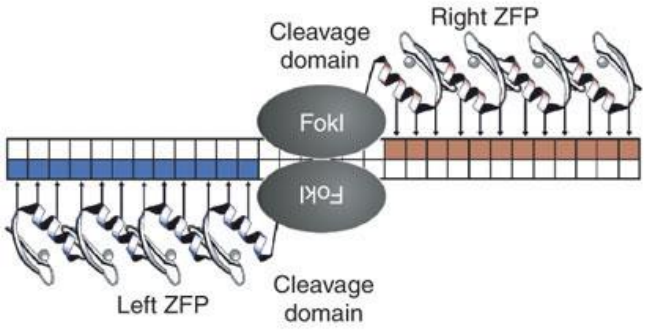

Figure 1.4 DNA recognition by zinc-finger nucleases (ZFNs). Scheme of a ZFN dimer bound to a typical, non-palindromic DNA target. Each ZFN consists of the cleavage domain of FokI fused to a zinc-finger domain protein (ZFP) that has been customized to specifically recognize either a left or right half-site (indicated by blue and red boxes), which are separated by a spacer of either 5 or 6 bp. Simultaneous binding by both ZFNs enables dimerization of the FokI nuclease domain and DNA cleavage. Adapted from Miller *et al*., 2007.

Because several zinc finger domains can tolerate mismatched DNA bases, off-target associated with ZFN usage has been reported (Gabriel *et al*., 2011; Pattanayak *et al*., 2012). Recognition of non-specific sites by ZFNs (or by any other nuclease used in genome editing approaches) generate undesired genomic DSB and might lead to unexpected mutations and chromosomal aberrations (Pattanayak *et al*., 2012; Radecke *et al*., 2010). ZFN strict DNA binding requirements greatly shortens the universe of genes that may be targeted (Ramirez *et al*., 2008). Despite these limitations, which hinder their wider application, ZFNs have been used for targeted genome editing in various organism like *Arabidopsis thaliana* (Zhang *et al*., 2010a), *Nicotiana tabacum* (Townsend *et al*., 2009; Marton *et al*., 2010), *Petunia hybrida* (Marton *et al*., 2010), *Glycine max* (Curtin *et al*., 2011), *Caenorhabditis elegans* (Wood *et al*., 2012) and medaka (Chen *et al*., 2012).

1.5.2 Transcription activator-like effectors nucleases

TALENs have rapidly emerged as an alternative genome editing tool to ZFNs. Similar to ZFNs, TALENs use the non-specific FokI domain as the DNA cleavage module. In TALENs, however, this domain is fused with a TALE (Transcription activator-like effector) protein (Christian *et al*., 2010; Li *et al*., 2011).

TALEs are a class of DNA binding proteins produced by plant pathogenic bacteria, *Xanthomonas sp*. The native function of these proteins is to directly modulate host gene expression to allow pathogen survival therein (Bonas *et al*., 1989). Upon delivery into host cells via the bacterial type III secretion system, TALEs enter the nucleus, bind specific sequences located within the host promoter regions and activate the transcription of the corresponding genes (Van Den Ackerveken *et al*., 1996). TALEs have in common an N-terminus required for type III secretion, a C-terminus containing nuclear localization signals (NLS) (Van Den Ackerveken *et al*., 1996) and an acidic activation domain (AAD) typical of transcription factors (Zhu *et al*., 1998).

The DNA specificity of TALEs is determined by a central domain composed of a tandem array of repeated segments, each consisting of 33-35 amino acid residues, followed by a single

truncated repeat of 20 amino acids. Each repeat is nearly identical except for two variable amino acid residues located at positions 12 and 13, called repeat variable di-residues (RVD) (Bonas *et al*., 1989). The RVD specifies the target sequence: each RVD recognizes one different nucleotide, and each of the four DNA bases is recognized by different RVDs (Figure 1.5 (A)). The knowledge of this correspondence (the code RVD-base) enabled the prediction of TALE targets. TALENs target sites are always preceded by a thymine (T) that is required for TALE activity (Moscou & Bogdanove, 2009; Boch *et al*., 2009).

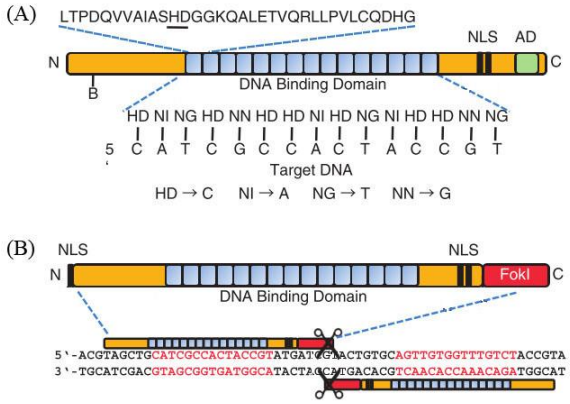

Figure 1.5 TAL effector and TALEN structure. (A) TALE structure. A consensus repeat sequence is shown with the repeat-variable di-residue (RVD) underlined. The sequence of RVDs determines the target nucleotide sequence. (B) TALEN structure. Two TALENs are required to bind the target site to enable FokI to dimerize and cleave DNA. NLS, nuclear localization signal(s); AD, transcriptional activation domain. Adapted from Cermak *et al*., 2011.

As mentioned before, since FokI cleaves as a dimer, TALENs function in pairs (left and right TALEN). This pairs are arranged in an opposing orientation on opposite sides of doubled-strand DNA (Figure 1.5 (B)) with an optimal spacer sequence between them (usually of 15 – 30 nt) (Cermak *et al*., 2011). When left and right TALENs bind to DNA, an active FokI domain is reconstituted and DNA is cleaved within the spacer, generating a DSB.

Unlike ZFNs, TALENs can be easily and rapidly constructed to target essentially any DNA sequence. In addition, TALENs exhibit significantly reduced off-target effects and cytotoxicity compared with ZFNs, making them an efficient genome editing tool (Mussolino *et al*., 2011). For this reasons, TALEN mediated Genome Editing was listed as one of the ten breakthroughs of 2012 by the journal Science (Alberts, 2012).

TALENs usage in plants was first reported in rice (Li *et al*., 2012). Since then, TALEN technology has been used in dicot species such as tobacco and Arabidopsis and in monocot species such as *Brachypodium*, barley and maize (Mahfouz *et al*., 2011; Christian *et al*., 2013; Shan *et al*., 2013a; Wendt *et al*., 2013; Liang *et al*., 2014).

1.5.3 CRISPR/ Cas9 system

Recently, Clustered regularly interspaced short palindromic repeats (CRISPR)-mediated DNA cleavage has been applied for genome editing and in 2013 was considered a major breakthrough in the field, by the journal *Science* (Coontz, 2013).

CRISPRs are loci encompassing several short repeats functioning as an adaptive microbial immune system, that prevents the cellular entry of unwanted virus and foreign DNA molecules (Barrangou *et al*., 2007). Several types of CRISPR-associated proteins (Cas) are encoded by *cas* genes located in the vicinity of CRISPRs. Cas proteins are required for the multistep defense against intruder genetic element. Their number, identity, and the corresponding operon organization appear to be extremely variable (Makarova *et al*., 2011). The CRISPR system has three basic components, the Cas endonuclease and two RNA molecules that guide Cas protein to the target sequence. By generating a single-guide RNA (sgRNA) that combines the function of the two RNA molecules in a chimeric molecule, CRISPR/Cas system became a useful tool for gene editing and silencing (Jinek *et al*., 2012). The sgRNA is complementary to a specific region, called protospacer adjacent motif (PAM), that recruits Cas protein (Wiedenheft *et al*., 2012). The type II CRISPR/Cas system most widely used for gene editing uses the monomeric enzyme Cas9 (Jinek *et al*., 2012; Cong *et al*., 2013) and has the advantage of possessing a PAM recognition sequence of only two nucleotides in length (GG) (Jinek *et al*., 2012). This allows the generation of a numerous 'guide' RNAs for any particular gene due to GG sequence abundance.

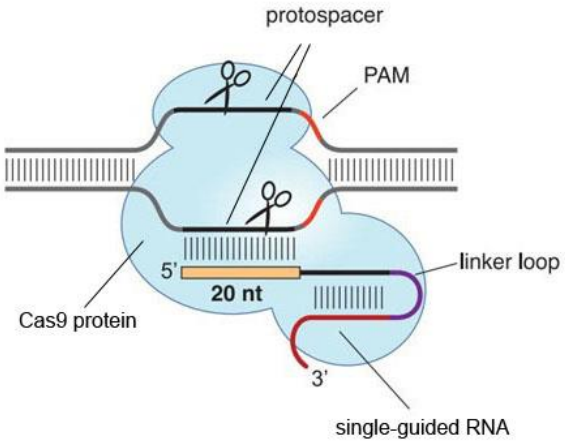

Figure 1.6 CRISPR/Cas9 system. The system consists of a single-guided RNA which guides the Cas9 nuclease to the genomic target site. Genomic target site has usually 20 base pairs (bp) of homology with the gRNA and a PAM sequence. Cleavage of both strands of target DNA occurs near the PAM. Adapted from Jinek *et al*., 2012.

Cas9 nuclease is functional as a transgene in most, if not all, cells in which it has been tested: rice (Shan, *et al*., 2013b; Mao *et al*., 2013; Xie & Yang, 2013), wheat (Shan, *et al*., 2013b), Arabidopsis (Mao *et al*., 2013; Li *et al*., 2013) and tobacco (Li *et al*., 2013; Nekrasov *et al*., 2013).

The great advantage of genome editing using CRISPR/Cas system relies on its simplicity - cells need only to be transformed with one plasmid containing the Cas9 gene and a gene encoding a sgRNA (Jiang *et al*., 2013). Moreover this technology allows targeting of several genes at the same time (Cong *et al*., 2013).

1.6 Objectives

The current global socio-economic and environmental scenario places the development of novel strategies to increase food production, maintaining or diminishing water consumption, at forefront of modern agriculture research.

In this work we intend to set up the tools that in the future will certainly enable us to study the relationship between plant stress resistance and epigenetic regulation mediated by siRNAs in the legume plant model *Medicago truncatula*. To this end, we will focus on DCL3 and RDR2, enzymes of the RdDm pathway. Growing evidence suggests the involvement of DCL3 in drought response (Capitão *et al*., 2011). As RDR2 is a component of the RdDm pathway upstream DCL3 (Xie *et al*., 2004) that substantially impacts on hc-siRNAs formation, it would be also interesting to study its relevance on stress response.

The generation of *rdr2* and *dcl3 Medicago truncatula* mutants, will allow the future identification of epialleles related to drought resistance and may provide novel strategies to enhance legume productivity.

In this context, the main objectives of this work are to:

1) Design and construct TALENs targeting two crucial enzymes of the RdDM pathway - DCL3 and RDR2;

2) Test those TALENs effectiveness;

3) Design CRISPRs/Cas systems to target those genes.

2 Materials and Methods

2.1 TALENs construction

2.1.1 Bioinformatic analysis

Medicago truncatula gene sequences were identified using John Craig Venter Institute (JCVI) database for *Medicago* genome (v.4.0). Alignments were performed using DIALIGN (Morgenstern, 2004), a web tool available at http://bibiserv.techfak.uni-bielefeld.de/dialign/submission.html, using the following parameters: Sequence mode = DNA w/o translation; Threshold= 0; Regions max. sim.= 9. The software used to design TALENs is available online – SAPTA: Scoring Algorithm for Predicting TALE(N) Activity (http://baolab.bme.gatech.edu/Research/BioinformaticTools/TAL_targeter.html). DNA target sequences (RDR2 and DCL3) were then scanned for suitable TALEN binding sites based on TALEN design guidelines described by (Lin *et al*., 2014). The spacer length was set to 14-19 bp and TAL arrays lengths were adjusted to 15-17. The SAPTA output data consists of a Table containing the sequence to be recognized by the left and right TALENs (target sites). TALEN pairs are ranking according a score (composite score) that estimates TALEN pair activity (high scores indicate higher activity). Since naturally occurring TAL effector recognition sites are preceded by a T, which is required for TAL effector activity, only TALEN recognition sites preceded by a T are included.

