Study Course AgriGenomics

Master Thesis

Knock-out of the phytic acid synthesis gene *BnIPK2β* in rapeseed by CRISPR/Cas9

submitted by

Noelle Blanc

Reviewer 1: Prof. Dr. Christian Jung

Reviewer 2: Dr. Hans Joachim Harloff

Institute of Crop Science and Plant Breeding

Faculty of Agricultural and Nutritional Sciences

Christian-Albrechts-University of Kiel

List of abbreviations.

2-PGK	2-phosphoglycerate kinase
ABA	Abscisic acid
BAR	Bialaphos resistance
blastn	Nucleotide-nucleotide BLAST
BLAT	BLAST-like alignment tool
BnA06_IPK2β	BnaA06g31750D
BnA09_IPK2β	BnaA09g05790D
BnA10_IPK2α	BnaA10g23820D
BnC09_IPK2α	BnaC09g48570D
BnC09_IPK2β	BnaC09g05390D
BnCnn_IPK2β	BnaCnng33290D
Ca.	Circa
CaM	Calmodulin
Cas9	CRISPR associated protein 9
CPK4	Calcium-dependent protein kinase 4
CRISPR	Clustered regulatory interspaced short palindromic repeats
crRNA	CRISPR RNA
CTAB	Cetyltrimethylammonium bromide
DSB	Double strand break
E-value	Expect value
ExPASy	Bioinformatics resource portal
Genoscope	French National Sequencing Center
GFP	Green fluorescent protein
HR	Homologous recombination
IKMB	Institute for Clinical Molecular Biology
IMP	Inositol monophosphate phosphatase
InDels	Insertion or deletion
Ins	Inositol
Ins(1,3,4,5)P_4	Inositol 1,3,4,5-tetrakisphosphate
Ins(1,3,4,5,6)P_5	Inositol 1,3,4,5,6-pentakisphosphate
Ins(1,4,5)P_3	Inositol 1,4,5-trisphosphate
Ins(3)P_1	*myo*-inositol 3-phosphate
Ins(3,4,6)P_3	Inositol 3,4,6-trisphosphate
InsP	Inositol phosphate
InsP$_2$	Inositol bisphosphate
InsP$_3$	Inositol tri-phosphate
InsP$_3$ - IP$_3$	Inositol trisphosphate
InsP$_4$	Inositol tetrakisphosphate
InsP$_5$	Inositol pentakisphosphate

$InsP_6$ - Ins(1,2,3,4,5,6)P_6	Phytic acid
IP - $InsP_1$	Inositol monophosphate
IPK1	Inositol polyphosphate 2-kinase
IPK2	Inositol polyphosphate multikinase 2
IPMK	Inositol polyphosphate multikinase
ITPK	Inositol 1,3,4-trisphosphate 5-/6-kinase
lpa	*low phytic acid*
lpa1-1	*low phytic acid 1-1*
lpa2-1	*low phytic acid 2-1*
MIK	*myo*-inositol kinase
MIPS	D-*myo*-inositol 3-phosphate synthase
MRP	Multidrug-resistance-associated protein
MSA	Multiple sequence alignment
NHEJ	Non-homologous end joining
NLS	Nuclear localization signal
***OsLpa*1**	*Oryza sativa low phytic acid 1*
PAM	Protospacer-adjacent motif
PEG	Polyethylene glycol
PFAM	Protein family database
Phytase	myo-inositol (1,2,3,4,5,6) hexakisphosphate phosphohydrolase
Pi	Inorganic phosphate
PLC	Phospholipase C
PPT	Phosphinothricin
PSV	Protein storage vacuole
PtdIns	Phosphatidylinositol
PtdIns kinase	Phosphatidylinositol kinase
PtdIns(4,5)P_2	Phosphatidylinositol 4,5-bisphosphate
PtdIS	Phosphatidylinositol synthase
Rn$I_3$3k-A	IPK2 rat orthologue
sgRNA	Single-guide RNA
SpCas9	*Streptococcus pyogenes* Cas9
TAIR	The *Arabidopsis* information resource
TALEN	Transcription activator-like effector nuclease
TIR1	Transport inhibitor response 1
tracrRNA	*trans*-activating CRISPR RNA
TS_1	Target site 1
TS_2	Target site 2
ZFN	Zinc finger nuclease

1 Table of contents

2 Introduction 8
2.1 Oilseed rape 8
2.1.1 Phytic acid 8
2.2 CRISPR/Cas9 technology 13
3 Aims and Questions 14
3.1 Aims 14
3.2 Questions 14
4 Materials and Methods 14
4.1 *In silico* analysis 14
4.1.1 Retrieve *AtIPK2β* homologs of *B. napus* 14
4.1.2 Functional domain prediction 14
4.1.3 CRISPR/Cas9 target site design 15
4.2 Plant material, bacteria strains, and cloning vectors 15
4.3 CRISPR/Cas9 target site genotype confirmation and construct development. 15
4.4 Plant transformation 17
4.5 Transgene and gene edition detection 17
4.5.1 DNA isolation 17
4.5.2 Transgene detection and edition of the target region 18
5 Results 19
5.1 *In silico* analysis 19
5.1.1 *AtIPK2β* homologs of *B. napus* 19
5.1.2 Functional domain prediction and target site design 20
5.1.3 CRISPR/Cas9 target site region genotyping 23
5.2 Construct development 24
5.2.1 pChimera assembly 24
5.2.2 pCas9-TPC assembly 24
5.3 Plant transformation 25
5.4 Transgene detection and gene editing 28
6 Discussion 29

6.1 *In silico* analysis 29
6.2 *AtIPK2β* homologs from *B. napus* 29
6.3 Functional domain prediction and target site design 29
6.4 Construct development 31
6.5 Plant Transformation 32
6.6 Transgene detection and gene editing 34
6.7 Prospects 34
7 Summary 35
8 References 36
9 Supplementary data 42
10 Declaration 57

List of Tables

Table 1: Genes analyzed in this study. 19
Table 2: CRISPR/Cas9 target sites for genome edition of *BnIPK2β* paralogs. Where TS_1 is the target site aiming to knock-out *BnA06_IPK2β* and *BnCnn_IPK2β* paralogs and, TS_2 is the target site aiming to knock-out *BnA09_IPK2β* and *BnC09_IPK2β* 22
Table 3: Rapeseed hypocotyl transformation experiments performed in this study. 27

List of Figures

Figure 1: $InsP_6$ biosynthesis pathway in *A .thaliana*. The early pathway supplies substrates that are converted via inositol lipid-dependent (PtdIns) and lipid-independent (InsP) pathways to $InsP_3$. Various inositol phosphatases kinases and phosphatases complete the conversion to $InsP_6$. Modified after Raboy, 2003, 2009. 10
Figure 2: Cas9/sgRNA binding and cleavage model. A) The crystal structure of SpCas9 in complex with guide RNA and target DNA. The sgRNA is tightly bound with the Cas9 protein. The first 20-nt of sgRNA (red) are base paired with DNA target (orange), and the rest of the sgRNA nucleotides (green) have rich interactions with the Cas9 protein. B) The competition between base stacking within the DNA and DNA-RNA hybrid base pairing/stacking results in the different DNA/RNA bound structures. In the unbound state, the target site maintains its original DNA-DNA base pairing within the chromatin. In the bound state, the sgRNA invades into the DNA duplex and forms the R-loop structure with the target DNA. The three-base-pair DNA helix stretches on both ends of the R-loop are shown in blue. S denotes the length (number of base pairs) of the hybrid helix (Xu et al., 2017). 13
Figure 3: Multiple protein sequence alignment of IPK2α and IPK2β from *A. thaliana, B. napus, B. oleracea,* and *B. rapa.* Residues highlighted in dark blue corresponds to IP3 binding domain (PxxxDxKxG), residues underlined by a black bar correspond to CRISPR/Cas9 target site, while residues highlighted in red are phosphorylation sites by CPK4, residues highlighted in dark green are critical for ATP/Mg^{+2} binding for β paralogs, and residues highlighted in light blue are critical for substrate specificity in AtIPK2α. 21
Figure 4: Graphical representation of *BnIPK2β* gene structure, CRISPR/Cas9 target sites, and motifs. Where the grey bar represents the sole exon of the paralogs, the red arrows represent the target site 1, the green arrow the target site 2, the black bars important nucleotides for ATP/Mg^{+2} binding, and the blue bars represent the important nucleotide for IP_3 binding. 22
Figure 5: Sequence alignment of genotyping sequencing, where the grey arrow shows the target site position for CRIPSR/Cas9. A. *BnA06_IPK2β*. B. *BnCnn_IPK2β*. C. *BnA09_IPK2β*. D. *BnC09_IPK2β*. 23
Figure 6: Colony PCR from pCAS9-TPC:TS_1:TS_2. Where L corresponds to Scientific O'GeneRuler 1 kb Plus DNA Ladder, 1 to 17 and 19 to 35 corresponds to colonies samples, 18 and 36 corresponds to the negative control. 25
Figure 7: Binary vector system used in this study. A. pChimera empty vector. B. pCas9-TPC empty vector. C. pChimera:TS_1 and pChimera:TS_2 obtained after cloning sgRNAs. D. pCas9-TPC:TS_1-TS_2 obtained after subcloning sgRNA:TS_1 and sgRNA:TS_2. 26

Figure 8: Cas9 and spectinomycin PCR from tissue culture leaf samples. A) Cas9 PCR, where L corresponds to 1 kb plus ladder, 1 to 16 to leaf samples, 17 to pCas9-TPC, 18 to pCas9-TPC:TS_1, 19 to pCas9-TPC:TS_1:TS_2, and 20 to the negative control. B) Spectinomycin PCR, where L corresponds to 1 kb plus ladder, 1 to 16 to leaf samples, 17 to pCas9-TPC, 18 to pCas9-TPC:TS_1, 19 to pCas9-TPC:TS_1:TS_2, and 20 to the negative control. 28

List of Supplementary Tables

Supplementary Table S. 1: Primers used in this study. 42
Supplementary Table S. 2: Digestion of pChimera.....42
Supplementary Table S. 3: Oligo annealing.....42
Supplementary Table S. 4: Oligo ligation into pChimera 43
Supplementary Table S. 5: Bacterial medium and antibiotics 43
Supplementary Table S. 6: PCR and Colony PCR, reagents and conditions. 43
Supplementary Table S. 7: sgRNA restriction digestion from pChimera. 44
Supplementary Table S. 8: Digestion of pCas9-TPC.....44
Supplementary Table S. 9: pCas9-TPC dephosphorylation. 44
Supplementary Table S. 10: sgRNA:TS_1 ligation into pCas9-TPC 44
Supplementary Table S. 11: Digestion of pCas9-TPC:TS_1.....44
Supplementary Table S. 12: sgRNA:TS_2 ligation into pCas9-TPC:TS_1 44
Supplementary Table S. 13: DNA isolation by CTAB method 45
Supplementary Table S. 14: putative off-target genes found by blast. PAM site of target sequences are underlined, and mismatches between the target sequences and putative off-target sequences are in bold.....45
Supplementary Table S. 15: Identity matrix of protein sequences analyzed in this study. 47

List of Supplementary Figures

Supplementary Figure S. 1: pChimera colony PCR, where L corresponds to 50 bp ladder, 1 to 13 colonies samples from pChimera:TS_1, 15 to 28 colonies samples from pChimera:TS_2, 14 and 28 are negative controls.....48
Supplementary Figure S. 2: Sequence alignment of PCR product from pChimera with the sgRNA (TS_1) annealed between the promoter and the sgRNA.....49
Supplementary Figure S. 3: Sequence alignment of PCR product from pChimera with the sgRNA(TS_2) annealed between the promoter and the sgRNA.....50
Supplementary Figure S. 4: Colony PCR after cloning sgRNA:TS_1 into pCAS9-TPC, yielding pCAS9-TPC:TS_1. Where L corresponds to 50 bp ladder, 1 to 13 colonies used as a template, and 14 to the negative control. 51
Supplementary Figure S. 5: Sequence alignment of PCR products from pCas9-TPC:TS_1 and of pCas9-TPC:TS_2. 52
Supplementary Figure S. 6: Sequence alignment of vector pCas9-TPC:TS_1:TS_2, where the sgRNAs are shown separately for the ease of understanding. 54
Supplementary Figure S. 7: Sequence alignment of PCR product of pCas9-TPC:TS_1:TS_2 against the reference. 54

Supplementary Figure S. 8: Cas9 PCR from bacterial stock, where L corresponds to 1 kb ladder, 1 to 3 corresponds to pCas9-TPC:TS_1 samples, 4 to 8 to pCas9-TPC:TS_2 samples, 9 to 15 to pCas9-TPC:TS_1:TS_2, and 16 to the negative control. 55
Supplementary Figure S. 9: Colony Cas9 PCR from *A. tumefaciens*, where L corresponds to 1 kb ladder, sample 1 to pCas9-TPC:TS_1, sample 2 to pCas9-TPC:TS_2, and sample 3 to pCas9-TPC:TS_1:TS_2, and sample 4 to the negative control. 55
Supplementary Figure S. 10: Cas9 PCR using the surrounding medium from tissue culture shoots as a template. Where L corresponds to 1 kb ladder, 1 to 17 corresponds to tissue culture medium samples, and 18 to the negative control. 55
Supplementary Figure S. 11: 4 weeks old hypocotyl explants of Express 617 in regeneration medium O_4, without a selective agent. 56
Supplementary Figure S. 12: Pink bacteria (black arrows) growing in hypocotyls explants four weeks after *A. tumefaciens* infection. 56
Supplementary Figure S. 13: Gene editing PCR. Where L corresponds to 1 kb ladder, 1 to 11 to leaf samples, and 12 to the negative control. 56

2 Introduction

2.1 Oilseed rape

Oilseed rape (*Brassica napus* L.) is a major crop due to its oil and protein content. This species was originated near 7,500 years ago by a process called allopolyploidy on which the genome of two ancestors, *Brassica oleracea* (n=9, genome C_oC_o) and *Brassica rapa* (n=10, genome A_rA_r), where merged and doubled, giving birth to the new species *B. napus* (n=19, genome $A_nA_nC_nC_n$) (Chalhoub et al., 2014). This crop is grown principally for the production of oil for human consumption, because of its high seed oil content, which ranges between 40 to 50% of dry matter. After oil extraction, the remaining meal contains about 40% of protein. Therefore, it is used as animal feed (Hüsken et al., 2005). Another compound which is present in substantial amounts of the seed is phytic acid ranging from 2 to 4% for the whole dry seed (El-Batal et al., 2001; Uppström et al., 1980).

