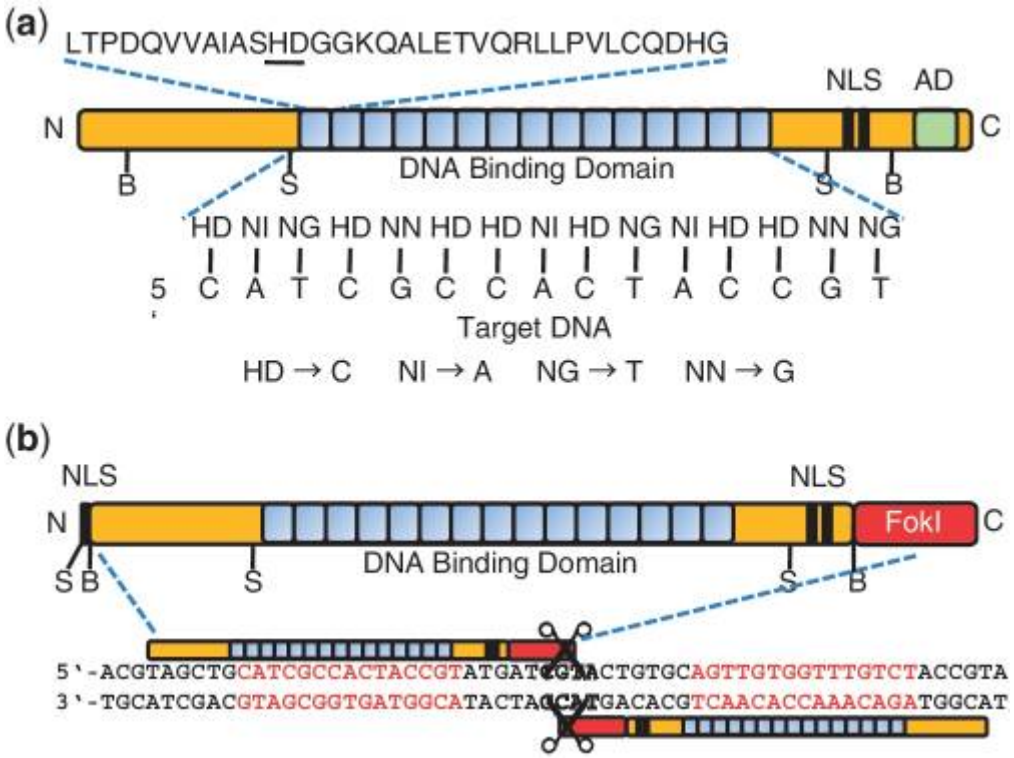
Towards TALEN-mediated Targeting and Exclusion of Immunogenic Epitopes of Gluten Proteins in Wheat Species Causing Celiac Disease



Jarst van Belle (author)

dr. ing. Jan G. Schaart (supervisor)

MSc Thesis Wageningen UR Plant Breeding

14 February 2014

Index

Abstract

Celiac disease (CD) is an autoimmune disorder that affects the small intestine of the patient causing it to become chronically inflamed when gluten containing immunogenic epitopes are digested. Patients have to follow a gluten-free diet which is difficult because a broad range of consumable products contains gluten, mainly from wheat. Gluten proteins (gliadins and glutenins) are important for dough and baking quality and so far studies have not managed to obtain gluten-free wheat without reducing its quality. A molecular technique using TALENs (transcription activator-like effector nucleases) may offer a solution as TALENs are able to specifically target and engineer DNA sequences, and may therefore exclude immunogenic epitopes from gluten proteins while conserving the industrial properties of wheat. TALENs have been designed targeting genes encoding immunogenic α-gliadins, located on the short arms of the group 6 chromosomes of *Triticum aestivum*. However, eventually a more specific TALENs than intended was constructed targeting a limited number of α-gliadin genes on chromosome 6A only. The construct is ready for wheat transformation and can still be used for further research, although other molecular constructs targeting at least chromosome 6D, containing the most immunogenic epitopes and in the highest frequencies, should be constructed as well.

Abbreviations: CD, celiac disease; TALEN, TAL (transcription activator-like) effector nuclease; GS, glutenin subunit; HMW, high molecular weight; LMW, low molecular weight; RVD, repeat-variable di-residue; ZFN, zinc finger nuclease.

Introduction

Celiac disease (CD) is an autoimmune disorder that affects the small intestine of the patient causing it to become chronically inflamed when gluten (mainly from wheat, *Triticum spp.*, but also barley and rye) is digested. The immune system generates a response to gluten proteins (certain gliadins and glutenins) and attacks the intestinal tissue of the patient, which leads to a decreased surface area of the small intestine causing malabsorption of nutrients and therefore malnutrition (Verdu et al. 2009). Patients with CD may have a broad range of symptoms (e.g. indigestion, constipation and fatigue) and have also an increased chance of getting intestinal cancer (Catassi et al. 2005). Both genetic and environmental factors contribute to disease development (Stepniak & Koning 2006). The prevalence of CD is approximately 1% of the general population (Niewinski 2008). Besides CD, there exists non-celiac gluten sensitivity, which differs from CD in that this sensitivity is not an autoimmune disorder and the intestinal tissue is not damaged, or damaged to a smaller extent. The inflammation, as well as the symptoms that may appear, will disappear when gluten is not digested by the patient anymore (Verdu et al. 2009).

In order to deal with CD, patients must follow a gluten-free diet, or in case of a mild form of non-celiac gluten sensitivity a diet low in gluten. This means that food products based on wheat

(bread, pasta, cake, etc.), but also barley and rye must be avoided. Many diets in the world rely for the major part on wheat-based products and although it is relatively easy to find replacements for bread and pasta, it must be mentioned that most commonly consumed products are in many cases processed with wheat starch containing gluten (soups, sauces, candies, etc.). Therefore, it is especially difficult for patients with CD or non-celiac gluten sensitivity to find products that do not contain gluten. Gluten-free products must also have adequate nutritional values that compensate for the alternative diet. Moreover, products like gluten-free bread will have a reduced quality, since gluten proteins are very important for dough and bread-baking quality. Gliadins give the dough the ability to expand, while glutenins also give elasticity (Belderok et al. 2000). The balance of both gluten proteins is considered to be important, whereas a too low elasticity leads to a lower volume of the product and a too high elasticity interferes with gas expansion, also leading to a lower volume (Barak et al. 2012).

The aim of the current research is to obtain gluten-free wheat, in which gluten genes that encode proteins with immunogenic epitopes are deleted, while industrial properties are maintained. Wheat flours contain approximately 10% gluten proteins (gliadins and glutenins). The gliadins can be further divided into α/β/γ/ω-gliadins and the glutenins into HMW-GS and B/C/D-type LMW-GS (high and low molecular weight glutenin subunits, respectively) (Van den Broeck et al. 2009/2011; Gilissen 2013). Gluten proteins are encoded by 15 major loci. In *T. aestivum*, an allohexaploid wheat species consisting of three diploid genomes (A, B and D), loci encoding γ/ω-gliadins (*Gli-A1*/*-B1*/*-D1* and *Gli-A3*/*-B3*/*-D3* respectively) and LMW-GS (*Glu-A3/-B3/-D3*) are tightly linked and located on the short arms of group 1 chromosomes (Singh & Shepherd 1988), while loci encoding HMW-GS (*Glu-A1*/*-B1*/*-D1*) are located on the long arms of group 1 chromosomes (Harberd et al. 1986). Most α/β-gliadins are encoded by loci located on the short arms of group 6 chromosomes (*Gli-A2*/*B2*/*D2*) (Marino et al. 1996), of which the α-gliadins encoded by chromosome 6D (*Gli-D2*) contain the most epitopes and in the highest frequencies (Van Herpen et al. 2006). α-gliadin epitopes are considered the most immunogenic (Arentz-Hansen et al. 2000a/2000b/2002; Camarca et al. 2009; Janatuinen et al. 2002; Maiuri et al. 2003; Molberg et al. 2003; Schuppan et al. 2003; Vader et al. 2002), of which the Glia-α9 is a major immunodominant epitope and part of a 33-mer (Shan et al. 2002/2005) with a high T-cell stimulatory effect. The 33-mer sequence (LQLQPFPQPQLPYPQPQLPYPQPQLPYPQPQPF) is only present in D-genome protein sequences of α/β-gliadins (Van den Broeck et al. 2011). An overview of gluten proteins in *T. aestivum* is shown in **Table 1**.

Table 1: *Overview of gluten proteins in T. aestivum with subtypes, genomic positions and loci. LMW-glutenins and γ/ω-gliadins are located on the short arms of group 1 chromosomes in genomes A, B and D and are closely linked. α/β-gliadins are located on the short arms of group 6 chromosomes. α-gliadins encoded by Gli-D2 (chromosome 6D) contain the most (immunogenic) epitopes and in the highest frequencies.*

Gluten Protein	Subtype	Chromosome Group	Loci
Glutenins	HMW-GS	1 (long arms)	*Glu-1* (A/B/D)
	LMW-GS	1 (short arms)	*Glu-3* (A/B/D)
Gliadins	γ		*Gli-1* (A/B/D)
	ω		*Gli-3* (A/B/D)
	α	6 (short arms)	*Gli-2* (A/B/D)
	β		

Before, studies have been performed where parts of chromosome arms containing entire loci for gluten proteins were deleted (Van den Broeck et al. 2009/2011). Deletion of loci encoding ω/γ-gliadins and LMW-glutenins reduces the number of T-cell stimulatory epitopes and causes minor deterioration of dough quality while deleting loci encoding α-gliadins results in a significant decrease in T-cell stimulatory epitopes, but also significantly alters dough quality (Van den Broeck et al. 2009/2011). These findings imply that it is not feasible for α-gliadins to delete entire parts of chromosome arms due to the great loss in quality. However, the highly immunogenic α-gliadins on the short arms of group 6 chromosomes are grouped together in a cluster of duplicated genes (Mecham et al. 1978), which is located between the centromeres and the remaining chromosomal parts of the arms (Van den Broeck et al. 2009/2011).

Van den Broeck et al. (2011) has described that focusing on the removal of α-gliadin genes from the short arm of chromosome 6D in wheat may be an important strategy toward the development of CD-safe wheat and food products, as approximately 50% of CD-patients respond to the previously described 33-mer (Shan et al. 2002/2005) that contains the immunodominant epitopes Glia-α2 and Glia-α9 (Camarca et al. 2009, Vader et al. 2002) encoded by the *Gli-D2* locus, and this locus may be deleted from wheat while maintaining its industrial properties. Preferably, the *Gli-D2* locus needs to be deleted without eliminating larger parts of the chromosomes as this may affect the quality and the fitness of the crop (Van den Broeck et al. 2011). During the current research an attempt will be made to elaborate this strategy by focusing on the deletion of the cluster of highly immunogenic α-gliadin genes on the short arm of chromosome 6D, but also 6A and 6B. Due to close linkage the cluster will always be inherited together with the remaining chromosomal material, which means that plant breeding holds no possibilities (Mecham et al. 1978). Even with a beneficial random genetic mutation in an epitope of an α-gliadin induced by mutagenesis, there are still many other functional immunogenic α-gliadins present. As a conclusion, gene technology may offer possibilities for targeting and cutting out specific genomic sequences containing only immunogenic α-gliadins.

This report focusses on the application of a recent methodology that involves TALENs (transcription activator-like effector nucleases), which are considered important and useful tools for genome engineering (Cermak et al. 2011; Zhang et al. 2013). Basically, a TALEN consists of a TAL effector of plant pathogenic *Xanthomonas spp.* which functions as a DNA-binding domain, and a *Fok*I nuclease that functions as a DNA cleavage domain. Two monomeric TALENs are required to bind the target site to enable the *Fok*I nucleases to dimerize and cleave DNA (see also **Figure 1**; Cermak et al. 2011). The target nucleotide sequence is determined by the sequence of RVDs (repeat-variable di-residues) in the TAL effector repeat array of the DNA-binding domain, and can be custom-designed. Compared to ZFNs (zinc finger nucleases), a previously developed method for site-specific genome manipulation, a major advantage of TALENs is that the DNA-binding domain can be easily engineered to recognize virtually any DNA sequence (Cermak et al. 2011; Reyon et al. 2012), while the possibilities with ZFNs are more limited (Cermak et al. 2011). This is due to the fact that each RVD in TALENs can recognize a single nucleotide, while the equivalent ZFNs recognize triplets, and not every triplet can be recognized by a ZFN yet. Some RVDs can even recognize multiple nucleotides.

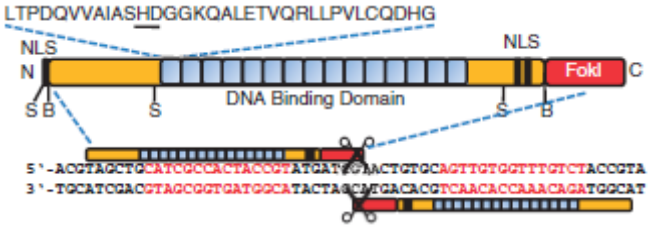


Figure 1: *Structure of a TALEN. Two monomeric TALENs are required to bind the target site to enable FokI to dimerize and cleave DNA. A consensus repeat sequence is shown with the RVD (repeat-variable di-residue) underlined. The sequence of RVDs (the TAL effector repeat array) determines the target nucleotide sequence. In this consensus repeat, the RVD in the first repeat unit is HD (underlined; amino acids Histidine and Aspartic Acid), which is most frequently associated with C nucleotides. NLS, nuclear localisation signal(s); B, BamHI; S, SphI. The figure is derived from Cermak et al. (2011) and edited for convenience.*

Since the highly immunogenic α-gliadin genes are grouped together as a large repeat (Mecham et al. 1978) and the α-gliadin protein family is highly conserved (Anderson & Greene 1997), it would be possible to target many α-gliadins at once with only up to a few constructs and possibly only one ligation would be needed to ligate the chromosome arms remainders (with the remaining genes) with the corresponding chromosomes. Concluding, using TALENs (further elaborated in *Methods*) may be of great aid of obtaining wheat with greatly reduced numbers of immunogenic epitopes.

Methods

In silico TALENs design

TALENs targeting α-gliadins of *Triticum aestivum* were designed using TALE-NT (Cermak et al. 2011; Doyle et al. 2012). Using this program, two TALEN pairs were designed of which one would target conserved regions of highly immunogenic α-gliadins located on the short arms of group 6 chromosomes and one would target α-gliadins located on chromosomes 6A and 6D but not 6B. The latter TALEN pair was designed to possibly circumvent lethal genotypes due to the high amount of deleted chromosomal material, while still targeting the most immunogenic α-gliadins. In total 32 sequences were used (provided by Salentijn et al. 2013), of which 17 represented α-gliadin sequences of the A-genome, 10 of the B-genome and 5 of the D-genome of which 2 sequences contained the 33-mer sequence with a high T-cell stimulatory effect; Shan et al. (2002/2005). **Figures 2** and **3** describe the sequences (**Figure 2**) and alignment (**Figure 3**) that were used in MEGA v5.10 (Tamura et al. 2011) for designing the TALENs, as well as the resulting target sites (**Figure 3**).

The preset architectures of the TALENs included RVD lengths targeting 15 to 20 bases. Spacer lengths of 15 to 20 bases were chosen since they were intermediate lengths considering the available options offered by TALE-NT. G nucleotides were primary targeted by NH, an RVD that binds G more specifically than NN (Cong et al. 2012; Streubel et al. 2012). Moreover, at least 1 out of 4 bases were C or G and runs of 6 or more A or T nucleotides were not allowed as suggested by Streubel et al. (2012) in order to obtain more reliable TALENs with sufficient 'strong' RVDs. The bases that preceded the target sites on the 5' ends were T nucleotides as recommended by TALE-NT (Boch et al. 2009; Moscou and Bogdanove 2009).

The α-gliadin sequences on the short arms of group 6 chromosomes to be targeted were (T)AAGGAACCTAGCCCTacagacgctacctgcAATGTGCAATGTCTAY(A) for chromosomes 6A, 6B and 6D and (T)ASAGACGCTACCTGCAATGTgcaatgtctacatccctccaTATTGCACCATCGCKCC(A) for chromosomes 6A and 6D, but not 6B (see **Figure 3**), of which the bases in upper case represent the actual DNA-binding domains of the TALENs and the bases in lower case represent the DNA target domains. The bases between brackets represent the bases upstream of the target sites. Note that the second TAL array of each TALEN pair binds to the negative strand, while only the positive strand sequence is described above. Reverse complement sequences should be interpreted for the design of the second TAL arrays; this is automatically done by TALE-NT.

The designed TAL arrays resulting from the target sites for the first TALENs were NI NI NH NH NI NI HD HD NG NI NH HD HD HD NG for the first DNA-binding domain (TAL array 1) and NN NG NI NH NI HD NI NG NG NH HD NI HD NI NG NG for the second DNA-binding domain (TAL array 2); see also *Results* for a comprehensive view. Note that the first RVD of TAL array 2 (NN) was chosen

after the actual TALE-NT design to allow binding to both A and G as found in the α-gliadin sequences of chromosome 6D and chromosomes 6A and 6B respectively (see **Figure 3**). It should be mentioned that unintentionally a more specific TALENs was constructed than intended due to focusing on an α-gliadin sequence with specific variation. Base 13 of the originally designed TALENs included a G instead of a C, which would result in the target site AAGGAACCTAGCGCT and TAL array 1 of the first TALENs would then be NI NI NH NH NI NI HD HD NG NI NH HD NH HD NG instead (see also *Results*). The actually constructed TALENs will be compatible with 3 of the A-genome α-gliadins used in the alignment (see **Figure 3**).

The designed TAL arrays resulting from the target sites for the second TALENs were NI HD NI NH NI HD NH HD NG NI HD HD NG NH HD NI NI NG NH NG for TAL array 1 and NH NH NP NH HD NN NI NG NH NH NG NH HD NI NI NG NI for TAL array 2. Again, NN was chosen after the actual TALE-NT design to allow binding to A (chromosome 6D) and G (chromosome 6A). Regarding TAL array 2, the novel RVD NP can be used to allow binding to A (chromosome 6D) and C (chromosome 6A) as it shows specificity for A, C and T (Streubel et al. 2012). Regarding TAL array 1, there currently is no RVD that allows binding to both C and G. The RVD NS, which previously was designated as nonspecific (Boch et al. 2009), has showed a clear preference for A and G in more recent studies (Streubel et al. 2012) and can therefore not be used. Until a novel RVD has been discovered, the RVD HD can be used, which is specific for C and will recognize all A-genome α-gliadins and still 2 out of 5 D-genome α-gliadins (see **Figure 3**).

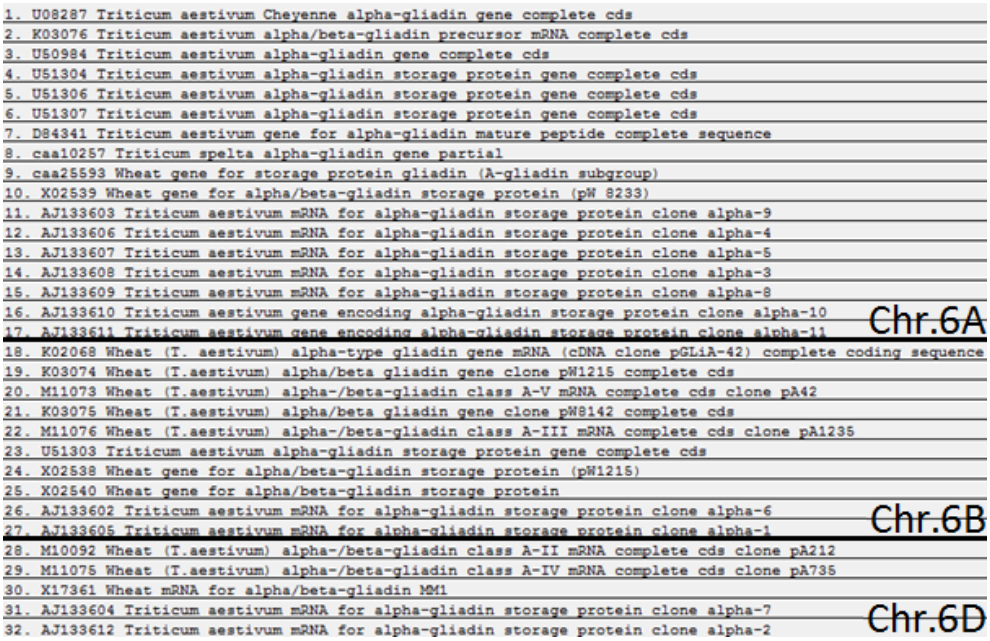


***Figure 2**: Origin of the α-gliadin sequences (provided by Salentijn et al. 2013) that were used for designing the TALENs using MEGA v5.10 (Tamura et al. 2011). The DNA sequences were aligned in the above order (see **Figure 3**). 1-17, 18-27 and 28-32 represent α-gliadin sequences located on the short arms of wheat chromosomes 6A, 6B and 6D respectively, of which 30 and 32 contained the 33-mer with a high T-cell stimulatory effect for CD-patients as described by Shan et al. (2002/2005; not shown).*