2.1.2 TALEN assembly via Golden Gate

TALEN repeat arrays recognizing the defined target sites (see above) were constructed using the Golden Gate kit (Addgene) according to the assembly method described by (Cermak *et al*., 2011) TALENs were assembled in two sequential steps: (i) assembly of repeat modules into intermediary arrays of 1-10 repeats and (ii) cloning of the intermediary arrays into a backbone to make the final construct. TALENs were assembled following the Golden Gate Cloning strategy, a recently developed method of assembling multiple DNA fragments in an ordered fashion in a single reaction (Engler *et al*., 2008). The Golden Gate method uses Type IIS restriction endonucleases, which cleave outside their recognition sites to create unique 4bp overhangs (sticky ends). Cloning is facilitated by digesting and ligating in the same reaction mixture because correct assembly eliminates the enzyme recognition site (Cermak *et al*., 2011).

2.1.2.1 Assembly of intermediate arrays

Plasmids encoding RVDs 1-10 were cloned into array plasmid pFUS_A. Modules for RVDs 11-16 were selected and cloned into pFUS_B plasmid. There are 10 different pFUS_B plasmids and the correct pFUS_B should be selected according to the number of modules to be assembled. Modules and array plasmids (150 ng of each) were submitted to digestion and ligation in a single 20µL reaction containing 10 U of BsaI (10 000 U/mL, New England Biolabs ®), 400 U of T4 DNA ligase (400 000

U/mL, New England Biolabs®) and 1x T4 DNA ligase buffer (New England Biolabs®). The reaction was incubated in a thermocycler for 10 cycles of 5 min at 37°C and 10 min at 16°C, then heated to 50°C for 5 min and to 80°C for 5 min. Then, 6.25 mM ATP (Epicentre ®) and 100 U of Plasmid-Safe™ ATP-Dependent DNase (10 000 U/mL, Epicentre®) were added to the reaction. The mixture was incubated at 37°C for 1h. Treatment with Plasmid Safe DNase is crucial to prevent linear DNA fragments, including partial arrays, from recombining.

2.1.2.2 Transformation of Escherichia coli

The ligation mixture was used to transform *Escherichia coli* XL1 Blue competent cells. Approximately 200 µL of chemically competent cells, prepared according to Inoue *et al*. 1990, were mixed with 10 µL of Golden Gate mix reaction and kept on ice for 30 min. Then, the competent/DNA mixture was heated at 42°C in a water bath for 45 seconds and quickly chilled in ice for 2 minutes. After SOC medium (Supplementary Table S1) (800 µL) addition, the mixture was incubated in a 37°C oven, for 1 hour. Samples were centrifuged 5 minutes at 5000 rpm, ressuspended in 250 µL of SOC medium and then plated on LB Agar (Supplementary Table S1) containing 50µg/mL spectinomycin (Nzytech), 20 mg/mL X-gal (Nzytech) and 0,1 M IPTG (Nzytech) and incubated at 37°C overnight, to allow blue/white screening of recombinants.

2.1.2.3 Selection of positive clones

The selection of positive clones (white colonies) was performed by colony PCR. A small amount of colony mass was introduced, with the help of sterile toothpick, into a PCR tube, paced in a microwave for 5 minutes (900W) to break cell wall and immediately chilled on ice. Afterwards, 0.4 µM primer – pCR8_F1 and pCR8_R1 (Supplementary Table S2) – provided by Golden Gate TALEN kit (Addgene), 1.5 mM $MgCl_2$ (Promega), 0.3 mM dNTPs (Nzytech), 1X GoTaq® reaction buffer (Promega), 0.6 U GoTaq® polymerase (Promega) and miliQ H_2O (up to 20 µL). PCR reactions were carried out in a thermocycler using the following temperature parameters: 2 minutes at 98°C (1 cycle) followed by 35 cycles of 45 seconds at 95°C, 30 seconds at 55°C (annealing step) and 105 (pFUS_B) or 70 (pFUS_B) seconds at 72°C (extension step).

PCR products were loaded onto a 1% or 2% agarose (SeaKem®) gel (pFUS_A and pFUS_B reactions, respectively) stained with ethidium bromide (EtBr) (Sigma-Aldrich) (0.05µL/mL) and gel electrophoresis was carried out (5V per cm). DNA fragments were visualized using Gel Doc XR+ system (Biorad). Agarose was dissolved in TAE 1X (Supplementary Table S3).

After positive clone identification, cell cultures carrying those clones were grown overnight in LB liquid medium (Supplementary Table S1) supplemented with 50µg/mL spectinomycin (Sigma-Aldrich®). Plasmid DNA was extracted using ZR Plasmid Miniprep™ - Classic (Zimo Research), according to manufacturer's instructions, and then quantified by Epoch Microplate Spectrophotometer (Biotek ®). Plasmid DNAs were next sequenced (at StabVida®) using 10 µM pCR8_F1 and 10 µM pCR8_R1 (Supplementary Table S2) primers provided by Golden Gate TALEN Kit. Sequencing

results were analysed through a quick alignment tool available at http://baolab.bme.gatech.edu/Research/BioinformaticTools/assembleTALSequences.html, which allowed the confirmation of the correct position of RVDs in the array.

2.1.2.4 Assembly and selection of final arrays

Intermediary arrays (into pFUS_A and pFUS_B), along with a last repeat were cloned into the backbone plasmids pTAL3 (left TALEN) and pTAL4 (right TALEN), which contain the nuclease FokI domain (Cermak *et al*., 2011). Intermediary arrays were also cloned into pZHY500 (left TALEN) and pZHY501 (right TALEN) backbones, which are truncated versions of pTAL3 and pTAL4 exhibiting a higher nuclease activity (Zhang *et al*., 2013). pZHY500 and pZHY501 were a kind gift of Dr. Daniel Voytas. Reactions were prepared as follows: 150 ng each of pFUS_A and pFUS_B plasmids containing the intermediary repeat arrays, 75 ng of pTAL3 (or pTAL4, pZHY500, pZHY501), 150 ng of last repeat plasmid, 10 U of Esp3I (10000 U/mL, ThermoScientific), and 400 U of T4 DNA ligase (400 000 U/mL, New England Biolabs®) in 1X T4 DNA ligase buffer (New England Biolabs®). Reactions were placed in thermocycler as in section 2.1.2.1. Plasmid Safe DNase treatment was omitted at this stage. The ligation mixture was used to transform *E.coli* as described in section 2.1.2.2, cultures were plated in ampicillin (100µg/mL) LB Agar plates (Supplementary Table S1) for selection of the transformants.

Several colonies were next chosen for plasmid DNA extraction. Plasmid DNA was extracted using ZR Plasmid Miniprep™ - Classic (Zimo Research), according to manufacturer's instructions. Selection of positive clones was performed by PCR. Correct constructs were expected to have around 2kb in length. A 20 µL PCR reaction was prepared by mixing: 0.4 µM primer – TAL_F1 and TAL_R2 (Supplementary Table S2) – provided with Golden Gate TALEN kit (Addgene), 1.5 mM MgCl2 (Promega), 0.3 mM dNTPs (Nzytech), 1X GoTaq® reaction buffer (Promega), 0.6 U GoTaq® polymerase (Promega) and miliQ H_2O (up to 20µL). The reaction mixture was placed for 2 minutes at 98°C and then subjected to 35 cycles of: 45 seconds at 95°C, 30 seconds at 55°C and 3 minutes at 72°C. The fragments amplified were then analysed by gel electrophoresis using a 1% agarose (SeaKem®) gel, prepared as described above (section 2.1.2.3). Putative positive clones were next confirmed by sequencing using 10 µM SeqTALEN 5-1 and 10 µM TAL_R2 primers (Supplementary Table S2) provided by Golden Gate TALEN Kit. Sequencing results were analysed as described in section 2.1.2.3.

2.2 TALEN's functionality: Yeast based-assay

2.2.1 *Medicago truncatula* DNA extraction

DNA was extracted from the leaves of M9-10a line of *M. truncatula* Gaertn cv. Jemalong developed by Neves *et al*., 1999 and Santos & Fevereiro, 2002, which has the capacity to form

somatic embryos through a regeneration process described previously (Araújo *et al*., 2004; Santos & Fevereiro, 2002). Plants were gown in vitro under the following regimes: 16h of light/8h of dark in growth cabinets maintained at 24 °C during the day and at 22°during the night with 50 % to 60 % humidity and a light intensity of 100 µmol $m^{-2}s^{-1}$. Leaf DNA extraction was performed according to Santos & Fevereiro 2002.

2.2.2 Constructs

The left TALEN of DCL3 and RDR2 were cloned into both pTAL3 and pZHY500, two Golden Gate compatible yeast expression vectors, containing *HIS3* as a (yeast) selectable marker. The right TALENs were cloned into pTAL4 and pZHY501, also two compatible yeast expression vectors, containing *LEU2* as a selectable marker. TALEN target sites for DCL3 or RDR2 were cloned into pRS416 (Sikorski & Hieter, 1989), a yeast vector containing *URA3* as a selectable marker. TALENs targets sites were amplified by PCR using as a template *M. truncatula* genomic DNA, previously digest with a rare cut restriction enzyme, and using the polymerase Phusion®, a proofreading enzyme, that originates blunt-ends fragments.

Each PCR reaction contained 50ng of digested DNA template, 0.6 µM of primer (DCL3_Fwd and DCL3_Rev or RDR2_Fwd and RDR2_Rev) (Supplementary Table S2) (StabVida), 0.3 mM dNTPs (Nzytech), 1X Phusion ® CG buffer, 0.4 U of Phusion® polymerase (New England Biolabs) and miliQ H_2O (up to 20 ul). The reaction mixture was placed in a thermocycler and subjected to the following temperature cycling parameters: 1 cycle at 98°C for 2 minutes and 35 cycles at 95°C for 45 seconds, 58°C for 30 seconds and at 72°C for 15 seconds. Amplified fragments were electrophoretically separated in a 2% agarose gel (SeaKem®) and recovered using Zymoclean™ Gel DNA Recovery Kit (Zymo Research) according to manufacturer's instructions. Purified fragments were cloned into pRS416 previously digested with SmaI (1U/µL) (Fermentas). The ligation step was performed overnight at 16ºC in a 20 µL reaction volume using 5 ng of pRS416, 100 ng of insert (DCL3 target or RDR2 target), 400 U of T4 DNA ligase (400 000 U/mL, New England Biolabs®) and 1X T4 DNA ligase buffer (New England Biolabs®).

2.2.3 Yeast competent cells

Yeast colonies from BY4742 (MATα *his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*) strain were grown in YPD medium (Supplementary table S1) overnight at 30 °C with shaking at 200 rpm. This culture was used to inoculate 20 mL of YPD to an OD_{600} of 0.2. Cells were grown in the same conditions until the culture reached an OD_{600} between 0.7-1. Cells were harvested by centrifugation at 4000 rpm for 2 min and washed with 0.5 volumes of sterile H_2O. The centrifugation step was repeated, cells were ressuspended in 0.01 volumes of sterile H_2O and transferred to a suitable centrifuge tube and pellet at 5000 rpm for 1 min. Cell pellet was washed in 0.01 volumes of LiAc/TE solution (Supplementary Table S3) and centrifuged again 5000 rpm for 1 min. Cells were finally ressuspended in 400 µL of LiAc/TE solution and stored at 4°C.

2.2.4 Yeast cells triple transformation

A volume of 50 µL of yeast competent cells was mixed with 1 mg/mL salmon sperm denature DNA (Sigma-Aldrich), 300 ng of plasmid DNA (pTAL3 (or PZHY500), pTAL4 (or PZHY501), pRS416 with DCL3 target and pRS416 with RDR2 target) and 300µL of transformation solution (Supplementary Table S3) were added. Cells were incubated at 30°C, for 30 minutes, 200 rpm and next heat-shocked at 42°C for 15 minutes. After addition of 800µL of sterile water, cells were centrifuged at 5000 rpm for 1 minute and plated onto selective agar plates (Synthetic drop out (SD) agar medium (Supplementary Table S1) supplemented with histidine (0.2mL/L) (Sigma-Aldrich) and leucine (0.2mL/L) (Sigma-Aldrich)). Growth was recorded after 3 days at 30ºC.

The above transformed yeast cells were made competent for transformation according to the above described protocol (section 2.2.3). SD medium supplemented with histidine and leucine was used instead of YPD. Competent cells were double transformed with the left and right TALENs and plated in SD plates without histidine, uracil and leucine.

2.2.5 Polyacrylamide gel electrophoresis (PAGE) analysis

Colony PCR of the transformed yeast cells was performed as described in section 2.1.2.3 using the primers (DCL3_Fwd, DCL3_Rev, RDR2_Fwd and RDR2_Rev) listed in Supplementary Table S2 and an extension time of 30 seconds. This PCR was followed by a second (nested) PCR using primers (DCL3In_Fwd, DCL3In_Rev, RDR2In_Fwd and RDR2In_Rev) listed in Supplementary Table S2 and an extension step of 15 seconds. The final fragments had an expected size of 60 bp.