2.1.1 Phytic acid

Phytic acid (*myo*-inositol-1,2,3,4,5,6-hexa*kis*phosphate; $InsP_6$) is ubiquitous in eukaryotic cells and the most abundant form of phosphorus in plants; it is involved in signal transduction and regulation, energy transfer and ATP generation, RNA export, DNA repair, and DNA recombination.

In plants, it is the major form of phosphorus in seeds, playing a major role in regulatory processes and signaling and is stored in protein storage vacuoles in the form of globoids as mixed salts of different minerals such as calcium, iron, magnesium, potassium, and zinc, in tissues and organs such as seeds, pollen, roots, and tubers.

During germination, phytate (the salt of phytic acid) is released from globoids and hydrolyzed by phytases and other phosphatases, releasing phosphate, micronutrients, and inositol, which are used as the main source of energy for the emerging seedling, until it reaches a stage of development advanced enough to absorb essential minerals by itself. Interestingly, this process of breaking down

the molecule happens in a way that the levels of inorganic phosphate (Pi) and cellular P remains constant, maintaining the homeostasis in the cell (Hatch et al., 2010; Raboy, 2003; Sparvoli et al., 2015).

2.1.1.1 Role of phytic acid in plants

Major efforts have been made to elucidate the role of $InsP_6$ in plants and its production and accumulation in seeds have been intensely studied. $InsP_6$ content in mature dry seeds of *B. napus* varies between 2 to 4% approximately (Uppström et al., 1980). This form of phosphorus comprises about 60 to 80% of the total phosphorus content in seeds, and it has been estimated that more than 50% of the phosphorus applied as fertilizer to crops ends stored as $InsP_6$ in seeds. Therefore seed $InsP_6$ signifies the main pool in P fluctuation through the world agriculture (Raboy et al., 2000).

It has been suggested that $InsP_6$ plays a role in pollen tube elongation in the style of pollinated plants by providing precursors for cell wall synthesis (Jackson, Jones, & Linskens, 1982). Other plant related functions involve the regulation of the K^+ channels that modulate stomatal guard cell closing induced by abscisic acid (ABA), and being a cofactor of the auxin TIR1 receptor, which is involved in regulation of auxin-mediated gene expression (Chen et al., 2017; Kim et al., 2011; Lemtiri-Chlieh et al., 2000).

Further cellular roles of $InsP_6$ that have been reviewed in eukaryotic organisms include DNA repair, stimulating nonhomologous end-joining repair, and regulation of nuclear mRNA export. Even though it remains unclear, how exactly it is involved in this process, it has been proposed that it might regulate the conformation of the nuclear pore complex, or by enabling the removal of an export inhibitor. Other roles comprise the activation of protein kinase C, cellular antioxidant, neurotransmission, apoptosis, and inhibition of protein phosphatase activities (Malabanan et al., 2016; Seeds et al, 2007; Shears, 2001).

2.1.1.2 Phytic acid synthesis

Phytic acid synthesis has been well studied in diverse organisms, including human (*Homo sapiens*), yeast (*Caenorhabditis elegans*), flies (*Drosophila melanogaster*) and mouse-ear cress (*Arabidopsis thaliana*). Here, we will use *A. thaliana* as reference organism due to its genetic proximity with *B. napus*, because they belong to the same family (Brassicaceae), and they share the same common evolutionary ancestor (Chalhoub et al., 2014).

In plants, phytic acid synthesis occurs through two different routes (Figure 1): a lipid-independent pathway, mostly occurring in seeds and a lipid-dependent pathway that is ubiquitous in all plant tissues.

First, *myo*-inositol 3-phosphate ($Ins(3)P_1$) is synthesized from D-glucose-6-phosphate by D-*myo*-inositol 3-phosphate synthase (MIPS). Then, $Ins(3)P_1$ is dephosphorylated to free Ins by inositol monophosphate phosphatase (IMP), an enzyme that possesses a dual activity (it can also hydrolyze L-galactose 1-phosphate). This last reaction can be reversed by *myo*-inositol kinase (MIK). From here, the pathway of synthesis it is divided into the lipid-dependent and lipid-independent, and the main difference is the way that they produce the higher inositol tri-phosphates ($InsP_3$).

The generation of $InsP_6$ in the lipid-independent pathway, as its name indicates, is completely independent of inositol lipid synthesis and is performed by serial phosphorylation of all Ins

hydroxyl groups, catalyzed by several inositol phosphate kinases. After the production of $InsP_1$, it is most likely that 2-phosphoglycerate kinase (2-PGK) will phosphorylate it producing $InsP_2$, the codifying gene has been well characterized in *A. thaliana* and rice (*Oryza sativa*), being denominated as *OsLpa*1 in rice, *At3g45090*, and *At5g60760* in *Arabidopsis.* The later is highly expressed during silique development and was demonstrated to be necessary for $InsP_6$ synthesis. There is no consensus about the enzyme that catalyzes the phosphorylation of $InsP_2$, producing $Ins(3,4,6)P_3$, but some authors reviewed that is very likely to be performed by an Ins polyphosphate kinase, and the product of the reaction can be further phosphorylated by inositol 1,3,4-trisphosphate 5-/6-kinase (ITPK), generating $Ins(1,3,4,5,6)P_5$. In this step lipid-independent and lipid-dependent pathway will converge, the last remaining phosphorylation step performed by an inositol polyphosphate 2-kinase (IPK1) that will yield $Ins(1,2,3,4,5,6)P_6$.

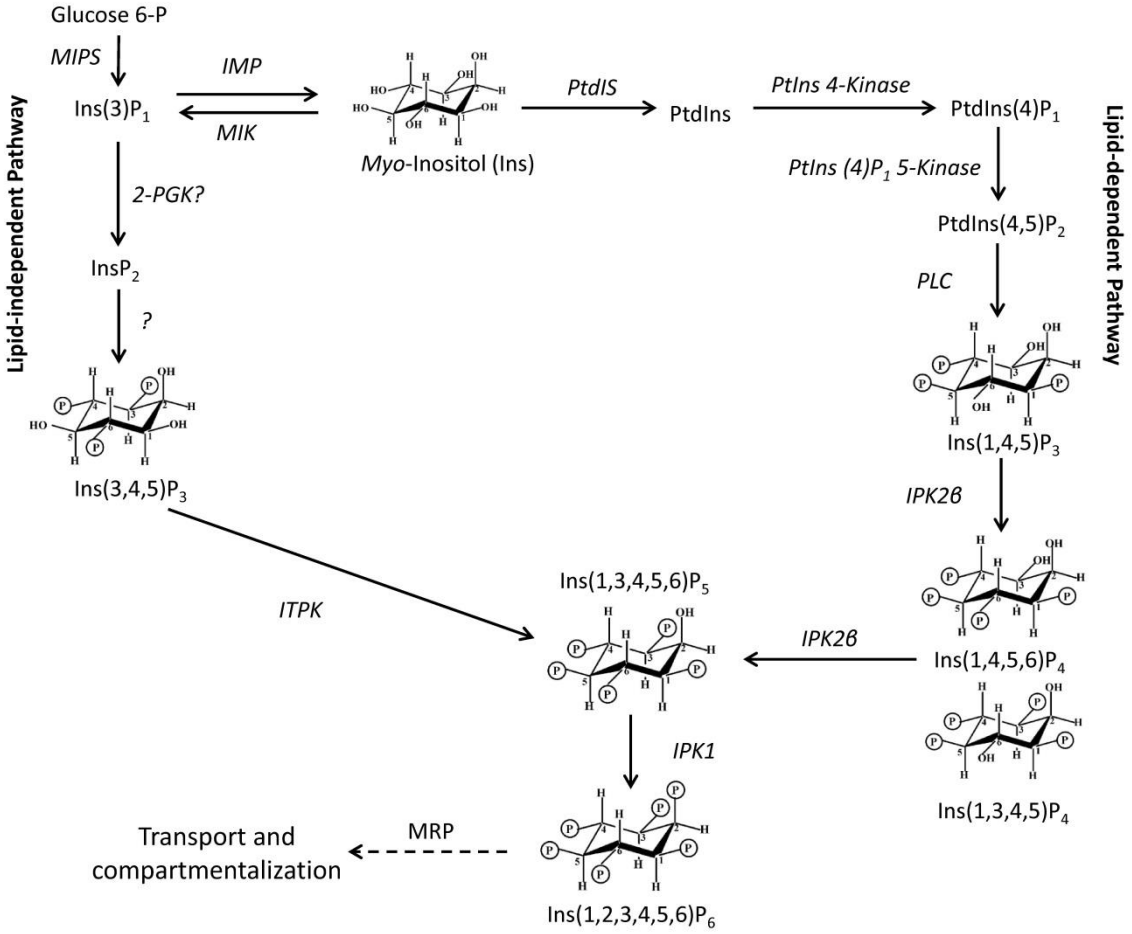


Figure 1: $InsP_6$ biosynthesis pathway in *A .thaliana*. The early pathway supplies substrates that are converted via inositol lipid-dependent (PtdIns) and lipid-independent (InsP) pathways to $InsP_3$. Various inositol phosphatases kinases and phosphatases complete the conversion to $InsP_6$. Modified after Raboy, 2003, 2009.

In the lipid-dependent pathway, phosphatidylinositol (PtdIns) is obtained by the conversion of Ins by a phosphatidylinositol synthase (PtdIS), then PtdIns is sequentially phosphorylated by phosphatidylinositol kinases (PtdIns 4-Kinase and $PtdIns(4)P_1$ 5-kinase), producing $PtdIns(4,5)P_2$. After that, the molecule serves as a substrate of Phospholipase C (PLC), an enzyme that performs sequential steps of hydrolysis, releasing diacylglycerol and $Ins(1,4,5)P_3$, second messenger molecules, relevant for signal transduction. The production of higher phosphorylated $InsP_4$ and $InsP_5$ is modulated by the highly conserved inositol polyphosphate multikinase 2 (IPK2β), an

enzyme that possesses a promiscuous 6-/3-kinase activity, yielding either Ins(1,3,4,5)P_4 or Ins(1,3,4,5,6)P_5. As mentioned above, this is the step where the lipid-independent and lipid-dependent pathway converge, before the last specific phosphorylation step in the second position to Ins(1,2,3,4,5,6)P_6. $InsP_6$ is actively transported into phytate inclusions found within protein storage vacuoles (PSVs), called globoids, by a multidrug-resistance-associated protein (MRP), belonging to the ATP-binding cassette family (Endo-Streeter et al., 2012; Raboy, 2003; Seeds et al., 2007; Sparvoli et al., 2015; Stevenson-Paulik et al., 2005; Hui-Jun Xia et al., 2005).

2.1.1.3 Nutritional value of phytic acid and environmental considerations

Since the $InsP_6$ molecule is highly phosphorylated, it acts as a polyanion at physiological pH, and it is considered as a strong chelator of metal ions, granting an antioxidant function by preventing the formation of free hydroxyl radicals. Other beneficial properties of $InsP_6$ consumption have been reported, like protection against different types of cancer such as colon cancer, mammary carcinoma, hepatocellular carcinoma, pancreatic cancer, blood and bone marrow cancer, and reduction of incidence of common diseases, for instance; coronary heart disease, fatty liver, and diabetes mellitus (Kumar et al., 2010).

The chelator ability of the $InsP_6$ molecule due to its unique structure makes it the perfect hijacker of important plant nutrients such as calcium, magnesium, iron, and zinc, being the same case when humans and animals are fed with seeds that have a high content of $InsP_6$. Therefore, this molecule is also considered as an antinutritional compound, because it decreases the bioavailability of such elements, causing at the end mineral deficiency.