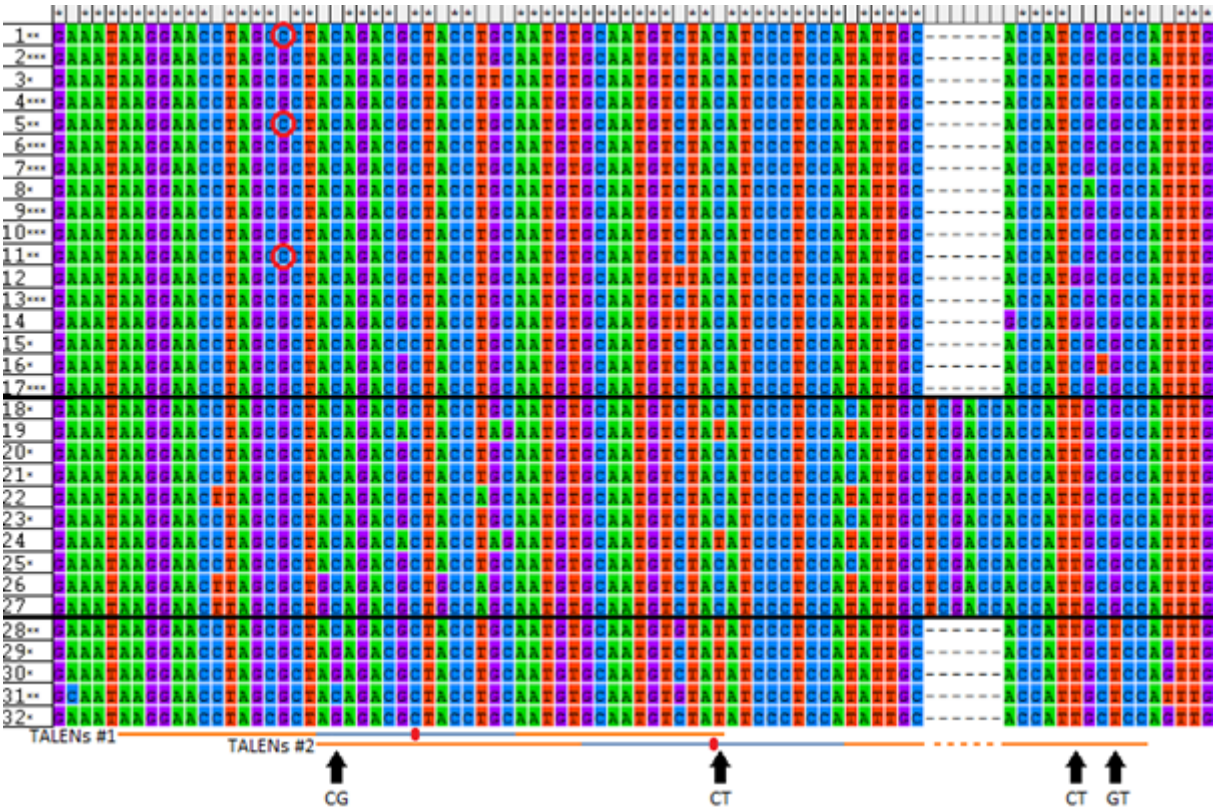


Figure 3*: Alignment (partly shown) of coding sequences of α-gliadins (provided by Salentijn et al. 2013; see* ***Figure 2*** *for the origin of the sequences) using MEGA v5.10 (Tamura et al. 2011), and the resulting target sites for the in silico designed TALEN pairs (using TALE-NT; Cermak et al. 2011) targeting α-gliadins on the short arms of wheat chromosomes 6A (1-17), 6B (18-27) and 6D (28-32) (TALENs #1) and 6A and 6D, but not 6B (TALENs #2). The orange parts of the TALENs represent the DNA-binding domains of which each RVD recognizes and binds a specific base (the dotted line indicates untargeted bases), while the blue parts represent the spacer DNA between these domains, where the FokI nucleases of each TAL array dimerize and cleave the DNA (red dots indicate predicted cutting sites). The second TALEN pair will not be able to target B-genome α-gliadins on as these α-gliadins contain a genome-specific insert of 6bp which disallows recognition. Of all 32 sequences, 20 were compatible with TALENs #1 and 13 with TALENs #2, of which 8 sequences were compatible with both TALENs (indicated by * and ** or *** respectively). Note that B-genome α-gliadins do not count towards sequences compatible with TALENs #2 as the B-genome is not targeted. Sequence 2 (any sequence indicated by ***) could be used as a consensus for designing both TALEN pairs, although before sequence 1 was used which contains an uncommon SNP (single nucleotide polymorphism) variant (indicated by red circles), which resulted in a constructed TALENs #1 that is only compatible with 3 A-genome α-gliadins and will therefore be more specific than intended. Other important SNPs present at the DNA-binding domains of the TALENs are indicated by arrows. The C/T-SNPs (actually A/G-SNPs: reverse complement sequences should be interpreted for the design of the second TAL arrays as these arrays bind to the negative strand, while only the positive strand sequence is shown), which discriminate between α-gliadins located on different genomes, were overcome by using the RVD NN that allows binding to both A and G. The G/T-SNP (actually an A/C-SNP) of the second TAL array of TALENs #2 can be overcome by using the novel RVD NP, but no appropriate RVD has been found yet for the C/G-SNP of the first TAL array of TALENs #2. Until a novel RVD has been discovered, sequences 29-30 and 32 are not compatible with TALENs #2 and the RVD HD can be used, which is specific for C and will recognize all A-genome α-gliadins and still 2 out of 5 D-genome α-gliadins.*

In vitro TALENs assembly

The TALEN pair (targeting a limited number of α-gliadins on the short arm of chromosome 6A) was constructed *in vitro* based on the Golden Gate TALEN assembly as described by Cermak et al. (2011) using module, array, last repeat and backbone plasmids with a few alterations as described by Zhang

et al. (2013) to further enable efficient plant genome engineering. TALEN backbone plasmids pTAL3/pTAL4 (**Figure 4**) include different yeast selection markers *HIS*3 and *LEU*2 respectively, but these markers were not used in this study which makes these plasmids practically identical. Moreover, deletions in the coding sequence result in NΔ152/C63 truncations of the TAL portion of the protein resulting into plasmids pZHY500/pZHY501 (derived from TAL3/TAL4 respectively; **Figure 4**; Zhang et al. 2013), of which pZHY500 was used for TALEN assembly. TAL arrays were constructed by assembling single repeat modules into intermediate arrays of up to 10 repeats and joining the intermediate arrays together with a last repeat for each TAL array into a backbone. Full details are (visually) described by Cermak et al. (2011). See *appendix i* for the complete eventually performed Golden Gate TALEN assembly protocol as described by Cermak et al. (2011), including alterations by Zhang et al. (2013) and additional alterations.

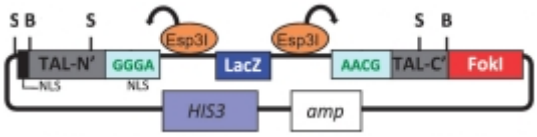


Figure 4*: TALEN backbone plasmid as described by Cermak et al. 2011 (pTAL3 is shown in the figure). pTAL4 differs from pTAL3 as pTAL4 contains the yeast selection marker LEU2 instead of HIS3. The plasmids are practically identical as these markers were not used in this study. pZHY500 and pZHY501 are derived from pTAL3 and pTAL4 respectively which contain truncations in the N-terminal and C-terminal portions of the TAL protein (TAL-N' and TAL-C' respectively) to further enable plant genome engineering (Zhang et al. 2013). GGGA and AACG represent cohesive overhangs for insertion of a complete assembled TAL array which will replace the LacZ gene that can be used as a marker to verify transformed clones with Blue White screening (without insertion of the TAL array LacZ is still present in the plasmid and expressed resulting in blue colonies, while successfully transformed clones contain the TAL array instead of LacZ resulting in white colonies). The restriction enzyme Esp3I is used to cut open the backbone plasmid for insertion of the TAL array. B (BamHI) and S (SphI), useful for subcloning custom repeat arrays; NLS, nuclear localization signal(s); amp, ampicillin resistance; FokI, FokI nuclease coding sequence.*

Prior to the actual TAL arrays assembly, stocks were made of each plasmid including pZHY013 and pZHY051 (**Figure 5**; pZHY013 is a Gateway-compatible entry plasmid with two heterodimeric *Fok*I nuclease domains (Miller et al. 2007) separated by a T2A translational skipping sequence (Halpin et al. 1999), while pZHY051 is the destination vector which contains a 35S promoter and a NOS terminator to drive expression of the TALEN pair; Zhang et al. 2013) and the quality of each plasmid was verified by measuring its concentration (see *appendix iii*) and by sequencing (see *appendix iv*). Module plasmids pHD/pNG/pNI/pNN/pNK/pNH1-10 were sequenced with primers pRVD-seq FW/REV (5'-gtgagtcgtattacatggtc-3' and 5'-tctgacgctcagtggaacga-3' respectively). Last repeat plasmids pLR-HD/NG/NI/NH (pLR-NK was not available) were sequenced with primer pT14_F1 (5'-cctactcaggagagcgttca-3') and backbone plasmids pZHY013/051/500/501 were sequenced with primers M13 F/R. Finally, array plasmids pFUS_A/A30A/A30B, pLR-FUS_B1/B2 and pFUS_B3-B10 were sequenced with primers pCR8_F1/R1 (5'-ttgatgcctggcagttccct-3' and 5'-cgaaccgaacaggcttatgt-3'

respectively). The obtained sequences were analyzed and verified using the free software program ApE v2.0.45 (A plasmid Editor; M. Wayne Davis, University of Utah), which can be used as a sequence analysis tool.

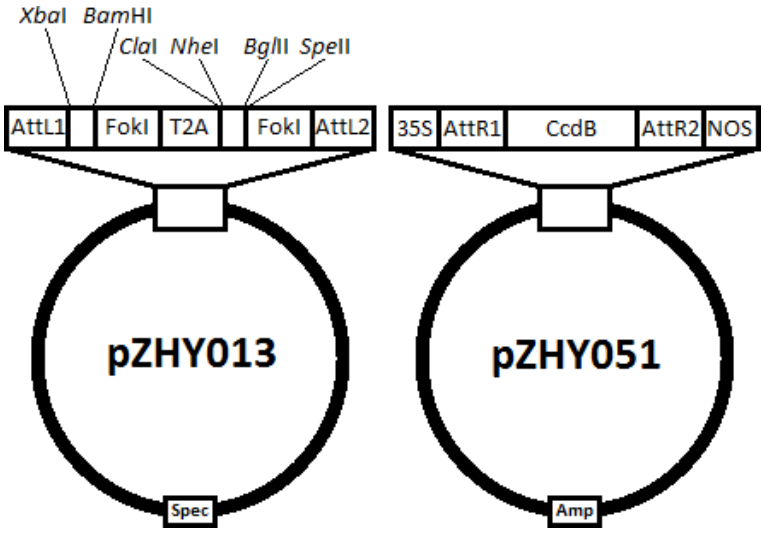


***Figure 5**: Schematic overview of plasmids pZHY013 and pZHY051. pZHY013 is a Gateway-compatible entry plasmid with two heterodimeric FokI nuclease domains (Miller et al. 2007) separated by a T2A translational skipping sequence (Halpin et al. 1999), while pZHY051 is the destination vector which contains a 35S promoter and a NOS terminator to drive expression of the TALEN pair (Zhang et al. 2013). As described in detail below, complete TAL arrays can be cloned one by one into pZHY013 and the TALENs can subsequently be moved into pZHY051 with a Gateway reaction (see **Figure 6** for a complete cloning scheme). XbaI/BamHI, restriction sites for insertion of a first TAL array; NheI/BglII, restriction sites for insertion of a second TAL array; ClaI/SpeII, restriction sites for restriction digestion for verification of the second insert as the NheI and BglII restriction sites are not present anymore after insertion; FokI, FokI nuclease coding sequence; Att, recognition sites for Gateway cloning; CcdB, a negative selection marker prohibiting the growth of the plasmid without the TALENs construct; Spec, spectinomycin resistance; Amp; ampicillin resistance. Read the text below under 'Preparing TALENs for efficient plant genome engineering' for full comprehension.*

For the construction of the TALENs eventually 4 intermediate TAL arrays were assembled which were TAL1A and TAL1B (belonging to TAL array 1 or TAL1) and TAL2A and TAL2B (belonging to TAL array 2 or TAL2). The 4 corresponding Golden Gate reactions included the following plasmids: pNI1, pNI2, pNH3, pNH4, pNI5, pNI6, pHD7, pHD8, pNG9, pNI10, pFUS_A (TAL1A); pNH1, pHD2, pHD3, pHD4, pFUS_B4 (TAL1B); pNN1, pNG2, pNI3, pNH4, pNI5, pHD6, pNI7, pNG8, pNG9, pNH10, pFUS_A (TAL2A); pHD1, pNI2, pHD3, pNI4, pNG5, pFUS_B5 (TAL2B).

In short, the first series of performed Golden Gate reactions aimed at generating intermediate TAL arrays containing ten repeats, was followed by a Plasmid-Safe nuclease treatment which destroyed all unligated linear dsDNA fragments including incomplete ligation products with lower number of repeats fused, and cut and linearized vectors. The incomplete, shorter fragments would be cloned into the destination vector *in vivo* by recombination in the bacterial cell, if not removed (Cermak et al. 2011). The Plasmid-Safe nuclease treatment was followed by transformation

of chemically competent TOP10 (*Escherichia coli*) cells according to the TOPO TA Cloning Kit for Sequencing protocol (Invitrogen). Overnight cultures were obtained from white colonies and used for bacterial glycerol stocks and isolating plasmid DNA for further TAL arrays assembly. Prior to the second series of Golden Gate reactions to assemble the intermediate TAL arrays with the last repeat and the backbone, the plasmids were verified by double restriction digestion with enzymes *Bsp*TI and *Xba*I (Fermentas) and by PCR and sequencing with primers pCR8_F1/R1, of which the obtained sequences were analyzed and verified using ApE. See *appendix v* for sequence analyses during the Golden Gate TALEN assembly protocol. The 2 Golden Gate reactions included the following plasmids: TAL1A, TAL1B, pLR-NG, pZHY500 (TAL1); TAL2A, TAL2B, pLR-NG, pZHY500 (TAL2).

Again, the performed Golden Gate reactions were followed by transformation (a Plasmid-Safe nuclease treatment was in this case not necessary according to the Golden Gate TALEN assembly protocol; Cermak et al. 2011), overnight cultures, bacterial glycerol stocks, isolating DNA and eventually analyzing and verifying the obtained plasmids. Enzymes *Stu*I and *Aat*II (Fermentas) were used for double restriction digestion. Primers TAL_F1/R2 (5'-ttggcgtcggcaaacagtgg-3' and 5'-ggcgacgaggtggtcgttgg-3' respectively) were used for PCR, while primers SeqTALEN 5-1 (5'-catcgcgcaatgcactgac-3') and TAL_R2 were used for sequencing, of which the obtained sequences were analyzed and verified using ApE (see *appendix v*), but also SeqMan Pro (DNASTAR) since ApE failed to align sequences obtained with SeqTALEN 5-1 well.

Preparing TALENs for efficient plant genome engineering

After the construction of the complete TAL arrays, TAL1 and TAL2 were to be cloned into pZHY013, a Gateway-compatible entry plasmid with two heterodimeric *Fok*I nuclease domains (Miller et al. 2007) separated by a T2A translational skipping sequence (Halpin et al. 1999), followed by a Gateway reaction to move the TALENs to destination vector pZHY051 which contains a 35S promoter and a NOS terminator to drive expression of the TALEN pair (Zhang et al. 2013); see **Figures 5 and 6**. See *appendix ii* for the complete protocol.

First TAL1 and TAL2 were released from pZHY500 using Fast Digest (FD) enzymes *Xba*I and *Bam*HI (Fermentas) and gel purified according to the QIAquick Gel Extraction Kit Protocol (using a microcentrifuge) by Qiagen. pZHY013 was also gel purified for removal of the insert at the first insertion site using these enzymes, followed by a ligation reaction with TAL1. Chemically competent TOP10 cells (Invitrogen) were transformed for overnight cultures and DNA isolation (bacterial glycerol stocks were forgotten). pZHY013/TAL1 (PT1) clones were verified by double restriction digestion with FD *Xba*I and *Bam*HI.

The PT1 clones were further gel purified for removal of the insert at the second insertion site using FD enzymes *Nhe*I and *Bgl*II (Fermentas) with *Xba*I and *Bam*HI compatible ends. A ligation

reaction with TAL2 was performed, followed by transformation for an overnight culture (one colony was found), a glycerol stock and DNA isolation. The pZHY013/TAL1/TAL2 (PT12) clone was verified by double restriction digestion with FD enzymes *Cla*I and *Spe*I (Fermentas) since the prior recognition sites were not present anymore. The PT12 clone was also sequenced to verify the TAL1 and TAL2 inserts with primers TAL1FW/REV (5'-aggaccacgatattgactac-3' and 5'-atcagctcgatgtactcgtg-3' resp.) and TAL2FW/REV (5'-acaaggatgacgatgacaag-3' and 5'-cctgagtggaatttctggca-3' resp.) respectively and analyzed using ApE v2.0.45 (A plasmid Editor, M. Wayne Davis, University of Utah). See *appendix vi* for sequence analyses during the preparation of the TALENs for plant genome engineering.

After verification of the clone, an LR Gateway reaction (Invitrogen) was performed with entry PT12 and pZHY051 (**Figures 5** and **6**) followed by transformation, overnight cultures and DNA isolation (bacterial glycerol stocks were initially forgotten but created later). The inserts were verified by sequencing the inserts from the inserts with primer TAL1REV and from pZHY051 with primers P-35S FW and T-NOS R1 and analyzed using ApE (see *appendix vi*). Prior to the actual transformation experiments, highly concentrated DNA was obtained by using the Plasmid Midi Kit (Qiagen); see *appendix viii*. A schematic overview of preparing the TALENs for efficient plant genome engineering is shown in **Figure 6**.

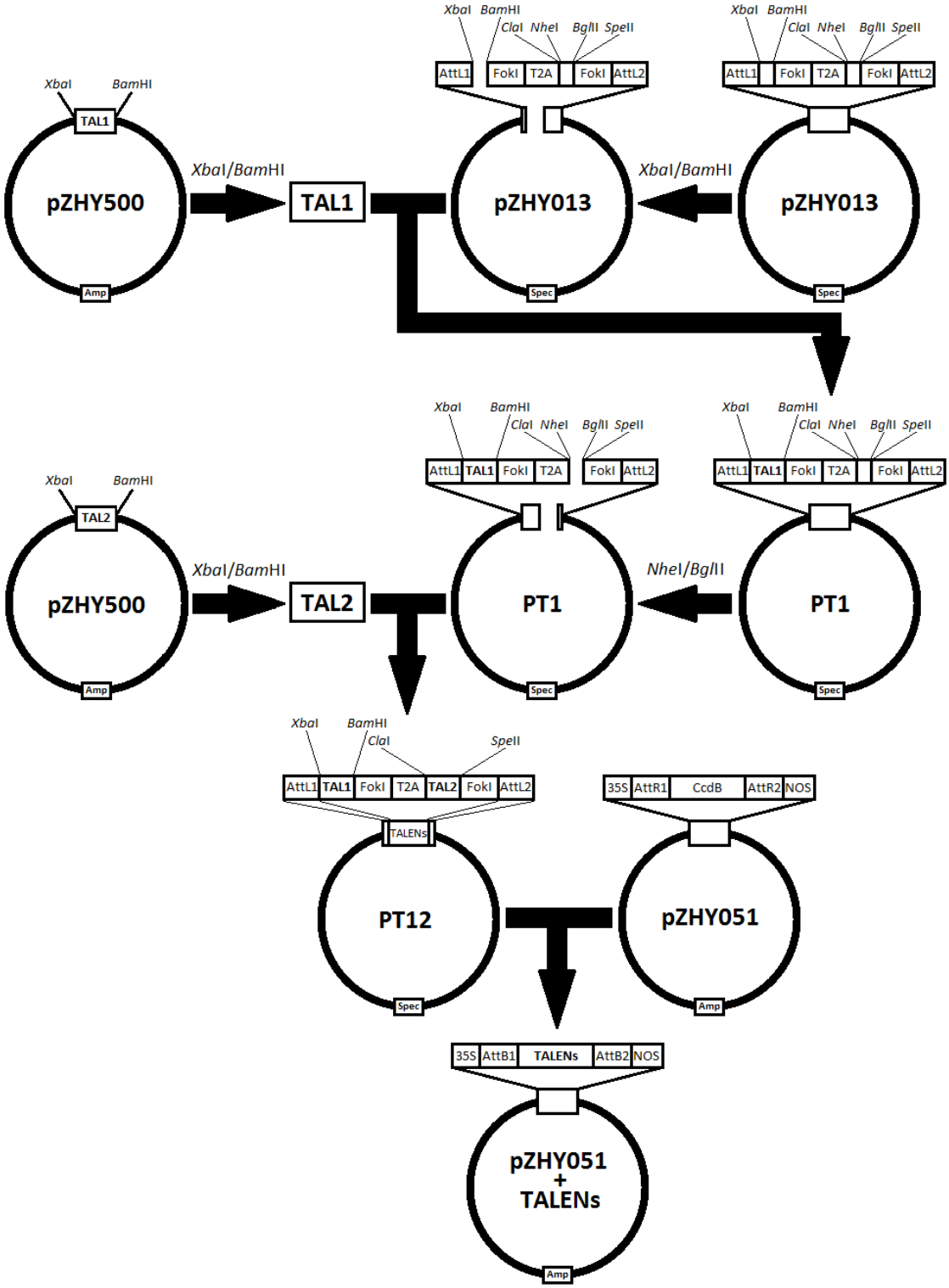


***Figure 6**: Schematic overview of preparing the TALENs for expression in plants. TAL1 and TAL2 were released from pZHY500 using (FD) XbaI/BamHI and ligated into pZHY013 (containing two FokI domains and a T2A translational skipping sequence) one after the other. pZHY013 was purified using XbaI/BamH1 before ligating TAL1, and the resulting PT1 was purified using (FD) NheI/BglII (with XbaI/BamHI compatible ends) before ligating TAL2. PT1 was checked for TAL1 insertion by XbaI/BamHI double restriction, but PT12 was checked for TAL2 insertion by (FD) ClaI/SpeII, since prior restriction sites were not present anymore. PT12 was also checked for TAL1/TAL2 by sequencing from the TAL arrays. Finally, an LR Gateway reaction was performed to move the TALENs to pZHY051 which contains a constitutive promoter (35S) and terminator (NOS) to drive expression in plants. pZHY051 was verified by sequencing from the TALENs itself (from TAL1) and from pZHY051 (from the promoter and the terminator).Amp, ampicillin resistance; Spec, spectinomycin resistance; Att, recognition sites for Gateway cloning; CcdB, a negative selection marker prohibiting the growth of the plasmid without the TALENs construct.*

Wheat seeds sterilization and wheat protoplasts isolation and transformation

The next step after having obtained the TALENs in a plant genome engineering vector is to test functionality of the constructed TALENs by transformation of the constructs to wheat protoplasts. Protoplasts will first be generated from shoot tissue from ~1.5 weeks (preferably ~10 days) old wheat seedlings grown under sterile conditions. A protocol still had to be developed for the sterilization of wheat seeds and a series of experiments was performed in the process. The basic setup of the experiments included optional pre-incubations with sterile water and/or 70% ethanol varying in length, incubations with disinfectant NaOCl (sodium hypochlorite) varying in length and concentration and with or without surfactants tween 20 or Triton X-100, and several washing steps with sterile water. Pre-incubating the wheat seeds with ethanol would effectively reduce surface contamination (Sauer and Burroughs 1986). The seeds were initially grown in large groups of 15 seeds, but since it was not possible to obtain exclusively disinfected germinating seeds during the performed experiments, while under optimal conditions at least 80% remained free from infections (12 out of 15 seeds), seeds were grown in smaller groups of 3 during later experiments to minimize the risk of cross-contamination of seeds. Infection rates of the seeds and the germination capacity were observed towards the development of an optimal protocol. A complete report on the performed experiments can be found in *appendix vii*; the eventually adopted wheat seeds sterilization protocol is as follows:

Under sterile conditions, *Triticum monococcum* (einkorn; 2n=2x=14, AA) (Shi et al. 1998) seeds were pre-incubated with 70% ethanol for 2min in sieves holding 15 seeds each and consequently incubated with 4% NaOCl (without any surfactant) for 5min, followed by 5 washing steps in sterile water for 3min per step. During each step the sieves were shaken continuously. Seeds were grown in small groups of 3 in sterile tubes with MS20 medium (4.4g·L^{-1} MS with vitamins, 20g·L^{-1} sucrose, 9g·L^{-1} agar) in a climate room for ~1.5 weeks (preferably ~10 days) before harvesting shoot tissue.

Wheat protoplasts were generated from shoot tissue and transformed with highly concentrated DNA according to the protocol as described by Shan et al. (2013), which is based on protocols from Yoo et al. (2007) and Zhang et al. (2011). Fresh green einkorn wheat shoots (preferably ~10 days old) were cut into fine strips and vacuum-infiltrated in the dark, followed by a digestion step with an enzyme solution. Protoplasts were isolated by filtration and eventually transformed using a PEG solution and harvested and cultured at room temperature in the dark for two days. The complete protocol, including additions and alterations, can be found in *appendix viii*. Although wheat protoplasts were successfully isolated, no transformation experiments were carried out yet.

Additional research: determining the mutation level of ZFN-mutated potato by restriction analysis

Next to the main research of this report, some additional experiments were performed on *Solanum tuberosum* (potato; 2n=4x=48) (Ortiz and Peloquin 1994) *cv. Karnico* plants that were transformed with a specific ZFNs construct targeting DNA of all four alleles of the starch-binding enzyme II-gene (*Sbe*II), which is involved in potato tuber starch synthesis. As has already been indicated in the *Introduction*, ZFNs and TALENs are highly similar molecular tools for genome engineering. The restriction analysis that has been performed on ZFN-mutated potato plants to determine the levels of deletions (as described below) can later be performed on TALEN-mutated wheat plants as well in a similar way.