Two types of polyacrylamide gels were assembled: 12% polyacrylamide gel with Acrylamide (Fluka)/Bisacrylamide (Sigma-Aldrich) ratio of 19:1 according to Biorad instructions; and a 20% polyacrylamide gel with Acrylamide/bisacrylamide (Roth) ratio of 29:1 according to Roth instructions. These types of gels have a separation range between 20-150 bp and 6-100 bp, respectively. PCR products (10 µL) were loaded onto polyacrylamide gels that run at 100 V during 3 hours in TBE 0.5x (Supplementary Table S3). DNA staining was performed by soaking gels into a TBE solution containing EtBr (Sigma-Aldrich) (0.1µL/mL). DNA fragments were visualized using Gel Doc XR+ system (Biorad).

2.2.6 Single Strand Conformation Polymorphisms (SSCP) analysis

Nested PCR products (12 µL) were mixed with 36 µL of formamide dye (Supplementary Table S3) and denatured at 95°C for 10 minutes. After denaturation, samples were immediately kept in -20°C deep freeze for 10 minutes.

Samples were analysed in a 12% polyacrylamide gel 49:1 Acrylamide (Fluka)-bisacrylamide (Sigma-Aldrich), TBE 0.5 X, 8,7% glycerol (PanReac), 0.04% APS (Sigma-Aldrich), 0.16% TEMED (Sigma-Aldrich) and MiliQ water up to 10 mL. Gels were previously pre-run in TBE 0.5X at 100V for 30 minutes. Gel wells were next flushed using the buffer and samples were loaded. Electrophoresis

was performed at 100V for 2 hours. After electrophoresis, gel was stained with silver as described by Byun *et al*. 2009 to visualize band patterns.

Briefly, after electrophoresis, SSCP gels are fixed and stained in a stain solution for 20 min, then, the gels are rinsed with distilled water once and are developed with a develop solution (Supplementary Table S3) until dark staining bands appear on the yellow background of the SSCP gels (varies between 5 and 10 min). Development is then stopped with a stop solution for 1 minute.

Putative positive clones (PCR products exhibiting a different gel migration pattern) were sequenced using M13 Forward primer (Supplementary Table S2) to detect possible indel mutations. Sequences were compared using Blast2Seq, a web tool available at http://blast.ncbi.nlm.nih.gov/Blast.cgi, to detect differences in the sequence comprising the spacer region of each TALEN target.

2.3 CRISPR-Cas9 system target sites identification

Sequences from DCL3 and RDR2 genes were analysed by means of CHOPCHOP (Montague *et al*., 2014), an online tool (https://chopchop.rc.fas.harvard.edu/) that designs potential CRISPRs target sequences. This tool efficiently predicts off-target binding of single-guide RNAs in several species, including plants, but not in *M. truncatula* (genome not available in CHOPCHOP). As such, to avoid the possibility of off-target effects CRISPR sequences designed by CHOPCHOP were blasted against the *M. truncatula* genome using http://blast.ncbi.nlm.nih.gov/Blast.cgi and checked in DIALIGN alignments used for TALENs design (see section 2.1.1). CRISPRs exhibiting homology with other genomic regions were excluded.

3 Results

3.1 Design of DCL3 and RDR2 TALENs

3.1.1 Identifying the region to be targeted

In endonuclease-based genome editing approaches, recognition of non-specific sites is often a problem, which lead to undesired genomic DNA breaks that might result in mutations and chromosomal aberrations (Pattanayak *et al*., 2012). As such, the precise identification of the DNA sequence to be targeted is one the most important steps when constructing TALENs.

The genome of *Medicago truncatula* is being sequenced and genomic data is available for the whole scientific community at John Craig Venter Institute (JCVI) homepage (Version 4.0 - http://www.jcvi.org/medicago/).

We started by searching the genome of *M. truncatula* for sequences annotated as 'dicer'. This search retrieved 10 possible dicer-sequences within *M. truncatula* genome (Table 3.1). Dicer-like 3 (MtDCL3) appeared associated with the Locus ID Medtr3g105390. Likewise, sequences annotated as 'RNA dependent RNA polymerase' were searched within *M. truncatula* genome, which resulted in seven RNA-dependent RNA polymerase family proteins (Table 3.2). RNA-dependent RNA polymerase 2 (MtRDR2) appeared associated with the locus ID Medtr4g106660.

Table 3.1: *M. truncatula* Dicer-like proteins

Locus ID	**Product name**	***A. thaliana* Ortholog**
Medtr7g118350	endoribonuclease dicer-like protein	dicer-like 1
Medtr2g030490	endoribonuclease dicer-like protein	dicer-like 2
Medtr3g105390	endoribonuclease dicer-like protein	dicer-like 3
Medtr4g116860	endoribonuclease dicer-like protein	dicer-like 4
Medtr1g063400	dicer-like protein, putative	n.a.
Medtr1g063410	endoribonuclease dicer-like protein	n.a
Medtr8g077780	endoribonuclease dicer-like protein	n.a
Medtr1g060740	endoribonuclease dicer-like protein	n.a
Medtr8g069975	endoribonuclease dicer-like protein	n.a
Medtr8g077770	dicer-like protein, putative	n.a

n.a. not assigned

Table 3.2: *M. truncatula* RNA-dependent polymerase family proteins

Locus ID	**Product name**	***A.thaliana* Ortholog**
Medtr4g106660	RNA-dependent RNA polymerase family protein	RNA-dependent RNA polymerase 2
Medtr3g107390	RNA-dependent RNA polymerase family protein	RNA-dependent RNA polymerase 6
Medtr8g064670	RNA-dependent RNA polymerase family protein	n.a.
Medtr6g088660	RNA-dependent RNA polymerase family protein	n.a.
Medtr2g059620	RNA-dependent RNA polymerase family protein	n.a.
Medtr1g032710	RNA-dependent RNA polymerase family protein	n.a.
Medtr0212s0050	mitovirus RNA-dependent RNA polymerase	n.a.

n.a. not assigned

We next, aligned MtDCL3 and MtRDR2, hereafter designated as DCL3 and RDR2, respectively, with each of the retrieved sequences (Tables 3.1 and 3.2) using DIALIGN (Morgenstern, 2004). DIALIGN is a software tool that combines local and global alignment features without gap penalties and is far more versatile than traditional alignment approaches. After DIALIGN analysis, the first 500 bp of DCL3 and RDR2 genes (corresponding to the first exon of the genes) emerged as the suitable regions for TALENs targeting. In fact, these regions exhibit low similarity with other Dicer-like or RDR-like genes present in *Medicago* genome (alignment not shown). Small stretches of sequence homology, however, were found within these 500 bp regions (Tables 3.3 and 3.4). These sequences were therefore taken into account for TALENs design.

Table 3.3 Local homologies between DCL3 and other dicer proteins

Locus	**Homologies with Medtr3g105390**
Medtr1g060740	156nt-189nt/219nt-258nt/310nt-348nt/373nt-413nt
Medtr1g063400	No Homologies
Medtr1g063410	1nt-8nt/122nt-137nt/203nt/240nt
Medtr2g030490	161nt-187nt/218nt-251nt/309nt-348nt/455nt-485nt
Medtr4g116860	144nt-177nt/224nt-261nt/311nt-349nt/370nt-399nt/433nt-451nt
Medtr7g118350	30nt-59nt/167nt-207nt/231nt-267nt/310nt-348nt/453nt-472nt
Medtr8g069975	155nt-194nt/221nt-259nt/310nt-348nt/454nt-477nt
Medtr8g077770	72nt-107nt/134nt-144nt/154nt-174nt
Medtr8g077780	No homologies

Table 3.4 Local homologies between RDR2 and other RNA-dependent RNA polymerases

Locus	Homologies with Medtr4g106660
Medtr0212s0050	10nt-29nt/148nt-160nt
Medtr1g032710	No Homologies
Medtr2g059620	93nt-132nt/175nt-184nt
Medtr3g107390	No Homologies
Medtr6g088660	27nt-66nt/235nt-259nt
Medtr8g064670	397nt-419nt

3.1.2 Choosing the best TALENs

The 500 bp region of DCL3 and RDR2 were next screened for possible TALEN target sites, using the design tool SAPTA (Lin *et al*., 2014), available online. The output of this analysis consists of a scored list of target sequences, with high scores indicating higher TALEN activity (Table 3.5 for DCL3 and Table 3.6 for RDR2). For DCL3, the four best-scored TALENs were excluded due to homologies with the above referred sequence stretches (Table 3.3). The TALEN pair positioned in the ranking position 5 was then chosen for further analysis. This TALEN was hereafter designated as T-DCL3 (Table 3.5).

Table 3.5 SAPTA output data for TALEN targeting of DCL3

Left TALEN		Right TALEN		Spacer Size	Composite Score
Binding sequence	**Size**	**Binding sequence**	Size		
T-CCCACTGTTCACCTTGT	17	T-CTGTATTCAGCTTTAT	16	19	49.01
T-TCTCCAGCTCAGGAAG	16	T-CCCTTCTAGGAACCAAG	17	16	40.37
T-CAGGAAGAAACCCTCAT	17	T-CCAACTGATCCCTTCT	16	16	39.37
T-CCAGCTCAGGAAGAA	15	T-CCCTTCTAGGAACCAAG	17	14	39.02
T-GTATGCAAAACCCAA	15	T-CTTCCTGAGCTGGAG	15	14	38.72

For RDR2, the first best-scored TALEN (Table 3.6) was accepted since the target region does not display similarities with any other RDR sequences (Table 3.4). Besides, this TALEN (T-RDR2) has the advantage of having a restriction enzyme site within the spacer region (Supplementary Table S4).

Table 3.6 SAPTA output data for TALEN targeting RDR2

Left TALEN		Right TALEN		Spacer Size	Composite Score
Binding sequence	**Size**	**Binding sequence**	**Size**		
T-TTCCCATCGGTCCTGAT	17	T-CGAACACCTTCCCATGA	17	18	41.75
T-TCGAGACATTGGAAGCT	17	T-CTTGTTCTCGGCGAGAG	17	16	35.76
T-ATGCTGGTTTTCCCAT	16	T-CCCATGATTGAAGCACG	17	18	34.82

3.1.3 T-DCL3 and T-RDR2 construction

The Golden Gate method was used for TALENs construction. With this method TALENs are assembled in two steps. In an initial step, the first 10 RVDs are cloned in the correct order into the vector pFUS_A and the remaining RVDS are cloned in the corresponding pFUS_B vector. In a second step both inserts (intermediate arrays) of pFUS plasmids are joined and cloned into the final vector backbone. The RVD order within T-DCL3 and T-RDR2 was established according to the retrieved TALEN binding sequences (Tables 3.5 and 3.6) and the known RVD-base correspondence (Supplementary Table S4).

At the end of the first step, positive clones were evaluated by PCR amplification using vector primers, followed by electrophoresis in 1% (pFUS_A) or 2% (pFUS_B) agarose gels. The expected size for correct clones was 1.2 Kb (pFUS_A, DCL3 and RDR2), 700 bp (pFUS_B5 DCL3) and 900 bp (pFUS_B6, RDR2).

PCR amplification of T-DCL3 intermediate clones is shown in Figure 3.1. In panel A, lanes 2 and 5 indicate positive clones for pFUS_A constructs. The smearing effect indicates a positive clone – the 1.2 kb expected fragment is almost undetectable. Lanes 3 and 6 correspond to empty vector clones since they have the same fragment size of the pFUS_A control lane (nr10). Lanes 4, 7, 8, and 9 are negative pFUS_A clones, which probably correspond to clones with one or two RVDs. Positive pFUS_B clones are shown in panel B in lanes 2-4.