This issue may not be considered as important in developed countries, where there is access to commodities and a variety of food that allows people to have a diverse diet, but in developing countries where many phytic acid containing cereals and vegetables are considered a staple food, it might have an effect on mineral nutrition. In the case of human and other monogastric organisms, when seeds are consumed, $InsP_6$ stored in the globoids is released into the gut, and because of the small amount of myo-inositol (1,2,3,4,5,6) hexa*kis*phosphate phosphohydrolases (phytases) produced by intestinal bacteria, it is not properly digested, being excreted in the manure. This is cause of phosphorus contamination of soil and main water sources, causing erosion, eutrophication, and other environmental problems, besides the waste of phosphorus utilized in soil fertilization (Josefsen et al., 2007; Kim et al., 2011; Raboy, 2002; Yamaji et al., 2016).

Among other negative effects that have been reported of foods containing $InsP_6$ it also affects protein digestibility by forming strong complexes with proteins, affecting their structure or solubility, or disrupting enzymatic activity, and it also affects carbohydrate and lipid utilization, by forming complexes with carbohydrates, reducing their solubility and impeding the proper digestion and absorption of glucose, or by forming lipophytin complexes. In poultry, these complexes with lipids and its derivatives may form metallic soaps in the gut lumen, restraining energy utilization derived from lipid sources (Kumar et al., 2010).

2.1.1.4 Low phytic acid (*lpa*) mutants

Due to the environmental and nutritional implications of $InsP_6$ consumption, major efforts have been performed by the scientific community to develop crops with a lower $InsP_6$ content without affecting their performance. Unfortunately, some negative effects have been observed regarding

agronomic traits in *lpa* mutants like dwarf vegetative growth, reduced seed development and weight and low germination rates (Raboy et al., 2001).

In the beginning, research was focused on natural variation in the $InsP_6$ content between lines of a crop species, leading to the identification and isolation of mutants in maize (*Zea mays* L.), rice (*O. sativa*), and barley (*Hordeum vulgare* L.). These non-lethal mutants presented altered P and Ins phosphate phenotypes in seeds, but did not present major alterations in total P content (Raboy et al., 2000).

The first mutations of this type identified in maize were called *low phytic acid 1-1* (*lpa1-1*) and *low phytic acid 2-1* (*lpa2-1*). In comparison to wild-type seeds, homozygous *lpa1-1* (first mutant allele of *lpa 1* locus) resulted in a 66% reduction of $InsP_6$, and even though it represented a decrease of $InsP_6$ content, Pi content was increased, resulting in the maintenance of total P content. On the other hand, homozygosity for *lpa2-1* resulted in a reduction of 50% of $InsP_6$, but this decrease was accompanied in part by an increase of Pi content and an increase of InsP with fewer P esters than $InsP_6$ (e.g., $InsP_3$, $InsP_4$, $InsP_5$). Similar mutations were isolated in barley, were *lpa* mutants showed a reduction of 50% of $InsP_6$ in seeds, with constant total P content. In rice, only one mutation was identified, corresponding to *lpa1-1*, showing a reduction of total P content and an increase of Pi again. Here, one non-lethal *lpa1-1* mutation was isolated and displayed a similar phenotype as in maize and barley (Bregitzer et al., 2006; Raboy et al., 2000).

Although the isolation and characterization of these mutations elucidated some of the unknown steps of $InsP_6$ synthesis and accumulation in seeds, there are still some important issues to improve, some of the mutations isolated showed to be lethal to the plant as homozygote or in the case to be viable, the plant showed impairment of seed germination, major losses in yield or reduced stress tolerance (Raboy et al., 2001). On the other hand, mapping some of these *lpa* mutations in the early steps in inositol metabolism may explain the pleiotropic phenotypes and may indicate that targeting later steps in the $InsP_6$ synthesis may be prudent (Raboy, 2009).

Other efforts farther than the *lpa* mutants were made with the aim of lower $InsP_6$ content in seeds. In *A. thaliana* specifically, the knockout of diverse genes was reported from the upper part of the metabolic pathway, the center, and the end as well, obtaining diverse results. Phenotypes obtained by insertional mutagenesis of *MIPS* and *IMP* genes (early pathway) did not have any effects on $InsP_6$ content on seeds, while mutations on *MIK* showed reductions of 62 to 66%. The *IPK2* gene (central pathway) has been described as a family in *A. thaliana*, formed by *AtIPK2α* and *AtIPK2β,* where insertional mutagenesis in the 5' and 3' UTR regions of *AtIPK2α* did not show any phenotype, while mutations on the first exon of *AtIPK2β* resulted in a decrease of 43 to 48%. The *ITPK* gene comprises several subfamilies, and mutations in *AtITPK1* (orthologue of maize *lpa2*), resulted in a 46% decrease of $InsP_6$, other mutations in other gene members of this family such as *AtITPK4* resulted in a 51% reduction. Finally, a major decrease of $InsP_6$ was reported when *MRP5* gene (late pathway), coding for the transporter of $InsP_6$ into the globoids, was mutated on its first and ninth exon, resulting in a reduction of a 73 and 80% respectively (Kim et al., 2011).

Other observations include a comparison of the synthesis of $InsP_6$ via the lipid-dependent or lipid-independent pathway, where the authors conclude that the lipid-dependent pathway has a more relevant regulatory role in mediating stress response, pointing out that mutations that affect inositol phosphate kinases activity are more relevant for abiotic stress sensitivity than mutations in genes of the early, lipid-independent pathway or the phytic acid transporter (Sparvoli et al., 2015).

2.2 CRISPR/Cas9 technology

Genome editing has enabled targeted genome modifications in any organism so far reported and is revolutionizing the field of genetics. The mechanism to repair DNA double-strand breaks (DSBs) has been known for more than 20 years and involves two different pathways, homologous recombination (HR) or non-homologous end joining (NHEJ). In somatic plant cells it was described that the NHEJ mechanism surpasses HR, and very often results in random mutations such as sequence insertions, deletions, duplications, translocations, or replacements, and it is also possible to insert or excise a specific piece of DNA from genomic sequences (Puchta, 2017; Zhang et al., 2017).

One important prerequisite to apply these mechanisms of repair as a technology for the genomic edition is the ability to produce DNA DSBs at specific genomic sites. To accomplish this, synthetic nucleases like zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) have been engineered, both widely used but presenting some restrictions; the construction of these enzymes is time-consuming, expensive and their specificity is limited.

Therefore, the genome editing field was revolutionized with the characterization of the CRISPR/Cas9 system (Figure 2), where Cas9 endonuclease from *Streptococcus pyogenes* acts as an easily reprogrammable nuclease from the Clustered Regulatory Interspaced Short Palindromic Repeats (CRISPR), sequences that play a key role in the bacterial defence system (Jinek et al., 2012; Puchta, 2017; Yin et al., 2017).

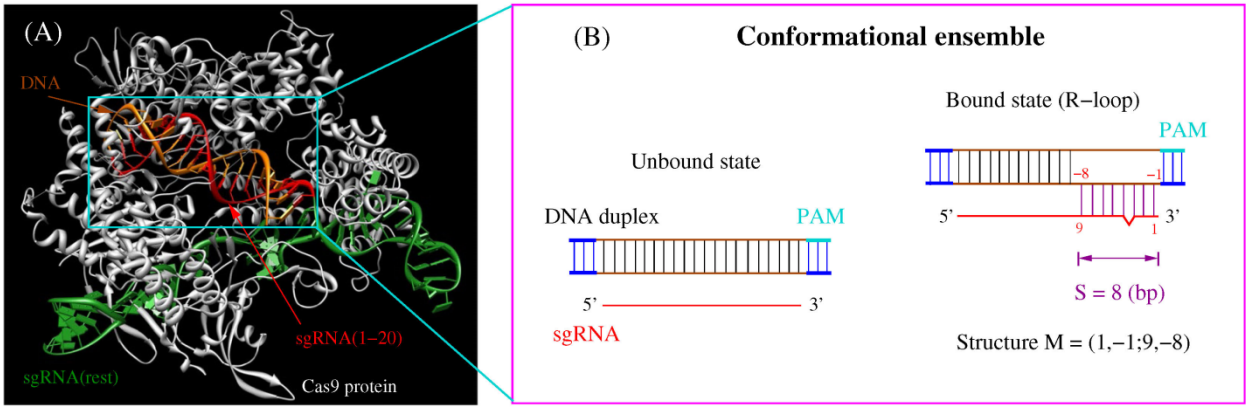


Figure 2: Cas9/sgRNA binding and cleavage model. **A)** The crystal structure of SpCas9 in complex with guide RNA and target DNA. The sgRNA is tightly bound with the Cas9 protein. The first 20-nt of sgRNA (red) are base paired with DNA target (orange), and the rest of the sgRNA nucleotides (green) have rich interactions with the Cas9 protein. **B)** The competition between base stacking within the DNA and DNA-RNA hybrid base pairing/stacking results in the different DNA/RNA bound structures. In the unbound state, the target site maintains its original DNA-DNA base pairing within the chromatin. In the bound state, the sgRNA invades into the DNA duplex and forms the R-loop structure with the target DNA. The three-base-pair DNA helix stretches on both ends of the R-loop are shown in blue. S denotes the length (number of base pairs) of the hybrid helix (Xu et al., 2017).

The Cas9 protein is an endonuclease that cleaves the double-helix of DNA by two nuclease domains, the HNH motif, and RuvC-like domain, each one cleaving one of the DNA strands. The specificity of this system is directed by a CRISPR RNA (crRNA) that binds to a 20 nucleotide (nt) sequence on the protospacer (target DNA), and that is directly adjacent to the 5' of a three nt element, called protospacer-adjacent motif (PAM). The PAM sequence is located downstream to the protospacer, is composed by the NGG sequence, and its presence is mandatory for Cas9

recognition, binding, and nuclease activity, meaning that this system can potentially target any 23 nt region ending in GG. The *trans*-activating CRISPR RNA (tracrRNA) interacts with crRNA facilitating the recruitment of the Cas9 protein and consequently, the cleavage of the target DNA. The fusion of these two RNA elements leads to the formation of a chimeric single-guide RNA (sgRNA), which can recruit and guide the Cas9 protein to the DNA target sequence without losing cleavage efficiency (Fauser et al., 2014).

Targeted genome editing by the CRISPR/Cas9 system has been succesfully reported in different plant species and crops, including members of the *Brassicaceae* family such as; *Arabidopsis*, rapeseed, and wild cabbage (Braatz et al., 2017; Yang et al., 2017; Yang et al., 2017a).

3 Aims and Questions

The objective of this study is to knock-out the phytic acid synthesis gene *BnIPK2β* in rapeseed by CRISPR/Cas9.

3.1 Aims

- Identification of a *BnIPK2β* target site for CRISPR/Cas9 *in silico*.
- Cloning the target sequence into pChimera and pCas9.
- *Agrobacterium*-mediated transformation into rapeseed cultivars Haydn and Express 617.
- Selection of *BnIPK2*β knock-out mutants.

3.2 Questions

- Is there a common target region in *BnIPK2*β for a simultaneous knock-out of all paralogs?
- Is there a similar regeneration and transformation rate in Haydn and Express 617?
- Is there a prediction of any off-target effects?
- How is the CRISPR/Cas9 system edition efficiency for multiple target sites?

4 Materials and Methods

4.1 *In silico* analysis

4.1.1 Retrieve *AtIPK2β* homologs of *B. napus*

A. thaliana IPK2β genomic sequence, accession number (AT5G61760.1), obtained from Stevenson-Paulik (2002), was retrieved from The Arabidopsis Information Resource (TAIR) database, and used to run a BLAT alignment (Kent, 2002) against the genomic sequence from *B. napus* (Darmor-*bzh*, version 4.1) (Chalhoub et al., 2014).

4.1.2 Functional domain prediction

Functional domains of protein sequences were analyzed by PFAM (Finn et al., 2016) setting hidden Markov models with an E-value of 1.0, and molecular protein weight was computed by ExPASy ProtScale online tool (Gasteiger et al., 2005).

Protein domain annotation and multiple sequence alignment (MSA) were performed with CLC Main Workbench 7.8.1 (QIAGEN Aarhus, Denmark), with the following parameters; gap open

cost: 10, gap extension cost: 1, end gap cost: as any other and very accurate alignment (slow). Others features from the sequences were annotated according to the literature.

4.1.3 CRISPR/Cas9 target site design

Putative CRISPR/Cas9 target sites for the *BnIPK2α* and *BnIPK2β* paralogs were screened utilizing the CRISPR-P 2.0 online tool (Liu et al., 2017). 20 bp target regions adjacent to a PAM (5'-NGG) sequence, which were located on the functional domain and were common between *BnIPK2α* and *BnIPK2β* paralogs, were picked to perform off-target analysis by blastn (Johnson et al., 2008). Based on literature (Cong et al., 2013; Paul et al., 2016), a threshold of 10 bp after PAM sequence was set to be fully conserved between paralogs.

4.2 Plant material, bacteria strains, and cloning vectors

The *B. napus* winter type genotype Express 617 (seed code 131612), and Haydn spring type genotype (seed codes 090060 and 161110) were chosen to perform this study.

Competent *Escherichia coli* XL1/Blue were utilized for cloning of the vectors and *Agrobacterium tumefaciens* GV3101::pMP90 to perform rapeseed hypocotyl transformation.