First DNA was isolated from several potato plants containing the ZFN constructs (based on the DNeasy Plant Mini Kit protocol by Qiagen) and PCRs using two primer sets were performed (*Sbe*II 227F/782R and 409F/936R). Each primer set sequences the *Sbe*II gene but the forward and reverse primer recognition sites of the one set were located upstream compared to the other set as can be derived from the primer codes (182bp and 154bp for the forward and reverse primers respectively). Multiple primer sets were developed as DNA amplification of the *Sbe*II gene was possibly more successful depending on the primer recognition sites. Full information on the primer sets is available from Jan Schaart upon request. The level of ZFN-induced mutations was determined by performing a restriction digestion on the PCR products with enzyme *Bfm*I (Fermentas). Successful ZFN-activity would have deleted the *Bfm*I restriction site, which is located at the ZFN target site in the *Sbe*II gene (see also **Figure 7**) resulting into one single gel band after gel electrophoresis for PCR products in which *Bfm*I is deleted, while unsuccessful deletion would result in *Bfm*I restriction of the PCR products and in this case into two different gel bands.

From plants that seemed to contain (almost) fully deleted *Sbe*II-sequences PCR products were gel purified according to the QIAquick Gel Extraction Kit Protocol (using a microcentrifuge) by Qiagen and eventually sequenced to determine which of the four alleles were deleted after all. The PCR products were cloned into TOPO vector and transformed to TOP10 cells (Invitrogen) and subjected to Blue White screening to prevent empty vectors. A high number of correct (white) clones was verified for inserts by PCR using primers M13 F/R and agarose gel analysis and PCR products were subsequently sequenced using primer M13 F. Potato plants with a high level of deleted *Sbe*II sequences were used in successive experiments aimed at further, complete enrichment for deleted *Sbe*II-sequences.

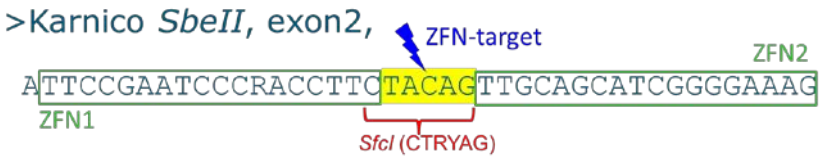


***Figure 7**: ZFN target site of the SbeII alleles in potato cv. Karnico. The DNA-binding domains of the ZFNs (as comparable to the DNA-binding domains of TALENs) are indicated by green boxes. The ZFN-target where restriction and deletion takes place is marked yellow. Successful ZFN-activity would have deleted the BfmI restriction site (indicated with the restriction enzyme SfcI identical to BfmI), which is located at the ZFN target site in the SbeII gene resulting into one single gel band after gel electrophoresis for PCR products in which BfmI is deleted, while unsuccessful deletion would result in BfmI restriction of the PCR products and in this case into two different gel bands.*

Results

TALENs design and assembly

Two TALEN pairs were designed (Cermak et al. 2011; Doyle et al. 2012) to target α-gliadins of *Triticum aestivum* causing celiac disease, of which one pair was designed to target highly conserved regions of highly immunogenic α-gliadins located on the short arms of group 6 chromosomes and one would target α-gliadins located on the chromosomes 6D (containing most immunogenic α-gliadins and in the highest frequencies; Van Herpen et al. 2006) and 6A but not 6B. Deletion with the latter TALENs would possibly circumvent lethal genotypes, which may be expected if all α-gliadin genes are deleted. Only one designed TALEN pair was subjected to assembly (Cermak et al. 2011; Zhang et al. 2013). It should be mentioned that unintentionally a more specific TAL array with a different RVD was constructed due to focusing on an α-gliadin sequence with specific variation, which rendered a more specific TALENs than intended (see *Methods*). **Table 2** describes the designed and assembled TAL arrays of the first TALENs, including a correction to the less specific TAL array.

Table 2: *Designed and assembled TAL arrays of the TALENs that was designed to target highly immunogenic α-gliadin of Triticum aestivum located on the short arms of group 6 chromosomes. Each TAL array consisted of RVDs (NI, HD, NH and NG) recognizing one nucleotide each (A, C, G and T respectively). TAL array 1 was unintentionally constructed with an different RVD (HD, marked red), whereas originally a G would be targeted and the supposed RVD would then be NH. This rendered a more specific TALENs than intended (see Methods). The first 10 RVDs of TAL1 and TAL2 represent intermediate TAL arrays TAL1A and TAL2A respectively, whereas the remaining RVDs represent intermediate TAL arrays TAL1B and TAL2B respectively.*

TAL array 1 (TAL1)	NI NI NH NH NI NI HD HD NG NI + NH HD HD HD NG
TAL array 2 (TAL2)	NN NG NI NH NI HD NI NG NG NH + HD NI HD NI NG NG

The plasmids required to assemble the TALENs (see *Methods*) were verified by sequencing and showed successful assembly. See *appendix iv* for sequencing results.

Intermediate TAL arrays were assembled into pFUS destination vectors by a first series of Golden Gate reactions and verified by PCR, restriction digestion and sequencing. According to Cermak et al. (2011), after PCR a correct clone should give a band around the expected size (1.2kb for vectors with 10 repeats), but also smearing and a ladder of bands are signs of a correct clone. Restriction digestion should result in arrays of 1048bp for vectors with 10 repeats. PCR and restriction digestion results are shown in **Figures 8** and **9** respectively. Note that for all gel electrophoresis experiments loading buffer with GelRed was used which caused 1kb ladders to possibly run faster than other loaded samples, leading to less accurate estimations of band sizes.

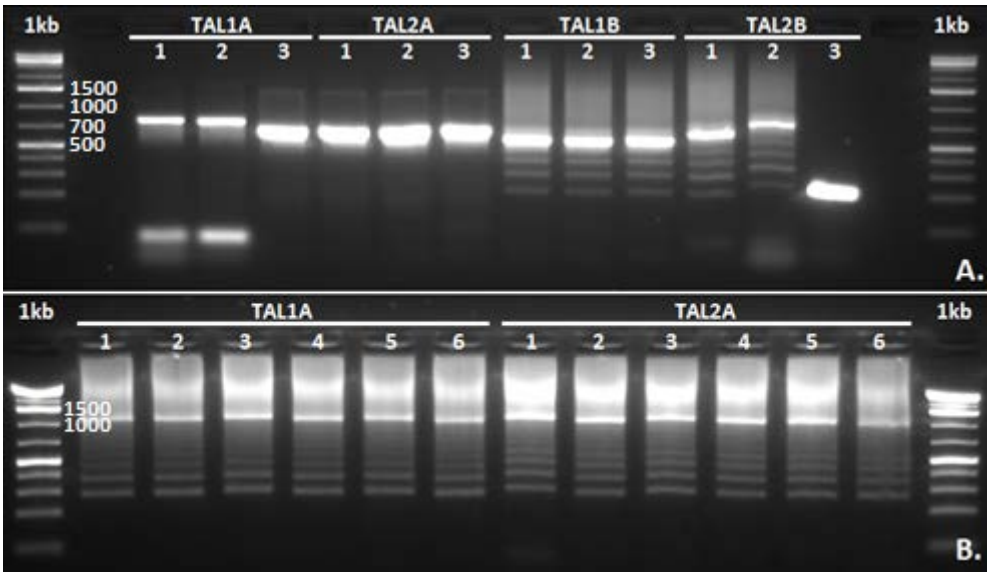


***Figure 8**: PCR results of intermediate TAL arrays after Golden Gate #1 reactions. **(A.)** TAL1A and TAL2A clones were incorrect as band sizes were not ~1.2kb and smearing and a ladder effect were not evident. TAL1B clones seemed to be correct; the smear and ladder effect were evident and the sizes of the bands were ~480bp (based on ~120bp per repeat). Regarding TAL1B clones, clone 3 was empty and although clones 1 and 2 both showed smearing and a ladder effect, the bands should be of identical size (~600bp) and clone 2 was probably incorrect. **Figure 9** indicates that TAL1B clone 1 was correct and clone 2 was incorrect. Although TALEN assembly should be highly efficient, many clones were incorrect which can be explained due to the fact that it is critical to heat shock transform cells at 42°C for 30s, while this step was carried out for 45s. This would especially result in unsuccessful transformation for plasmids with larger inserts (TAL1A and TAL2A), as can be seen in the figure. **(B.)** Intermediate arrays TAL1A and TAL2A were used for (correct) heat shock transformation once more and all clones seemed to be correct; in all cases the band sizes were ~1.2kb and smearing and a ladder effect were evident.*

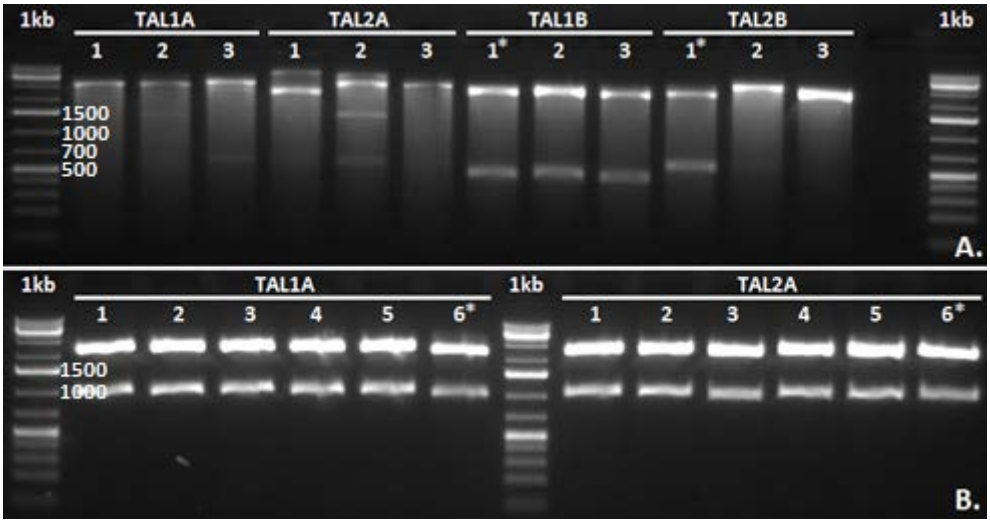


***Figure 9**: Restriction digestion results of intermediate TAL arrays (using enzymes BspTI and XbaI) after Golden Gate #1 reactions. Upper bands represent pFUS plasmids, while lower bands of correct clones represent TAL arrays. Based on restriction digestion results and sequencing results (see appendix v), the clones indicated by an asterisk were eventually chosen for further TALEN assembly. **(A.)** TAL1B clones and clone TAL2B-1 contained TAL arrays of correct sizes (~420bp for TAL1B and ~524bp for TAL2B), which corresponds with the PCR results as shown in **Figure 8**. TAL2B-2 showed a faint band and restriction digestion was unsuccessful. TAL1A-2 and TAL2A-2 showed faint bands at ~1048bp, but also at ~500bp, which other TAL1A and TAL2A clones showed as well. As has been explained in **Figure 8**, many clones were incorrect probably due to an incorrectly performed heat shock transformation. **(B.)** Intermediate arrays TAL1A and TAL2A were used for (correct) heat shock transformation once more and all digested clones seemed to be correct. In all cases the TAL arrays were ~1048bp as expected.*

See *appendix v* for sequence analysis of the intermediate TAL arrays. Sequencing results for TAL1B are partly shown below as an example. TAL1B was constructed *in silico* (marked yellow) using the free software program ApE v2.0.45 (A plasmid Editor; M. Wayne Davis, University of Utah) and used as a reference sequence in an ApE alignment with sequences of the TAL1B clones 1, 2 and 3 (see **Figures 8** and **9**). IJ55-57 and IJ67-69 represent TAL1 clones 1-3 sequenced with forward and reverse primers respectively. The sequences between the restriction sites *Bsp*TI (marked green) and *Xba*I (marked blue) should be identical to the *in silico* assembled TAL array, as these sequences represent the intermediate TAL arrays. This was the case for each TAL1B clone. The sequencing results supported the previous PCR and restriction digestion results as shown in **Figures 8** and **9** respectively. Note that in the partly shown alignment below 300bp between the restriction sites were left out.

```
Wed Nov 06, 2013 14:57 +0100#1 TALEN#1 TAL1B.str from 1 to 676 to
IJ55.seq--  Matches:675; Mismatches:0; Gaps:417; Unattempted:0
IJ56.seq--  Matches:674; Mismatches:1; Gaps:413; Unattempted:0
IJ57.seq--  Matches:674; Mismatches:1; Gaps:403; Unattempted:0
IJ67.seq--  Matches:552; Mismatches:1; Gaps:590; Unattempted:0
IJ68.seq--  Matches:551; Mismatches:2; Gaps:540; Unattempted:0
IJ69.seq--  Matches:546; Mismatches:3; Gaps:696; Unattempted:0

    16>GCCAGTCTTAAGCGTCTCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAATCACGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTT>115
    31>GCCaGTCTTAAGCGTCTCATGGCCTGACCCCGGACCAAGTGGTGGctATCGCCAGCAATCACGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTT>130
    30>GCCaGTCTTAAGCGTCTCATGGCCTGACCCCGGACCAAGTGGTGGctATCGCCAGCAATCACGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTT>129
    22>GCCaGTCTTAAGCGTCTCATGGCCTGACCCCGGACCAAgtgGTGGctATCGCCAGCAATCACGGCGGCAAGCAAGCGCTCGAAACGgtGCAGCGGCTGTT>121
  538<GCCAGTCTTAAGCGTCTCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAATCACGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTT<439
  538<GCCAGTCTTAAGCGTCTCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAATCACGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTT<439
  534<GCCAGTCTTAAGCGTCTCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAATCACGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTT<435

  416>gcTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGAGACGTCTAGACCAGCCAGGACAGAAATGCCTCGACTTCGC>515
  431>GCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGAGACGTCTAGACCAGCCAGGACAGAAATGCCTCGACTTCGC>530
  430>GCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGAGACGTCTAGACCAGCCAGGACAGAAATGCCTCGACTTCGC>529
  422>GCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGAGACGTCTAGACCAGCCAGGACAGAAATGCCTCGACTTCGC>521
  138<GCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGAGACGTCTAGACCAGCCAGGACAGAAATGCCTCGACTTCGC<39
  138<GCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGAGACGTCTAGACCAGCCAGGACAGAAATGCCTCGACTTCGC<39
  134<GCTGTTGCCGGTGCTGTGCcagGaccATGGCCTGacCCCGGACcAAGTGGTGGCTATCGAGACGTCTaGaccAGCcagGACAGAAATGCctCGACTTCGC<35
```

After verification of the intermediate TAL arrays, a second series of Golden Gate reactions was performed to assemble the complete TAL arrays into destination vector pZHY500. The complete TAL arrays were verified by PCR, restriction digestion and sequencing. After PCR a correct clone should again give smear and a ladder effect. PCR and restriction digestion results are shown in **Figures 10** and **11** respectively. See *appendix v* for sequence analysis of the complete TAL arrays. TAL1 and TAL2 were constructed *in silico* using ApE v2.0.45 (M. Wayne Davis, University of Utah) and used as reference sequences in ApE alignments with sequences of the TAL1 and TAL2 clones (see **Figures 10** and **11**). The sequences should be identical to the *in silico* assembled TAL arrays between the *Stu*I and *Aat*II restriction sites, as these sequences represent the complete TAL arrays. The sequencing results supported the previous PCR and restriction digestion results as shown in **Figures 10** and **11** respectively. Note that sequences resulting from primer SeqTALEN 5-1 were aligned using SeqMan Pro (DNASTAR) instead, as ApE v2.0.45 gave poor alignments.

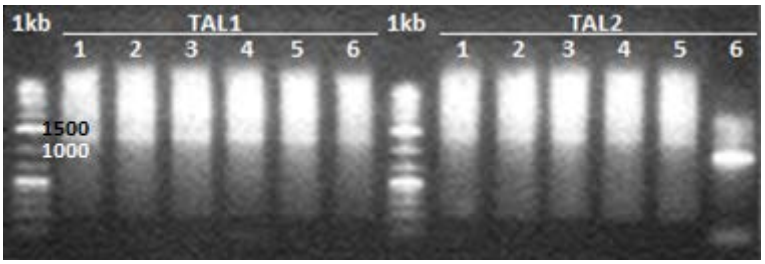


***Figure 10**: PCR results of complete TAL arrays after Golden Gate #2 reactions. Faint bands were visible between 1-1.5kb, and all clones except TAL2-6 were probably correct because of smearing and a ladder effect (which cannot be seen as the figure is of poor quality due to practical reasons).*

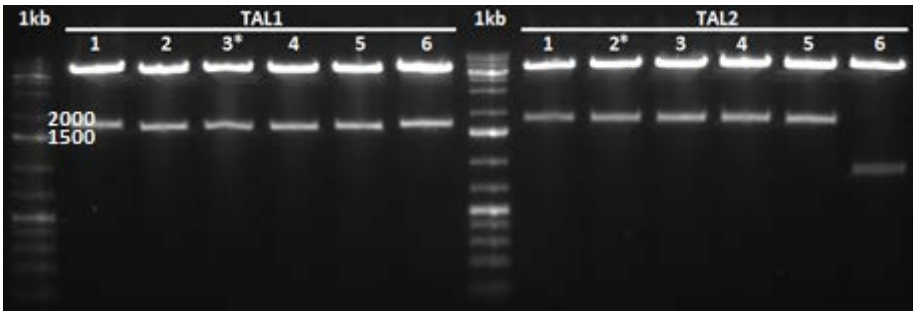


***Figure 11:** Restriction digestion results of complete TAL arrays (using enzymes StuI and AatII) after Golden Gate #2 reactions. Upper bands represent pZHY500 plasmids, while the lower bands of correct clones represent TAL arrays. As expected, the arrays were between 1.5-2kb, since TAL1 with 15 repeats should be ~1560bp and TAL2 with 16 repeats should be ~1664bp (based on ~104bp per repeat), of which the difference in size between the arrays was visible. Again, results indicate that clone TAL2-6 was incorrect. Based on restriction digestion results and sequencing results (see appendix v), the clones indicated by an asterisk were eventually chosen for further experiments.*

Preparing TALENs for efficient plant genome engineering

The complete TAL arrays TAL1 and TAL2 were cloned one after the other into pZHY013, a Gateway-compatible entry plasmid with two heterodimeric *Fok*I nuclease domains (Miller et al. 2007) separated by a T2A translational skipping sequence (Halpin et al. 1999), followed by a Gateway reaction to move the TALENs to destination vector pZHY051 which contains a 35S promoter and a NOS terminator to drive expression of the TALEN pair (Zhang et al. 2013). See **Figure 6** in *Methods*.

For this, TAL1 and TAL2 were first released from pZHY500 using applicable restriction enzymes and gel purified. pZHY013 was also digested and gel purified for removal of the insert at the first insertion site. Results are shown in **Figure 12**. The results of the restriction digestion after the ligation reaction of pZHY013 and TAL1 (PT1) to verify the insert are shown in **Figure 13**.

***Figure 12**: Gel purification of Gateway-compatible entry plasmid pZHY013 and complete arrays TAL1 and TAL2 from pZHY500 using fast digest (FD) enzymes XbaI and BamHI. As the insert at the first restriction site of pZHY013 consists of only a small number of bp, the insert had run off and was not visible. The single present band was cut out to obtain purified pZHY013. TAL1 and TAL2 were gel purified by cutting out the lower bands, as the upper bands represent pZHY500 which is of a larger size than the arrays.*

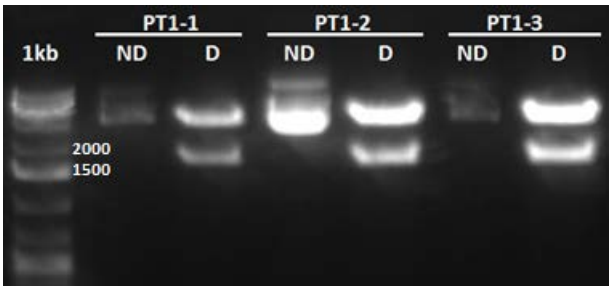


***Figure 13**: Restriction digestion results (using FD enzymes XbaI and BamHI) of 3 clones containing pZHY013/TAL1 (PT1) including digested (D) clones and not-digested (ND) controls. Bands of undigested controls appeared to be lower (not higher) compared to pZHY013 without inserts (the upper bands), because the control plasmids were partly supercoiled and therefore ran faster on the gel. Supercoiling was avoided later. Regarding the digested PT1 clones, all TAL1 arrays (the lower bands) were of expected size (between ~1.5-2kb). Clone PT1-1 was chosen for further experiments.*

pZHY013/TAL1 (PT1) was opened using a second set of restriction enzymes and gel purified for removal of the insert (**Figure 14**) and ligated with TAL2. The resulting pZHY013/TAL1/TAL2 (PT12) clone was verified by restriction digestion (see **Figure 15**). The PT12 clone was also sequenced to verify the TAL1 and TAL2 inserts; see *appendix vi* for the sequence analysis. The *in silico* constructed complete TAL arrays TAL1 and TAL2, which have been used for sequence analysis before to verify successful assembly of complete TAL arrays, were used as reference sequences again for the sequence alignments.

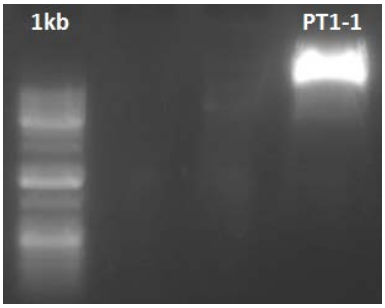


***Figure 14:** Gel purification of pZHY013/TAL1, clone 1 (PT1-1) using FD enzymes NheI and BglII. As the insert at the second restriction site of pZHY013 consists of only a small number of bp, the insert had run off and was not visible. The single present band was cut out to obtain purified pZHY013/TAL1. For a next time, the gel should be run longer to be able to estimate band sizes.*

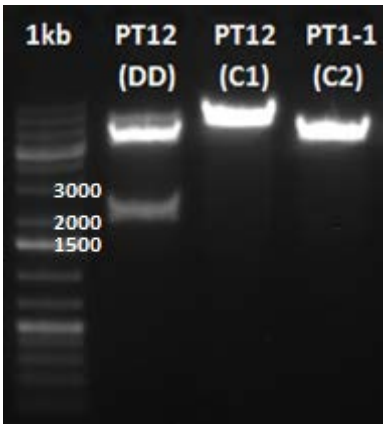


***Figure 15**: Restriction digestion results of pZHY013/TAL1/TAL2 (PT12) using FD enzymes ClaI and SpeI and two undigested controls of PT12 and pZHY013/TAL1 (clone 1; PT1-1). PT12 and PT1-1 were linearized using only FD enzymes ClaI and FD SpeI respectively to avoid supercoiling of the plasmids. As can be seen in the figure, PT1-1 is smaller than PT12 but has the same size as the double digest (DD) of PT12 as expected. PT12 without TAL2 should have the same size as PT1-1. The array TAL2 (the lower band) appeared to have a size of ~2kb, which can be expected due to the size of the array (~1664bp, based on ~104bp per repeat) and the restriction sites that were more distant from the insert (~78bp in total compared to the prior restriction sites), but especially due to the two present FokI nuclease domains (and the T2A translational skipping sequence) that were cut out of pZHY013 together with the TAL arrays upon this restriction digestion. See **Figure 6** in Methods for a comprehensive view.*

Eventually an LR Gateway reaction (Invitrogen) was performed to move the TALENs, including the *Fok*I nuclease domains and the T2A translational skipping sequence, into pZHY051 (see also **Figure 6** in *Methods*). The inserts were verified by sequencing from the inserts (using the *in silico* constructed complete TAL array TAL1 as a reference sequence in the sequence alignment which has been used for sequence analyses before) and from pZHY051; sequence analysis is shown in *appendix vi*.