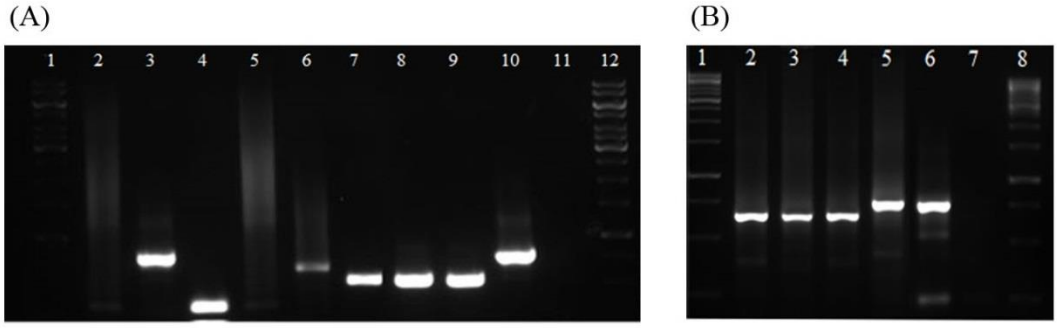

Figure 3.1 Intermediate constructs of T-DCL3. (A) pFUS_A clones were amplified by PCR and analysed in a 1% agarose gel: lanes 2 and 5 correspond to positive clones (lane 2 is the left array and lane 5 is the right array), lanes 3 and 6 indicate empty vector clones, lanes 4, 7, 8 and 9 indicate negative clones, lane 10 is the control for pFUS_A vector alone (empty vector) and lane 11 is the negative control. The ladder Gene Ruler^{TM} 1Kb DNA was used to indicate fragments size

(lanes 1 and 12). (B) pFUS_B clones were amplified by PCR and analysed in a 2% agarose gel. Lanes 2, 3 and 4 indicate positive constructs, lane 5 corresponds to a negative clone, lane 6 is the control for pFUS_B5 vector alone and lane 7 is the negative control. Lanes 1 and 8 contain Gene Ruler^{TM} 1Kb DNA ladder.

The results of PCR amplification of T-RDR2 intermediate clones are presented in Figure 3.2. In panel A, lane 6, 9 and 12 indicate positive clones for pFUS_A intermediate constructs. Positive pFUS_B clones are shown in panel B: in lanes 5, 8 and 9, appear the 'ladder effect' and bands with around 900bp. Selected clones were further confirmed by sequencing.

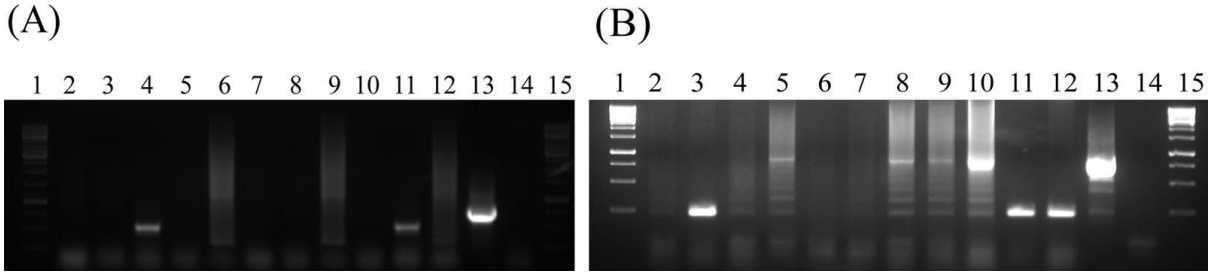

Figure 3.2 Intermediate constructs of T-RDR2. (A) pFUS_A clones were amplified by PCR and analysed in a 1% agarose gel: lanes 6, 9 and 12 correspond to positive clones (lane 6 is the left array and lanes 9 and 12 are the right array), lanes 2-5, 7, 8, 10 and 11 indicate negative clones, lane 13 is the control for pFUS_A vector alone (empty vector) and lane 14 is the negative control. The ladder Gene Ruler^{TM} 1Kb DNA was used to indicate fragments size (lane 1 and 15). (B) pFUS_B6 clones were amplified by PCR and analysed in a 2% agarose gel. Lanes 5, 8 and 9 indicate positive constructs (lane 5 is the left array and lanes 8 and 9 are the right arrays), lanes 2, 4, 6 and 7 correspond to a negative clones, , lanes 3, 11, 12 indicate empty vectors lane 13 correspond to the control for pFUS_B6 vector and lane 14 is the negative control. Lanes 1 and 15 contain Gene Ruler^{TM} 1Kb DNA ladder.

Intermediate arrays were then assembled into the backbone plasmids: pTAL 3 (left TALEN) and pTAL4 (right TALEN). Final constructs (pTAL3-DCL3 and pTAL4-DCL3; pTAL3-RDR2 and pTAL4-RDR2) were evaluated by PCR followed by agarose gel electrophoresis (Figure 3.3 and 3.4). As mentioned above, a smear or ladder effect indicates positive clones. All the tested clones were positive and were further confirmed by sequencing.

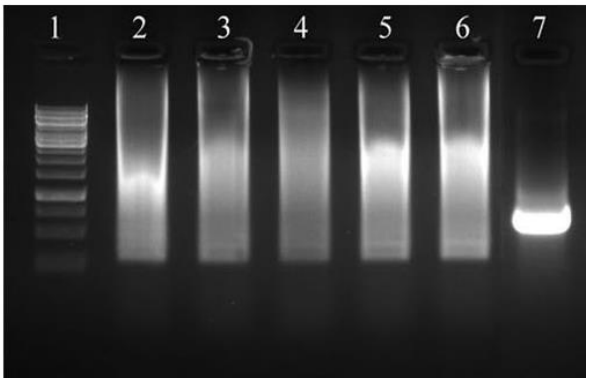

Figure 3.3 Final constructs for T-DCL3. pTAL3-DCL3 and pTAL4-DCL3 clones were amplified by PCR and analysed in a 1% agarose gel: lanes 2 – 6 correspond to positive clones (lanes 2, 3 and 4 indicate pTAL3-DCL3 and lane 5 and 6 indicate pTAL4-DCL3), lane 7 corresponds to the amplification of pTAL empty vector. The ladder Gene Ruler^{TM} 1Kb DNA was used to indicate fragments size (lane 1).

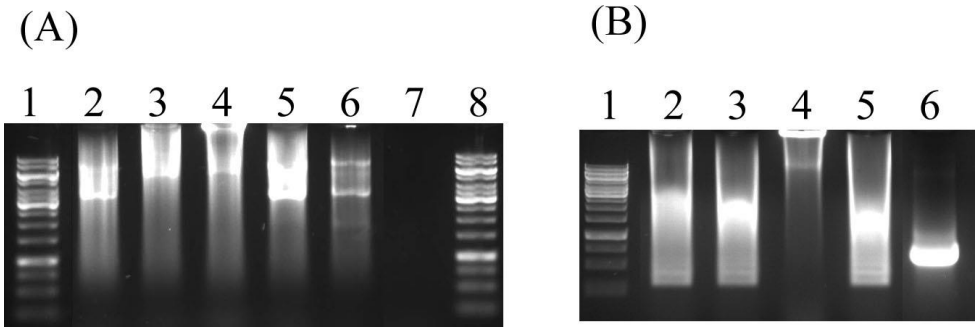

Figure 3.4 Final constructs for T-RDR2 (A) pTAL3-RDR2 clones were amplified by PCR and analysed in a 1% agarose gel: lanes 3 and 4 correspond to positive clones, lanes 2 and 5 indicate negative clones, lane 6 corresponds to the amplification of pTAL empty vector and lane 7 is the negative control. The ladder Gene Ruler^{TM} 1Kb DNA was used to indicate fragments size (lane 1and 8); (B) pTAL4-RDR2 clones were amplified by PCR and analysed in a 1% agarose gel: lanes 2 corresponds to a positive clone, lanes 3- 5 indicate negative clones, lane 6 corresponds to the amplification of pTAL empty vector. The ladder Gene Ruler^{TM} 1Kb DNA was used to indicate fragments size (lane 1)

3.1.4 Testing TALENs effectiveness – Yeast based assay

After successful assembly, TALENs can be cloned into plant expression vectors and used for plant transformation. However, previous validation of TALENs in simpler organisms although not being mandatory, will assure TALEN effectiveness prior to hard laborious downstream protocols, such as plant transformation. For this reason, the eukaryotic model yeast *Saccharomyces cerevisiae* has been widely used to evaluate TALENs functionality before their introduction in plants (Chames *et al*. 2005; Townsend *et al*. 2009; Cermak *et al*. 2011 and Beurdeley *et al*. 2013).

With this purpose, we have also setup an yeast-based assay to test TALENs activity. Yeast cells (BY4742 strain) were triple transformed with pTAL3-DCL3, pTAL4-DCL3 and with another construct containing T-DCL3 target site (pTS_DCL3). For T-RDR2, yeast cells were triple-transformed with pTAL3-RDR2, pTAL4-RDR2 and with a vector containing T-RDR2 target site (pTS_RDR2).

Yeast triple transformed colonies were the template of a nested PCR reaction using primers flanking the spacer region (Supplementary Table S2) (within TALENs target region) contained in pTS_DCL3 and pTS_RDR2

3.1.4.1 PAGE analysis of TALEN activity

PCR products where next analysed by polyacrylamide gel electrophoresis (PAGE) using a 20% polyacrylamide gel (range of separation between 6-100 bp, for T-DCL3) or a 12% polyacrylamide gel (range of separation 20-150 bp, for T-RDR2) in order to detect small differences in amplicons size, which would reflect possible indel mutations within the spacer region of pTS_DCL3 or of pTS_RDR2.

The fragment of 60 bp in lane 5 (Figure 3.5) corresponds to the amplification of pTS_DCL3 target region (single yeast transformation) and acts as a control – similar bands would indicate lack of TALEN activity. As depicted in Figure 3.5 we could not detect any triple transformant (lanes 2, 4 and 8) whose amplification of pTS_DCL3 by PCR generated different fragment sizes as compared to the

control (lane 5). We have screened by PCR thirty colonies, but we could not detect amplicons of different sizes (data not shown).

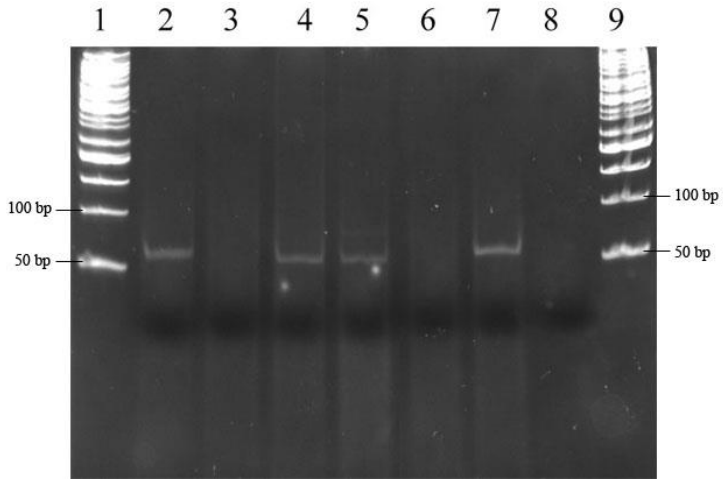

Figure 3.5 PAGE analysis of T-DCL3 activity *in vivo*. TALEN target region of yeast triple-transformants were PCR amplified and evaluated on a 20% polyacrylamide gel: lane 5 corresponds the control for pTS_DCL3, lanes 2, 4, 5 and 7 correspond to triple-transformants clones, lane 3 present any amplification, lane 8 indicates the negative control and lanes 1 and 9 contain Gene Ruler^{TM} 50 bp DNA ladder.

Regarding T-RDR2, PCR amplification of uncut pTS_RDR2 would retrieve a fragment of 100 bp. All the clones tested (twenty-one in total, data not shown), by PCR generated a fragment with a size identical to that obtained after amplification of uncut pTS_RDR2 (lane 8 that corresponds to the amplification by PCR of a yeast colony single transformed with pTS_RDR2).

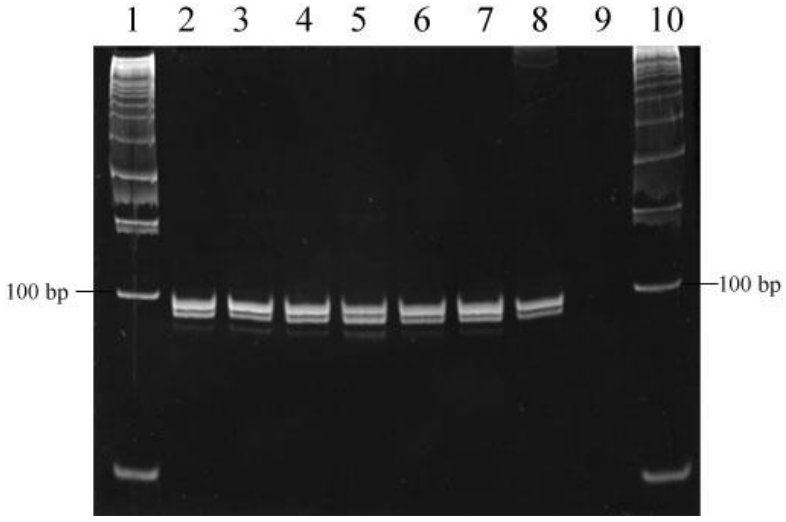

Figure 3.6 PAGE analysis of T-RDR2 activity *in vivo*. TALEN target region of yeast triple-transformants were PCR amplified and evaluated on a 12% polyacrylamide gel: lane 8 corresponds the control for pTS_RDR2, lanes 2-7 correspond to triple-transformants clones, lane 9 indicates the negative control and lanes 1 and 10 contain Gene Ruler^{TM} 50 bp DNA ladder.