Binary vector system, pChimera (Figure 7.A) and pCas9-TPC (Figure 7.B) vectors were used in this study as described by Fauser et al., 2014, with some modifications to subclone two sgRNA into one pCas9-TPC vector (Figure 7.C and 7.D).

4.3 CRISPR/Cas9 target site genotype confirmation and construct development.

Primers were designed and available primers (Supplementary Table S. 1, target region genotyping) designed by Niharika Sashidhar were picked to amplify the target region by PCR and perform Sanger sequencing at the Institute for Clinical Molecular Biology (IKMB, Christian-Albrechts-Universität, Kiel). The wild-type (wt) DNA template from Express 617 and Haydn was provided by Nirosha Lakmali. The resulting sequences were assembled to the reference (Darmor-*bzh*, version 4.1) with CLC Main Workbench 7.8.1, setting the following parameters; minimum aligned read length: 50, alignment stringency: high.

For construct development, the binary vector system, pChimera (Figure 7.A) and pCas9-TPC (Figure 7.B) vectors were used in this study as described by Fauser et al., 2014, with modifications to subclone two sgRNA into one pCas9-TPC vector (Figure 7.C and 7.D).

4.3.1.1 Oligo annealing, ligation into pChimera and bacterial transformation

The pChimera empty vector was isolated from bacterial stock grown on LB/Ampicillin (100 mg l^{-1}) overnight, using Macherey-Nagel, NucleoSpin® Plasmid Columns, Collection Tubes (2 mL), buffers, RNase A, according to manufacturer instructions, and quality of the isolation was checked by electrophoresis on an agarose gel, running at 60V, during 60 minutes in 1x TA buffer. The vector was digested (Supplementary Table S. 2) by BbsI (Thermo Scientific), by incubating one hour at 37°C in the water bath, and then it was purified utilizing PCR clean-up & gel extraction kit (Macherey-Nagel), eluting samples in 15 µl of NE buffer. After that, the concentration was determined by NanoDrop photometer and adjusted to 5 ng/µl.

The two sgRNAs (designed on 4.1.3 section) containing 5'-ATTG on the forward oligo and 5'-AAAC on the reverse oligo overhangs, were annealed (Supplementary Table S. 3) by heating to 95°C during 5 minutes and then gradually cooled down to room temperature to allow oligo hybridization. The annealed product was ligated (Supplementary Table S. 4) separately within BbsI restriction sites of pChimera, under the U6-26 promoter from *A. thaliana*, by incubating it at room temperature for 1 hour. The products obtained where named pChimera:TS_1 and pChimera:TS_2 (Figure 7.C), and used to transform competent *E. coli* XL1/Blue (Supplementary Table S. 5) by adding 2 µl of ligation product to a competent cell vial thawed and incubated on ice for 5 minutes. Then, a heat shock was performed at 42°C for 45 seconds and incubated again on ice for 5 minutes. After that, 500 µl of fresh LB medium was added, and tubes were incubated on a shaker at 110 rpm and 37°C for 1 hour to allow the recovery of the bacteria. 200 µl of recovered bacteria were plated on LB/Ampicillin (100 mg l^{-1}) plates and incubated at 37°C overnight. Screening of positive colonies was performed by colony PCR (Supplementary Table S. 1, pChimera colony PCR, and Supplementary Table S. 6, colony PCR setup), 13 colonies were selected, and products were sent for Sanger sequencing at IKMB for confirmation, and subsequently assembled to the reference genome as described above.

4.3.1.2 sgRNA restriction

The positive colonies were grown overnight on LB/Ampicillin (100 mg l^{-1}) medium and plasmids were isolated performing the same procedure as above.

The sgRNAs were restricted (Supplementary Table S. 7) by XmaJI (Thermo Scientific) (AvrII isoschizomer), incubating samples at 37°C during 2 hours, and the reaction was heat inactivated by incubating at 80°C for 20 minutes. The restriction products were run on a 2% NuSieve GTG agarose gel, at 60V, for 40 minutes in 1x TA buffer, cut under 70% UV light, and cleaned utilizing PCR clean-up & gel extraction kit (Macherey-Nagel), according to manufacturer instructions.

4.3.1.3 sgRNA:TS_1 and sgRNA:TS_2 subcloning into pCas9-TPC

The pCas9-TPC empty vector was isolated from bacterial stock grown on LB/Spectinomycin (100 mg l^{-1}) overnight at 37°C, using Macherey-Nagel, NucleoSpin® Plasmid Columns, Collection Tubes (2 mL), buffers, RNase A, eluting samples two times in 15 µl of NE buffer at 70°C, and digested (Supplementary Table S. 8) with XmaJI (Thermo Scientific) by incubating 2 hours at 37°C. Then, it was dephosphorylated (Supplementary Table S. 9) to prevent recircularization of the vector, with FastAP thermosensitive alkaline phosphatase (Thermo Scientific), by incubating during 2 hours at 37°C; the reaction was heat inactivated by incubating at 75°C for 5 minutes. The dephosphorylated product was directly cleaned utilizing PCR clean-up & gel extraction kit (Macherey-Nagel), according to manufacturer instructions.

The sgRNAs were sub-cloned into pCas9-TPC in two separated reactions, first, sgRNA:TS_1 was ligated into AvrII restriction site in a molar ratio 3:1 (insert:vector), the reaction (Supplementary Table S. 10) was mediated by T4 DNA ligase (Thermo Scientific), and incubated first, during 2 hours at room temperature, and then, overnight at 4°C to increase efficiency of ligation. The enzyme was heat inactivated by incubating at 65°C during 10 minutes, and 2 µl of the product (pCas9-TPC:TS_1) were used to transform competent *E. coli* XL1/Blue, following the procedure as described above. Then, 100 and 400 µl of the reaction were plated on LB/Spectinomycin (100 mg l^{-1}) plates, and incubated overnight at 37°C. 13 colonies were tested by PCR, as described above, and one positive result was sent for Sanger sequencing at IKMB using Cas9 primers

(Supplementary Table S. 1, pCas9-TPC colony PCR single target) and assembled to the reference genome as mentioned above.

Positive colonies were grown on LB/ Spectinomycin (100 mg l^{-1}), and the plasmid was isolated in the same way than empty pCas9-TPC before. Then it was digested (Supplementary Table S. 11) with BcuI (Thermo Scientific) by incubating 2 hours at 37°C, dephosphorylated and cleaned, following the procedure described above. Afterwards, sgRNA:TS_2 was ligated (Supplementary Table S. 12) with BcuI (produces same overhangs as AvrII but recognizes another sequence) restriction site, in the same way sgRNA:TS_1, obtaining pCas9-TPC:TS_1-TS_2 (Figure 7.D).
The ligation reaction product was utilized to transform competent *E. coli* XL1/Blue, following the same procedure mentioned above, plating 100 µl and 400 µl of the reaction on LB/Spectinomycin (100 mg l^{-1}) plates, and incubated overnight at 37°C. After that, 34 colonies were tested, and the vector was isolated as described above, sent for Sanger sequencing at IKMB using Cas9 primers (Supplementary Table S. 1, pCas9-TPC colony PCR multiplex target), and assembled to the reference genome in the same way as before.

4.3.1.4 pCas9-TPC:TS_1-TS_2 transformation into *Agrobacterium tumefaciens*

Competent *A. tumefaciens* GV3101::pMP90, were prepared and transformed following the protocol described by Hofgen & Willmitzer, 1988. Then, 200 µl of the reaction were plated on YEB/Rifampicin (25 mg l^{-1})/Gentamycin (50 mg l^{-1})/Spectinomycin (100 mg l^{-1}) plates and incubated at 28°C during 48 hours. Positive colonies were identified by PCR as described above.

Positive colonies of *A. tumefaciens* prepared in the previous step, were grown in YEB medium (Supplementary Table S. 5) supplemented with rifampicin (25 mg l^{-1}) and gentamicin (50 mg l^{-1}) to select bacteria and helper plasmid respectively, and spectinomycin (100 mg l^{-1}) to select pCas9-TPC vector as well, and used for rapeseed hypocotyl transformation.

4.4 Plant transformation

Rapeseed hypocotyl transformation was performed following the protocol described by Zarhloul et al., 2006 with the following modifications; the seeds were surface sterilized in 70% ethanol for two minutes, then in 3% NaOCl supplemented with 0.01% of Tween 20 during 10 minutes. The seeds were grown during five days in dark and co-cultivation of hypocotyls was performed in a solid medium during 48 hours. The subculture of explants was performed every four weeks. 400 mg l^{-1} of carbenicillin and 50 mg l^{-1} of cefotaxime were used to deplete *A. tumefaciens*, and selection of transgenic plants was made by adding 5 mg^{-1} of phosphinothricin (PPT) to the medium, because pCas9-TPC vector contains a bialaphos resistance (*BAR*) cassette, which has been widely used as a positive selection marker, granting resistance to the herbicide on transgenic plants.

4.5 Transgene and gene edition detection

4.5.1 DNA isolation

DNA was isolated from tissue culture leaves by a CTAB method; samples were taken during subculture, placed in a sterile 2 mL tube and freeze-dried for 48 hours. Two metallic beads (4 mm) were placed in the tube to grind the tissue, and then placed on a Retsch MM2 mill during 1 minute at maximum speed. Then 1 mL of 1.2x CTAB buffer (supplemented with 3 µl of β-mercaptoethanol before using) (Supplementary Table S. 13) prewarmed to 65°C was added and the mix incubated

for 30 minutes at 65°C in the water-bath. After incubation, samples were cooled down during 5 minutes on ice, 500 µl of chloroform/isoamyl alcohol (24:1) was added and incubated for 10 minutes in the shaker.

The samples were centrifuged for 15 minutes at 15°C and 14,000 RPM; the supernatant was transferred to a new sterile 2 mL tube, and 700 µl of isopropanol (100%) was added. The tubes were mixed by careful inversion and incubated overnight at -20°C to precipitate the DNA.

After precipitation, samples were centrifuged for 10 minutes at 4°C and 13,000 RPM. The supernatant was discarded, and 500 µl of Wash Buffer I (Supplementary Table S. 13) was added to the tube without disrupting the pellet, incubated during 5 minutes and then the buffer was discarded and the procedure was repeated with 500 µl of Wash Buffer II (Supplementary Table S. 13). In the end, the pellet was dried at room temperature and resuspended in 50 to 100 µl of sterile ddH_2O, supplemented with 1 µl of RNase.

Finally, the DNA quality was checked by electrophoresis on a 1% agarose gel, following the same procedure as above.

4.5.2 Transgene detection and edition of the target region

To determine the presence or absence of the T-DNA in the genome of the rapeseed regenerating shoots, a PCR was performed using primers (Supplementary Table S. 1, pCas9-TPC colony PCR single or multiplex target and spectinomycin detection) to detect Cas9 and spectinomycin genes. This primer combination allowed to discriminate between transgenic plants and the false positives due to *A. tumefaciens* presence.

The target region was amplified by PCR (Supplementary Table S. 6) in samples showing positive results for Cas9 and negative for spectinomycin and sent for Sanger sequencing at IKMB using specific primers (Supplementary Table S. 1, target region genotyping) to sequence the target region from each of the four paralogs. Sequencing results were assembled to the reference genome in the same way as before.

5 Results

5.1 *In silico* analysis

5.1.1 AtIPK2β homologs of *B. napus*

The nucleotide sequences of two paralogs from the *BnIPK2α* and four paralogs from the *BnIPK2β* family were retrieved from The French National Sequencing Center[1], Genoscope (Chalhoub et al., 2014), annotated as BnaC09g48570D, BnaA10g23820D, and BnaA06g31750D, BnaA09g05790D, BnaC09g05390D and BnaCnng33290D, respectively. The Table 1 and Supplementary Table S. 15 show detailed information about the gene structure, protein length, identities and gene nomenclature used in this study. For the ease of understanding, the nomenclature of gene names were changed, assigning the first two letters, corresponding to genus and species respectively, then a letter and number corresponding to the subgenome (*B. oleracea*, subgenome C_oC_o, and *B. rapa*, subgenome A_rA_r) and chromosome number, followed by an underscore symbol to separate the gene name (e.g., for paralog BnaA09g05790D the nomenclature will be *BnA09_IPK2β*).

Gene family	Gene product	TAIR/ Genoscope Id	Nomenclature used in this study	mRNA (bp)	DNA (bp)	Protein (aa)	Molecular Weight (kDa)	Exon/intron No.	
								exons	introns
IPK2α	Inositol (1,4,5) P3_3/6 kinase	At5g07370	*IPK2α*	1,168	1,897	286	31.945	1	0
		BnaC09g48570D	*BnC09_ IPK2α*	852	1,848	284	31.727	1	0
		BnaA10g23820D	*BnA10_ IPK2α*	852	1,774	284	31.755	1	0
IPK2β	Inositol (1,4,5) P3_3/6 kinase	At5g61760	*IPK2β*	1,201	1,715	300	33.486	2	1
		BnaA09g05790D	*BnA09_ IPK2β*	861	2,222	287	31.985	1	0
		BnaC09g05390D	*BnC09_ IPK2β*	858	1,322	286	31.961	1	0
		BnaA06g31750D	*BnA06_ IPK2β*	846	1,049	282	31.513	1	0
		BnaCnng33290D	*BnCnn_ IPK2β*	846	1,046	282	27.055	1	0

Table 1: Genes analyzed in this study.