Wheat seeds sterilization

The next step after having obtained the TALENs in a plant genome engineering vector is to test functionality of the constructed TALENs by transformation of the constructs to wheat protoplasts. Protoplasts will first be generated from shoot tissue from ~1.5 weeks (preferably ~10 days) old wheat seedlings grown under sterile conditions. A protocol for wheat seeds sterilization was developed and adopted during the current research and is described in *Methods*. The aim was to develop an optimal protocol for complete disinfected seeds with maximum germination capacity.

In short, it was found that it was essential to pre-incubate the seeds with 70% EtOH to reduce the amount of infected seeds. An additional overnight pre-incubation in water, however, resulted in increased infections. An incubation in 4% NaOCl seemed to work well and applying surfactants were of no advantage; when using Triton-X 100 as a surfactant the germination capacity even seemed to decrease. Also, it was essential to grow the seeds in smaller groups of 3 instead of larger groups of 15, since none of the wheat seeds sterilization experiments succeeded in the complete exclusion of infections. Growing seeds in smaller groups reduced loss due to cross-

contamination. The eventually adopted protocol resulted in ~80% infection-free seeds, of which ~25% of the seeds (20% of the total) yielded harvestable shoots after ~10 days. A complete report on the performed experiments for the development of the wheat seeds sterilization protocol can be found in *appendix vii*.

Wheat protoplasts were isolated twice from fresh green einkorn wheat shoots resulting from the abovementioned wheat seeds sterilization protocol. The first time a high number of intact protoplasts was observed under a light microscope (no figure available) which indicated successful protoplast isolation. However, the second time mainly broken protoplasts were observed, possibly due to vigorous pipetting which damages the vulnerable protoplasts. Too few intact protoplasts were counted for transformation experiments and these were therefore not performed anymore.

Additional research: determining the mutation level of ZFN-mutated potato by restriction analysis

Next to the main research of this report, some additional experiments were performed on *Solanum tuberosum* (potato; 2n=4x=48) (Ortiz and Peloquin 1994) *cv. Karnico* plants that were transformed with a specific ZFNs construct targeting DNA of all four alleles of the starch-binding enzyme II-gene (*Sbe*II), which is involved in potato tuber starch synthesis. As has already been indicated in the *Introduction*, ZFNs and TALENs are highly similar molecular tools for genome engineering. The restriction analysis that has been performed on ZFN-mutated potato plants to determine the levels of deletions can later be performed on TALEN-mutated wheat plants as well in a similar way. The level of ZFN-induced mutations was determined by performing a restriction digestion on the PCR products with enzyme *Bfm*I (Fermentas). Successful ZFN-activity would have deleted the *Bfm*I restriction site, which is located at the ZFN target site in the *Sbe*II gene resulting into one single gel band of roughly ~550bp after gel electrophoresis for PCR products in which *Bfm*I is deleted, while unsuccessful deletion would result in *Bfm*I restriction of the PCR products and in this case into two different gel bands of roughly ~200bp and ~350bp depending on the primer set used. These results are shown in **Figure 16**.

From plants that seemed to contain (almost) fully deleted *Sbe*II sequences (plant 25 in **Figure 16**), PCR products were gel purified (Qiagen) and eventually sequenced to determine which of the four alleles were deleted after all. The sequencing results are available upon request from Jan Schaart and showed that multiple *Sbe*II alleles, but not all were deleted by the ZFNs construct. Potato plants with a high level of deleted *Sbe*II sequences were used in successive experiments aimed at further, complete enrichment for deleted *Sbe*II sequences. These results are shown in **Figure 17**.

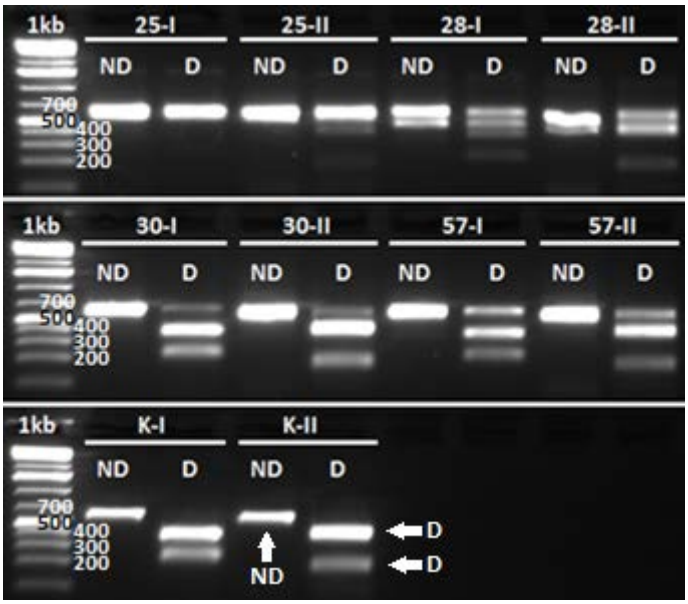


***Figure 16**: PCR analysis on untransformed (K) and ZFNs transformed (25, 28, 30, 57) Karnico potato plants using two different primer pairs both amplifying the target site (SbeII) and flanking DNA. Untransformed K resulted into two DNA fragments upon restriction digestion with BfmI (ND: not-digested; D: digested) as expected, since the SbeII alleles were present and thus the BfmI restriction sites. DNA from plants 30 and 57 was digested for a major part indicating low levels of deletions. DNA from plant 28, but especially 25 remained mainly undigested. Primer pair I showed no restriction digestion products for plant 25 upon amplification. Later research (performed by Jan Schaart) showed that one primer of pair I was located at a position with SNPs and even when taking this variation into account, primer pair I had amplification preferences for certain alleles. In general primer pair I resulted in better amplification than primer pair II; later a novel primer pair was used that resulted in better amplification of all alleles. Still, the results indicated that plant 25 contained almost fully deleted SbeII sequences and that this plant was an appropriate candidate to determine how many out of the four SbeII alleles were deleted after all, and to use in further experiments aimed at further, complete enrichment for deleted SbeII sequences.*

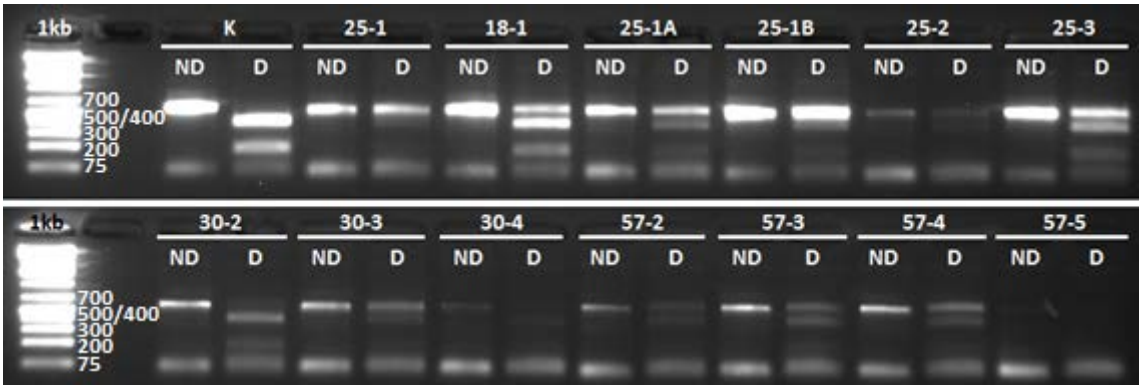


***Figure 17**: PCR analysis on SbeII alleles of untransformed (K) and ZFNs transformed (18, 25, 30, 57) Karnico potato plants (continued). Plants 25, 30 and 57 were analyzed before (see **Figure 16**). For each of these plants, individual plants were generated from shoot tissue (e.g. 25-1, 25-2 and 25-3 were derived from 25). 25-1A and 25-1B represent plants derived from 25-1. Successful ZFN-activity would have deleted the BfmI restriction site, located at the ZFN target site in the SbeII gene, resulting into one band, while unsuccessful deletion would result into two bands. In each lane, the lowest bands represent primer dimers. None of the plants seemed to be completely deleted. Moreover, the level of deletions of plants regenerated from different shoot tissue samples was different (e.g. 25-1 and 25-2 versus 25-3) and indicated that the original plant (e.g. 25) was chimeric. Plant 25-1 was possibly chimeric itself as well, as results showed that 25-1A showed slightly different restriction digestion results compared to 25-1 and 25-1B.*

Discussion

The ultimate goal of the current research is to obtain common wheat (*Triticum aestivum*) that is safe for patients with celiac disease (CD) by deleting genes that code for gluten proteins with immunogenic epitopes causing an autoimmune response, while the dough quality is not deteriorated. So far studies have not succeeded in obtaining gluten-free wheat without reducing its quality traits, which has led to the need of the emergence of a novel approach. Van den Broeck et al. (2011) has described that focusing on the removal of α-gliadin genes from the short arm of chromosome 6D in wheat may be an important strategy toward the development of CD-safe wheat and food products, as approximately 50% of CD-patients respond to a 33-mer (Shan et al. 2002, Shan et al. 2005) that contains the immunodominant epitopes Glia-α2 and Glia-α9 (Camarca et al. 2009, Vader et al. 2002) encoded by the *Gli-D2* locus, and this locus may be deleted from wheat varieties while maintaining its industrial properties. Preferably, the *Gli-D2* locus needs to be deleted without eliminating larger parts of the chromosomes as this may affect the fitness of the crop (Van den Broeck et al. 2011). During the current research an attempt has been made to elaborate this strategy by focusing on the deletion of a cluster of highly immunogenic α-gliadin genes encoding the 33-mer on the short arm of chromosome 6D, but also 6A and 6B.

Two TALENs have been designed *in silico*, of which one targets α-gliadin genes on the short arms of group 6 chromosomes and one on the short arms of chromosomes 6A and 6D, but not 6B. Eventually one TALENs targeting all group 6 chromosomes has been constructed *in vitro*, although it must be mentioned that one incorrect RVD (repeat-variable di-residue), namely HD (His-Asn) which recognizes a C nucleotide has been used whereas a G should be recognized. Unfortunately, HD specifically binds to C and is not specific for G (Boch et al. 2009). Because this RVD is located on the proximal end of the TAL array (i.e. 2 RVDs from the spacer DNA between the two TAL arrays to which the *Fok*I nucleases bind, dimerize and cleave the DNA), it contibutes largely to the specificity of the TALEN. Moreover, each RVD is part of a repeat of typically 34 amino acids (Moscou & Bogdanove 2009) and the size of the repeats will make it very unlikely for the earlier constructed TALENs to allow dimerization of the *Fok*I nucleases and therefore DNA cleavage of α-gliadin sequences located on all group 6 chromosomes. On the other hand, this holds opportunities for further research as the acquired TALENS can still be tested for activity in wheat.

The incorrectly assembled TALENs will still be able to recognize and target a few α-gliadin genes located on the short arm of chromosome 6A (see **Figure 3** in Methods). Although this TALENs is more specific than intended, the construct is ready for wheat transformation and it may still be worthwhile to test this construct for its TALENs activity *in vivo*. It is still likely to delete at least a part of the cluster with α-gliadin genes on chromosome 6A, if the specific genes are located between the actually targeted sequences. Moreover, it may still be possible that TALENs activity is

evident in the α-gliadin clusters located on chromosomes 6B and more importantly 6D. If the TALENs works as expected, this holds promising results for further research as these results would indicate that TALENs are applicable to wheat. Moreover, it may be possible to ligate the remaining chromosomal material to the chromosome to improve or restore vitality of the crop, as this may be affected by deleting larger parts of chromosomes. TALENs have already been applied successfully in other organisms like zebrafish (*Danio rerio*) where large chromosomal deletions have been performed (Lim et al. 2013; Xiao et al. 2013). Research that is currently being performed on another plant species potato (*Solanum tuberosum*) using ZFNs, a molecular technique that is very similar to TALENs, also indicates that this kind of molecular tools can be applied for the engineering of plant genomes of crops and more importantly can target and delete multiple genes at once. Moreover, the restriction analysis that has been performed on ZFN-mutated potato plants to determine the levels of deletions can later be performed on TALEN-mutated wheat plants as well in a similar way.

Einkorn wheat (*Triticum monococcum*), that has been used for the development of a convenient wheat seeds sterilization protocol, can be used for first (protoplast) transformation experiments with the available TALENs construct. An advantage of using einkorn wheat is that it is diploid and coincidently only contains the A-genome, while common wheat is allohexaploid and contains three genomes (A, B and D). Yet, for further transformation experiments research should also be performed on common wheat, preferably on cultivars that contain relatively many α-gliadin genes that can be targeted with the transformation-ready TALENs construct. *T. aestivum cv. 'Chinese Spring'*, which has been investigated by Van den Broeck et al. (2011) may be an appropriate candidate in general. Although it is probably impossible to use the construct for genomes B and D due to the assembly of the TALENs with a highly specific RVD at a critical location, its TALENs activity can still be tested for all wheat genomes.

If the first protoplast transformation experiments turn out to be successful, also the earlier described correct TALENs constructs should be assembled and tested *in vivo*. Several protocols have been optimized which can be found in the appendix and one should be able to perform transformation experiments within four weeks: one week (five days) is required for assembling TAL arrays (assuming sufficient plasmid material is initially available), two weeks are required to prepare the TALENs for plant transformation, and during the fourth week protoplasts can be isolated from 9 days old (einkorn) wheat seedlings (~10 days is preferred) and actually transformed. Depending on how long it will take to regenerate plant tissue from the transformed protoplasts, it is possible to analyze the transformed plants in approximately one month from TALENs design *in silico*.

Although using TALENs is an efficient approach, recently an even more efficient gene editing tool based on the bacterial clustered interspaced short palindromic repeats (CRISPR)-associated protein (Cas) adaptive immune systems has been developed, in which DNA can be targeted and

cleaved using synthetic-guide RNA (sgRNA) and Cas9, an endonuclease derived from *Streptococcus pyogenes* (Jinek et al. 2012). In other words, Cas9 is a DNA endonuclease guided by two RNAs. This sgRNA:Cas9 technology is advantageous compared to ZFNs and TALENs in that ZFNs and TALENs demand elaborate design and assembly of individual DNA-binding proteins for each DNA target site, while using sgRNA:Cas9 technology is considerably less elaborate, also for plants (Li et al. 2013). Even though this novel technology may also be applied in order to attempt to obtain wheat with deleted loci encoding immunogenic α-gliadin, the research with TALENs should not be neglected and it may be possible that the currently adopted approach is eventually more efficient for (wheat) genome editing after all.

Concluding, although a TALENs construct has been designed that should target and delete highly immunogenic α-gliadin genes from the short arms of group 6 chromosomes in common wheat, especially chromosome 6D, but eventually has been constructed more specific than initially intended and will only target a few α-gliadin genes from the short arm of chromosome 6A, this TALENs can still be of use for further research. Successively, the actual designed TALENs constructs that are described in the report should be assembled and used for further transformations of which the first results could be obtained in a month. Also, parallel experiments using the novel sgRNA:Cas9 technology could be performed. Eventually, new wheat lines may be obtained that can be used for the production of numerous products that are derived from or processed with wheat with maintained traits for industrial properties, while it is also safe for CD-patients to consume these food products.

References

- Anderson OD and Greene FC (1997). *The α-gliadin gene family. II. DNA and protein sequence variation, subfamily structure, and origins of pseudogenes*. Theoretical and Applied Genetics 95(1-2):59-65.
- Arentz-Hansen EH, Korner R, Molberg Ø, Quarsten H, Vader W, Kooy YMC, Lundin KEA, Koning F, Roepstorff P, Sollid LM, and McAdam SN (2000a). *The intestinal T cell response to α-gliadin in adult celiac disease is focused on a single deamidated glutamine targeted by tissue transglutaminase*. Journal of Experimental Medicine 191(4):603-612.
- Arentz-Hansen EH, McAdam SN, Molberg Ø, Kristiansen C, and Sollid LM (2000b). *Production of a panel of recombinant gliadins for the characterisation of T cell reactivity in coeliac disease*. Gut 46(1):46-51.
- Arentz-Hansen EH, McAdam SN, Molberg Ø, Fleckenstein B, Lundin KEA, Jorgensen TJD, Jung G, Roepstorff P, and Sollid LM (2002). *Celiac lesion T cells recognize epitopes that cluster in regions of gliadins rich in proline residues*. Gastroenterology 123(3):803-809.
- Barak S, Mudgil D, and Khatkar BS (2012). *Relationship of gliadin and glutenin proteins with dough rheology, flour pasting and bread making performance of wheat varieties*. LWT - Food Science and Technology 51(1):211-217.
- Belderok B, Mesdag J, and Donner DA (2000). *Bread-making quality of wheat: A century of breeding in Europe*. p57-62. ISBN 978-94-017-0950-7.
- Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, and Bonas U (2009). *Breaking the code of DNA binding specificity of TAL-type III effectors*. Science, 326(5959):1509-1512.
- Broeck HC, Herpen TJWM, Schuit C, Salentijn EMJ, Dekking L, Bosch D, Hamer RJ, Smulders MJM, Gilissen LJWJ, and Meer IM (2009). *Removing celiac disease-related gluten proteins from bread wheat while retaining technological properties: a study with Chinese Spring deletion lines*. BMC Plant Biology 9:41.
- Broeck HC, Gilissen LJWJ, Smulders MJM, VanDerMeer IM, and Hamer RJ (2011). *Dough quality of bread wheat lacking α-gliadins with celiac disease epitopes and addition of celiac-safe avenins to improve dough quality*. Journal of Cereal Science 53(2):206-216.
- Camarca A, Anderson RP, Mamone G, Fierro O, Facchiano A, Constantini S, Zanzi D, Sidney J, Auricchio S, Sette A, Troncone R, and Gianfrani C (2009). *Intestinal T cell responses to gluten peptides are largely heterogeneous: implications for a peptide-based therapy in celiac disease*. Journal of Immunology 182(7):4158-4166.
- Catassi C, Bearzi I, and Holmes GKT (2005). *Association of celiac disease and intestinal lymphomas and other cancers*. Gastroenterology 128(4):S79-S86.
- Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, Schmidt C, Baller JA, Somia NV, Bogdanove AJ, and Voytas DF (2011). *Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting*. Nucleic Acids Research 39(12):e82.

- Cong L, Zhou R, Kuo YC, Cunniff M, and Zhang F (2012). *Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains*. Nature Communications 24(3):968.
- Doyle EL, Booher NJ, Standage DS, Voytas DF, Brendel VP, Dyk, JK, and Bogdanove AJ (2012). *TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction*. Nucleic Acids Research 40:W117-122.
- Gilissen LJWJ (2013). *Allergy, allergens and allergy management* [PowerPoint slides]. Retrieved from: not available. Contact: http://www.wageningenur.nl/en/Persons/dr.-LJWJ-Luud-Gilissen.htm.
- Halpin C, Cooke SE, Barakate A, El Amrani A, and Ryan MD (1999). *Self-processing 2A-polyproteins: a system for co-ordinate expression of multiple proteins in transgenic plants*. The Plant Journal 17(4):453-459.
- Harberd NP, Bartels D, and Thompson RD (1986). *DNA restriction-fragment variation in the gene family encoding high molecular weight (HMW) glutenin subunits of wheat*. Biochemical Genetics 24(7-8):579-596.
- Herpen TWJM, Goryunova SV, Schoot J, Mitreva M, Salentijn E, Vorst O, Schenk MF, Veelen PA, Koning F, Soest LJM, Vosman B, Bosch D, Hamer RJ, Gilissen LJWJ, and Smulders MJM (2006). *Alpha-gliadin genes from the A, B, and D genomes of wheat contain different sets of celiac disease epitopes*. BMC Genomics 7(1):1.
- Janatuinen EK, Kemppainen TA, Julkunen RJK, Kosma VM, Maki M, Heikkinen M, and Uusitupa MIJ (2002). *No harm from five year ingestion of oats in coeliac disease*. Gut 50(3):332-335.
- Jinek M, Chylinksi K, Fonfara I, Hauer M, Doudna JA, and Charpentier E (2012). *A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity*. Science 337(6096):816-821.
- Li T, Huang S, Jiang WZ, Wright D, Spalding MH, Weeks DP, and Yang B (2011). *TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain*. Nucleic Acids Research 39(1):359-372.
- Li JF, Norville JE, Aach J, McCormack M, Zhang D, Bush J, Church GM, and Sheen J (2013). *Multiplex and homologous recombination–mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9*. Nature Biotechnology 31(8):688-691.
- Lim S, Wang Y, Yu X, Huang Y, Featherstone MS, and Sampath K (2013). *A simple strategy for heritable chromosomal deletions in zebrafish via the combinatorial action of targeting nucleases*. Genome Biology 14(7):R69.
- Maiuri L, Ciacci C, Ricciardelli I, Vacca L, Raja V, Auricchio S, Picard J, Osman M, Quaratino S, and Londei M (2003). *Association between innate response to gliadin and activation of pathogenic T cells in coeliac disease*. The Lancet 362(9377):30-37.
- Marino CL, Tuleen NA, Hart GE, Nelson JC, Sorrells ME, Lu YH, Leroy P, and Lopes CR (1996). *Molecular genetic maps of the group 6 chromosomes of hexaploid wheat (Triticum aestivum L. em. Thell.)*. Genome 39(2):359-366.

- Mecham DK, Kasarda DD, and Qualset CO (1978). *Genetic aspects of wheat gliadin proteins*. Biochemical Genetics 16(7-8):831-853.
- Miller JC, Holmes MC, Wang J, Guschin DY, Lee YL, Rupniewski I, Beausejour CM, Waite AJ, Wang NS, Kim KA, Gregory PD, Pabo CO, and Rebar EJ (2007). *An improved zinc-finger nuclease architecture for highly specific genome editing*. Nature Biotechnology 25(7):778-785.
- Molberg Ø, Uhlen AK, Jensen T, Flæte NS, Fleckenstein B, Arentz-Hansen H, Raki M, Lundin KEA, and Sollid LM (2005). *Mapping of gluten T-cell epitopes in the bread wheat ancestors: Implications for celiac disease*. Gastroenterology 128(2):393-401.
- Moscou MJ and Bogdanove AJ (2009). *A simple cipher governs DNA recognition by TAL effectors*. Science 326(5959): 1501.
- Mussolino C, Morbitzer R, Lütge F, Dannemann N, Lahaye T, and Cathomen T (2011). *A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity*. Nucleic Acids Research 39(21):9283-9293.
- Niewinski MM (2008). *Advances in celiac disease and gluten-free diet*. Journal of the American Dietetic Association 108(4):661-672.
- Ortiz R and Peloquin SJ (1994). *Potato genetics*. p133-154. ISBN 0-85198-869-5.
- Reyon D, Tsai SQ, Khayter C, Foden JA, Sander JD, and Joung JK (2012). *FLASH assembly of TALENs for high-throughput genome editing*. Nature Biotechnology 30(5):460-465.
- Salentijn EMJ, Esselink DG, Goryunova SV, Meer IM, Gilissen LJWJ, and Smulders MJM (2013). *Quantitative and qualitative differences in celiac disease epitopes among durum wheat varieties identified through deep RNA-amplicon sequencing*. BMC Genomics 14:905.
- Sauer DB and Burroughs R (1986). *Disinfection of seed surfaces with sodium hypochlorite*. Phytopathology 76(7):745-749.
- Schuppan D, Esslinger B, and Dieterich W (2003). *Innate immunity and coeliac disease*. The Lancet 362(9377):3-4.
- Shan L, Molberg Ø, Parrot I, Hausch F, Filiz F, Gray GM, Sollid LM, and Khosla C (2002). *Structural basis for gluten intolerance in celiac sprue*. Science 297(5590):2275-2279.
- Shan L, Qiao SW, Arentz-Hansen H, Molberg Ø, Gray GM, Sollid LM, and Khosla C (2005). *Identification and analysis of multivalent proteolytically resistant peptides from gluten: Implications for celiac sprue*. Journal of Proteome Research 4(5):1732-1741.
- Shan Q, Wang Y, Li J, Zhang Y, Chen K, Liang Z, Zhang K, Liu J, Xi JJ, Qiu JL, and Gao C (2013). *Targeted genome modification of crop plants using a CRISPR-Cas system*. Nature Biotechnology 31(8):686-688.
- Shi AN, Leath S, and Murphy JP (1998). *A major gene for powdery mildew resistance transferred to common wheat from wild einkorn wheat*. Phytopathology 88(2):144-147.
- Singh NK and Shepherd KW (1988). *Linkage mapping of genes-controlling endosperm storage proteins in wheat.* Theoretical and Applied Genetics 75(4):628-641.