Our results raised several hypotheses: (i) PAGE is not sensitive enough to detect small indels, (ii) PAGE is not sensitive enough to detect low TALEN activity (which would result in different populations of amplicons within the same PCR reaction) or (iii) the tested TALENs are not effective.

To examine the first two hypotheses, we have undertaken an alternative approach based on SSCP analysis of the nested PCR products.

3.1.4.2 SSCP analysis of TALEN

SSCP is based on DNA mobility due to sequence and not to size, which allows that a single nucleic acid change affects the migration of a single-stranded DNA fragment (Orita *et al*., 1989). Sequence differences generate different band patterns in non-denaturing polyacrylamide gels.

Fragments resulting from nested PCR amplification (see section 3.14) of yeast colonies triple transformed with TALENs and plasmids carrying the corresponding target sites were analysed by SSCP. In Figure 3.7, lane 7 corresponds to the amplification of pTS_DCL3 target region (single yeast transformation) and acts as a control – similar band patterns would indicate lack of TALEN activity. Lanes 1 to 5 appear to have a distinct band pattern, which might indicate the presence of mutations.

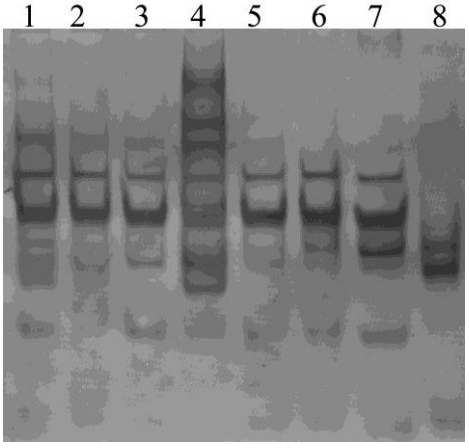

Figure 3.7 SSCP analysis to test T-DCL3 activity *in vivo*. TALEN target region of yeast triple-transformants were PCR amplified and evaluated by SSCP. Lane 7 corresponds to the control for pTS_DCL3 target site and, lanes1-5 indicate putative positive triple transformants clones, lane 6 corresponds to a negative clone, lane 8 corresponds to the negative control of the PCR reaction.

In Figure 3.8, lane 8, indicates the PCR amplified fragment of pTS_RDR2. Lanes 2, 3, 4 and 5 appear to have a distinct band pattern from lane 8, which suggests the presence of mutations in these clones.

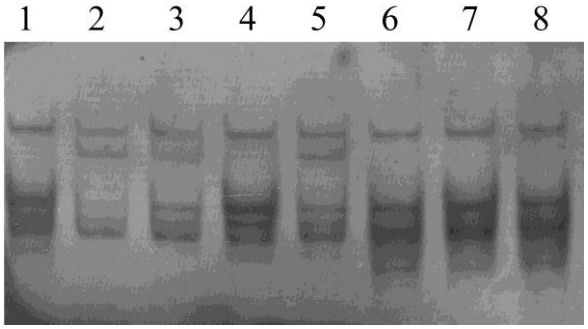

Figure 3.8 SSCP analysis to test T-RDR2 activity *in vivo*. TALEN target region of yeast triple-transformants were PCR amplified and evaluated by SSCP. Lane 8 corresponds to the control pTS_RDR2, lanes 2-5 indicate putative positive triple-transformants clones, lanes 1, 6 and 7 indicate negative clones, lane 8 indicates the negative control of the PCR reaction.

In order to confirm if the putative positive clones were indeed carriers of indel mutations, the corresponding PCR fragments were cloned into pRS416 and used to transform *E.coli* cells. Transformants were subjected to colony PCR and the resulting fragments were again analysed by SSCP. As an example for T-DCL3, the results regarding clone 3 of Figure 3.7 are shown in Figure 3.9. Lanes 1, 2 and 7 have different band patterns when compared to lane 9 (amplification of *E.coli* cells transformed with pTS_DCL3) and were therefore selected for further confirmation by sequencing.

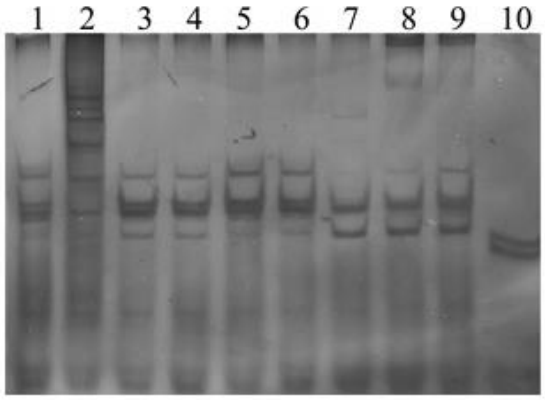

Figure 3.9 Second SSCP analysis to confirm T-DCL3 activity. TALEN target region of *E. coli* transformants were PCR amplified and evaluated by SSCP. Lanes 1, 2 and 7 indicate putative positive clones, lane 9 correspond to the pTS_DCL3 and lane 10 is the negative control of the PCR reaction.

As an example for T-RDR2, the results regarding the second SSCP analysis of clone 2 (Figure 3.8) are shown in Figure 3.10. Lanes 1, 3, 5 and 7 have different band patterns when compared to lane 9 (amplification of *E.coli* cells transformed with pTS_RDR2) and were, therefore, selected for sequencing confirmation.

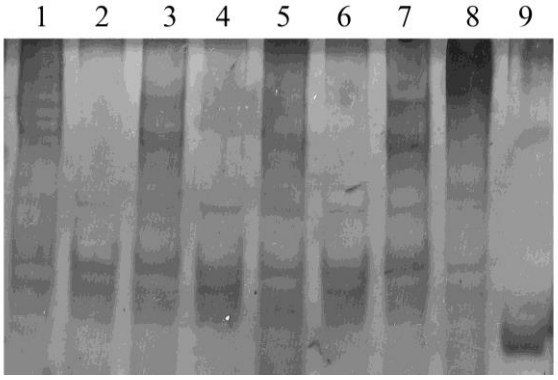

Figure 3.10 Second SSCP analysis to confirm T-RDR2 activity TALEN target region of *E. coli* transformants were PCR amplified and evaluated by SSCP. Lanes 1, 3,5 and 7 indicate putative positive clones, lane 8 indicate the pTS_RDR2 amplification, lane 9 is the negative control of the PCR reaction.

Contrary to our expectations, however, sequencing results clearly showed that the different patterns observed in Figure 3.9 (lanes 1, 2 and 7) and in Figure 3.10 (lanes 1, 3, 5 and 7) did not reflect indel mutations within the spacer region.

Our exhaustive analyses strongly suggest that TALENs chosen might not be effective *in vivo* and that SSCP is not a good method to detect sequence variability under our experimental conditions.

3.1.5 T2-DCL3 and T2-RDR2 construction

In the current scenario, we decided to contact Professor Daniel Voytas who has developed the Golden Gate TALEN strategy. Professor Voytas strongly advise us to use the newly improved versions of pTAL3 and pTAL4, pZHY500 and pZHY501 respectively, which have been proved to have higher efficiencies in inducing DSB (Zhang *et al*., 2013).

We have repeated the second step of Golden Gate reaction using as vector backbones pZHY500 (pZHY500-DCL3 and pZHY500-RDR2) and pZHY501 (pZHY501-DCL3 and pZHY501-RDR2). Final constructs were designated T2-DCL3 and T2-RDR2. Selection of positive clones was made as described above (see section 3.1.3).

Confirmation of T2_DCL3 final constructs is shown Figure 3.11. In panel A, lanes 2 and 3 indicate positive clones for left T2-DCL3 and in panel B, lane 2 corresponds to positive clone for right T2-DCL3.

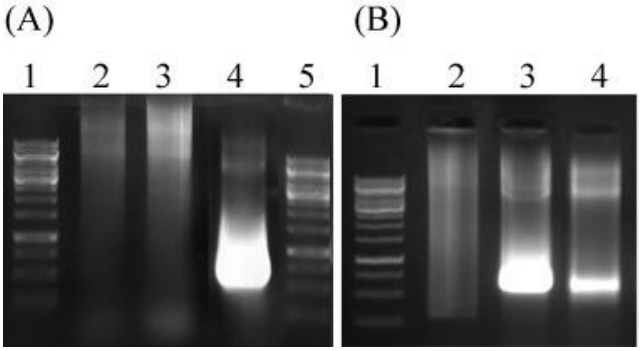

Figure 3.11 Final constructs of T2-DCL3. (A) pZHY500-DCL3 constructsts were amplified by PCR and analysed in a 1% agarose gel: lanes 1 and 5 contain Gene Ruler^{TM} 1Kb DNA ladder, lanes 2 and 3 indicate positive clones, lane 4 correspond to the empty vector control (B) pZHY501-DCL3 constructs were amplified by PCR and analysed in a 1% agarose gel: lane 1 contains the Gene Ruler^{TM} 1Kb DNA ladder, lane 2 indicates positive clone, lane 3 corresponds to a negative clone, lane 4 indicates the empty vector control.

T2-RDR2 final constructs were also confirmed by PCR as depicted in Figure 3.12. Lane 3 (panel A) and lane 6 (panel B) indicate positive clones. These clones were confirmed by sequencing.

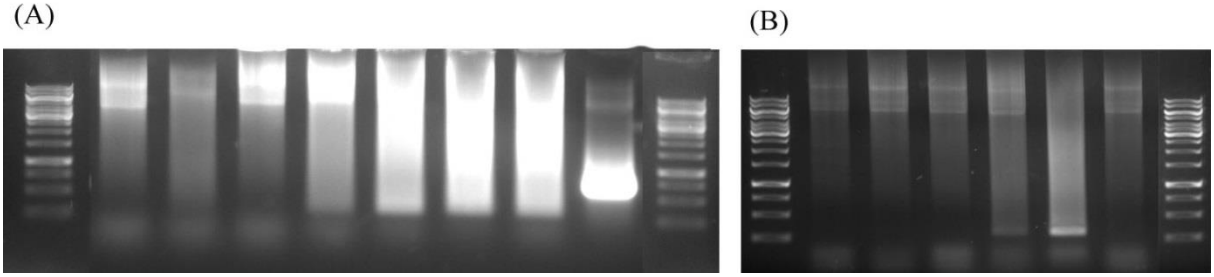

Figure 3.12 Final constructs for T2-RDR2. (A) pZHY500-RDR2 were amplified by PCR and analysed in a 1% agarose gel: lane 1 contains Gene Ruler^{TM} 1Kb DNA ladder, lane 3 indicates a positive clone, lane 2 and 4-8 correspond to negative clones, lane 9 indicates the control for pZHY500 empty vector; (B) pZHY501-RDR2 were amplified by PCR and analysed in a 1% agarose gel: lane 1 and 8 contain the Gene Ruler^{TM} 1Kb DNA ladder, lane 6 indicates a positive clone, lane 2-5 correspond to negative clones, lane 4 indicates the control for pZHY501 empty vector.

3.2 CRISPR/Cas9 system target sites identification

Recently, the CRISPR/Cas9 system has been successfully applied for genome editing (Nekrasov *et al*., 2013). We have also design CRISPRs targeting the same genes used for TALEN design – DCL3 and RDR2 – in order to investigate in the future which genome editing tool is most suitable to use in *M. truncatula*. The full sequences of RDR2 and DCL3 were screened for possible CRISPR target sites using the software CHOPCHOP (Montague et al., 2014), available online. The program output consists of a list of target sequences ranked according to the position within the gene.

Ten CRISPR/Cas9 system target sites found by CHOPCHOP software are shown in Table 3.9 for DCL3 gene and Table 3.10 for RDR2 gene. Target sequences were compared to DIALIGN alignments used in TALEN target selection and blasted into *M. truncatula* genome to avoid possible off-target effects.