As is shown in Table 1, all the *IPK2α* paralogs have only one exon, while *BnIPK2β* paralogs also present only one exon, in contrast with *AtIPK2β* which possesses two exons and one intron. All the *B. napus* paralogs presented a similar mRNA length when compared with *A.*

[1] http://www.genoscope.cns.fr/brassicanapus/

thaliana, and the correspondent protein molecular weight is also similar between *B. napus* and *A. thaliana* orthologues, except for *AtIPK2β* and *BnCnn_IPK2β.*

5.1.2 Functional domain prediction and target site design

The prediction of the functional domains was performed by PFAM, and an Inositol Polyphosphate Multikinase domain was predicted for the two BnIPK2α and four BnIPK2β paralogs. A multiple sequence alignment (MSA) was constructed for the ease of analysis, including IPK2α and IPK2β amino acid sequences from *B. napus, A. thaliana, B. rapa* and *B. oleracea,* common putative target regions for CRISPR/Cas9 were identified (Figure 3, residues underlined with a black bar), and functional domains were annotated.

Other features from the sequences were annotated according to the literature. The critical phosphorylation sites for the AtIPK2β function were described by Wang et al., (2017), pointing out that Tyr46, Thr128, and Ser147 (Figure 3, residues highlighted in red) are critical amino acids. On the MSA only Tyr46 and Thr128 are fully conserved among IPK2α and IPK2β orthologues, whereas Ser147 is conserved in IPK2β of all species and in *A. thaliana* IPK2α, but showing an amino acid substitution (S147L) in Br_IPK2α, Bo_IPK2α, BnC09_IPK2α, and BnA10_IPK2α.

The IP_3 binding domain (PxxxDxKxG) (Figure 3, residues highlighted in dark blue) has been well characterized and is conserved among kinases and inositol polyphosphate kinases, being a trademark of this protein family, and is present on all sequences analyzed (Cheek et al., 2005; Seeds et al., 2007; Shears, 2004; Stevenson-Paulik et al., 2002; Hui-Jun Xia et al., 2005; Zhan et al., 2015)

Other important residues that have been described as critical for substrate specificity are Lys102, Lys117, and Lys121 (Figure 3, residues highlighted in light blue). Where Lys102 (from PxxxDxKxG domain) was described as critical amino acid in IPK2β from human, nematode (*Caenorhabditis elegans*), Arabidopsis and yeast (*Saccharomyces cerevisiae*) (Hui-Jun Xia et al., 2003), and Lys117 and Lys121 were described as critical amino acids at the catalytic region of IPK2α from *A. thaliana* (Endo-Streeter et al., 2012). The MSA shows that this catalytic region described on IPK2α is also conserved among orthologues, except IPK2β from *B. oleracea, B. rapa* and paralogs A06 and Cnn from *B. napus* which present in all cases a K117Q and K121N substitution.

For AtIPK2β, Arg135 and Asp252 (in red) were described as critical amino acids for ATP/Mg^{+2} binding site (Hui-Jun Xia et al., 2003), while the same author reviewed that the zone comprised between Leu250 and His255 (in dark green) are the critical ones for ATP/Mg^{+2} binding (Hui -Jun Xia et al., 2005). In the case of the sequences analyzed, the amino acids described as critical are fully conserved among orthologues, except residue 254, which presents a substitution from alanine to threonine in IPK2β from *B. oleracea, B. rapa* and in BnA06_IPK2β and BnCnn_IPK2β.

No nuclear localization signal (NLS) or calmodulin (CaM) binding domain were predicted for any of the sequences analyzed, even though they have been described in yeast and mammalian IPK2 orthologues.

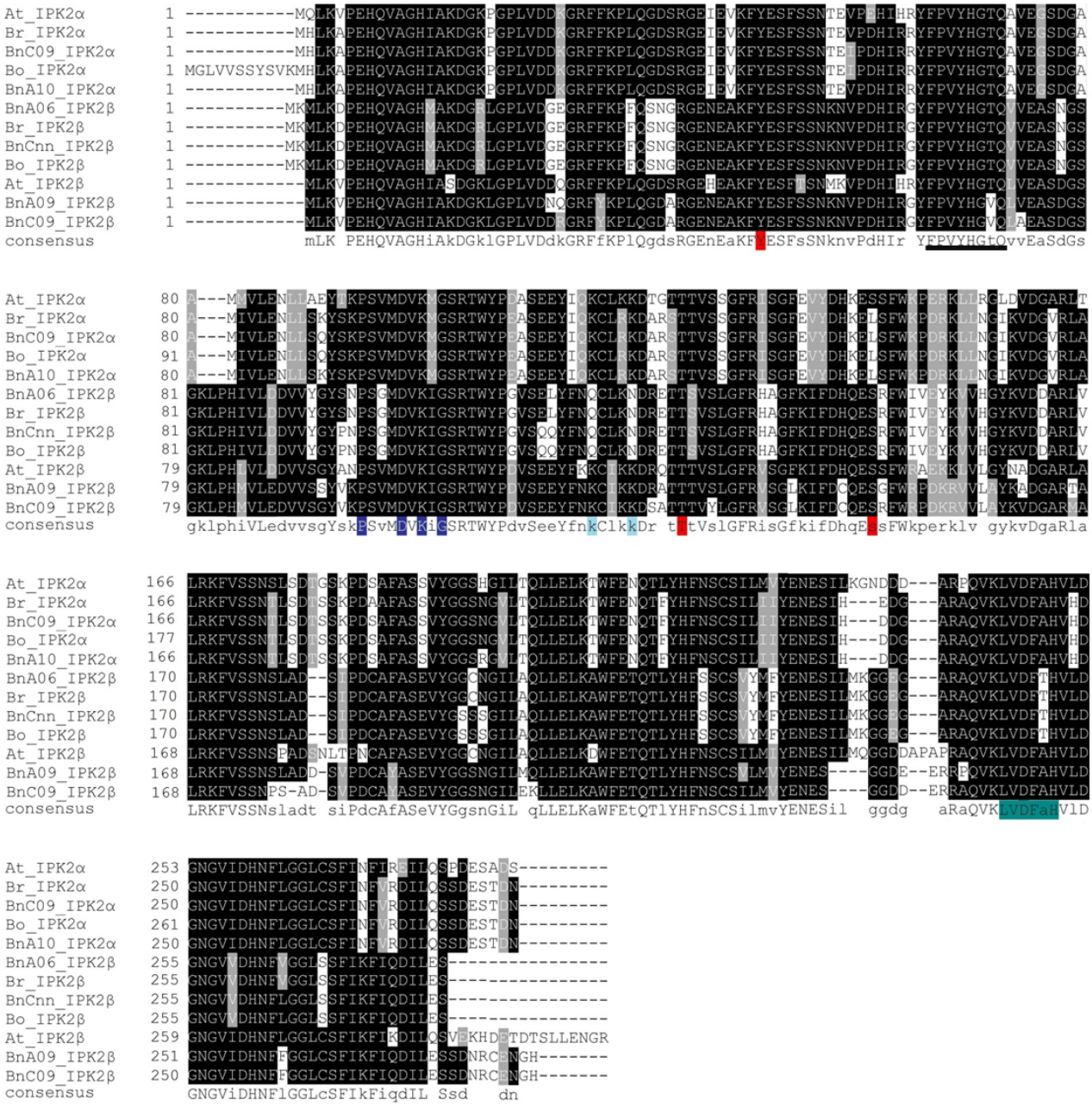


Figure 3: Multiple protein sequence alignment of IPK2α and IPK2β from *A. thaliana, B. napus, B. oleracea,* and *B. rapa.* Residues highlighted in dark blue corresponds to IP3 binding domain (PxxxDxKxG), residues underlined by a black bar correspond to CRISPR/Cas9 target site, while residues highlighted in red are phosphorylation sites by CPK4, residues highlighted in dark green are critical for ATP/Mg^{+2} binding for β paralogs, and residues highlighted in light blue are critical for substrate specificity in AtIPK2α.

Common putative target sites for *BnIPK2α* and *BnIPK2β* were searched to target the six paralogs with a single construct, but no conserved region was big enough to reach the threshold of 10 bp after PAM sequence with 100% conservation, therefore common regions were searched separately for α and β paralogs.

In the case of *BnIPK2α* paralogs, all the sequences located next to a PAM site (-NGG) were selected as putative target sites, off-target analysis was performed, and it was not possible to identify a

common region without having an off-target effect on important genes (like MYB transcription factor or elongation factor-1 α).

In the case of *BnIPK2β*, after filtering 72 putative CRISPR/Cas9 target sites with off-target prediction, there were no sites without off-target effect on important genes for survival of the plant, but one common region (Figure 3, residues underlined by a black bar). It was identified as highly conserved and was selected to design sgRNA (Table 2), named target site 1 (TS_1); targeting *BnA06_IPK2β* and *BnCnn_IPK2β* paralogs, and target site 2 (TS_2); targeting *BnA09_IPK2β* and *BnC09_IPK2β* respectively.

In the off-target effects analysis (Supplementary Table S. 14), made by blastn, according to the rule of 100% nucleotide conservation on the 10 bp upstream region of the PAM sequence, the sgRNAs designed did not show any putative off-target effect in other genes. In the case of the target site 1, there are mismatches in the PAM site. Therefore, no edition of off-target genes is expected, and in the case of the target site 2, the mismatches are located after the 10th position.

Name	Target sequence	Target paralog
TS_1	CGTGTATCACGGCACTCAAG	*BnA06_IPK2β* *BnCnn_IPK2β*
TS_2	TGTGTACCACGGCGTTCAGC	*BnA09_IPK2β* *BnC09_IPK2β*

Table 2: CRISPR/Cas9 target sites for genome edition of *BnIPK2β* paralogs. Where TS_1 is the target site aiming to knock-out *BnA06_IPK2β* and *BnCnn_IPK2β* paralogs and, TS_2 is the target site aiming to knock-out *BnA09_IPK2β* and *BnC09_IPK2β*

The sgRNAs designed are aiming at the complete knock-out (via frameshift) of the four *BnIPK2β* paralogs, they are targeting the paralogs in the same position (Figure 4), and the *in silico* analysis predicts that the genetic modification will be done in a region located upstream the PxxxDxKxG IP_3 binding domain (Figure 4, blue bars), and the critical amino acids for ATP/Mg^{+2} binding (black bars).

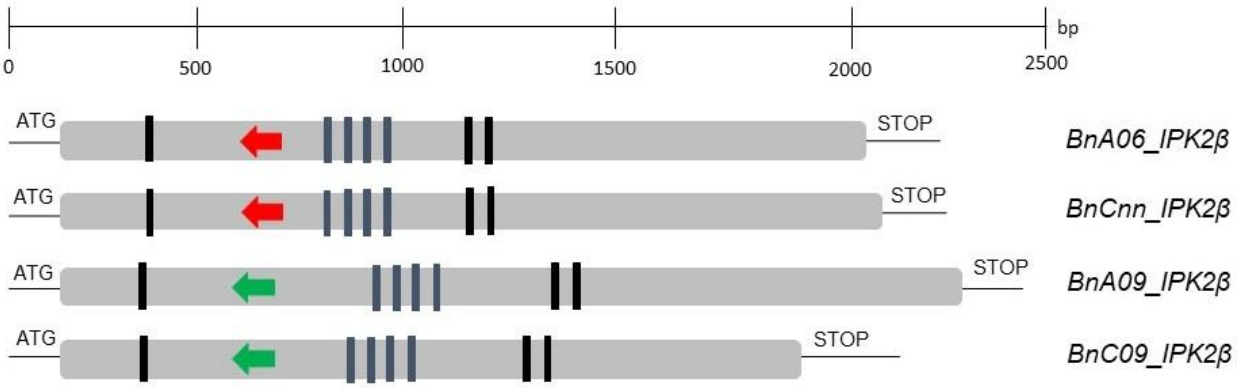


Figure 4: Graphical representation of *BnIPK2β* gene structure, CRISPR/Cas9 target sites, and motifs. Where the grey bar represents the sole exon of the paralogs, the red arrows represent the target site 1, the green arrow the target site 2, the black bars important nucleotides for ATP/Mg^{+2} binding, and the blue bars represent the important nucleotide for IP_3 binding.

5.1.3 CRISPR/Cas9 target site region genotyping

The target site region was amplified by PCR using target region genotyping primers (Supplementary Table S. 1, target region genotyping), and the sequencing output was assembled to the genomic reference sequence (Darmor-*bzh*, version 4.1) to compare structural variations that are present between the different genotypes such as SNPs, insertions or deletions.

Both Haydn and Express 617 showed to be conserved at the target region (Figure 5) with100% conservation at the PAM and target sites.