- Stepniak D and Koning F (2006). *Celiac disease: sandwiched between innate and adaptive immunity*. Human Immunology 67(6):460-468.
- Streubel J, Blücher C, Landgraf A, and Boch J (2012). *TAL effector RVD specificities and efficiencies*. Nature Biotechnology 30(7):593-595.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S (2011). *MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods*. Molecular Biology and Evolution 28(10):2731-2739.
- Vader LW, Kooy YMC, VanVeelen PA, DeRu A, Harris D, Benckhuijsen W, Peña S, Mearin L, Drijfhout JW, and Koning F (2002). *The gluten response in children with celiac disease is directed toward multiple gliadin and glutenin peptides*. Gastroenterology, 122(7):1729-1737.
- Verdu EF, Armstrong D, and Murray JA (2009). *Between celiac disease and irritable bowel syndrome: the "no man's land" of gluten sensitivity*. The American Journal of Gastroenterology 104(6):1587-1594.
- Xiao A, Wang Z, Hu Y, Wu Y, Luo Z, Yang Z, Zu Y, Li W, Huang P, Tong X, Zhu Z, Lin S, and Zhang B (2013). *Chromosomal deletions and inversions mediated by TALENs and CRISPR/Cas in zebrafish*. Nucleic Acids Research 41(14):e141.
- Yoo SD, Cho YH, and Sheen J (2007). *Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis*. Nature Protocols 2(7):1565-1572.
- Zhang Y, Su J, Duan S, Ao Y, Dai J, Liu J, Wang P, Li Y, Liu B, Feng D, Wang J, and Wang H (2011). *A highly efficient rice green tissue protoplast system for transient gene expression and studying light/chloroplast-related processes*. Plant Methods 7:30.
- Zhang Y, Zhang F, Li X, Baller JA, Qi Y, Starker CG, Bogdanove AJ, and Voytas DF (2013). *Transcription activator-like effector nucleases enable efficient plant genome engineering*. Plant Physiology 161(1):20-27.

Appendix

Appendix i – Golden Gate TALEN assembly protocol

This appendix describes the methods of the Golden Gate TALEN assembly protocol as described by Cermak et al. (2011), including alterations by Zhang et al. (2013) and additional alterations. It is recommended to start the practical work on a Monday to be able to finish in one working week.

DAY1

1. Choose your TALEN RVD sequence: N = number of RVDs (12-31). Visit https://tale-nt.cac.cornell.edu/ for sequence analysis tools. The correctly designed TALENs (targeting α-gliadins of *Triticum aestivum* chromosomes 6A/6B/6D) with TAL arrays TAL1 and TAL2 are used as example:

TAL1	NI	NI	NH	NH	NI	NI	HD	HD	NG	NI	NH	HD	NH	HD	NG
TAL2	NN	NG	NI	NH	NI	HD	NI	NG	NG	NH	HD	NI	HD	NI	NG	NG

If the TALEN length is 12-21 (15-20 RVDs are recommended by Cermak et al. (2011), Miller et al. (2011), Mussolino et al. (2011) and Li et al. (2011) and are included in the preset architectures offered by the sequence analysis tools):

2. Pick plasmids for the RVDs 1-10 (*e.g. pNI1, pNI2, pNH3, pNH4, pNI5, pNI6, pHD7, pHD8, pNG9 and pNI10 for intermediate TAL1 array TAL1A*) + destination vector pFUS_A
3. Pick plasmids 11 up to N-1 (*e.g. pNH1, pHD2, pNH3 and pHD4 for intermediate TAL1 array TAL1B, but **NOT** pNG5*) + destination vector pFUS_B#N-1

If the TALEN length is 22-31 (not recommended; larger TAL arrays will be more difficult to assemble):

2. Pick plasmids for the RVDs 1-10 + destination vector pFUS_A30A, pick plasmids for the RVDs 11-20 + destination vector pFUS_A30B
3. Pick plasmids 21 up to N-1 + destination vector pFUS_B#N-1 (pFUS_B plasmids are labeled B1-B10 but they are used for RVDs 11-30 – if the RVD #N-1 is 19 or 29, use the same destination vector pFUS_B9)

4. Mix Golden Gate reaction **#1** – for each set of vectors separately:

1-10 + pFUS_A (*e.g. resulting in intermediate TAL arrays TAL1A and TAL2A*)

11-(N-1) + pFUS_B(N-1) (*e.g. resulting in intermediate TAL arrays TAL1B and TAL2B*)

or

1-10 + pFUS_A30A

11-20 + pFUS_A30B

21-(N-1) + pFUS_B(N-1)

a) 150ng of each module vector + 150ng of pFUS vector (see *appendix iii* for the DNA concentrations of all plasmids supported by the protocol and pZHY013 and pZHY051)
b) 1µl *Bsa*I restriction enzyme (10u) (NEB; New England BioLabs)
c) 1µl T4 DNA ligase (20u) (NEB)
d) 2µl 10X T4 DNA ligase buffer (to a final concentration of **1X**)
e) H_2O up to **20µl total reaction volume** (10µl reaction volumes are reliably effective but not as effective as 20µl reaction volumes; upscaling may be needed)

Run cycle: **10x** (37˚C/5min + 16˚C/10min) + 50˚C/5min + 80˚C/5min (+ 10˚C/∞)

With this cycle you will get hundreds of white colonies with 90-100% efficiency.

Plasmid-Safe nuclease treatment: this destroys all unligated linear dsDNA fragments including incomplete ligation products with lower number of repeats fused; and cut and linearized vectors. The incomplete, shorter fragments would be cloned into the destination vector in vivo by recombination in the bacterial cell, if not removed (the start of the first repeat and the end of the last repeat are in the destination vector backbone, so the backbone has homology to each repeat module as they differ only in RVDs).

To each of your Golden Gate reactions #1 add:

a) 1µl 10mM ATP
b) 1µl Plasmid-Safe nuclease (10u) (Epicentre)

Incubate at 37˚C/1h

Note: The Plasmid-Safe nuclease manual says you should inactivate the enzyme by heating the reaction to 70˚C for 30min, but inactivation is not necessary for bacterial transformation.

Chemically competent cells are transformed according to the TOPO TA Cloning Kit for Sequencing protocol (Invitrogen) by adding 5µl of the Golden Gate reactions to (50µl) TOP10 (*Escherichia coli*) cells that are thawed on ice and mixing without pipetting. Leave the cells on ice for 10min, perform a heat shock transformation of 42˚C for 30s (**this is a very critical step**) and immediately place the cells back on ice. Add 250µl S.O.C. medium that is at room temperature and incubate the cells on a shaker

platform (200rpm) at 37˚C for 1h with the tubes being placed on their sides (tightly capped and ensured).

5. Plate on $Spec^{50}$ plates + 40µL of 20mg/mL X-gal + 40µL of 0.1M IPTG (in our labs, $Spec^{50}$, X-Gal and IPTG are added to LB medium with the respective ratios 1:1000, 1:1000 and 1:10000). When plating transformations of the pFUS_B vectors that have fewer repeats (especially less than 6 repeats), be careful to not plate all the cells as the efficiency is so high you cannot pick single colonies on day 2. Plating 10% (30µl) and consequently 90% (centrifuge the cells at 9000xg for 1min in a microcentrifuge, discard the supernatants and suspend the pellets in 100µl LB medium) of the cells seems effective. The plates are incubated upside down at 37˚C overnight.

DAY2

6. Optionally store the plates in a cold room (4˚C) for the development of the blue color of colonies. Pick 3 white colonies of each plate and start overnight cultures (10ml tubes with 3ml LB medium + 1:1000 $Spec^{50}$, on the shaker platform at 37˚C).

DAY3

Note: Due to the full schedule on day 3 it is recommended to already mix Golden Gate reaction **#2** after miniprepping the plasmids, assuming there is sufficient DNA of each required vector. You may choose to continue with one clone as the cloning step is highly efficient. An initial check on the colonies can be performed by PCR (~2h program) during the Golden Gate reaction (~3h program); definitive checks will have to be performed on day 4.

7. Miniprep the plasmids according to the QIAprep Spin Miniprep Kit protocol (Qiagen) and measure the DNA concentrations.
    1. Centrifuge the cultures in 10ml tubes at 2800xg, discard the supernatants, suspend the pellets in 250µl Buffer P1 (with 2.5µl RNase A solution and 0.25µl LyseBlue reagent) and transfer to 1.5ml Eppendorf tubes.
    2. Add 250µl Buffer P2 and mix thoroughly (**do NOT mix vigorously**) by inverting the tubes 4-6 times. If using LyseBlue reagent, the solutions turn blue.
    3. Add 350µl Buffer N3 and mix immediately and thoroughly by inverting the tubes 4-6 times. If using LyseBlue reagent, the solutions turn colorless.
    4. Centrifuge for 10min at 17900xg in a microcentrifuge.
    5. Apply the supernatants to the QIAprep spin columns by decanting.

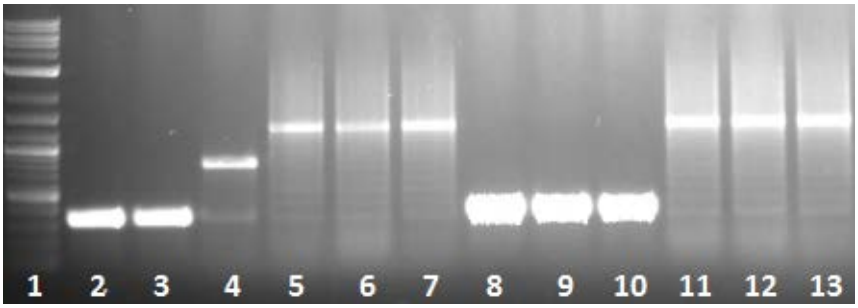
6. Centrifuge for 60s. Discard the flow-throughs.
7. Wash the QIAprep spin columns by adding 0.5ml Buffer PB and centrifuging for 60s. Discard the flow-through.
8. Wash the QIAprep spin columns by adding 0.75ml Buffer PE (with ethanol) and centrifuging for 60s.
9. Discard the flow-through, and centrifuge for an additional 60s to remove residual wash buffer.
10. To elute DNA, place the QIAprep columns in clean 1.5ml Eppendorf tubes. Add 50µl Buffer EB to the center of each QIAprep spin column, let stand for 1min, and centrifuge for 1min.

8. Perform PCRs using primers pCR8_F1 (5'-ttgatgcctggcagttccct-3') and pCR8_R1 (5'-cgaaccgaacaggcttatgt-3'); primers are the same for each vector.
   a) 0.3µl 10µM pCR8_F1
   b) 0.3µl 10µM pCR8_R1
   c) 1.5µl DT (DreamTaq) Buffer
   d) 0.6µl dNTPs
   e) 0.075µl DT (Dreamtaq) Polymerase
   f) 0.5µl DNA
   g) H_2O up to **15µl total reaction volume**

Run cycle: 95˚C/2min (place the tubes after 40˚C) + **35x** (95˚C/30s + 55˚C/30s + 72˚C/1.75min) + 72˚C/5min (+ 10˚C/∞)

You should get a band around your expected size (~1.2kb for vectors with 10 repeats), but you will also get smearing and a 'ladder' of bands. This is the sign of a correct clone and is the result of the repeats in the clone.

Example of (Colony) PCR results for pFUS vectors (Cermak et al. 2011):



Note: Lanes 2 and 3 are negative pFUS clones (empty). Lanes 4, 5, 6, 7, 11, 12 and 13 contain a 'correct' clone for this pFUS, of which lane 4 contains a clone with fewer repeats. pFUS clones that only contain 1 or 2 repeats are very similar in size to empty pFUS clones – check the size carefully. Lanes 5, 6, 7, 11, 12, 13 show smearing and the 'ladder' effect well. DNA ladder (lane 1) is NEB's 2Log. In our labs, a 2% gel is used for this step, with the DNA ladder 1kb. When using a medium tray, run any gel of the protocols at ~5.2V/cm (in our labs, run gels in medium and small trays at 90V and 60V respectively).

9. Mix Golden Gate reaction **#2** – for each set of vectors separately (see *appendix iii* for the DNA concentrations of all plasmids supported by the protocol and pZHY013 and pZHY051):
   a) 150ng of each pFUS vector (*e.g. TAL1A + TAL1B and TAL2A + TAL2B for complete TAL arrays TAL1 and TAL2 respectively*)
   b) 150ng of respective pLR vector (*e.g. pLR-NG for complete TAL array TAL1*)
   c) 75ng of destination vector (pZHY500 or pZHY501; both vectors are practically indifferent, although pZHY500 with 6561bp is recommended above pZHY501 with 7498bp)
   d) 1µl FD *Esp*3I restriction enzyme (10u) (Fermentas)
   e) 1µl T4 DNA ligase (20u) (NEB)
   f) 2µl FD buffer + 1µl 10mM ATP + 2µl 10mM DTT (**do NOT use 2µl 10X T4 DNA ligase buffer** – our experience is that FD *Esp*3I is not compatible with this buffer for the Golden Gate reaction and will result in no colonies)
   g) H_2O up to **20µl total reaction volume** (upscaling may be needed)

Run cycle: **10x** (37˚C/5min + 16˚C/10min) + 37˚C/15min + 80˚C/5min (+ 10˚C/∞) (for hundreds of white colonies according to the protocol; recommended) OR 37˚C/10min + 16˚C/15min) + 37˚C/15min + 80˚C/5min (+ 10˚C/∞) (for tens of white colonies according to the protocol; not recommended). One cycle should be sufficient for the second Golden Gate reaction, but in our experience (assembling TAL arrays of 15 and 16 RVDs) 10 cycles are needed for tens of colonies instead.

Note: Plasmid-Safe nuclease treatment is not necessary in this case, because the final destination vector has no homology with the inserted repeats.

Chemically competent cells are transformed as before (step 4). Use 5µl of the Golden Gate reactions.

10. Plate on Amp^{50} plates (**NOT** $Spec^{50}$) + X-Gal + IPTG as before (step 5).

DAY4

11. Pick 3 white colonies of each plate and start overnight cultures (10ml tubes with 3ml LB medium + 1:1000 Amp^{50}, on the shaker platform at 37˚C). In our experience (assembling TAL arrays of 15 and 16 RVDs) the colonies will be very small; you may choose to incubate the plates for a longer time than until the start of the day.

12. In the meantime perform definite checks on the clones from the first Golden Gate reactions by restriction digestion (to cut out the array of fused repeats: 1048bp for pFUS_A vectors, different sizes depending on the number of repeats clones for pFUS_B vectors) **and** sequencing.

    Restriction digestion:

    a) 0.5µl *Bsp*TI (5u) (Fermentas)
    b) 0.5µl *Xba*I (5u) (Fermentas)
    c) 4µl 10X Buffer Tango (to a final concentration of **2X**)
    d) 500ng DNA
    e) H_2O up to **20µl total reaction volume** (upscaling may be needed)

    Incubate the samples at 37˚C for 1h and run the samples on a (1.5%) gel.

    Sequencing:

    a) 2.5µl 10µM primer (pCR8_F1 **OR** pCR8_R1)
    b) DNA (optimal amounts are estimated upon plasmid sizes; **empty** pFUS plasmids are 2447bp)
    c) H_2O up to **10µl total reaction volume** (do **NOT** upscale reactions)

Sequencing is performed by GATC Biotech AG (Jakob-Stadler-Platz 7, 78467 Konstanz, Germany). Prepare the sequencing reactions as described in the protocol (use safe 1.5ml tubes and barcode label the tubes correctly). In our labs, sequencing reactions can be put in a small plastic bag and stored in a GATC collection box in the fridge. If the tubes are stored before 14:00 on a Tuesday or a Thursday (if the practical work is started on a Monday, day 4 will be on a Thursday), someone will bring the samples to the mail room and place the samples in a second GATC collection box. Otherwise, you may choose to bring the samples to the mail room yourself on any day (Monday to Friday) before 15:00. Be sure to not only place the tubes in a plastic bag, but also in an envelope. Assuming that day 4 will be performed on a Thursday, sequence analysis can be performed the week after with ApE v2.0.45 (A plasmid Editor; M. Wayne Davis, University of Utah) that can be used as a

sequence analysis tool without any problems. Guidelines for aligning with ApE can be found in *appendix v*.

DAY5

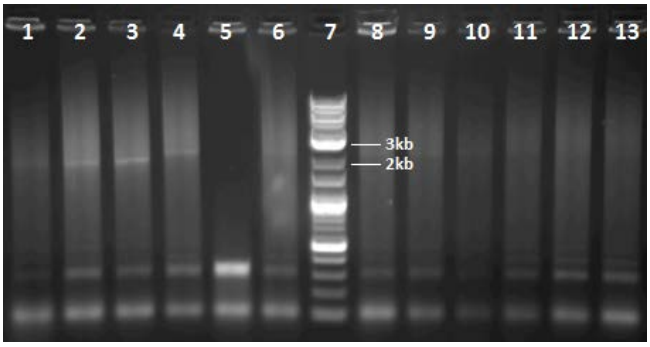
13. Miniprep the plasmids (step 7) and measure the DNA concentrations.

14. Perform PCRs using primers TAL_F1 (5'-ttggcgtcggcaaacagtgg-3') and TAL_R2 (5'-ggcgacgaggtggtcgttgg-3') as an initial check on the colonies.
    a) 0.3µl 10µM pCR8_F1
    b) 0.3µl 10µM pCR8_R1
    c) 1.5µl DT (DreamTaq) Buffer
    d) 0.6µl dNTPs
    e) 0.075µl DT (DreamTaq) Polymerase
    f) 1µl DNA (since large plasmids are used it is recommended to dilute the DNA to <1ng·µl)
    g) H_2O up to **15µl total reaction volume**

Run cycle: 95˚C/2min (place the tubes after 40˚C) + **35x** (95˚C/30s + 55˚C/30s + 72˚C/3min) + 72˚C/5min (+ 10˚C/∞)

Very often, you cannot see the band of the size you expect, but instead see a smear and the 'ladder' effect again – again, this is the sign of a correct clone. Run on a (1%) gel.

Example of (Colony) PCR results (Cermak et al. 2011):



Note: Lane 5 does not contain very much 'smear' (Ladder is NEB's 2Log) which indicates that this clone is NOT correct. In most other cases you can see faint bands around 2-3kb, which is the correct length for the completed TALs in the picture above. For TALs with >22 repeats, it is common to fail to amplify enough full-length TAL, however if you can see the 'smear', the clones are almost always correct. The 'ladder' effect is evident in some of the lanes (4, 11, 12 and 13).

15. In the meantime perform definite checks on the clones from the second Golden Gate reactions by restriction digestion (to cut out the final array of repeats) **and** sequencing.

    Restriction digestion:

    a) 0.5µl *Stu*I (5u) (Fermentas)
    b) 0.5µl *Aat*II (5u) (Fermentas)
    c) 2µl 10X Buffer Tango (to a final concentration of **1X**)
    d) 500ng DNA
    e) H_2O up to **20µl total reaction volume** (upscaling may be needed)

    Incubate the samples at 37˚C for 1h and run the samples on a (1%) gel.

    Sequencing:

    a) 2.5µl 10µM primer (SeqTALEN 5-1 **OR** TAL_R2)
    b) DNA (optimal amounts are estimated upon plasmid sizes; **empty** pFUS plasmids are 2447bp)
    c) H_2O up to **10µl total reaction volume** (do **NOT** upscale reactions)

    Note: Primer SeqTALEN 5-1 (5'-catcgcgcaatgcactgac-3') is used instead of TAL_F1. Sequence analysis can be performed the week after with ApE without any problems for sequences with primer TAL_R2. Sequences with primer SeqTALEN 5-1 seem to not align well in ApE; SeqMan Pro (DNASTAR) should be used instead. Guidelines for aligning with ApE can be found in *appendix v*.

You can now prepare your TALENs for efficient plant genome engineering as described in *appendix ii*.

Appendix ii – Preparing TALENs for efficient plant genome engineering

This appendix describes the cloning methods of the complete TAL arrays (*e.g. TAL1 and TAL2*; see *appendix i*) into pZHY013, a Gateway-compatible entry plasmid which contains two heterodimeric *Fok*I nuclease domains (Miller et al. 2007) separated by a T2A translational skipping sequence (Halpin et al. 1999), followed by a Gateway reaction to move the TALENs to destination vector pZHY051 which contains a 35S promoter and a NOS terminator to drive expression of the TALEN pair (Zhang et al. 2013). The protocol below is adjusted to working weeks of 5 days and starts at a Monday.

DAY6

1. Release TAL1 and TAL2 from pZHY500 and gel purify using Fast Digest (FD) enzymes *Xba*I and *Bam*HI (Fermentas). FD *Bam*HI is preferred above normal *Bam*HI since the latter will result in star activity. Also gel purify pZHY013 for removal of the insert at the first insertion site using these enzymes.
   a) 0.5µl FD *Xba*I (5u)
   b) 0.5µl FD *Bam*HI (5u)
   c) 2µl 10X FD Buffer (to a final concentration of **1X**)
   d) >1mg DNA
   e) H_2O up to **20µl total reaction volume** (upscaling may be needed)

   Incubate the samples at 37˚C for 30min and run the samples on a (1%) gel. Thick gel combs are preferred for gel purification.

   Gel purification is based on the QIAquick Gel Extraction Kit Protocol (using a microcentrifuge) by Qiagen (not described in this appendix). Cut out as much of the desired DNA as possible and try to avoid taking too much gel. Weigh each 2ml Eppendorf tube before and after adding the gel slice to determine its weight. Note that a water bath should be preheated to 50°C for this gel purification protocol.

2. Perform a ligation reaction of pZHY013 (~4300bp) and TAL array 1 (*e.g. ~1700bp*)
   a) v ng gel purified pZHY013 (*for the current example v = 40ng*)
   b) i ng TAL array 1 (*for the current example i = 60ng*)
   c) 1µl 10X T4 DNA ligase buffer (to a final concentration of **1X**)
   d) 0.3µl T4 DNA ligase (6u) (NEB; New England BioLabs)
   e) H_2O up to **10µl total reaction volume** (upscaling may be needed)

Even though the ligation reaction may be successful after several hours an overnight incubation at 16˚C is preferred. Note that the amount of DNA of vector pZHY013 (v) and insert TAL array 1 (i) depends on the size of the vector and the insert in bp. The insert:vector ratio should be 3:1. When using a total amount of 100ng DNA, the following formula can be used:

(3 * *size insert*) / *size vector* = i (100 - i); v = 100 – i

DAY7

Note: Sequence analysis of the sequencing reactions on days 4 and 5 (Golden Gate TALEN assembly protocol; *appendix i*) may be performed on days 7, 8 and 10.