Table 3.9 CHOPCHOP output data for CRISPR targeting DCL3.

Ranking	Target sequence	Genomic location	GC content (%)
1	GGTCAAGCTATTAGGTCCAGTGG	Medtr3g105390.1:272	52
2	GCATCACCAGTTGGTAAAAGAGG	Medtr3g105390.1:644	48
3	AGCCAAAAATTGAAGCCTCGTGG	Medtr3g105390.1:834	48
4	TGGACTGTTGTGTGCTTATGAGG	Medtr3g105390.1:982	48
5	GCAGAGGAGTACTTGTCTCGTGG	Medtr3g105390.1:1799	57
6	GTATCGCTGCCTATGGAGGGAGG	Medtr3g105390.1:1835	61
7	GTTTGGAAAGCTGCATTAGGTGG	Medtr3g105390.1:1883	48
8	GTGTAACTCCTTTCTTTTCGTGG	Medtr3g105390.1:2094	43
9	GCCAGTGTTCGAGCATTATGTGG	Medtr3g105390.1:2123	52
10	CGAGCATTATGTGGTTCATGGGG	Medtr3g105390.1:2132	52

CRISPRs targeting DCL3 in ranking positions nr 2,4,5, 6, 7, 9 and 10 (Table 3.9) were excluded because they have homologies in the PAM region with MtDCL4 (Table 3.1). CRISPRs in ranking positions 1, 3 and 8 have no homologies with other Dicer-like proteins and a genome blast search revealed no significant homologies with other regions.

For CRISPRs targeting RDR2 gene, the three top ranked sequences were chosen since they neither have homologies with RNA-dependent RNA polymerases nor with other genomic regions. The three CRISPRs will introduce a double strand break in the first exon, ensuring the disruption of a functional protein.

Table 3.10 CHOPCHOP output data for CRISPR targeting RDR2.

Ranking	Target sequence	Genomic location	GC content (%)
1	GAGAAGGTCGTTGGCGGTGGCGG	Medtr4g106660.1:57	70
2	GTATCGGAGAAGGTCGTTGGCGG	Medtr4g106660.1:63	61
3	GACGGAGGAACGGCCAACAGTGG	Medtr4g106660.1:93	65
4	GATTTCGAGAGCGAAGACGGAGG	Medtr4g106660.1:108	57
5	CGGAGGAAGCGTCAAGACAGAGG	Medtr4g106660.1:253	61
6	TGAAATGGCCCCTCAAGATGTGG	Medtr4g106660.1:1576	52
7	GCATTTCAAATTCGGTACGGTGG	Medtr4g106660.1:1733	48
8	GGCGCTTTTGGCATTGCAAGAGG	Medtr4g106660.1:1936	57
9	GATGTTGACTGATAAGGAAGCGG	Medtr4g106660.1:1981	48
10	AGTATTTGTCCGCATAACTGTGG	Medtr4g106660.1:2230	43

4 Discussion

A crucial step in designing TALENs is the choice of the region to be targeted. In this work, RDR2 and DCL3 genes were carefully inspected for homologies with other related and non-related genomic sequences. After this analysis, the first exon of both genes emerged as good candidates for the design of TALENs with lower off-target rates. Moreover, disruption of the coding frame of the first exon will ensure early protein truncation which will guarantee a complete functional lost due to the abrogation of all functional domains of the protein (Wright *et al*., 2014).

In this work, we have also attempted to establish a quick and affordable method to evaluate TALENs functionality in yeast, prior to plant transformation. The eukaryotic model yeast *Saccharomyces cerevisiae* has been extensively used in TALENs functionality evaluation assays (Chames *et al*. 2005; Townsend *et al*. 2009; Cermak *et al*. 2011 and Beurdeley *et al*. 2013). The strategy used usually requires yeast mating to combine both left and right TALENs with the TALEN target site. The first step of that strategy is to introduce left and right TALENs in a α mating type yeast strain and to transform a yeast strain of mating type a with the vector carrying the TALEN target site. Mating is then promoted and at the end of this process, diploid cells with the three plasmids are selected. The triple-transformation approach presented here avoids the mating process by successfully introduce three plasmids in a haploid yeast cell. This strategy is straightforward and easy to apply, avoiding the use of a multiple step protocol and the time-consuming selection of cells carrying the three plasmids.

After yeast transformation, TALENs activity needs to be assessed. To this end, two approaches are often adopted: restriction fragment length polymorphism (RFLP) analysis (Miller *et al*., 2011; Wendt *et al*., 2013) and *lacZ* recovery/disruption assay (Hisano *et al*., 2013).

RFLP analysis requires the presence of an appropriate restriction site within the spacer region (Miller *et al*., 2011; Wendt *et al*., 2013). The spacer region is amplified by PCR and PCR products are digested with the suitable restriction enzyme. However, this method might not assure the detection of all TALEN-induced mutations since DBS can be introduce within the spacer but not modify the restriction enzyme (RE) site. Furthermore, by requiring the existence of a RE site within the spacer region, it greatly limits the genomic regions susceptible to be targeted.

In the *lacZ* recovery assay, the TALEN target site are inserted into the *lacZ* gene out of frame (white colonies). When DSBs are generated at the target site by TALENs, TALEN-induced frameshifts lead to the recovery of *lacZ* activity because of in-frame fusion (blue colonies). Blue colonies will thus contain sequences with frameshift mutations in the TALEN target site (positive clones). In the *lacZ* disruption assay, TALEN target site are inserted in-frame with the *lacZ* gene (blue colonies). TALEN-mediated frameshifts cause the disruption of *lacZ* activity because of out-of-frame fusion (white colonies) (Hisano *et al*., 2013). This method however is not very efficient and leads to several false negatives.

In this work we have tried to circumvent the limitations of the above approaches, by using two strategies: (1) PAGE and (2) SSCP analyses of small PCR products.

The triple yeast transformants were subjected to a nested PCR, in order to amplify specific amplicons of small size that were further analysed by PAGE or SSCP.

With PAGE analyses we were not able to detect PCR products of different sizes. These results led us to hypothesize that PAGE was not sensitive enough to detect small indels or to detect low TALEN activity. To overcome this possible limitation of PAGE we tested a SSCP strategy. SSCP is widely used to detect DNA mutations since it is more sensible than other approaches, such as RFLP (Orita *et al*., 1989). Our results showed that, although different band patterns were observed in SSCP analyses, they did not correlate with a mutation within the target sequence of both TALENs. The significant ratio of false positives obtained with SSCP analyses may be due to the fact that we have performed a colony-PCR followed by silver staining and this procedure also stains proteins (Chevallet *et al*., 2006). As such it is possible that the different patterns obtained were merely caused by gel loading of non-purified PCR fragments contaminated with proteins from cellular debris. In any case we could not test TALENs efficiency and this data together with the advice of Prof. Voytas to change the final vector backbone to improve TALENs activity, strongly suggests that the TALENs designed are not suitable for plant transformation.

5 Conclusions and Future Work

In this work we have set up in the laboratory several bioinformatics and molecular methodologies allowing the design of CRISPR-Cas9 systems and the assembly of TALENs, which in the future will certainly enable the successful disruption of DCL3 and RDR2 genes and shed the light on the relationship between plant stress resistance and epigenetic regulation mediated by siRNAs in the legume plant model *Medicago truncatula*.

Overall, our findings suggest that:

1. TALENs construction using pTAL vectors is not very efficient and the alternative use of pZHY500 and pZHY501 vectors is strongly advisable;
2. Yeast-based assays might be used to test TALENs efficacy prior to plant transformation;
3. Yeast-based assays using yeast triple transformation are a good and rapid alternative to laborious yeast mating strategies;
4. PAGE analysis of TALENs functionality might be a valuable tool to test TALENs efficiency *in vivo* if we could increase TALENs activity (for instance by cloning TALENs into the pZHY500 and pZHY501 backbones);
5. SSCP-based approach to test TALENs activity generates several false positives possibly due to protein contamination of the DNA fragments analysed.

In the future, TALENs T2-DCL3 and T2-RDR2, assembled in pZHY500 vectors, have to be tested in order to see if these new TALENs are, in fact, more efficient than the previous ones (T-DCL3 and T-RDR2). Yeast triple transformation and PAGE analysis might be used as a fast and rapid screening to test those TALENs activity. Alternatively, a DNA mismatch cleavage approach might be assayed. This assay is based on heteroduplexes formation and cleavage by T7 endonuclease. Wild type and putative mutant PCR products are joined, denatured and allowed to anneal, a mismatch sensitive endonuclease (T7 endonuclease) is then used to digest heteroduplexes, which are eventually formed. Nuclease activity is estimated by electrophoresis of the digested fragments (Miller *et al*., 2007; Kim *et al*., 2009)

TALENs will need next to be inserted in plants. One limiting step in plant genome editing using TALENs is the introduction of two plasmids into plant cells since co-transformation has a lower success rate compared to single-transformation (Wright *et al*., 2014). Two strategies might be used to solve this drawback. One is to clone left and right TALENs in the same vector, but under the control of different promoters (Li *et al*., 2012). The other option is to construct a single expression unit in which both left and right TALEN coding sequences are separated by a T2A translation skipping sequence (Kim *et al*. 2011) and expressed under the control of a single promoter (usually 35S promoter) (Zhang *et al*., 2013). The latter strategy is more advantageous since it minimizes the size of the plasmid containing TALEN constructs (Zhang *et al*., 2013). In order to clone left and right

TALENs separated by the T2A peptide, a gateway compatible entry plasmid, pZHY013 (Zhang *et al*., 2013), will be used. A gateway LR reaction will be performed to clone TALENs into the destination vector, pK7WG2D (Karimi *et al*., 2002), downstream the 35S promoter.

We also intend to test the designed CRISPR-Cas9 Systems. A comparative study between the two types of engineered nucleases –TALENs and CRISPRs – will allow us to establish the best method for *Medicago truncatula* genome editing.

6 References

Alberts, B., 2012. The breakthroughs of 2012. Science 326, 1589.

Araújo, S.D.S., Duque, A.S., Santos, D. & Fevereiro, M.P., 2004. An efficient transformation method to regenerate a high number of transgenic plants using a new embryogenic line of Medicago truncatula cv . Jemalong. Plant Cell, Tissue and Organ Culture 78, 123–131.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A. & Horvath, P., 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science 315, 1709–12.

Benjamin, J.G. & Nielsen, D.C., 2006. Water deficit effects on root distribution of soybean, field pea and chickpea. Field Crops Research 97, 248–253.

Beurdeley, M., Bietz, F., Li, J., Thomas, S., Stoddard, T., Juillerat, A., Zhang, F., Voytas, D.F., Duchateau, P. & Silva, G.H., 2013. Compact designer TALENs for efficient genome engineering. Nature communications 4, 1762.

Bitinaite, J., Wah, D.A., Aggarwal, A.K. & Schildkraut, I., 1998. FokI dimerization is required for DNA cleavage. Proceedings of the National Academy of Sciences of the United States of America 95, 10570–10575.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A. & Bonas, U., 2009. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science (New York, N.Y.)*, 326(2009), 1509–1512.

Bonas, U., Stall, R.E. & Staskawicz, B., 1989. Genetic and structural characterization of the avirulence gene avrBs3 from Xanthomonas campestris pv. vesicatoria. Molecular & general genetics 218, 127–136.

Boyko, A. & Kovalchuk, I., 2010. Transgenerational response to stress in Arabidopsis thaliana. Plant signaling & behavior 5, 995–998.

Britt, A.B. & May, G.D., 2003. Re-engineering plant gene targeting. Trends in Plant Science 8, 90–95.

Brodribb, T.J. & Holbrook, N.M., 2003. Stomatal closure during leaf dehydration, correlation with other leaf physiological traits. Plant physiology 132, 2166–2173.

Byun, S.O., Fang, Q., Zhou, H. & Hickford, J.G.H., 2009. An effective method for silver-staining DNA in large numbers of polyacrylamide gels. Analytical Biochemistry 385, 174–175.

Capitão, C., Paiva, J. a P., Santos, D.M. & Fevereiro, P., 2011. In *Medicago truncatula*, water deficit modulates the transcript accumulation of components of small RNA pathways. BMC plant biology 11, 79.

Cathomen, T. & Joung, J.K., 2008. Zinc-finger nucleases: the next generation emerges. Molecular therapy : the journal of the American Society of Gene Therapy 16, 1200–1207.

Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J. a, Somia, N. V, Bogdanove, A.J. & Voytas, D.F., 2011. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic acids research 39, e82.