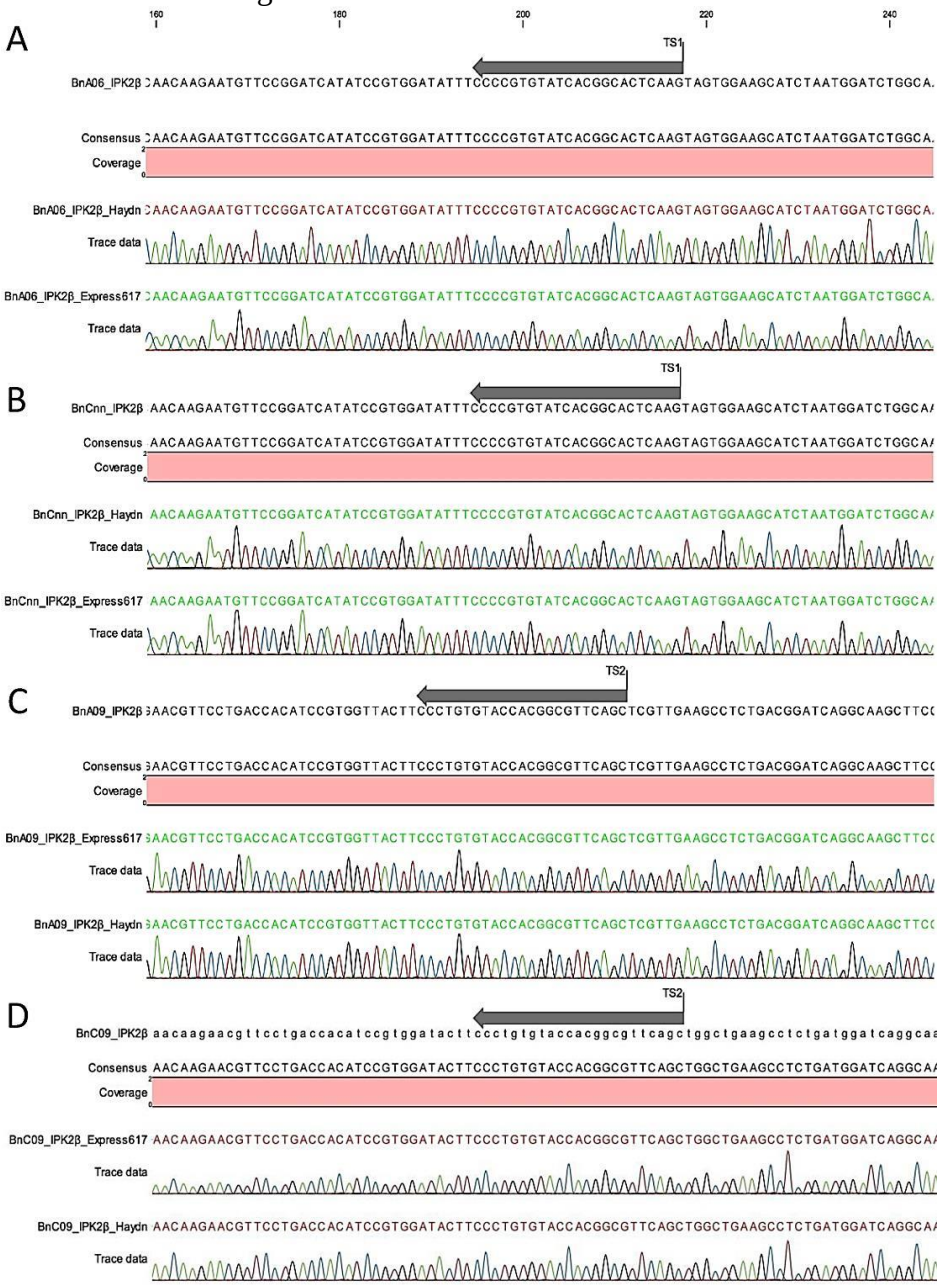


Figure 5: Sequence alignment of genotyping sequencing, where the grey arrow shows the target site position for CRIPSR/Cas9. **A.** *BnA06_IPK2β*. **B.** *BnCnn_IPK2β*. **C.** *BnA09_IPK2β*. **D.** *BnC09_IPK2β*.

Therefore, the sgRNAs previously (section 4.1.3) designed are suitable for the aim of this study.

5.2 Construct development

5.2.1 pChimera assembly

Two sgRNAs, TS_1 and TS_2, (designed on 4.1.3 section) containing 5'-ATTG on the forward oligo and 5'-AAAC on the reverse oligo overhangs, were annealed and ligated separately within BbsI restriction sites of pChimera, yielding pChimera:TS_1 and pChimera:TS_2. The ligation products were used to transform competent *E. coli* XL1/Blue, and screening of positive colonies was performed by colony PCR (Supplementary Table S. 1, pChimera colony PCR), where 13 colonies were picked for each target site (TS_1 and TS_2), resulting in 4 positives for TS_1, and 12 positives for TS_2 (Supplementary Figure S. 1). The 228 bp products indicating the insertion of the sgRNA were sent to sequencing to IKMB, and the resulting output was assembled to the reference sequence of the vector (Supplementary Figure S. 2 and Supplementary Figure S. 3).

Confirmed colonies containing the corresponding sgRNA were grown in LB/Ampicillin (100 mg l^{-1}), pChimera:TS_1 and pChimera:TS_2 were isolated and the corresponding sgRNA:TS_1 and sgRNA:TS_2 were restricted from the vector by XmaJI.

5.2.2 pCas9-TPC assembly

The sgRNA:TS_1 was cloned into AvrII restriction site of pCAS9-TPC yielding pCAS9-TPC:TS_1 and transformed into competent *E. coli* XL1/Blue. Positive insertions were screened by colony PCR (Supplementary Table S. 1, pCas9-TPC colony PCR single target), where 3 of the 13 colonies screened resulted in being positive for the insertion (Supplementary Figure S. 4). One of the products indicating the existence of the vector and the insert (expected size 1,019 bp) was sent to sequencing at IKMB, and the resulting output was assembled to the reference sequence of the vector (Supplementary Figure S. 5), corroborating the correct insertion of the sgRNA to the vector. The colony containing the vector plus the insert was grown in LB/Spectinomycin (100 mg l^{-1}), and the plasmid was isolated. After that, sgRNA:TS_2 was cloned into BcuI restriction site of pCAS9-TPC:TS_1, yielding pCAS9-TPC:TS_1-TS_2, and transformed into competent *E. coli* XL1/Blue. Positive insertions were screened by colony PCR (Figure 6, Supplementary Table S. 1, pCas9-TPC colony PCR multiplex target), were 34 colonies where picked, resulting in 31 positives. The colony indicating the positive insertion of the second sgRNA was grown, the vector isolated and sent to Sanger sequencing for confirmation. The sequencing result was assembled to the reference sequence of the vector (Supplementary Figure S. 6) indicating the correct insertion of both, sgRNA:TS_1 and sgRNA:TS_2 into the vector.

Intriguingly, the PCR product (Figure 6) obtained after cloning both sgRNA:TS_1 and sgRNA:TS_2 into the same vector (obtaining pCas9-TPC:TS_1-TS_2), showed to be unspecific regarding the product size retrieved. In addition to the expected 1,545 bp product indicating the presence of both sgRNA on the vector, another band of ca. 1,000 bp could be observed, a size corresponding to the vector containing only one sgRNA. Surprisingly, one case showed (Figure 6, lane 29) the same "pattern of double band", but with the difference that there is one band at ca. 1,000 bp, a size corresponding to the vector containing only one sgRNA (1,019 bp), and another band of ca. 500 bp corresponding to the empty pCAS9-TPC vector.

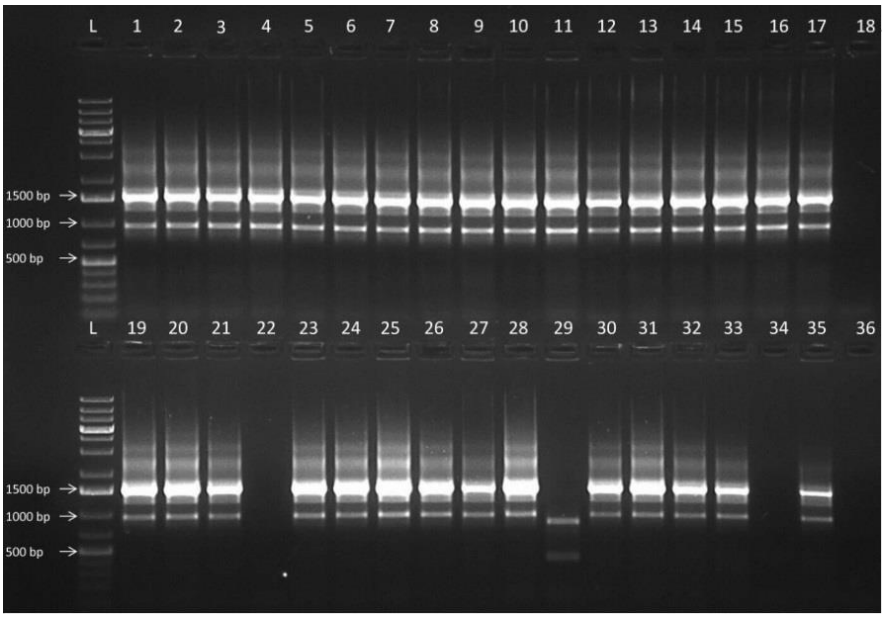


Figure 6: Colony PCR from pCAS9-TPC:TS_1:TS_2. Where L corresponds to Scientific O'GeneRuler 1 kb Plus DNA Ladder, 1 to 17 and 19 to 35 corresponds to colonies samples, 18 and 36 corresponds to the negative control.

Since this pattern of double band might be caused by an intraplasmid rearrangement, the 1,000 bp band was isolated and sent to IKMB for Sanger sequencing (Supplementary Figure S. 7) to determine if one sgRNA was deleted from pCas9-TPC:TS_1-TS_2 during bacterial transformation, and in this case, we could only partially determine that the sgRNA:TS_2 suffered modifications. As is shown in Supplementary Figure S. 7, the product from the forward primer was assembled at the region corresponding to the sgRNA:TS_1, while the product from the reverse primer was assembled in the region corresponding to sgRNA:TS_2. However, there was no overlapping of the sequences. Therefore it was not possible to discriminate if there is a real deletion. Nevertheless the 5' region of the forward and reverse sequences align perfectly to the vector, indicating the presence of the sgRNA corresponding to sgRNA:TS_1 and the AtU6-26 promoter from *A. thaliana* corresponding to sgRNA:TS_2.

5.3 Plant transformation

In total 5,054 rapeseed hypocotyls from Haydn and Express 617 were transformed using *A. tumefaciens* (Table 3) in nine different experiments, with three different vectors (Table 3 and Figure 7.D) from those 1,165 were from the Express 617 genotype and 3,889 were from the Haydn genotype.

The seed germination rates greatly varied for Haydn (seed code 0900060), oscillating between 92 and 100%, while Express 617 always presented a 100% of germination.

Unexpectedly, in both of the experiments (Table 3) performed with Express 617 (transformation experiment number I and VIII), all the hypocotyls died, in the case of transformation experiment number I, most of the explants were developing poorly, turning brown before transferring them to selection medium (O_5), where all died. In the transformation experiment number VIII, 99.74% of the hypocotyls died in O_4 medium, except three explants, which started to regenerate shoots and were transferred to selection medium (O_5). From those, two died, and one was transferred to rooting medium (O_6) and died.

Hypocotyls from Haydn died at a lower rate (between 16.3 and 0.7% in O_4) and responded positively to the regeneration protocol, producing calli and shoots (from 4 weeks until the delivery of this manuscript) that were transferred to selection medium (between 11.5 and 45.6% of shoots transferred to O_5 containing phosphinothricin were surviving).