3. Transform chemically competent cells according to the TOPO TA Cloning Kit for Sequencing protocol (Invitrogen) by adding 2µl of the ligation reaction to (50µl) TOP10 (*Escherichia coli*) cells that are thawed on ice and mixing without pipetting. Leave the cells on ice for 10min, perform a heat shock transformation of 42˚C for 30s (**this is a very critical step**) and immediately place the cells back on ice. Add 250µl S.O.C. medium that is at room temperature and incubate the cells on a shaker platform at 37˚C for 1h (200rpm) with the tubes being placed on their sides (tightly capped and ensured).

4. Plate on $Spec^{50}$ (in our labs, $Spec^{50}$ is added to LB medium with the ratio 1:1000).

Note: Plating 10% (30µl) and consequently 90% (centrifuge the cells at 9000xg for 1min in a microcentrifuge, discard the supernatants and suspend the pellets in 100µl LB medium) can in our experience be **replaced by plating all cells at once**, since there will only be a few colonies. The plate is incubated upside down at 37˚C overnight.

DAY8

5. Pick 3 colonies of each plate and start overnight cultures (50ml tubes with 10ml LB medium + 1:1000 $Spec^{50}$, on the shaker platform at 37˚C).

DAY9

6. Make bacterial glycerol stocks of the overnight cultures (0.4ml LB medium + 30% glycerol, 0.4ml culture; store at -80˚C and code and label properly). Miniprep the plasmids according

to the QIAprep Spin Miniprep Kit protocol (Qiagen) and measure the DNA concentrations. See also day 3 (*appendix i*). Note that larger volumes are used:

- Start with 2ml Eppendorf tubes and apply double volumes of Buffers P1, P2 and N3
- Mixing with Buffers P2 and N3 will require a higher number of tube inversion
- Applying supernatants to the QIAprep spin columns needs to be done in two times

7. Verify the pZHY013/TAL1 (PT1) clones by double restriction digestion with FD enzymes *Xba*I and *Bam*HI (Fermentas); a sequencing check will be performed later.
   a) 0.5µl FD *Xba*I (5u)
   b) 0.5µl FD *Bam*HI (5u)
   c) 2µl 10X FD Buffer (to a final concentration of **1X**)
   d) 500ng DNA
   e) H_2O up to **20µl total reaction volume** (upscaling will not be needed)

   Incubate the samples at 37˚C for 30min and run the samples on a (1%) gel.

8. Gel purify one PT1 clone further for removal of the insert at the second insertion site using FD enzymes *Nhe*I and *Bgl*II (Fermentas) with *Xba*I and *Bam*HI compatible ends.
   a) 0.5µl FD *Nhe*I (5u)
   b) 0.5µl FD *Bgl*II (5u)
   c) 2µl 10X FD Buffer (to a final concentration of **1X**)
   d) >1mg DNA
   e) H_2O up to **20µl total reaction volume** (upscaling will not be needed)

   Incubate the samples at 37˚C for 30min and run the samples on a (1%) gel. Thick gel combs are preferred for gel purification.

   Gel purification is based on the QIAquick Gel Extraction Kit Protocol by Qiagen. See step 1.

DAY10

9. Perform a ligation reaction of PT1 (*e.g. ~6000bp*) and TAL array 2 (*e.g. ~1800bp*)
   a) v ng gel purified PT1 (*for the current example v = 55ng; see step 2*)
   b) i ng TAL array 1 (*for the current example i = 45ng; see step 2*)
   c) 1µl 10X T4 DNA ligase buffer (to a final concentration of **1X**)
   d) 0.3µl T4 DNA ligase (6u) (NEB)
   e) H_2O up to **10µl total reaction volume** (upscaling may be needed)

Even though the ligation reaction may be successful after several hours an overnight incubation at 16˚C is preferred. In this case, there will only be time for an overnight incubation and due to the following steps it is even more convenient to perform an overweekend incubation.

DAY11

10. Transform and plate chemically competent cell as described in steps 3 and 4. Again, plating all cells at once is recommended since there will be very few colonies (during this step, you might even expect one single colony). **Note: it is advised to sterilize and grow wheat seeds for transformation during this day, as the seedlings will be 9 days old and a certain number of shoots (~25) will be ready for harvesting**.

DAY12

11. Pick (if possible) 3 colonies of each plate and start overnight cultures (50ml tubes with 10ml LB medium + 1:1000 $Spec^{50}$, on the shaker platform at 37˚C).

DAY13

Note: Due to the full schedule on day 13 it is recommended to already start the LR Gateway reaction (~5h) after miniprepping. You may choose to continue with one clone as the cloning step is highly efficient. Checks on colonies before the LR Gateway reaction can be performed in the meantime.

12. Make bacterial glycerol stocks of the overnight cultures (0.4ml LB medium + 30% glycerol, 0.4ml culture; store at -80˚C and code and label properly). Miniprep the plasmids according to the QIAprep Spin Miniprep Kit protocol (Qiagen) and measure the DNA concentrations. See also day 3 (*appendix i*). Note that larger volumes are used (see step 6).

13. Verify the pZHY013/TAL1/TAL2 (PT12) clones by double restriction digestion with FD enzymes *Cla*I and *Spe*I (Fermentas) as prior recognition sites are not present anymore.
    a) 0.5µl FD *Cla*I (5u)
    b) 0.5µl FD *Spe*I (5u)
    c) 2µl 10X FD Buffer (to a final concentration of **1X**)
    d) 500ng DNA
    e) H_2O up to **20µl total reaction volume** (upscaling will not be needed)

    Incubate the samples at 37˚C for 30min and run the samples on a (1%) gel.

14. Verify the PT12 clones for the TAL1 and TAL2 inserts with primers TAL1FW/REV (5'-aggaccacgatattgactac-3' and 5'-atcagctcgatgtactcgtg-3' respectively) and TAL2FW/REV (5'-acaaggatgacgatgacaag-3' and 5'-cctgagtggaatttctggca-3' respectively). Sequences can be analyzed using ApE v2.0.45 (A plasmid Editor, M. Wayne Davis, University of Utah) the week after; see *appendix vi* for sequencing results.
    a) 2.5µl 10µM primer (TAL1FW **OR** TAL1REV **OR** TAL2FW **OR** TAL2REV)
    b) DNA (optimal amounts are estimated upon plasmid sizes)
    c) H_2O up to **10µl total reaction volume** (do **NOT** upscale reactions)

15. Perform an LR Gateway reaction (Invitrogen) with entry PT12 and pZHY051 in a 1.5ml Eppendorf tube.
    a) 150ng entry clone (PT12)
    b) 150ng vector (pZHY051)
    c) TE buffer up to **8µl total reaction volume** (upscaling will not be needed)

    Add 2µl LR clonase II Enzyme mix (keep it stored at -20˚C **as long as possible**; before adding thaw on ice for 2min and mix for 2s only) and mix for 2s twice, then centrifuge for 2s in a microcentrifuge. Incubate the samples at 25˚C for 1h. Add 1µl proteinase K Solution to terminate the reaction and incubate at 37˚C for 10min.

16. Transform chemically competent cells according to the TOPO TA Cloning Kit for Sequencing protocol (Invitrogen) by adding 1µl of the LR Gateway reaction to (50µl) TOP10 (*Escherichia coli*) cells that are thawed on ice and mixing without pipetting. See step 3 for further steps.

17. Plate on Amp^{50} (**NOT** $Spec^{50}$) (in our labs, Amp^{50} is added to LB medium with the ratio 1:1000).

DAY14

18. Pick 3 colonies of each plate and start overnight cultures (50ml tubes with 10ml LB medium + 1:1000 Amp^{50} (**NOT** $Spec^{50}$), on the shaker platform at 37˚C). All cells can be plated at once as few colonies are expected.

DAY15

19. Make bacterial glycerol stocks of the overnight cultures (0.4ml LB medium + 30% glycerol, 0.4ml culture; store at -80˚C and code and label properly). Miniprep the plasmids according

to the QIAprep Spin Miniprep Kit protocol (Qiagen) and measure the DNA concentrations. See also day 3 (*appendix i*). Note that larger volumes are used (see step 6).

20. Verify the insert by sequencing the inserts from the inserts with primer TAL1REV and from pZHY051 with primers P-35S FW and T-NOS R1. Sequences can be analyzed using ApE the week after; see *appendix vi* for sequencing results.
    a) 2.5µl 10µM primer (TAL1REV **OR** P-35S FW **OR** T-NOS R1)
    b) DNA (optimal amounts are estimated upon plasmid sizes)
    c) H_2O up to **10µl total reaction volume** (do **NOT** upscale reactions)

Appendix iii – DNA concentrations (Golden Gate TALEN Kit and other pZHY plasmids)

This appendix describes the DNA concentrations of the plasmids (including antibiotic resistance) obtained which are required for the Golden Gate TALEN assembly protocol (see *appendix i*) and pZHY013 and pZHY051 for preparing the TALENs for efficient plant genome engineering (see *appendix ii*). Results are shown in **Table A1**. 260/280 and 260/230 ratios of approximately 1.8 and 2.0-2.2 respectively verified the quality (purity) of the DNA samples. If no sufficient DNA is available for the assembly of specific TALENs, glycerol stocks are available from Jan Schaart upon request.

Table A1*: DNA measurements on plasmids for TALEN assembly including pZHY013 and pZHY051. Lower DNA yields were possibly caused by vigorous shaking during DNA isolation with Buffer P2. Higher yields of pZHY plasmids were obtained by starting DNA isolations with 10ml cultures instead of 3ml.*

Plasmid (antibiotic)	DNA (ng/µl)	260/280	260/230	Plasmid (antibiotic)	DNA (ng/µl)	260/280	260/230
pHD1 (tet)	150.8	1.84	1.69	**pNK4 (tet)**	65.4	1.94	2.27
pHD2 (tet)	106.2	1.88	2.13	**pNK5 (tet)**	46.5	1.88	2.11
pHD3 (tet)	91.3	1.93	2.30	**pNK6 (tet)**	62.0	1.83	2.18
pHD4 (tet)	74.0	1.96	2.37	**pNK7 (tet)**	51.6	1.81	2.14
pHD5 (tet)	98.1	1.93	2.33	**pNK8 (tet)**	51.8	1.85	2.14
pHD6 (tet)	71.6	1.95	2.56	**pNK9 (tet)**	61.3	1.85	2.06
pHD7 (tet)	115.5	1.86	2.06	**pNK10 (tet)**	63.4	1.83	2.12
pHD8 (tet)	51.0	1.90	2.16	**pNH1 (tet)**	55.2	1.88	2.18
pHD9 (tet)	77.1	1.88	2.17	**pNH2 (tet)**	58.5	1.83	1.88
pHD10 (tet)	80.9	1.89	2.38	**pNH3 (tet)**	49.3	1.89	2.20
pNG1 (tet)	80.7	1.89	2.23	**pNH4 (tet)**	65.4	1.85	2.16
pNG2 (tet)	85.5	1.87	2.02	**pNH5 (tet)**	71.4	1.88	2.16
pNG3 (tet)	54.3	1.94	2.25	**pNH6 (tet)**	52.2	1.89	2.30
pNG4 (tet)	206.6	1.88	2.06	**pNH7 (tet)**	73.6	1.82	2.27
pNG5 (tet)	84.6	1.92	2.19	**pNH8 (tet)**	92.9	1.84	2.25
pNG6 (tet)	94.7	1.93	2.28	**pNH9 (tet)**	105.4	1.86	2.30
pNG7 (tet)	73.8	1.92	2.33	**pNH10 (tet)**	108.7	1.84	2.17
pNG8 (tet)	76.2	1.90	2.35	**pLR-HD (spec)**	47.0	1.89	2.06
pNG9 (tet)	83.8	1.91	2.32	**pLR-NG (spec)**	79.7	1.85	2.36
pNG10 (tet)	64.0	1.91	2.58	**pLR-NI (spec)**	63.7	1.81	2.17
pNI1 (tet)	83.8	1.90	2.26	**pLR-NN (spec)**	63.7	1.88	2.11
pNI2 (tet)	69.9	1.93	2.35	**pLR-FUS_B1 (spec)**	96.6	1.89	2.38
pNI3 (tet)	83.6	1.95	2.38	**pLR-FUS_B2 (spec)**	53.9	1.86	1.95
pNI4 (tet)	88.6	1.89	2.27	**pFUS_B3 (spec)**	54.0	1.73	2.01
pNI5 (tet)	31.7	1.84	1.97	**pFUS_B4 (spec)**	31.2	1.90	1.97
pNI6 (tet)	86.9	1.91	2.12	**pFUS_B5 (spec)**	91.1	1.86	2.22
pNI7 (tet)	70.2	1.90	2.01	**pFUS_B6 (spec)**	34.4	1.93	1.95
pNI8 (tet)	63.2	1.84	2.08	**pFUS_B7 (spec)**	59.0	1.87	2.04
pNI9 (tet)	65.1	1.97	2.26	**pFUS_B8 (spec)**	35.3	2.04	2.24
pNI10 (tet)	76.3	1.88	2.26	**pFUS_B9 (spec)**	89.3	1.90	2.11
pNN1 (tet)	62.7	1.87	2.22	**pFUS_B10 (spec)**	46.7	1.97	2.21
pNN2 (tet)	84.5	1.89	2.31	**pFUS_A (spec)**	145.1	1.89	2.20
pNN3 (tet)	65.4	1.88	2.24	**pFUS_A30A (spec)**	39.8	1.84	1.96
pNN4 (tet)	79.7	1.82	1.88	**pFUS_A30B (spec)**	110.7	1.90	2.31
pNN5 (tet)	95.1	1.93	2.24	**pLR-NH (spec)**	56.9	1.89	2.29
pNN6 (tet)	112.4	1.91	2.38	**pTAL1 (amp)**	116.6	1.84	2.28
pNN7 (tet)	65.6	1.90	2.23	**pTAL2 (amp)**	115.2	1.91	2.32
pNN8 (tet)	62.6	1.99	2.23	**pTAL3 (amp)**	118.8	1.88	2.32
pNN9 (tet)	50.8	1.90	2.25	**pTAL4 (amp)**	135.8	1.92	2.40
pNN10 (tet)	74.2	1.93	2.24	**pZHY013 (spec)**	361.5	1.85	2.38
pNK1 (tet)	54.9	1.87	2.30	**pZHY051 (amp)**	408.4	1.86	2.43
pNK2 (tet)	65.9	1.92	2.16	**pZHY500 (amp)**	363.4	1.87	2.43
pNK3 (tet)	61.9	1.96	2.15	**pZHY501 (amp)**	339.1	1.88	2.44

Appendix iv – Sequence analysis of Golden Gate TALEN Kit plasmids and pZHY013 and pZHY051

The Golden Gate TALEN Kit plasmids were sequenced and analyzed to verify the contents of the kit (i.e. each plasmid contains the correct sequence). The sequence analysis is described below. Note that analysis should be repeated after re-prepping; also look carefully at the RVD-encoding bases!

1. For the 60 module plasmids you can compare sets of 10 sequences with each other (see below; the numbers represent the file names of the samples). ApE v2.0.45 (A plasmid Editor, M. Wayne Davis, University of Utah) will take one of these sequences as the reference sequence. To verify all plasmids, it is advised to perform multiple alignments using different reference sequences. Because reverse primers have been used for Plate 2 (36CA56, available upon request), right overhangs are read from the right to the left. Because Plate 2 has been analyzed upside down, well H12 represents pHD1 (originally well A1) et cetera, as indicated below.

pHD1-10	pNG1-10	pNI1-10	pNN1-10	pNK1-10	pNH1-10
36CA75_A01	36CA75_A11	36CA75_B09	36CA75_C07	36CA75_D05	36CA75_E03
36CA75_A02	36CA75_A12	36CA75_B10	36CA75_C08	36CA75_D06	36CA75_E04
36CA75_A03	36CA75_B01	36CA75_B11	36CA75_C09	36CA75_D07	36CA75_E05
36CA75_A04	36CA75_B02	36CA75_B12	36CA75_C10	36CA75_D08	36CA75_E06
36CA75_A05	36CA75_B03	36CA75_C01	36CA75_C11	36CA75_D09	36CA75_E07
36CA75_A06	36CA75_B04	36CA75_C02	36CA75_C12	36CA75_D10	36CA75_E08
36CA75_A07	36CA75_B05	36CA75_C03	36CA75_D01	36CA75_D11	36CA75_E09
36CA75_A08	36CA75_B06	36CA75_C04	36CA75_D02	36CA75_D12	36CA75_E10
36CA75_A09	36CA75_B07	36CA75_C05	36CA75_D03	36CA75_E01	36CA75_E11
36CA75_A10	36CA75_B08	36CA75_C06	36CA75_D04	36CA75_E02	36CA75_E12
Overhang L (FW)	**Overhang L (FW)**	**Overhang L (FW)**	**Overhang L (FW)**	**Overhang L (FW)**	**Overhang L (FW)**
OR	OR	OR	OR	OR	OR
36CA56_H12	36CA56_H02	36CA56_G04	36CA56_F06	36CA56_E08	36CA56_D10
36CA56_H11	36CA56_H01	36CA56_G03	36CA56_F05	36CA56_E07	36CA56_D09
36CA56_H10	36CA56_G12	36CA56_G02	36CA56_F04	36CA56_E06	36CA56_D08
36CA56_H09	36CA56_G11	36CA56_G01	36CA56_F03	36CA56_E05	36CA56_D07
36CA56_H08	36CA56_G10	36CA56_F12	36CA56_F02	36CA56_E04	36CA56_D06
36CA56_H07	36CA56_G09	36CA56_F11	36CA56_F01	36CA56_E03	36CA56_D05
36CA56_H06	36CA56_G08	36CA56_F10	36CA56_E12	36CA56_E02	36CA56_D04
36CA56_H05	36CA56_G07	36CA56_F09	36CA56_E11	36CA56_E01	36CA56_D03
36CA56_H04	36CA56_G06	36CA56_F08	36CA56_E10	36CA56_D12	36CA56_D02
36CA56_H03	36CA56_G05	36CA56_F07	36CA56_E09	36CA56_D11	36CA56_D01
Overhang R (REV)	**Overhang R (REV)**	**Overhang R (REV)**	**Overhang R (REV)**	**Overhang R (REV)**	**Overhang R (REV)**

As an example to verify the module plasmids, below is a partly shown alignment of plasmids pNK1-10 with pNK9 as a reference sequence, which shows a clear alignment to verify the right overhang sequences (as a result upon *Bsa*I-digestion of the plasmids) and thus the plasmids. By verifying the overhangs, the components of the TALEN Kit can be verified as being the correct ones. The samples are from Plate 2 and represent pNK10-1 (wells D11-E8 respectively). The reference sequence is marked yellow, the right overhangs are marked red. The overhangs can be identified by taking the C nucleotide (marked green) as base 0, whereas the overhang bases are bases 2-5 ignoring gaps by poor aligning. Remember: a reverse primer (pRVD-seq REV; see *Methods*) has been used for the first 60 samples, so the overhang sequences are read from the right to the left (e.g. CGAC is the left overhang of pNK9). For the left and right overhang sequences of each plasmid, see Cermak et al. 2011.

```
1>t--GntctGGGTGncTCGAGGGTCTC----------------------------------------CGAGCGCTTGCTTGCCGCCCTTGTTGCTGGC>55 REF pNK9
1>~~~GntctGGGTGgcTCGAGGGTCTC----------------------------------------------------------CCGCCCTTGTTGCTGGC>40 D11 pNK10
1>t--GntctGGGTGncTCGAGGGTCTC----------------------------------------CGAGCGCTTGCTTGCCGCCCTTGTTGCTGGC>55 D12 pNK9
1>tttGgtctGGGTGGcTCGAGGGTCTCTT--------------------------------------CGAGCGCTTGCTTGCCGCCCTTGTTGCTGGC>59 E01 pNK8
1>~~~GntctGGGTGGcTCGAGGGTCTCGTTT------------------------------------CGAGCGCTTGCTTGCCGCCCTTGTTGCTGGC>58 E02 pNK7
1>~~~GgtctGGGTGGcTCGAGGGTCTCCaCCGTTT--------------------------------CGAGCGCTTGCTTGCCGCCCTTGTTGCTGGC>62 E03 pNK6
1>~~~GntctGGGTGGcTCGAGGGTCTCCTGCACCGTTT-----------------------------CGAGCGCTTGCTTGCCGCCCTTGTTGCTGGC>65 E04 pNK5
1>tt-GntctGGnTGGcTCGAGGGtCTCCaACAGCCGCTGCACCGTTT-------------------CGAGCGCTTGCTTGCCGCCCTTGTTGCTGGC>76 E05 pNK4
1>t--GntctGGGTGGcTCGAGGGTCTCCCTGGCACaGCACCGGCAACAGCCGCTGCACCGTTT-------CGAGCGCTTGCTTGCCGCCCTTGTTGCTGGC>91 E06 pNK3
1>t--GntctGGGTGgcTCGAGGgTCTCGgTCCTGGCACAGCACCGGCAACAGCCGCTGCACCGTTT----CGAGCGCTTGCTTGCCGCCCTTGTTGCTGGC>94 E07 pNK2
1>t--GgtctGGgTGGcTCGAGGGTCTCCCaTGGTCCTGGCACAGCACCGGCAACAGCCGCTGCACCGTTTCGAGCGCTTGCTTGCCGCCCTTGTTGCTGGC>98 E08 pNK1
```

2. For pFUS_B1-10 the found sequences will need to be compared to reference sequences found at www.addgene.org (September 13, 2013). The 10 sequences from the experiment can be aligned together, and the 10 reference sequences can be aligned together. Eventually you can compare the two alignments to conclude whether the found sequences correspond with their references. This is the case for pFUS_B1-10 considering Plate 1 (36CA75, available upon request); it will be unnecessary to do a second check with Plate 2. Wells F5-G2 represent pFUS_B1-10 respectively. As can be observed below, the pattern of the found and reference sequences is exactly the same, which seems to be a sufficient verification of these samples.

pFUS_B1-10

36CA75_F05

36CA75_F06

36CA75_F07

36CA75_F08

36CA75_F09

36CA75_F10

36CA75_F11

36CA75_F12

36CA75_G01

36CA75_G02

Sequences:

```
491>TTGCGCTCACTGCCCGCTTTCCACCGGTGGTCTCTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCT>590 REF B10
491>TTGCGCTCACTGCCCGCTTTCCACCGGTGGTCTC------------------------------------------------------------CCATGGCCT>533 F5  B1
492>TTGCGCTCACTGCCCGCTTTCCACCGGTGGTCTC----------------------------------------------------------AGGACCATGGCCT>538 F6  B2
491>TTGCGCTCACTGCCCGCTTTCCACCGGTGGTCTC--------------------------------------------------GCCAGGACCATGGCCT>540 F7  B3
491>TTGCGCTCACTGCCCGCTTTCCACCGGTGGTCTC------------------------------CTGTTGCCGGTGCTGTGCCAGGACCATGGCCT>556 F8  B4
491>TTGCGCTCACTGCCCGCTTTCCACCGGTGGTCTC----------------------GTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCT>565 F9  B5
500>TTGCGCTCACTGCCCGCTTTCCACCGGTGGTCTC--------------------ACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCT>577 F10 B6
501>TTGCGCTCACTGCCCGCTTTCCACCGGTGGTCTC----------------CGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCT>582 F11 B7
491>TTGCGCTCACTGCCCGCTTTCCACCGGTGGTCTC--------------CTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCT>574 F12 B8
492>TTGCGCTCACTGCCCGCTTTCCACCGGTGGTCTC------------CGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCT>577 G1  B9
491>TTGCGCTCACTGCCCGCTTTCCACCGGTGGTCTCTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCT>590 G2  B10
```

Reference sequences:

```
497>TCACTGCCCGCTTTCCACCGGTGGTCTCTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCC>596 REF B10

498>TCACTGCCCGCTTTCCACCGGTGGTCTC------------------------------------------------------------CCATGGCCTGACCCC>540 N/A B1
498>TCACTGCCCGCTTTCCACCGGTGGTCTC----------------------------------------------------------AGGACCATGGCCTGACCCC>544 N/A B2
495>TCACTGCCCGCTTTCCACCGGTGGTCTC--------------------------------------------------GCCAGGACCATGGCCTGACCCC>544 N/A B3
491>TCACTGCCCGCTTTCCACCGGTGGTCTC------------------------------CTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCC>556 N/A B4
491>TCACTGCCCGCTTTCCACCGGTGGTCTC----------------------GTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCC>565 N/A B5
495>TCACTGCCCGCTTTCCACCGGTGGTCTC--------------------ACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCC>572 N/A B6
500>TCACTGCCCGCTTTCCACCGGTGGTCTC----------------CGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCC>581 N/A B7
501>TCACTGCCCGCTTTCCACCGGTGGTCTC--------------CTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCC>584 N/A B8
497>TCACTGCCCGCTTTCCACCGGTGGTCTC------------CGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCC>582 N/A B9
497>TCACTGCCCGCTTTCCACCGGTGGTCTCTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCC>596 N/A B10
```

3. pFUS_A, pFUS_A30A and pFUS_A30B are verified using two alignments. First pFUS_A and pFUS_A30A are verified using samples with primer pCR8_F1 (see *Methods* and *appendix i*; Plate 1). Since it has been reported that well G5 is empty, pFUS_A30B was meant to be found in well H5 of Plate 2 (computer: A8; **the Plate has been analyzed upside down**), but this well appears to be empty as well. Therefore pFUS_A30B has to be verified with the sample with the primer pCR8_ R1 (see *Methods* and *appendix i*; Plate 2), and for convenience pFUS_A and pFUS_A30A are also included in the alignment. As before with pFUS_B1-10, the found sequences are compared with reference sequences from www.addgene.org (September 13, 2013; also available upon request). However, here sequences and their references are aligned together.