Chabaud, M., De Carvalho-Niebel, F. & Barker, D.G., 2003. Efficient transformation of Medicago truncatula cv. Jemalong using the hypervirulent Agrobacterium tumefaciens strain AGL1. Plant Cell Reports 22, 46–51.

Chames, P., Epinat, J.C., Guillier, S., Patin, A., Lacroix, E. & Pâques, F., 2005. In vivo selection of engineered homing endonucleases using double-strand break induced homologous recombination. Nucleic Acids Research 33, 1–10.

Chan, S.W.-L., Henderson, I.R. & Jacobsen, S.E., 2005. Gardening the genome: DNA methylation in Arabidopsis thaliana. Nature reviews. Genetics 6, 351–360.

Chaves, M.M., Maroco, J.P. & Pereira, J.S., 2003. Understanding plant responses to drought—from genes to the whole plant. Functional Plant Biology 30, 239–264.

Chen, J., Zhang, X., Wang, T., Li, Z., Guan, G. & Hong, Y., 2012. Efficient detection, quantification and enrichment of subtle allelic alterations. DNA Research 19, 423–433.

Chevallet, M., Luche, S. & Rabilloud, T., 2006. Silver staining of proteins in polyacrylamide gels. Nature protocols 1, 1852–1858.

Christian, M., Cermak, T., Doyle, E.L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A.J. & Voytas, D.F., 2010. Targeting DNA double-strand breaks with TAL effector nucleases. Genetics 186, 757–761.

Christian, M., Qi, Y., Zhang, Y. & Voytas, D.F., 2013. Targeted mutagenesis of Arabidopsis thaliana using engineered TAL effector nucleases. G3: Genes | Genomes | Genetics 3, 1697–1705.

Cong, L., Ann Ran, F., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al., 2013. Multiplex Genome Engineering Using CRISPR/Cas Systems. Science 339, 819–824.

Cook, D.R., 1999. *Medicago truncatula* - a model in the making! Current Opinion in Plant Biology 2, 301–304.

Coontz, R., 2013. Breaktrough of 2013. Science 342, 1434–1435.

Curtin, S.J., Zhang, F., Sander, J.D., Haun, W.J., Starker, C., Baltes, N.J., Reyon, D., Dahlborg, E.J., Goodwin, M.J., Coffman, A.P., et al., 2011. Targeted mutagenesis of duplicated genes in soybean with zinc-finger nucleases. Plant physiology 156, 466–473.

Daxinger, L., Kanno, T., Bucher, E., van der Winden, J., Naumann, U., Matzke, A.J.M. & Matzke, M., 2009. A stepwise pathway for biogenesis of 24-nt secondary siRNAs and spreading of DNA methylation. The EMBO journal 28, 48–57.

Decottignies, A., 2013. Alternative end-joining mechanisms: a historical perspective. Frontiers in genetics 4, 48.

Engler, C., Kandzia, R. & Marillonnet, S., 2008. A one pot, one step, precision cloning method with high throughput capability. PloS one 3, e3647.

Gabriel, R., Lombardo, A., Arens, A., Miller, J.C., Genovese, P., Kaeppel, C., Nowrouzi, A., Bartholomae, C.C., Wang, J., Friedman, G., et al., 2011. An unbiased genome-wide analysis of zinc-finger nuclease specificity. Nature biotechnology, 29, 816–823.

Graham, P.H. & Vance, C.P., 2003. Update on Legume Utilization Legumes : Importance and Constraints to Greater Use. Plant Physiology 131, 872–877.

Henderson, I.R., Zhang, X., Lu, C., Johnson, L., Meyers, B.C., Green, P.J. & Jacobsen, S.E., 2006. Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning. Nature genetics 38, 721–725.

Henderson, I.R. & Jacobsen, S.E., 2007. Epigenetic inheritance in plants. Nature, 447, 418–424.

Herr, a J., Jensen, M.B., Dalmay, T. & Baulcombe, D.C., 2005. RNA polymerase IV directs silencing of endogenous DNA. Science 308, 118–20.

Hisano, Y., Ota, S., Arakawa, K., Muraki, M., Kono, N., Oshita, K., Sakuma, T., Tomita, M., Yamamoto, T., Okada, Y., et al., 2013. Quantitative assay for TALEN activity at endogenous genomic loci. Biology open 2, 363–7.

Humbeck, K., 2013. Epigenetic and small RNA regulation of senescence. Plant Molecular Biology, 82, 529–537.

Inoue, H., Nojima, H. & Okayama, H., 1990. High efficiency transformation of Escherichia coli with plasmids. Gene 96, 23–28.

Jia, Y., Lisch, D.R., Ohtsu, K., Scanlon, M.J., Nettleton, D. & Schnable, P.S., 2009. Loss of RNA-dependent RNA polymerase 2 (RDR2) function causes widespread and unexpected changes in the expression of transposons, genes, and 24-nt small RNAs. PLoS genetics 5, e1000737.

Jiang, W., Zhou, H., Bi, H., Fromm, M., Yang, B. & Weeks, D.P., 2013. Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. Nucleic acids research 41, e188.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. & Charpentier, E., 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816–21.

Karimi, M., Inzé, D. & Depicker, A., 2002. GATEWAY^{TM} vectors for Agrobacterium-mediated plant transformation. Trends in Plant Science 7, 193–195.

Kasschau, K.D., Fahlgren, N., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S. a & Carrington, J.C., 2007. Genome-wide profiling and analysis of Arabidopsis siRNAs. PLoS biology 5, e57.

Kim, H.J., Lee, H.J., Kim, H., Cho, S.W. & Kim, J., 2009. Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly.Genome Research 19, 1279–1288.

Kim, J.H., Lee, S.R., Li, L.H., Park, H.J., Park, J.H., Lee, K.Y., Kim, M.K., Shin, B.A. & Choi, S.Y., 2011. High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. PLoS one 6, 1–8.

Kumar, S., Kumari, R., Sharma, V. & Sharma, V., 2013. Roles, and establishment, maintenance and erasing of the epigenetic cytosine methylation marks in plants. Journal of Genetics 92, 629–666.

Lauria, M. & Rossi, V., 2011. Epigenetic control of gene regulation in plants. Biochimica et biophysica acta 1809, 369–378.

Lawlor, D.W. & Tezara, W., 2009. Causes of decreased photosynthetic rate and metabolic capacity in water-deficient leaf cells: A critical evaluation of mechanisms and integration of processes. Annuals of Botany 103, 561–579.

Li, C.F., Pontes, O., El-Shami, M., Henderson, I.R., Bernatavichute, Y. V., Chan, S.W.L., Lagrange, T., Pikaard, C.S. & Jacobsen, S.E., 2006. An ARGONAUTE4-Containing Nuclear Processing Center Colocalized with Cajal Bodies in Arabidopsis thaliana. Cell 126, 93–106.

Li, J.-F., Norville, J.E., Aach, J., McCormack, M., Zhang, D., Bush, J., Church, G.M. & Sheen, J., 2013. Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. Nature biotechnology 31, 688–91.

Li, L., Wu, L.P. & Chandrasegaran, S., 1992. Functional domains in Fok I restriction endonuclease. Proceedings of the National Academy of Sciences of the United States of America 89, 4275–4279.

Li, T., Huang, S., Jiang, W.Z., Wright, D., Spalding, M.H., Weeks, D.P. & Yang, B., 2011. TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. Nucleic acids research 39, 359–372.

Li, T., Liu, B., Spalding, M.H., Weeks, D.P. & Yang, B., 2012. High-efficiency TALEN-based gene editing produces disease-resistant rice. Nature Biotechnology 30, 390–392.

Liang, Z., Zhang, K., Chen, K. & Gao, C., 2014. Targeted mutagenesis in Zea mays using TALENs and the CRISPR/Cas system. Journal of Genetics and Genomics 41, 63–68.

Lin, Y., Fine, E.J., Zheng, Z., Antico, C.J., Voit, R. a, Porteus, M.H., Cradick, T.J. & Bao, G., 2014. SAPTA: a new design tool for improving TALE nuclease activity. Nucleic acids research 42, e47.

Lu, C., Kulkarni, K., Souret, F.F., MuthuValliappan, R., Tej, S.S., Poethig, R.S., Henderson, I.R., Jacobsen, S.E., Wang, W., Green, P.J., et al., 2006. MicroRNAs and other small RNAs enriched in the Arabidopsis RNA-dependent RNA polymerase-2 mutant. Genome research, 16, 1276–1288.

Mahfouz, M.M., Li, L., Shamimuzzaman, M., Wibowo, A., Fang, X. & Zhu, J.-K., 2011. De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. Proceedings of the National Academy of Sciences 108, 2623–2628.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J.M., Wolf, Y.I., Yakunin, A.F., et al., 2011. Evolution and classification of the CRISPR-Cas systems. Nature reviews Microbiology 9, 467–477.

Mao, Y., Zhang, H., Xu, N., Zhang, B., Gou, F. & Zhu, J.-K., 2013. Application of the CRISPR-Cas system for efficient genome engineering in plants. Molecular plant 6, 2008–2011.

Marris, E., 2008. Water: more crop per drop. Nature 452, 273–277.

Marton, I., Zuker, A., Shklarman, E., Zeevi, V., Tovkach, A., Roffe, S., Ovadis, M., Tzfira, T. & Vainstein, A., 2010. Nontransgenic genome modification in plant cells. Plant physiology 154, 1079–1087.

Matzke, M., Kanno, T., Daxinger, L., Huettel, B. & Matzke, A.J.M., 2009. RNA-mediated chromatin-based silencing in plants. Current opinion in cell biology 21, 367–376.

Miller, J.C., Holmes, M.C., Wang, J., Guschin, D.Y., Lee, Y.-L., Rupniewski, I., Beausejour, C.M., Waite, A.J., Wang, N.S., Kim, K. a, et al., 2007. An improved zinc-finger nuclease architecture for highly specific genome editing. Nature biotechnology 25, 778–785.

Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J., et al., 2011. A TALE nuclease architecture for efficient genome editing. Nature biotechnology 29, 143–148.

Montague, T.G., Cruz, M., Gagnon, J.A., Church, G.M. & Valen, E., 2014. CHOPCHOP : a CRISPR / Cas9 and TALEN web tool for genome editing. Nucleic Acids Research 12, 401–407.

Morgenstern, B., 2004. DIALIGN: Multiple DNA and protein sequence alignment at BiBiServ. Nucleic Acids Research 32, 33–36.

Moscou, M. & Bogdanove, A., 2009. A Simple Cipher Governs DNA Recognition by TAL Effectors. Science 326, 1501.

Mussolino, C., Morbitzer, R., Lütge, F., Dannemann, N., Lahaye, T. & Cathomen, T., 2011. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. Nucleic Acids Research 39, 9283–9293.

Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J.D.G. & Kamoun, S., 2013. Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease. Nature biotechnology 31, 691–693.

Neves, L.., Duque, A.S., Almeida, J.S. & Fevereiro, P.., 1999. Repetitive somatic embryogenesis in Medicago truncatula ssp . Narbonensis. Plant Cell Reports 18, 398–405.

Nunes, C., Araújo, S.D.S., Silva, J.M., Fevereiro, M.P.S. & Silva, A.B., 2008. Physiological responses of the legume model. Environmental and Experimental Botany 63, 289–296.

Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K. & Sekiya, T., 1989. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. Proceedings of the National Academy of Sciences of the United States of America 86, 2766–2770.

Paszkowski, J. & Grossniklaus, U., 2011. Selected aspects of transgenerational epigenetic inheritance and resetting in plants. Current Opinion in Plant Biology 14, 195–203.

Pattanayak, V., Ramirez, C.L., Joung, J.K. & Liu, D.R., 2012. Revealing Off-Target Cleavage Specificities of Zinc Finger Nucleases by In Vitro Selection. Nature methods 29, 765–770.

Pennisi, E., 2010. Sowing the seeds for the ideal crop. Science 327, 802–803.

Pingoud, a, Fuxreiter, M., Pingoud, V. & Wende, W., 2005. Type II restriction endonucleases: structure and mechanism. Cellular and molecular life sciences 62, 685–707.

Porteus, M.H. & Carroll, D., 2005. Gene targeting using zinc finger nucleases. Nature biotechnology 23, 967–973.

Radecke, S., Radecke, F., Cathomen, T. & Schwarz, K., 2010. Zinc-finger nuclease-induced gene repair with oligodeoxynucleotides: wanted and unwanted target locus modifications. Molecular therapy : the journal of the American Society of Gene Therapy 18, 743–753.