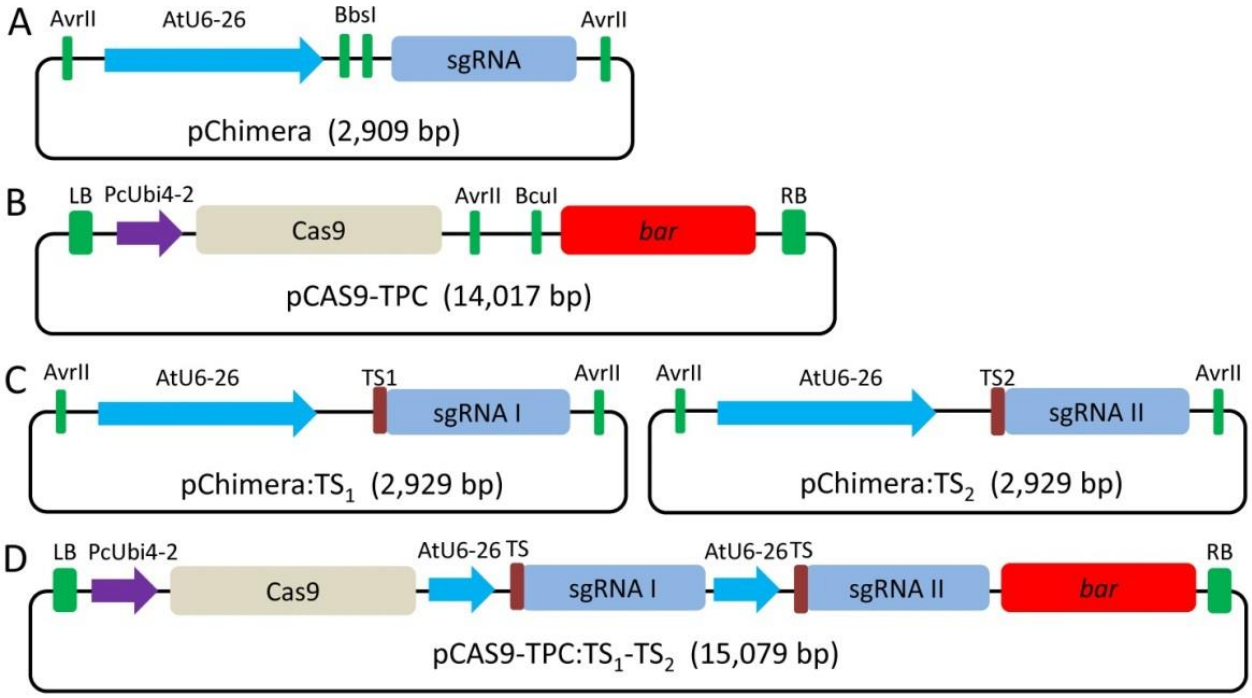


Figure 7: Binary vector system used in this study. **A.** pChimera empty vector. **B.** pCas9-TPC empty vector. **C.** pChimera:TS_1 and pChimera:TS_2 obtained after cloning sgRNAs. **D.** pCas9-TPC:TS_1-TS_2 obtained after subcloning sgRNA:TS_1 and sgRNA:TS_2.

Haydn responded better than Express 617 to the regeneration protocol followed. The transformation experiment number II yielded 348 regenerated explants that were transferred to O_5 medium and from those, 142 shoots were surviving, and one was transferred to O_6 rooting medium. The transformation experiment number III had 669 hypocotyl explants regenerating in O_4 medium, and 384 regenerated shoots that were transferred to the O_5 medium, and from those, 84 shoots were surviving. The transformation experiment number IV had 453 hypocotyl explants regenerating in O_4 medium, and 63 regenerated shoots that were transferred to the O_5 medium, and from those 61 were surviving. The transformation experiment number V had 626 hypocotyl explants regenerating in O_4 medium, and 355 regenerated shoots that were transferred to the O_5 medium, from those, 72 were surviving, and one was transferred to O_6 rooting medium.

Transformation No.	Genotype	Vector	Seeds Germinated	Germination Rate (%)	Hypocotyl Number	Regenerated shoots	Surviving shoots in O_5	Rooting medium
I	Express 617	pCas9-TPC:TS_1:TS_2	65	100	363	363	321	-
II	Haydn	pCas9-TPC:TS_1	100	100	475	348	142	
III	Haydn	pCas9-TPC:TS_2	100	97	669	412	84	
IV	Haydn	pCas9-TPC:TS_1:TS_2	100	92	453	63	61	
V	Haydn	pCas9-TPC:TS_1	104	97	626	355	72	1
VI	Haydn	pCas9-TPC:TS_2	104	100	654	653	3	-
VII	Haydn	pCas9-TPC:TS_1:TS_2	105	90	709	514	0	-
VIII	Express 617	pCas9-TPC:TS_1:TS_2	98	100	802	779	3	-
IX	Haydn	pCas9-TPC:TS_1:TS_2	64	100	303	301	0	-

Table 3: Rapeseed hypocotyl transformation experiments performed in this study.

The transformations number VII and IX were discarded because of unexpected contamination in all the plates. About 3 to 4 weeks after *A. tumefaciens* infection I observed bacteria forming pink colonies (Supplementary Figure S. 12) both, in the explants and the medium. The transformation number VI was contaminated with the same bacteria during subculture under the clean bench, leaving only three shoots in the O_5 medium at this time. An attempt to identify the bacteria was made, and it was cultured in LB medium and observed under the light microscope. Sadly, the growth rates in LB medium were prolonged, and it was only possible to determine that the organism contaminating the culture was some ampicillin resistant bacteria, and it was not possible to assess genus, species or the source of the contamination.

To rescue the contaminated explants, two strategies were tested. The first one consisted of washing the explants in 10% EtOH, dried on a sterile Whatman filter paper, then rinsing in sterile $_{dd}H_2O$, dried again and placed back on sterile O_4 medium. The second strategy followed the same as the first one; however, 2% NaOCl was utilized for washing the explants instead of 10% EtOH. The second strategy proved to be better to restrict the bacterial growth; however, the contamination was not eradicated.

Finally, other observations were made, for example, in the transformation experiments with the vector carrying the sgRNA:TS_1 (transformation number II and V) an important number (near 12%) of the explants developed roots, even though they were in shooting medium. Other observations were made on experiment number IV, where nearly 50% of the explants generated calli before shoots, and from those, 8% generated calli without developing shoots yet.

5.4 Transgene detection and gene editing

Tissue culture leaf samples were taken to screen for transgenes and gene editing by PCR using Cas9 and spectinomycin primers (Figure 8. A and B), and by amplifying the target region (Supplementary Figure S. 13) to find out if the shoots were edited.

As is shown in Figure 8. A, all the samples present a weak 500 bp band in Cas9 PCR, a signal that might be due to a vector contamination, with exception of leaf sample number 6, which presents two bands, from 1,000 bp and 500 bp. This signal could come from bacteria growing on the medium or the explant because it also showed an amplification in the spectinomycin PCR (Figure 8. B), indicating that it might not be transgenic. To clarify where the 500 bp product is coming from, the PCR product was sent to Sanger sequencing, but no data were obtained because of the weak product.

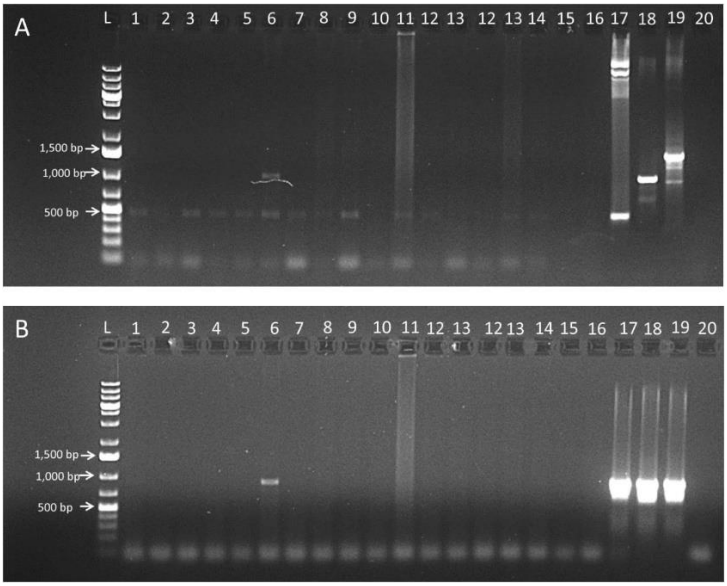


Figure 8: Cas9 and spectinomycin PCR from tissue culture leaf samples. **A)** Cas9 PCR, where L corresponds to 1 kb plus ladder, 1 to 16 to leaf samples, 17 to pCas9-TPC, 18 to pCas9-TPC:TS_1, 19 to pCas9-TPC:TS_1:TS_2, and 20 to the negative control. **B)** Spectinomycin PCR, where L corresponds to 1 kb plus ladder, 1 to 16 to leaf samples, 17 to pCas9-TPC, 18 to pCas9-TPC:TS_1, 19 to pCas9-TPC:TS_1:TS_2, and 20 to the negative control.

The sequences of the PCR products of the target region were assembled to the reference sequence for each paralog, and no editing in any of the paralogs could be found in the 19 plants screened.

6 Discussion

Several approaches have been used to modify seed phytate and total seed phosphorous, including targeting genes for Ins and $InsP_1$ synthesis, the conversion of $InsP_1$ to $InsP_6$, or aiming to disrupt intracellular compartmentation, transport, and storage. Some studies concluded that targeting genes from the early pathway (Ins and $InsP_1$ synthesis), may result in pleiotropic phenotypes, since the metabolites involved in this part of the pathway take an important role in other cellular processes and secondary metabolism, therefore targeting later steps in $InsP_6$ synthesis may be prudent (Raboy, 2009).

Kim & Tai (2011), aimed to knock out genes from the early and late pathway using a reverse genetics approach to identify genes necessary for wild-type $InsP_6$ levels in *Arabidopsis* seeds, finding that knocking out *AtIPK2α* did not produce any phenotypic effect related to $InsP_6$ content in seeds, while the disruption of the lipid-dependent branch of the pathway could be partially accomplished by knocking out *AtIPK2β*. In this case, two different insertional lines from SALK institute were analyzed, SALK_104995, previously described by Stevenson-Paulik (2005), and SALK_024846, leading to a reduction of 43 to 48% of the $InsP_6$ content.

In consequence of the previous information, *IPK2α* and *IPK2β* orthologues from *B. napus* were chosen to be knocked out by CRISPR/Cas9 to see if the phenotype produced correlated with the findings in *Arabidopsis*. *IPK2α* paralogs were included because the absence of a clear phenotype in *A. thaliana* is not reason enough to ensure that this will be the case in *B. napus*.

6.1 *In silico* analysis

6.2 *AtIPK2β* homologs from *B. napus*

A. thaliana IPK2α and *IPK2β* genomic sequences were retrieved from TAIR and used as a query to run a BLAT alignment against the genomic sequence from *B. napus* (Darmor-*bzh*, version 4.1).

Two genomic sequences corresponding to *BnC09_IPK2α* and *BnA10_IPK2α* were retrieved and their protein sequence used to compare identity with AtIPK2α. On the other hand, four genomic sequences corresponding to *BnA06_IPK2β*, *BnCnn_IPK2β*, *BnA09_IPK2β*, and *BnC09_IPK2β* were retrieved, and their amino acid sequences were used to compare identity against AtIPK2β as well. As is shown in Supplementary Table S. 15, sequences from *B. napus* α homologs share a high identity (above 80%) with *A. thaliana*, while sequences from *B. napus* β homologs corresponding to BnA09_IPK2β and BnC09_IPK2β share a high identity with *A. thaliana* (above 80%), but the other members of this gene family showed to be less conserved (above 77%). Nevertheless, the fact that two of the paralogs from the *B. napus* IPK2β resulted in being less conserved, 77% of identity between sequences is a high sequence similarity. Therefore *BnA06_IPK2β* and *BnCnn_IPK2β* paralogs were included in this study as members of the *IPK2β* gene family. Another reason to include these paralogs is that they presented high coverage and sequence similarity (above 95%) with *IPK2β* from *B. oleracea* and *B. rapa* (data not shown).

6.3 Functional domain prediction and target site design

A multiple protein sequence alignment (MSA) (Figure 3) including IPK2α and IPK2β amino acid sequences from *B. napus, A. thaliana, B. rapa*, and *B. oleracea*, was constructed and functional

domains, CRISPR/Cas9 target regions, and other features, like critical amino acids for IP or ATP/Mg^{+2} binding, were annotated.

Most of the enzymes belonging to the inositol polyphosphate kinase superfamily can be classified into three subgroups, inositol 3-kinases, inositol phosphate multikinases, and inositol hexakisphosphate kinases. These family members present several conserved motifs like the PxxxDxKxG motif (Figure 3, residues highlighted in dark blue), which is responsible for IP binding, and has been described by several authors as conserved among all inositol polyphosphates kinases (Shears, 2004; Stevenson-Paulik et al., 2002; Zhan et al., 2015).

Although these enzymes, except the conserved motifs previously mentioned, differ in their amino acid sequence, they are expected to have the same backbone folding, assuring their characteristic kinase activity. It has been reported that AtIPK2β 3-kinase activity requires conserved Lys residues at 102, 117 and 121 positions (Figure 3, residues highlighted in light blue). The first one, belonging to the PxxxDxKxG motif, providing kinase activity and the latter substrate specificity (Endo-Streeter et al., 2012).

The MSA constructed showed that the latter (Lys117 and 121) residues are not conserved either on *B. oleracea* and *B. rapa* homologs, nor BnA06_IPK2β and BnCnn_IPK2β, presenting in all cases a K117Q and K121N substitution. Therefore it is logical to hypothesize that these homologs might have lost their substrate specificity while retained this amino acid substitution during evolution, or that they might acquire a new function through a neofunctionalization process, or acquired a specialized function (Innan & Kondrashov, 2010).

In the case of the ATP/Mg^{+2} binding domain (Figure 3, residues highlighted in dark green), it has been demonstrated that the region between Leu250 and His255 is critical in AtIPK2β (Hui -Jun Xia et al., 2005). In the case of the sequences analyzed, all presented fully conserved residues, except residue 254, which have a substitution from alanine to threonine on IPK2β orthologues from *B. oleracea*, *B. rapa* and A06 and Cnn paralogs from *B. napus.* Taken together, this might indicate of reduction or loss of catalytic activity due to a poor or lack of ability to bind co-factors necessary for enzymatic activity.