Alignment 1 includes (besides references):		Alignment 2 includes (besides references):	
pFUS_A	: 36CA75_G3	pFUS_A	: 36CA56_B10
pFUS_A30A	: 36CA75_G4	pFUS_A30A	: 36CA56_B09
		pFUS_A30B	: 36CA56_B08

Alignment 1: Both pFUS_A30A sequences contain a TGAC that is lacking for both pFUS_A sequences. Regarding the experiment, this will be sufficient evidence. pFUS_A and pFUSA_30A are verified.

```
500>TCACTGCCCGCTTTCCACCGGTGGTCTCTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCC----GAG>595 REF A
501>TCACTGCCCGCTTTCCACCGGTGGTCTCTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCC----GAG>596 A
501>TCACTGCCCGCTTTCCACCGGTGGTCTCTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACGAG>600 A30A
500>TCACTGCCCGCTTTCCACCGGTGGTCTCTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCC----GAG>595 A
496>TCACTGCCCGCTTTCCACCGGTGGTCTCTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACGAG>595 A30A
```

Alignment 2 (see below): Both pFUS_A30B sequences are distinct from the other sequences as in they lack a series of bases compared to pFUS_A and pFUS_A30A. Moreover, pFUS_A30A shows to be different since both pFUS_A30A sequences contain a TG that is lacking for the other sequences. This alignment (and the previous one) will be enough evidence for verifying all three samples.

```
  1>~~~~~~CACGACGTTGTAAAACGACGGCCAGTCTTAAGCGTCTCCCCCTGAACCTGACCCCGGACCAAGTGGTGGCTATCGAGACCGGCGCCGCTACAGG>94    REF A

  1>~~~~~~CACGACGTTGTAAAACGACGGCCAGTCTTAAGCGTCTCCCCCTGAACCTGACCCCGGACCAAGTGGTGGCTATCGAGACCGGCGCCGCTACAGG>94   A
  1>~~~~~~~~~~ACGTTGTAAAACGACGGCCAGTCTTAAGCGTCTCCCCCTGAACCTGACCCCGGACCAAGTGGTGGCTATCGAGACCGGCGCCGCTACAGG>90   A30A
  8>------CACGACGTTGTAAAACGACGGCCAGTCTTAAGCGTCTCCC--------TGACCCCGGACCAAGTGGTGGCTATCGAGACCGGCGCCGCTACAGG>93   A30B
665<------CACGACGTTGTAAAACGACGGCCAGTCTTAAGCGTCTCCC--------TGACCCCGGACCAAGTGGTGGCTATCGAGACCGGCGCCGCTACAGG<580  A30B
681<CCCAGTCACGACGTTGTAAAACGACGGCCAGTCTTAAGCGTCTCCCCCTGAACCTGACCCCGGACCAAGTGGTGGCTATCGAGACCGGCGCCGCTACAGG<582  A30A
669<AGT---CACGACGTTGTAAAACGACGGCCAGTCTTAAGCGTCTCCCCCTGAACCTGACCCCGGACCAAGTGGTGGCTATCGAGACCGGCGCCGCTACAGG<573  A

495>TGCGCTCACTGCCCGCTTTCCACCGGTGGTCTCTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCC-->592  REF A
495>TGCGCTCACTGCCCGCTTTCCACCGGTGGTCTCTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCC-->592  A
491>TGCGCTCACTGCCCGCTTTCCACCGGTGGTCTCTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTG>590  A30A
494>TGCGCTCACTGCCCGCTTTCCACCGGTGGTCTCTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCC-->591  A30B
179<TGCGCTCACTGCCCGCTTTCCACCGGTGGTCTCTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCC--<82   A30B
181<TGCGCTCACTGCCCGCTTTCCACCGGTGGTCTCTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTG<82   A30A
172<TGCGCTCACTGCCCGCTTTCCACCGGTGGTCTCTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCC--<75   A
```

4. pLR-HD/NG/NI/NN/NH (pLR-NK does not exist) correspond with wells F1-4 of and G6 of Plate 1 respectively. Since it has been reported that well G6 is empty, pLR-NH could be found in well G6 of Plate 2 (computer: B7; **the Plate has been analyzed upside down**). For convenience of the alignments (i.e. an alignment of all found sequences versus an alignment of all reference sequences from www.addgene.org (September 13, 2013; also available upon request), like the verification of the pFUS_B1-10 samples) the file names of the reference genes have been adjusted so that the order of these sequences corresponds with the order of the found sequences. The alignments look exactly the same (see below); the pLR samples are verified.

pLR-HD : 36CA75_F1

pLR-NG : 36CA75_F2

pLR-NI : 36CA75_F3

pLR-NN : 36CA75_F4

pLR-HD : 36CA56_B7

Sequences:

```
36CA56_B07.seq from 1 to 977 to

36CA75_F01.seq--  Matches:918; Mismatches:4; Gaps:71; Unattempted:0
36CA75_F02.seq--  Matches:918; Mismatches:11; Gaps:54; Unattempted:0
36CA75_F03.seq--  Matches:964; Mismatches:12; Gaps:36; Unattempted:0
36CA75_F04.seq--  Matches:965; Mismatches:11; Gaps:8; Unattempted:0
36CA56_B07.seq--  Matches:977; Mismatches:0; Gaps:0; Unattempted:0

   82>CCCAGTCACGACGTTGTAAAACGACGGCCAGTCTTAAGCTCGGGCCCGTCTCGCTATCGCCAGC-AATC-A-CGGCGGCAAGCAAGCGCTCGAAAGCATT>178
   91>CCCAGTCACGACGTTGTAAAACGACGGCCAGTCTTAAGCTCGGGCCCGTCTCGCTATCGCCAGCCA--CGA-TGGCGGCAAGCAAGCGCTCGAAAGCATT>187
   87>CCCAGTCACGACGTTGTAAAACGACGGCCAGTCTTAAGCTCGGGCCCGTCTCGCTATCGCCAGC-AA-CGG-TGGCGGCAAGCAAGCGCTCGAAAGCATT>183
   87>CCCAGTCACGACGTTGTAAAACGACGGCCAGTCTTAAGCTCGGGCCCGTCTCGCTATCGCCAGC-AA-C-ATTGGCGGCAAGCAAGCGCTCGAAAGCATT>183
   84>CCCAGTCACGACGTTGTAAAACGACGGCCAGTCTTAAGCTCGGGCCCGTCTCGCTATCGCCAGC-AA-CAA-TGGCGGCAAGCAAGCGCTCGAAAGCATT>180
   82>CCCAGTCACGACGTTGTAAAACGACGGCCAGTCTTAAGCTCGGGCCCGTCTCGCTATCGCCAGC-AATC-A-CGGCGGCAAGCAAGCGCTCGAAAGCATT>178
```

Reference sequences:

```
6) pLR-NH.txt from 1 to 1006 to

1) pLR-HD.txt--  Matches:840; Mismatches:4; Gaps:164; Unattempted:0
2) pLR-NG.txt--  Matches:782; Mismatches:3; Gaps:222; Unattempted:0
3) pLR-NI.txt--  Matches:755; Mismatches:5; Gaps:247; Unattempted:0
4) pLR-NN.txt--  Matches:871; Mismatches:3; Gaps:133; Unattempted:0
6) pLR-NH.txt--  Matches:1006; Mismatches:0; Gaps:0; Unattempted:0

  101>TGTAAAACGACGGCCAGTCTTAAGCTCGGGCCCGTCTCGCTATCGCCAGC-AATC-A-CGGCGGCAAGCAAGCGCTCGAAAGCATTGTGGCCCAGCTGAG>197
   92>TGTAAAACGACGGCCAGTCTTAAGCTCGGGCCCGTCTCGCTATCGCCAGCCA--CGA-TGGCGGCAAGCAAGCGCTCGAAAGCATTGTGGCCCAGCTGAG>188
   88>TGTAAAACGACGGCCAGTCTTAAGCTCGGGCCCGTCTCGCTATCGCCAGC-AA-CGG-TGGCGGCAAGCAAGCGCTCGAAAGCATTGTGGCCCAGCTGAG>184
   95>TGTAAAACGACGGCCAGTCTTAAGCTCGGGCCCGTCTCGCTATCGCCAGC-AA-C-ATTGGCGGCAAGCAAGCGCTCGAAAGCATTGTGGCCCAGCTGAG>191
   90>TGTAAAACGACGGCCAGTCTTAAGCTCGGGCCCGTCTCGCTATCGCCAGC-AA-CAA-TGGCGGCAAGCAAGCGCTCGAAAGCATTGTGGCCCAGCTGAG>186
  101>TGTAAAACGACGGCCAGTCTTAAGCTCGGGCCCGTCTCGCTATCGCCAGC-AATC-A-CGGCGGCAAGCAAGCGCTCGAAAGCATTGTGGCCCAGCTGAG>197
```

5. pZHY013, pZHY051, pZHY500 and pZHY501. pZHY500 and pZHY501 do practically not differ from each other; they have a different selection marker (HIS and LEU respectively). The primers we have used (M13 F/W) do not reach the sequence part where pZHY500 and pZHY501 differ. The samples nicely align with pZHY500 and pZHY501 though. HIS3-R and Leu2-F could have been used as primers instead.

Data for alignments:

pZHY500/pZHY501 references from www.addgene.org

09IF20 pZHY500 with M13 F

09IF21 pZHY501 with M13 F

09IF22 pZHY500 with M13 R

09IF23 pZHY501 with M13 R

(Note: these are loose sequences; available upon request)

Regarding pZHY013 and pZHY051, reference sequence pZHY013 corresponds with 36CA56_A10 (originally well H3 with pZHY013 and M13 F) but not with 36CA56_A9 (originally well H4 with pZHY051 and M13 F) in an alignment. In another alignment, with reference sequence pZHY051 and the two samples, pZHY051 corresponds with 36CA56_A9 but not with 36CA56_A10 as expected. This means that samples pZHY013 and pZHY051 are verified.

Data for alignments:

pZHY013/pZHY051 references from www.addgene.org

36CA56_A9 pZHY051 with M13 F

36CA56_A10 pZHY013 with M13 F

Appendix v – Sequence analysis Golden Gate TALEN assembly (days 4 and 5)

DAY 4

Use vector plasmid and donor plasmids (www.addgene.org) for each Golden Gate reaction #1 *in silico* using ApE v2.0.45 (A plasmid Editor, M. Wayne Davis, University of Utah). In order to perform the Golden Gate reaction #1, the enzyme *Bsa*I (NEB; New England BioLabs) needs to be added to ApE; GGTCTC(1/5). Select the enzyme for the Golden Gate reaction #1 and select the plasmids in the right order (vector plasmid, donor plasmids 1-n). Be sure to make the vector plasmid circular in order to be able to perform the Golden Gate reaction #1. Align the *in silico* constructed Golden Gate reaction #1 product as the reference sequence to the sequenced samples, search for the *Bsp*TI and *Xba*I (Fermentas) restriction sites, which are CTTAAG(1/5) and TCTAGA(1/5), and pay attention to the parts of the sequences within these two restriction sites. Below results of the research are shown. Sequences, *in silico* constructed Golden Gate reactions #1 and alignments are available upon request and additional information can be found in the lab journal of Jarst van Belle.

TAL1A (p. 72 lab journal Jarst van Belle; 67AG81-86/93-98; available upon request)
Reaction: pFUS_A + pNI1 + pNI2 + pNH3 + pNH4 + pNI5 + pNI6 + pHD7 + pHD8 + pNG9 + pNI10
All 12 sequences (6 clones, sequenced with both forward and reverse primers pCR8_F1 and pCR8_R1 (see *Methods* and *appendix i*); see p. 72 lab journal) altogether align very well with the reference sequence. Forward sequences will generally not be able to reach the *Xba*I restriction site and reverse sequence will generally not be able to reach the *Bsp*TI restriction site, but taking the forward and reverse sequences of each clone together, the sequences in between the restriction sites are identical to the reference sequence. 1-1 1/10 forward (67AG84) lacks however ~400bp (very roughly), even though the corresponding reverse sequence does not lack this part and the size on the gel (p. 72 lab journal; *Results*: **Figures 8B**/**9B**, TAL1A clone 4) seems also correct.

TAL2A (p. 72 lab journal Jarst van Belle; 67AG87-92/99 and 67AH01-04; available upon request)
Reaction: pFUS_A + pNN1 + pNG2 + pNI3 + pNH4 + pNI5 + pHD6 + pNI7 + pNG8 + pNG9 + pNH10
See above; every clone seems correct, even though 2-3 9/10 reverse (67AH01) lacks ~350bp (very roughly) and 2-1 9/10 forward (67AG87) lacks ~100bp (*Results*: **Figures 8B/9B**, TAL2A clones 3 and 1 respectively).

TAL1B (p. 61 lab journal Jarst van Belle; 09IJ55-57/67-69; available upon request)

Reaction: pFUS_B4 + pNH1 + pHD2 + pHD3 + pHD4

As expected, all 3 clones are correct (see p. 60/62 for gel pictures; *Results*: **Figures 8A/9A**). Now all forward and reverse sequences reach both *Bsp*TI and *Xba*I (the sequencing goes even beyond the Golden Gate construct *in silico* due to a smaller amount of RVDs used than 10 like with TAL1A and TAL2A).

TAL2B (p. 61 lab journal Jarst van Belle; 09IJ58-60/70-72; available upon request)

Reaction: pFUS_B5 + pHD1 + pNI2 + pHD3 + pNI4 + pNG5

As expected, clone 4-3 (see p. 60/62 lab journal for gel pictures; *Results*: **Figures 8A/9A**, TAL2B clone 3) is empty: there is no insert in between the restriction sites. The forward sequencing of clone 4-2 (clone 2) has failed; there are still reverse sequencing results. According to the sequencing results, both clone 4-1 (clone 1) and 4-2 seem to be correct (identical with the reference sequence in between the restriction sites), even though the samples look different on the gel. The negative gel result for 4-2 is possibly linked to a possible lack of *Xba*I as has been reported. The restriction digestion will be repeated, together with an extra check (3-3; TAL1B clone 3).

DAY 5

Use vector plasmid and donor plasmids (www.addgene.org) for each Golden Gate reaction #2 *in silico* using ApE. However, DNASTAR should be used for alignments with clones that have been sequenced with primer SeqTALEN 5-1 (see *Methods*), since ApE will not result in appropriate alignments. In order to perform the Golden Gate reaction, the enzyme *Esp*3I (Fermentas) needs to be added to ApE; CGTCTC(1/5). Select the enzyme for the Golden Gate reaction #2 and select the plasmids/Golden Gate reaction #1 products in the right order (vector plasmid pZHY500, TAL1A/TAL2A (see day 3), TAL1B/TAL2B (see day 3), plasmid pLR-NG). Be sure to make the vector plasmid circular in order to be able to perform the Golden Gate reaction #2. Align the *in silico* constructed Golden Gate reaction #2 product as the reference sequence to the sequenced samples, search for the *Stu*I and *Aat*II (Fermentas) restriction sites, which are AGGCCT(3/3) and GACGTC(5/1), and pay attention to the parts of the sequences within these two restriction sites. **In this case the actual reference sequences were the *in silico* Golden Gate reaction #2 products without the remaining sequences outside the two restriction sites; in other words the double digested Golden Gate reaction #2 products**. Below results of the research are shown. Sequences, *in silico* constructed Golden Gate reactions and alignments are available upon request and additional information can be found in the lab journal of Jarst van Belle.

TAL1 (p.84 lab journal Jarst van Belle; 67AJ73-78/85-90)

Reaction: pZHY500 + TAL1A + TAL1B + pLR-NG

As has been described above, DNASTAR should be used for alignments with clones that have been sequenced with primer SeqTALEN 5-1 since ApE does not result in appropriate alignments in this case while DNASTAR does. Clone 1.3 (67AJ75; *Results*: **Figures 10/11**, TAL1 clone 3) does completely align with the reference sequence, even though this sequence stops 60bp earlier than sequences of other clones. On the contrary, the same clones sequenced with primer TAL_R2 (see *Methods*) do align well with the reference sequence in ApE, even though the restriction sites are not covered by the reads. Clone 1.6 (67AJ90) does also align well, although it is remarkable that its read sequence is approximately 350-420bp shorter than the read sequences of the other clones. Clone 1.6 can still be correct, especially when regarding the colony PCR and double restriction digestion checks (p.84/85 lab journal), but it is safer not to use this clone. Clone 1.2 (67AJ86) possibly has an A at location 1027 instead of a T, but this base and the directly neighboring nucleotides are poorly read exactly like it is the case for clone 1.1 (67AJ85), although for the latter clone no nucleotide is recognized. The recognized A at location 1027 in clone 1.2 is highly probably a misread. Clones 1.1-1.5 will all be correct choices for further experiments, although clone 1.3 has eventually been selected for further experiments.

TAL2 (p.84 lab journal Jarst van Belle; 67AJ79-84/91-96)

Reaction: pZHY500 + TAL2A + TAL2B + pLR-NG

Again, ApE does not result in an appropriate alignment for clones sequenced with primer SeqTALEN5-1 while DNASTAR does. It is remarkable that the sequence of clone 2.1 (67AJ79; *Results*: **Figures 10/11**, TAL2 clone 1) stops ~100bp earlier and the sequence of clone 2.3 (67AJ81) starts later compared to other clones. Although these two clones can still be correct, it is preferred to select another clone for further experiments. As before, the same clones sequenced with primer TAL_R2 do generally align well with the reference sequence in ApE, even though the restriction sites are not covered by the reads. As already shown by the colony PCR and double restriction digestion checks (p.84/85 lab journal), clone 2.6 (67AJ96) is indeed incorrect. The read partly aligns with the reference sequence, but the location of the alignment is shifted to the 5' site of the reference sequence compared to the other reads and the read goes even beyond the *Stu*I restriction site. Furthermore, clone 2.3 (67AJ93) has not been read and clone 2.5 (67AJ95) seems not to align with the reference sequence. At least clones 2.2 and 2.4 (but also 2.1) will all be correct clones for further experiments, although clone 2.2 has eventually been selected.

DAY 13

pZHY013/TAL1/TAL2 (PT12) sequences are analyzed using ApE v2.0.45 (A plasmid Editor, M. Wayne Davis, University of Utah) and aligned with the (double digested) *in silico* constructed Golden Gate reaction #2 products (see *appendix v*, day 5) in ApE. Reads from primers TAL1FW/REV and TAL2FW/REV (see *Methods* and *appendix ii*; in our example 67EI17, 67EI21, 67EI19 and 67EI22 respectively; see p. 97 lab journal Jarst van Belle) are aligned with different Golden Gate reaction #2 products that correspond with the correct TAL arrays. Sequences and alignments are available upon request. All reads of the clone turn out to be correct, even though the read resulting from primer TAL2FW includes a T insert. The clone was sequenced a second time (98GA18). However, when checking the quality of the reads with Chromas v1.45 (Conor McCarthy, Griffith University), a program that opens chromatogram files, it appears that there is no additional base but rather a small gap which means that the T insert has been 'made up' in both situations (see **Figure A1**).

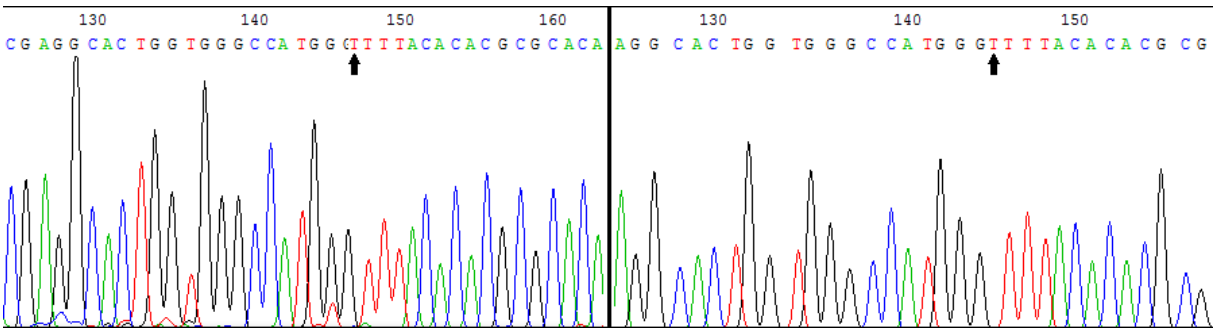


Figure A1*: Chromatogram files of a clone that has been sequenced twice with primer TAL2FW. Small, high peaks indicate high quality reads as can be seen in the figure. The indicated T nucleotides do not contain an own (red) peak, but are still read due to a gap between an adjacent G and T, of which the peak of the adjacent T seems to 'fill up' the gap, which means that there are no actual T inserts.*

DAY 15

pZHY051/TAL1/TAL2 reads from primers TAL1REV/P-35S FW are analyzed using ApE and aligned with the (double digested) *in silico* constructed Golden Gate reaction #2 product of TAL array 1, while reads from primer T-NOS R1 are analyzed with the (double digested) *in silico* constructed Golden Gate reaction #2 product of TAL array 2. In our example all clones seem to be correct, although clone 2 out of three (reads 98GA12-14) has been selected for protoplast transformation. Sequences and alignments are available upon request. Note that 98GA09-11 and 98GA15-17 represent clone 1 and 3 out of three respectively. See also p. 100/101 of the lab journal of Jarst van Belle.