Ramirez, C.L., Foley, J.E., Wright, D. a, Müller-Lerch, F., Rahman, S.H., Cornu, T.I., Winfrey, R.J., Sander, J.D., Fu, F., Townsend, J. a, et al., 2008. Unexpected failure rates for modular assembly of engineered zinc fingers. Nature methods 5, 374–375.

Richards, E.J., 2011. Natural epigenetic variation in plant species: A view from the field. Current Opinion in Plant Biology 14, 204–209.

Santos, D. & Fevereiro, P., 2002. Loss of DNA methylation affects somatic embryogenesis in Medicago truncatula. Plant Cell, Tissue and Organ Culture 70, 155–161.

Shan, Q., Wang, Y., Chen, K., Liang, Z., Li, J., Zhang, Y., Zhang, K., Liu, J., Voytas, D.F., Zheng, X., et al., 2013a. Rapid and efficient gene modification in rice and Brachypodium using TALENs. Molecular plant 6, 1365–1368.

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., Zhang, K., Liu, J., Xi, J.J., Qiu, J.-L., et al., 2013b. Targeted genome modification of crop plants using a CRISPR-Cas system. Nature biotechnology 31, 686–688.

Sijen, T., Vijn, I., Rebocho, a, van Blokland, R., Roelofs, D., Mol, J.N. & Kooter, J.M., 2001. Transcriptional and posttranscriptional gene silencing are mechanistically related. Current biology 11, 436–440.

Sikorski, R.S. & Hieter, P., 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. Genetics 122, 19–27.

Steimer, A., Schöb, H. & Grossniklaus, U., 2004. Epigenetic control of plant development: New layers of complexity. Current Opinion in Plant Biology 7, 11–19.

Talamè, V., Ozturk, N.Z., Bohnert, H.J. & Tuberosa, R., 2007. Barley transcript profiles under dehydration shock and drought stress treatments: A comparative analysis. Journal of Experimental Botany 58, 229–240.

Townsend, J.A., Wright, D.A., Winfrey, R.J., Fu, F., Maeder, M.L., Joung, J.K. & Voytas, D.F., 2009. High-frequency modification of plant genes using engineered zinc-finger nucleases. Nature 459, 442–5.

Trinh, T.H., Ratet, P., Kondorosi, E., Durand, P., Kamaté, K., Bauer, P. & Kondorosi, a., 1998. Rapid and efficient transformation of diploid Medicago truncatula and Medicago sativa ssp. falcata lines improved in somatic embryogenesis. Plant Cell Reports 17, 345–355.

Van Den Ackerveken, G., Marois, E. & Bonas, U., 1996. Recognition of the bacterial avirulence protein AvrBs3 occurs inside the host plant cell. Cell 87, 1307–1316.

Vanyushin, B.F. & Ashapkin, V. V, 2011. DNA methylation in higher plants: past, present and future. Biochimica et biophysica acta 1809, 360–368.

Weinthal, D., Tovkach, A., Zeevi, V. & Tzfira, T., 2010. Genome editing in plant cells by zinc finger nucleases. Trends in Plant Science 15, 308–321.

Wendt, T., Holm, P.B., Starker, C.G., Christian, M., Voytas, D.F., Brinch-Pedersen, H. & Holme, I.B., 2013. TAL effector nucleases induce mutations at a pre-selected location in the genome of primary barley transformants. Plant Molecular Biology 83, 279–285.

Wiedenheft, B., Sternberg, S.H. & Doudna, J. a., 2012. RNA-guided genetic silencing systems in bacteria and archaea. Nature 482, 331–338.

Wood, A.J., Lo, T., Zeitler, B., Pickle, C.S., Edward, J., Lee, A.H., Amora, R., Miller, J.C., Leung, E., Zhang, L., et al., 2012. Targeted Genome Editing Across Species Using ZFNs and TALENs. Science 333, 3–6.

Wright, D. a, Li, T., Yang, B. & Spalding, M.H., 2014. TALEN-mediated genome editing: prospects and perspectives. The Biochemical journal 462, 15–24.

Xie, K. & Yang, Y., 2013. RNA-guided genome editing in plants using a CRISPR-Cas system. Molecular plant 6, 1975–1983.

Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E. & Carrington, J.C., 2004. Genetic and functional diversification of small RNA pathways in plants. PLoS biology 2, 642-652.

Yamaguchi-Shinozaki, K. & Shinozaki, K., 2006. Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. Annual review of plant biology 57, 781–803.

Young, N.D., Cannon, S.B., Sato, S., Kim, D.J., Cook, D.R., Town, C.D., Roe, B. a & Tabata, S., 2005. Sequencing the genespaces of *Medicago truncatula* and *Lotus japonicus*. Plant Physiology 137, 1174–1181.

Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R.W., Steward, R. & Chen, X., 2005. Methylation as a crucial step in plant microRNA biogenesis. Science 307, 932–935.

Zhang, F., Maeder, M.L., Unger-wallace, E., Hoshaw, J.P., Reyon, D. & Christian, M., 2010a. High frequency targeted mutagenesis in Arabidopsis thaliana using zinc finger nucleases. Proceedings of the National Academy of Sciences of the United States of America 107, 12028–12033.

Zhang, J.-Y., Cruz de Carvalho, M.H., Torres-Jerez, I., Kang, Y., Allen, S.N., Huhman, D. V, Tang, Y., Murray, J., Sumner, L.W. & Udvardi, M.K., 2014. Global reprogramming of transcription and metabolism in Medicago truncatula during progressive drought and after rewatering. Plant, cell & environment 37, 2553–2576.

Zhang, M., Kimatu, J.N., Xu, K. & Liu, B., 2010b. DNA cytosine methylation in plant development. Journal of Genetics and Genomics 37, 1–12.

Zhang, X., Henderson, I.R., Lu, C., Green, P.J. & Jacobsen, S.E., 2007. Role of RNA polymerase IV in plant small RNA metabolism. Proceedings of the National Academy of Sciences of the United States of America 104, 4536–4541.

Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S.W.L., Chen, H., Henderson, I.R., Shinn, P., Pellegrini, M., Jacobsen, S.E., et al., 2006. Genome-wide High-Resolution Mapping and Functional Analysis of DNA Methylation in Arabidopsis. Cell 126, 1189–1201.

Zhang, Y., Zhang, F., Li, X., Baller, J.A., Qi, Y., Starker, C.G., Bogdanove, A.J. & Voytas, D.F., 2013. Transcription activator-like effector nucleases enable efficient plant genome engineering. Plant physiology 161, 20–27.

Zhu, W., Yang, B., Chittoor, J.M., Johnson, L.B. & White, F.F., 1998. AvrXa10 contains an acidic transcriptional activation domain in the functionally conserved C terminus. Molecular plant-microbe interactions 11, 824–832.

Zilberman, D., Gehring, M., Tran, R.K., Ballinger, T. & Henikoff, S., 2007. Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. Nature genetics 39, 61–69.

7 Supplementary data

Supplementary Table S1: Growth culture media for *E.coli* strains and *S. cerevisiae* strains

***E.coli* growth Media**	**LB liquid Medium**	10g/L Bacto Triptone (DIFCO™)	Autoclave 121°C 15 minutes
		5g/L Bacto Yeast Extract (DIFCO™)	
		10g/L NaCl (Panreac AppliChem)	
		dH_2O up to the desired volume	
	LB Agar Medium	LB liquid medium + 20g/L Agar (DIFCO™)	
		dH_2O up to the desired volume	
	SOC Medium [(a)]Add after autoclave.	20g/L Bacto Triptone (DIFCO™)	
		5g/L Bacto Yeast Extract (DIFCO™)	
		0.5g/L NaCl (Panreac AppliChem)	
		3M KCl (BDH®)	
		1M $MgSO_4$ (BDH®)	
		1M $MgCl_2$[(a)] (Merck S.A)	
		20mM Glucose[(a)] (VWR)	
		dH_2O up to the desired volume	
***S. cerevisiae* Growth Media**	**YPD liquid Medium**	20g/L Peptone (DIFCO™)	Autoclave 121°C 15 minutes
		10g/L Bacto Yeast Extract (DIFCO™)	
		20g/L Glucose (VWR)	
		dH_2O up to the desired volume	
	YPD Agar Medium	YPD liquid medium + 20g/L Agar (DIFCO™)	
		dH_2O up to the desired volume	
	SD Medium (2x)	200mL/L YNB (10x) (DIFCO™)	Solutions are previously autoclaved 115°C 20 minutes
		100mL/L Glucose 40% (VWR)	
		100mL/L Aminoacid mix (10x) (Arginine, isoleucine, lysine, methionine, fenilalanine, threonine, tyrosine, valine (all from Sigma-Aldrich). Without histidine, leucine and uracil.	
		600mL/L dH_2O	

Supplementary Table S2: Primers sequences

Primer name	Primer sequence	Use
DCL3_Fwd	CTAAAAATCTCAAGAGAAAATTCG	TALEN target site amplification by colony PCR
DCL3_Rev	CAAGTTGGCAATGAGGGTTTC	
RDR2_Fwd	CAATGGTGCCCTTTATGCTG	
RDR2_Rev	CCTTTCAGGCATAACCCAAC	
DCL3In_Fwd	GATCACTTGTATGCAAAACCC	TALEN target site amplification by nested PCR
DCL3In_Rev	CTCCAGCTCAGGAAGAAACC	
RDR2In_Fwd	GTTTTCCCATCGGTCCTGAT	
RDR2In_Rev	CATGGGAAGGTGTTCGAGG	
pCR8_F1	TTGATGCCTGGCAGTTCCCT	PCR amplification of pFUS construct and sequence pFUS vector
pCR8_R1	CGAACCGAACAGGCTTATGT	
TAL_F1	TTGGCGTCGGCAAACAGTGG	PCR amplification of pTAL and pZHY constructs
TAL_R2	GGCGACGAGGTGGTCGTTGG	PCR amplification of pTAL and pZHY constructs and Sequence completed TALEN array
TAL_Seq_5-1	CATCGCGCAATGCACTGAC	Sequence completed TALEN array
M13_Fwd	TGTAAAACGACGGCCAGT	Sequence pRS416 vector

Supplementary Table S3: Solutions and reagents

Develop solution (pre- warmed to 55 °C)	3% NaOH (Merck S.A)
	0.1% HCOH (Merck S.A)
	dH_2O up to the desired volume
Formamide dye	90% formamide (Merck S.A)
	10% loading dye (Bromophenol blue 0.2% (Sigma-Aldrich), Xylene Cyanol 0.2% (BDH®), Glycerol 50% (PanReac))
	dH_2O up to the desired volume
LiAc/TE	100mM LiAc (Sigma-Aldrich)
	TE1x
	dH_2O up to the desired volume
Stain Solution	10% ethanol (PanReac)
	0.5% acetic acid
	0.2% silver nitrate
	dH_2O up to the desired volume
Stop solution	10% ethanol
	0.5% acetic acid
	dH_2O up to the desired volume
TAE (10X)	40 mM Tris-base (Nzytech)
	20 mM Acetic acid (Sigma-Aldrich)
	2mM EDTA (Merck S.A.)
	dH_2O up to the desired volume
TBE (10X)	1M Tris-base (Nzytech)
	1mM Boric acid (Merck S.A.),
	2 mM EDTA(Merck S.A.).
	dH_2O up to the desired volume
TE (10X)	10mM Tris-base (Nzytech)
	1mM EDTA (Merck S.A)
	dH_2O up to the desired volume
Transformation solution	PEG 50% (Sigma-Aldrich)
	LiAc 100mM (Sigma-Aldrich)
	TE 1x
	dH_2O up to the desired volume

Supplementary Table S4 The gene target locus TALENs target sequence and binding site RVDs

Gene	Target Site[(a)]	Left Binding Site RVDs	Right BindingSite RVDs	RE sites in spacer
DCL3	GTATGCAAAACCCA AtgttgaacaagattCTCCA GCTCAGGAAG	NN NG NI NG NN HD NI NI NI NI HD HD HD NI NI	HD NG NG HD HD NG NN NI NN HD NG NN NN NI NN	N/A
RDR2	TTCCCATCGGTCCTG ATtgcatgagcgtgcttcaaTC ATGGGAAGGTGTTC G	NG NG HD HD HD NI NG HD NN NN NG HD HD NG NN NI NG	HD NN NI NI HD NI HD HD NG NG HD HD HD NI NG NN NI	HpyCH4 V TGCA

[(a)]Sequences recognized by the TAL effector repeat arrays are in uppercase letters and the spacer sequences are in lowercase letters