Appendix vii – Wheat seeds sterilization

Eventually wheat protoplasts can be used for transformation with TALENs. Protoplasts will be developed from shoot tissue of wheat seedlings. *Triticum monococcum* (einkorn; 2n=2x=14, AA) (Shi et al. 1998) seeds were sterilized according to different methods in order to obtain a convenient protocol for wheat seeds sterilization with at least infected seeds as possible without a reduced germination capacity. Eventually six experiments were performed of which the resulting protocol includes a pre-incubation with 70% EtOH for 2min with sieves holding 15 seeds each, an incubation with 4% NaOCl without surfactants, and five washing steps with sterile water of 3min each. During each step the sieves were shaken continuously. Seeds were grown in small groups of 3 in sterile tubes with MS20 medium with vitamins (4.4g·L^{-1} MS, 20g·L^{-1} sucrose, 9g·L^{-1} agar) in a climate room (at 24°C with a 16h light and 8h dark cycle) for ~1.5 weeks (preferably ~10 days) before harvesting shoot tissue.

Methods

Experiment 1: Groups of 15 seeds in sieves were pre-incubated in 70% EtOH or sterile H_2O for 2min or not pre-incubated at all. The seeds were then incubated in either 1% or 4% NaOCl with tween 20 (a surfactant) for 5min, followed by a washing series of three times 3min with sterile H_2O. For each of the 6 treatments, one group of 15 seeds was used. Note that during each experiment the sieves were shaken continuously. The seeds were grown per group on MS20 medium with vitamins and agar (4.4g·L^{-1} MS, 20g·L^{-1} sucrose, 9g·L^{-1} agar) in a growth room at 24°C with a 16h light and 8h dark cycle. During growth the seeds were regularly checked for possible fungal infections. The results of the first experiment, as described below, led to the need of the development of an improved protocol with at least some 70% EtOH pre-incubation step.

Experiment 2: A pre-incubation step in 70% EtOH was always included. The pre-incubation times were 2, 5 or 10min. Seeds were then incubated in 2, 4, 6 or 8% NaOCl for 5min and washed in sterile H_2O for five times 3min. For each of the 12 treatments one group of 15 seeds was used.

Experiment 3: 8 groups of 15 seeds were treated with 70% EtOH for 2min, 4% NaOCl for 5min and sterile H_2O for five times 3min as had been done before. However, this time the seeds were also incubated in water overnight without shaking the seeds since this additional incubation step might improve the exclusion of fungi from the seeds as these fungi might germinate before sterilization and therefore be removed more easily. In later experiments this step was excluded.

Experiment 4: Seeds were treated with 4% NaOCl for 10, 20 or 30min and washed with sterile H_2O for five times 3min. Since it was not clear whether NaOCl has a negative effect on growth rates, especially with longer incubation times, half of the treatments did not include a 2min 70% EtOH pre-incubation. Each of the 6 treatments was performed with groups of 15 seeds in duplicates.

Experiment 5: New treatments with 4% NaOCl and different concentrations of Triton X-100 were performed. Triton X-100 is another and more aggressive surfactant than tween 20, which was also included again. Due to its aggressiveness, pre-incubations with 70% EtOH for 2min were partly excluded, resulting in 8 different treatments. In total 8 groups of 15 seeds were treated with 4% NaOCl for 5min with 0.15% tween 20 or with 0.05%, 0.1% or 0.5% Triton X-100, either with or without a 70% EtOH pre-incubation for 2min. Again, the seeds were washed with sterile H_2O for five times 3min.

Experiment 6: Seeds were treated with 70% EtOH for 2min and 4% NaOCl for 5 or 10min with either 0.15% tween 20 or no surfactant at all and washed with sterile H_2O for five times 3min. Although research was already performed toward the effect of tween 20 and the incubation time in NaOCl, nevertheless seeds had to be grown. Moreover, in this case the seeds of each of the 4 treatments were grown in groups of 3 (in tubes with 15ml MS20 medium) instead of 15, which means that 10 groups of seeds were grown per treatment instead of 1 or 2 and trends of different effects were possibly clearer. During the experiment 4 sieves holding 30 seeds each were used.

Final protocol: 8 sieves holding 15 seeds were used and the treatment included a pre-incubation of 2min 70% EtOH, an incubation of 4% NaOCl for 5min and five 3min washing steps in sterile H_2O (like before, the sieves were shaken continuously). Again, seeds were grown in groups of 3, resulting in a total amount of 40 tubes. Results of the final protocol were tracked twice.

Results

Experiment 1: Results after 6 days. Seedlings with clear shoots were reported. Note that the amount of seeds treated with only 4% NaOCl + tween 20 was 14 instead of 15 due to loss. See also **Table A2**.

Table A2*: Infection rates and number of shoots for each wheat seeds sterilization treatment of experiment 1 after 6 days. The treatment with only 4% NaOCl and tween 20 included 14 seeds instead of 15 (indicated by an asterisk).*

Pre-incubation (2min)	% NaOCl + tween 20 (5min)	Infected seeds (n)	Shoots (n)
-	1	12	9
-	4	12*	5
Sterile H_2O	1	6	1
Sterile H_2O	4	9	2
70% EtOH	1	1	4
70% EtOH	4	1	7

Experiment 2*:* Results after 3 days included the disposal of trays of treatments 2-6 (2min 70% EtOH, 6% NaOCl), 2-8, 5-2, 5-4 and 10-2 as they included infected seeds. After 5 days results from the

remaining treatments were reported. Shoots that were approximately 1cm or longer were counted. See also **Table A3**.

***Table A3**: Infection rates and number of shoots (>1cm) for each wheat seeds sterilization treatment of experiment 2 after 5 days. Seeds of several treatments were disposed after 3 days as infected seeds were present. Each treatment included 15 seeds.*

70% EtOH (min)	% NaOCl (5min)	Infected seeds (n)	Shoots (n)
2	2	2	7
2	4	0	5
5	6	0	4
5	8	1	4
10	4	0	2
10	6	2	3
10	8	1	3

Experiment 3: Results after 4 days from the treatment with an overnight water incubation (and 2min 70% EtOH, 5min 4% NaOCl and five times 3min sterile H_2O). Infection rates were 2, 5, 6, 6, 7, 8, 9 and 10 out of 15 seeds.

Experiment 4: Results after 3 days. Fungal infections were partly spotted earlier than during prior experiments by holding the trays against a light source. Seeds that showed signs of germination were reported. One group of the 10min 4% NaOCl treatment without a pre-incubation turned out to be non-sterile as there were two infected spots on the medium that were not from any of the seeds. See also **Table A4**.

***Table A4**: Infection rates and germinated seeds for each wheat seeds sterilization treatment of experiment 4 after 3 days. Each treatment was performed in duplicates of 15 seeds. Infected seeds were not reported for one of the duplicates (10min 4% NaOCl) as the medium was non-sterile itself. Note that results were different after 6 days (see below).*

70% EtOH (2min)	4% NaOCl (min)	Infected seeds (n)		Germinated seeds (n)	
No	10	-	7	7	10
Yes	10	3	0	12	13
No	20	9	9	10	11
Yes	20	0	2	11	12
No	30	6	4	9	9
Yes	30	0	0	9	11

After 6 days, the prior supposedly sterile seeds with a 2min pre-incubation in 70% EtOH still included infections (10min 4% NaOCl had 2 infections, 20min 4% NaOCl had 1 infection, and one of 30min 4% NaOCl had 2 infections). Only one of the 30min 4% NaOCl duplicates (with 11 germinated seeds after 3 days) remained free of infections. After 7 days 8 out of 15 seeds of this duplicate showed shoots that were longer than 1cm, and 14 out of 15 seeds showed signs of germination.

Experiment 5: Results after 4 days are shown in **Table A5**. Signs of germination were reported, as well as the numbers and relative sizes of shoots. Apart from the shoots, treatments with

0.5% Triton X-100 resulted mainly in white tissue. In general, higher concentrations of Triton X-100 resulted generally in less shoots and lower growth rates (differences were visually clear), while infections still remained present and infection rates were comparable with other treatments of other experiments. However, infection rates were possibly negatively affected for treatments without 70% EtOH for 2min, although treatments without EtOH always resulted in high infection rates compared to treatments with EtOH and still did. In this experiment, the treatment with 0.05% Triton X-100 resulted in more shoots than the treatment with tween 20, yet the sizes of the shoots for tween 20 with an EtOH pre-incubation were remarkably larger after 7 days of growth.

Table A5: *Infection rates and germinated seeds as well as the numbers and relative sizes of shoots for each wheat seeds sterilization treatment of experiment 5 after 4 days. A pre-incubation step with 2min 70% EtOH was partly excluded due to the aggressiveness of surfactant Triton X-100. Any signs of germination were reported. Relative sizes of shoots other than small (medium and large) as shown in the table were based on the height of the trays in which the seeds were grown. Medium and large shoots approximately reached half or whole of the height respectively. Large shoots could even exceed this height. In some cases one seed was lost during sterilization (indicated by an asterisk).*

70% EtOH (2min)	4% NaOCl (5min) + tween 20 (%)	4% NaOCl (5min) + Triton X-100 (%)	Infected seeds (n)	Germinated seeds (n)	Shoots (n)	Relative sizes shoots
Yes	0.15%	-	2	13	5	3 Large
No	0.15%	-	12	13*	6	-
Yes	-	0.05%	2	13	9	4 Small/Medium
No	-	0.05%	12	14	7	4 Medium
Yes	-	0.1%	4	12	6	-
No	-	0.1%	4	14*	5	3-4 Medium
Yes	-	0.5%	1	12	1	-
No	-	0.5%	6	9	3	1 Large

Experiment 6: Results of experiment 6 after 4, 8, 11 days and 5 weeks and 4 days (due to vacation); see **Table A6**. Infected tubes and the numbers and sizes of shoots per treatment were determined after 4 and 6 days, as well as the harvestable shoots after 8 and 11 days. Infected tubes were disposed after 6 days. After 5 weeks and 4 days, it turned out that 8 tubes turned out not to be sterile after all after this time. Moreover, the material was considered too old for protoplast transformation due to the presence of leaf yellowing. Infections and growth rates seemed not to differ when comparing treatments with or without tween 20 and with different incubation times in 4% NaOCl.

***Table A6**: Results of experiment 6 (with 2min 70% EtOH, 5min 4% NaOCl with or without tween 20 and five 3min washing steps with sterile H_2O) 4, 6, 8 and 11 days after sterilization. After 4 days, shoots of at least 1cm were reported, while after 4 days shoots that reached half the height of the tubes were reported. After 8 and 11 days, shoots ready for harvesting (primarily shoots that reached the height of the tubes) were reported. Note that due to subjective, inaccurate measurements 2 shoots less were counted after 11 days than after 8 days.*

	t=4		**t=6**		**t=8**	**t=11**
4% NaOCl (min)	**Infected tubes (n)**	**Shoots (n) (>1cm)**	**Infected tubes (n)**	**Shoots (n) (~½h)**	**Shoots (n) (ready)**	**Shoots (n) (ready)**
5min (+ tween 20)	1	2	3	0	6	5
5min	2	4	3	4	7	7
10min (+ tween 20)	3	3	5	1	5	4
10min	3	6	4	3	6	6
Total (n)	9	15	15	8	24	22

Final protocol: The first time this protocol was performed, 3 weeks and 2 days after sterilization (due to vacation) 21 out of 40 tubes were left due to infections, and still 14 harvestable shoots were counted, but there was also much old material present. The second time 29 out of 40 tubes were left after 6 days and 12 harvestable shoots were present, of which 11 reached the height of the tubes. 3 days later, after 9 days, the number of harvestable shoots had doubled (26) and in total there were 6 additional small shoots present. Eventually 24 of 40 tubes were left; most tubes were disposed after 6 days, but a couple of tubes showed novel visible infections after 9 days. After 13 days 6 large shoots started to get too old for harvesting, probably caused by bruised shoot tissue due to shoots that reached lengths that were higher than the sealed tubes heights.

Conclusion

Experiment 1: A pre-incubation in EtOH will be necessary for excluding fungal infections from the seeds. There are no clear indications whether 1% and 4% NaOCl solutions affect the germination and infection rates of the seeds differently.

Experiment 2: A longer 70% EtOH pre-incubation than 2min might not improve seeds sterilization but there is a trend that it reduces the viability of the seeds and therefore a treatment with 70% EtOH will need to remain 2min. NaOCl solutions higher than 0.5-5% seem not to reduce the infection rates but are also possibly not affecting the germination rates.

Experiment 3: An overnight incubation in water will not reduce but rather increase the fungal infection rates instead and should therefore not be included in a seeds sterilization protocol.

Experiment 4: Once again the results show that it will be necessary for seeds sterilization to include some EtOH pre-incubation step, even with longer incubations in NaOCl. Germination and infection rates seem not to decrease with longer NaOCl incubations, although there is possibly a slight trend.

Experiment 5: Again, results show that a pre-incubation step with 2min 70% EtOH should be included in any seeds sterilization protocol, even when using an aggressive surfactant like Triton X-100. Triton X-100 and tween 20 seem not to reduce infection rates compared to treatments without surfactants of other experiments. However, Triton X-100 shows a clear visible trend in which the numbers of shoots and the growth rates are affected negatively when using higher concentrations of this surfactant. Moreover, when working with surfactants, especially Triton X-100 is not very convenient to work with due to its high viscosity. Therefore using Triton X-100 should be excluded from the protocol. Although using tween 20 also seems not to be advantageous, first a more extensive study should be conducted in a following experiment (experiment 6). Finally, it can be concluded that it will most likely be impossible to obtain a protocol that guarantees no infections at all. It is expected that there are also fungi presents inside the available seeds, rather than just on the surface.

Experiment 6: Infection and growth rates seem not to differ for treatments with different incubation times in 4% NaOCl and either with or without the addition of surfactant tween 20. Therefore, a 5min incubation in 4% NaOCl without tween 20 is recommended for sterilizing the available wheat seeds. Moreover, shoots could already be harvested for the generation of protoplasts approximately >1 week after sterilization (according to this experiment at least after 8 days, of which the amount of harvestable shoots is not different after 11 days). It is expected that most shoots will be harvestable approximately 1 to 2 weeks after sterilization, yet fresher shoots (after ~1.5 weeks, preferably ~10 days) are preferred. When there is the need to always be able to generate protoplasts from shoots, (120) seeds should be sterilized and grown at least once a week for sufficient fresh material (estimated up to 25 harvestable shoots).

Final protocol and overall conclusion: In earlier experiments, ~80% of the seeds remained free from infections and these optimal observed infection rates were achieved by including a pre-incubation of the seeds in 70% EtOH and not including an overnight pre-incubation in water. Other additions (longer incubation times in EtOH or NaOCl, surfactants) seemed not to decrease infection rates further but possibly had negative effects on growth rates of the shoots, and were therefore neglected. In later experiments, including experiments performed according to the final protocol, also ~80% of the tubes remained free of infections but by growing the seeds in smaller groups of 3 seeds instead of 15, the spread of pathogens from infected seeds was minimized. As infections still might occur later (according to experiment 6), it is recommended to not let the germinated seeds touch the shoots while harvesting. Seeds from the infection-free tubes (~80% x 40 tubes x 3 = ~96 seeds) resulted on average in ~25 harvestable shoots (enough for 10 transformation experiments; see *appendix viii*) after nearly 1.5 weeks (9-10 days), which is roughly 25% of the infection-free seeds and 20% of the total amount of seeds which is a convenient protocol for the current situation.

Appendix viii – Wheat protoplasts isolation and transformation

After the TALENs have been assembled (*appendix i*) and moved into an expression-ready vector (*appendix ii*), protoplasts can be isolated from ~1.5 weeks (preferably ~10 days) old seedlings of which the seeds have been sterilized and grown under optimal conditions (*appendix vii*) and subsequently transformed with large amounts of highly concentrated plasmid DNA. When taking into account that practical work is only performed during working weeks, wheat seeds can be sterilized during day 11 (*appendix ii*) and grown for 9 days prior to harvesting of shoots (~25). It is possible to assemble the TALENs and move the construct into the expression-ready vector in 3 working weeks, and during the fourth week (days 16-20) large amounts of highly concentrated plasmid DNA can be obtained and protoplasts can be isolated and transformed. This appendix describes the protocols for isolation and transformation of wheat protoplasts as described by Shan et al. 2013, including an optimized protocol for high yields of DNA based on the Plasmid Midi Kit protocol (Qiagen). It is recommended to start the experiments on a Monday in order to finish in one working week. Although sequencing results for verification of the plasmid DNA (pZHY051 with the TALENs construct) will not be able to be analyzed in time, LR reactions are highly efficient and the resulting clones will in principle always be correct in practice.

DAY16

1. Pick a sample from a bacterial glycerol stock of pZHY051 with the TALENs construct as has been obtained during day 15 (*appendix ii*) and incubate a starter culture (10ml tube 3ml LB medium + 1:1000 $Spec^{50}$) at 37°C for ~8h on the shaker platform (preferably at 250rpm). Autoclave (up to) 6 100ml Erlenmeyer flasks in the meantime. The volumes of the flasks should be at least 4 times the volumes of the cultures (25ml) that are incubated later.

2. Incubate (up to) 6 cultures (25ml LB medium + 1:1000 $Spec^{50}$ with 30µl starter culture) in the 100ml flasks at 37°C on the shaker platform (250rpm) for 12-16h. Optionally a new bacterial glycerol stock can be prepared from the starter culture (0.4ml culture with 0.4ml LB medium + 30% glycerol) that replaces the previously used one.

DAY17

3. Midiprep the cultures as follows:
    1. Decant the cultures to 50ml tubes and harvest the bacterial cells by centrifugation at 6,000xg (in our labs, 4,556xg is the maximum) for 15min at 4°C (pre-cooling the centrifuge takes ~1h).

2. Resuspend each bacterial pellet in 4ml Buffer P1 (with 40µl RNase A solution and 4µl LyseBlue reagent); the bacteria should be resuspended completely by vortexing or pipetting up and down until no cell clumps remain.
3. Add 4ml Buffer P2, mix thoroughly by vigorously inverting the sealed tubes 4-6 times (in our experience 6-10 times are needed to obtain homogeneously colored suspensions), and incubate at room temperature for 5min, but do not allow the lysis reaction to proceed for more than 5min. Do not vortex, as this will result in shearing of genomic DNA.
4. Add 4ml of chilled Buffer P3 (incubated on ice), mix immediately and thoroughly by vigorously inverting 4-6 times, and incubate on ice for 15min.
5. Decant the suspensions to 50ml tubes that each contain a funnel and micracloth. Optionally press on the micracloth after decanting for higher yields. This step is not described in the Plasmid Midi Kit protocol but it is more efficient and it makes one of the two 30min centrifugation steps redundant.
6. Centrifuge at 4,566xg at 4°C for 30min. Remove the supernatants promptly.
7. Equilibrate (up to) 6 QIAGEN-tip 100 by applying 4ml Buffer QBT, and allow the columns to empty by gravity flow.
8. Apply the supernatants from step 8 to the QIAGEN-tips and allow it to enter the resins (50ml tubes) by gravity flow. If the supernatants are left too long and become cloudy due to further precipitation of protein, they must be centrifuged again.
9. Wash the QIAGEN-tips with 2 x 10ml Buffer QC (be careful to not allow it to enter the resins).
10. Elute DNA with 5ml Buffer QF.
11. Precipitate DNA by adding 3.5ml (0.7 volumes) of (room-temperature) isopropanol to the eluted DNA. Mix, make 4 aliquots of each sample using 2ml microcentrifuge tubes, and centrifuge immediately at maximum g (≥15,000xg) at 4°C for 30min. When starting the protocol with 6 flasks, you will have a total of 24 aliquots that can be centrifuged at 4°C using one centrifuge only. Carefully decant the supernatants.
12. Wash the DNA pellets with 500µl of (room-temperature) 70% ethanol, and centrifuge at maximum g (≥15,000xg) for 10min. Carefully decant the supernatant without disturbing the pellet.
13. Air-dry the pellets overnight; place a tissue over the aliquots to avoid contamination. If necessary, this step can be performed overweekend as well without problems.

DAY18

4. Before eluting the DNA samples, start with the isolation of wheat protoplasts. Cut fresh green shoots of wheat seedlings (~25) into fine strips in a petri dish; try to avoid lifting the germinated seeds as they may contain 'hidden infections'. Cutting the strips can be facilitated by adding some fresh filter-sterilized enzyme solution (1.5% Cellulase R10, 0.75% Macerozyme R10, 0.6M mannitol, 10mM MES, 10mM $CaCl_2$, 0.1% BSA; pH 5.7, adjust with KOH). 40ml enzyme solution is more than sufficient for the whole protocol and can be prepared in a 50ml tube. Use a special balance while adding Cellulase R10 and Macerozyme R10 as these compounds are highly allergenic.

5. Start digesting the strips by adding enzyme solution (40ml is more than sufficient for one single petri dish) and perform a vacuum-infiltration for 30min in the dark using a vacuum pump at -15~20 (in Hg). Continue digestion of the tissue with gentle shaking (20-30rpm) for preferably 6-7h.

6. In the meantime, elute the air-dried DNA pellets of day 17. Add 50µl EB buffer to aliquots 1 of 4, dissolve the pellets, decant to the second aliquots, dissolve the pellets, etcetera. The 6 samples, each containing DNA dissolved in 50µl EB buffer, can be added together, resulting in a total volume of 300µl. The DNA concentration will be ~2.5µg·µl on average (based on previous results) while only 10µg is needed for a transformation experiment.

7. Prepare the following solutions:
   a) W5 washing solution (154mM NaCl, 125mM $CaCl_2$, 5mM KCl, 2mM MES; pH 5.7, adjust with KOH). 40ml is more than sufficient for the whole protocol and can be prepared in a 50ml tube. Filter-sterilize and store at room temperature.
   b) MMG solution (0.4M mannitol, 15mM $MgCl_2$, 4mM MES; pH 5.7, adjust with KOH). 8ml is more than sufficient for the whole protocol and can be prepared in a 10ml tube. Filter-sterilize and store at room temperature.
   c) PEG solution (40% (W/V) PEG 4000, 0.2M mannitol, 0.1M $CaCl_2$). 5ml is sufficient. Note: Start with only 3ml deionized water before adding PEG 4000 as this compound will increase the total volume to almost 5ml; wait at least 1h before filter-sterilization as the compounds should completely dissolve. Filter-sterilization can be performed during a later step.

8. Isolate protoplasts by filtration using sterilized beakers with filter holders containing two nylon filters of 250µm (upper filter) and 55µm (lower filter) and wash with W5 solution (1:1, ~25ml). Decant into round 10ml tubes, centrifuge twice at 80xg for 3min and discard as much of the supernatants as possible in a quickly and careful manner as the pellets are very loose. Add the suspensions together, centrifuge again and discard as much of the supernatant as possible. Keep the protoplasts on ice for 30min. PEG solution can be filter-sterilized during this step. Note: do not decant any protoplast suspension vigorously as this will result in broken protoplasts.

9. Resuspend the protoplasts in MMG solution. $1 \cdot 10^7$ protoplasts can be obtained from 50 shoots which is sufficient for 20 transformation experiments (using the current wheat seeds sterilization protocol, ~25 shoots will be available).

10. Transform the isolated protoplasts in PEG solution. Agitate transformation mixtures (10) gently (250µl PEG solution with 200µl protoplasts and 10µg plasmid DNA). Incubate at room temperature in the dark for 30 min. Harvest the protoplasts and wash with 800µl W5 solution. Centrifuge and resuspend in 2ml W5 solution and culture in 6-well plates at room temperature in the dark for 48h. Note: When starting the experiments of this appendix on a Monday as recommended, the 48h incubation step will be finished on a Friday, which is within the working week.