On the other hand, a recent study (Wang et al., 2017), determined the relation of AtIPK2β in ABA signaling, discovering that a calcium-dependent protein kinase 4 (CPK4), a key ABA signaling component, can interact with AtIPK2β phosphorylating Tyr46, Thr128 and Ser147 residues (Figure 3, residues highlighted in red). Later, in a yeast two-hybrid and biomolecular fluorescence complementation assay they confirmed the interaction between AtIPK2β and CPK4 and found these three residues to be significant phosphorylation sites in AtIPK2β, claiming that they could be critical for inositol phosphate kinase activity.

On the MSA only Tyr46 and Thr128 are fully conserved among IPK2α and IPK2β orthologues, whereas Ser147 is conserved on IPK2β from all species and only in *A. thaliana* IPK2α. This amino acid substitution (Ser147Leu) in Br_IPK2α, Bo_IPK2α, BnC09_IPK2α, and BnA10_IPK2α might affect the catalytic activity of the enzyme since it was described as a critical amino acid phosphorylation site. If the enzymatic activity is affected, mutations in α paralogs might not show a clear phenotype (Malabanan & Bilnd, 2016; Stevenson-Paulik et al., 2002).

The role of IPK2 in transcriptional regulation and nuclear inositide metabolism was recently reviewed, and there are some studies that demonstrate its relationship with transcriptional regulation, mRNA export from the nucleus, and chromatin remodeling in higher eukaryotes and yeast (Malabanan & Bilnd, 2016).

No nuclear localization signal (NLS) has been predicted on the amino acid sequence of BnIPK2α or BnIPK2β paralogs, either in AtIPK2α or AtIPK2β, although the human orthologue from this family presents this signal. It was demonstrated that AtIPK2α is localizing at the nucleus and in the plasma membrane by expressing a fusion protein AtIPK2α-green fluorescent protein (GFP) in epidermal cells of *Allium cepa* (onion) transiently (Shears, 2004; J. Xu et al., 2005).
On the other hand, a study established that AtIPK2β is a protein localizing in the nucleus, that can complement a yeast mutant lacking a functional Arg-Mcm1 transcription complex, and made a fusion *AtIPK2β-GFP* construct that was expressed in *Nicotiana tabacum* (tobacco) protoplasts, finding that the protein was localized in the nucleus of cells and at a lower rate in the cytosol (Hui-Jun Xia et al., 2003).

Mammalian orthologues of IPK2β, like $RnI_3$3k-A (rat orthologue), are regulated by calmodulin (CaM) in a calcium-dependent manner, where CaM binds the enzyme activating it, and in this case, either AtIPK2α and AtIPK2β do not bind CaM, nor has been predicted the consensus CaM-binding sites on the functional domain analysis performed in this study (Wang et al., 2017; Hui-Jun Xia et al., 2003; Hui-Jun Xia et al., 2005).

For the design of the CRISPR/Cas9 target site (Figure 3, residues underlined by a black bar), a threshold of 10 conserved nucleotides after the PAM site was set, this decision was made based on the literature available at the moment. Many of the CRISPR/Cas9 studies were made on mammal cell lines, and it has been reported that at least 7 to 12 nt after PAM site (seed region) are critical for target site recognition and DNA cleavage, therefore a 10 nt threshold was set to be fully conserved among targeted paralogs at the 5' end after the PAM site.

Setting a threshold of 10 nt conservation after the PAM site, allowed to reduce the number of putative off-targets (Supplementary Table S. 14). Nevertheless, there are putative off-targets that contain some SNPs between the 10 and 12 positions that might be edited, and it is necessary to remember that this is a new technology works in live systems, and this is not a proven rule. Therefore, it is not possible to discard the option of having any off-target effects in other genes, even though this rule was followed, and in the end, edition of off-target genes will depend on the GC content of the sgRNA, stoichiometry of the sgRNA after folding, and other factors that affect DNA recognition and cleavage that have not been described yet (Bortesi & Fischer, 2014; Wolt et al., 2016; X.-H. Zhang et al., 2015).

Following the parameters described above, two target sites were identified aiming to knock-out in a simultaneous way *BnA06_IPK2β* and *BnCnn_IPK2β* (TS_1), and *BnA09_IPK2β* and *BnC09_IPK2β* (TS_2).

6.4 Construct development

The colony PCR (Figure 6) performed after cloning sgRNA:TS_1 and sgRNA:TS_2 into pCas9-TPC, presented an unanticipated result, in addition to the expected 1,545 bp product, a second band at ca. 1,000 bp was found, which most probably represented the vector carrying only one sgRNA (1,019 bp). Since the Cas9 primers used (Supplementary Table S. 1) to perform the PCR reaction

were tested before, and they were working specifically and efficiently to perform a PCR or Sanger sequencing reaction, it was hypothesized that an intraplasmid rearrangement might happen after bacteria transformation, during the replication of the vector.

It has been reported that intraplasmid rearrangements can occur in *E. coli* and *A. tumefaciens,* leading to InDels, by many different mechanisms, being common RecA-dependent and RecA-independent pathways; the first one based on homologous recombination, and the second, is produced by replication misalignment. RecA-independent mechanism of recombination depends on the proximity and the perfect homology of the repeat sequence because any difference between the repeats will reduce deletion efficiency (Bočkor & Malenica, 2013; Bzymek & Lovett, 2001).

The pCas9-TPC vector, designed by the group of Professor Holger Puchta (Fauser et al., 2014), was engineered to work with only one sgRNA subcloned from pChimera. In the case of this study, it was required to use two sgRNAs to target the four *BnIPK2β* paralogs with only one construct (to reduce putative off-target effects), and to accomplish this, a strategy to subclone a second sgRNA from pChimera was developed. Taking advantage of the BcuI restriction site adjacent (58 bp) to AvrII, allowed cloning two sgRNAs in tandem, producing pCas9-TPC:TS_1:TS_2 (Figure 7.D). Since the promoter (AtU6-26) and the major part of the sgRNA sequence are 100% identical, the only difference between both sgRNAs constructed is the 20 nt region that possesses homology with the target DNA (adjacent to PAM site). Based on the previous information it is proposed that the pCas9-TPC:TS_1:TS_2 vector could suffer an intraplasmid rearrangement and lose one of the sgRNAs during vector replication, as the *E. coli* XL1/Blue carries a *recA1* mutation (preventing recombination). Therefore it is more likely that the rearrangement happened during replication. Based on this information, the unexpected 1,000 bp band was isolated from the gel and sent to sequencing.

After the sequencing results (Supplementary Figure S. 7) were assembled to the reference vector (pCas9-TPC:TS_1:TS_2), no overlapping was observed between the forward and reverse sequences, because they were too short. Therefore, it was not possible to conclude if the corresponding sgRNAs were intact or if they suffered modifications. Since the sequence alignment did not result as expected, the vector was sent to sequencing. However, the results showed that the vector was intact when the sequencing results were assembled to the reference (Supplementary Figure S. 6).

To help to solve this issue, further analysis like cloning the 1,000 bp PCR product into a pGEM®-T vector, and send it to be sequenced, or using the second sgRNA with a different promoter might help to elucidate and resolve this issue.

6.5 Plant Transformation

Several authors have reported transforming *B. napus* by different methods such as floral dip, cotyledon, and hypocotyl transformation, or polyethylene glycol (PEG) mediated transformation, but until today, the most common way to transform *B. napus* is by *A. tumefaciens* mediated transformation of hypocotyls by tissue culture, and it is well known that the success of regenerating a whole transgenic plant will depend on its genotype, the strain of the bacteria, the vector, and many other environmental factors like the medium utilized, solidifying medium agent, pH, and handling (Bhalla & Singh, 2008; Boszoradova et al., 2011; De Block et al., 1989; H. Yang et al., 2017; Y. Yang et al., 2017).

In total 5,054 rapeseed hypocotyls from Haydn and Express 617 were transformed using *A. tumefaciens* (Table 3), from those 1,165 were from the Express 617 genotype and 3,889 were from the Haydn genotype. The germination rates of the two Haydn seed codes used in this study, greatly varied, being between 92 and 100% for seed code 0900060 and 100% for seed code 161110. This variation in germination rates can be explained by the age of the seeds, in the case of 0900060 seed code, they were generated before 161110 seed code. Therefore, they might lose viability over time.

About the tissue culture performance of the two cultivars used in this study, Haydn proved to respond positively to the regeneration protocol, having in average at the time, near 190 shoots per experiment in selection medium O_5, representing between 11.5 and 45.6% of surviving shoots at the time. In contrast, Express 617 cultivar showed to be recalcitrant to regenerating, presenting nearly all (99.74%) the hypocotyls dead short after *A. tumefaciens* infection. The remaining ones (0.17%) died in O_5 medium or in rooting medium (0,086%) O_6.

In the plant transformation protocol used to perform this study, it was expected that the hypocotyls start to generate shoots soon (4 weeks) after *A. tumefaciens* infection, but in this study it was observed that shoots started to develop from the third week. Until the moment that this thesis was delivered (26 to 30 weeks after *A. tumefaciens* infection) some explants were still generating new shoots in shooting medium (O_4) without selective agent (phosphinothricin), many others generated calli and stayed like that until now, others generated calli and then started to shoot, others generated roots even though they were in shooting media, and root development was not reported on the protocol used to perform rapeseed hypocotyl transformation.

Plant orthologues of IPK2 have been well characterized, overexpression of *IPK2* from *Thellungiella halophila* in *B. napus*, and *IPK2β* from *A. thaliana* in *N. tabacum*, presented enhancement of abiotic stress tolerance (L. Yang et al., 2008; Zhu et al., 2009), while overexpression of *IPK2β* from *A. thaliana* in the wild-type and *A. thaliana* mutants resulted in more axillary shoot branches, through auxin signaling pathway (Z.-B. Zhang et al., 2007). On the other hand, single knock-out of *ipk2α* and *ipk2β* from *A. thaliana*, did not show any differences from the wild type, although the double mutant presented reduced fertility due to an impairment on the pollen tube guidance and an abnormal pollen development, furthermore the early embryo development was aborted (Zhan et al., 2015), same phenotypes as above were observed when overexpressing or silencing *IPK2* from *O. sativa* (Chen et al., 2017; Cheng et al., 2017).

Having in mind the above information, it can be expected that the regeneration rates might vary because the important role that IPK2β fulfill in the modulation of ABA and gibberellic acid signaling, in consequence, the effect of the plant growth regulators involved in hypocotyl regeneration might be affected. Nevertheless, there are *B. napus* TILLING single mutants (unpublished data from "DFG Phytic Acid TILLING" project) of *BnA09_IPK2β* and *BnC09_IPK2β* which have not been backcrossed yet. Therefore they carry an important load of background mutations, and none of them have shown the phenotypes mentioned above.

The CRISPR/Cas9 constructs designed in this study, are aiming to knock-out the four *BnIPK2β* paralogs. Therefore the phenotype obtained might be clearer than the one obtained by knocking-out only two *BnIPK2β* paralogs by TILLING, but this issue will be clarified in the future when phenotyping a T_2 population.

6.6 Transgene detection and gene editing

The leaf samples showing a positive signal for Cas9 and negative for spectinomycin PCR, are a strong indicator of transgenesis because the Cas9 gene is located within the left and the right border of the vector, while the spectinomycin gene is located outside of the T-DNA borders.

The Cas9 PCR reaction showed a 500 bp band (size expected for the empty pCas9-TPC vector), which might be an indication of contamination of the bacterial stock with bacteria carrying the empty vector. Therefore, an extensive PCR screening (Supplementary Figure S. 8, Supplementary Figure S. 9, and Supplementary Figure S. 10) was made to dismiss the possibility of contamination of the bacterial stock by performing a Cas9 PCR using as template the bacterial stock, *A. tumefaciens* colonies grown in LB plates with the corresponding antibiotics, and tissue culture medium samples taken from the surrounding of the shoots screened.

As is shown in Supplementary Figure S. 8, the PCR results from the bacterial stock showed in every case the expected band size for the pCas9-TPC:TS_1 and pCas9-TPC:TS_2 (1,000 bp) and the pCas9-TPC:TS_1:TS_2 (1,500 bp). While the colony Cas9 PCR (Supplementary Figure S. 9) also shows the expected band size for the vector with only one sgRNA (1,000 bp) and the vector for multiplexing (1,500 bp).

Finally, the PCR using the surrounding medium of shoots as template (Supplementary Figure S. 10), showed in its majority two or three bands, with sizes of 500, 1,000 or 1,500 bp, corresponding to the pCas9-TPC empty vector, the pCas9-TPC:TS_1 or pCas9-TPC:TS_2 vectors, or the pCas9-TPC:TS_1:TS_2 (depending on the experiment) respectively. Therefore bacterial contamination is very likely. As is shown in Figure 8. A, the weak signal from the Cas9 PCR when compared with the PCR product obtained from the target site region genotyping (Supplementary Figure S. 13), might indicate that this signal is coming either from chimeric plants, from non transgenic ones that have a bacterial contamination, or from a vector contamination, rather than the T-DNA inserted in the plant. With the aim to establish the origin of the 500 bp band, this PCR product was sent to sequencing, but since the signal was too weak, no data were obtained.

From the 19 shoots screened for targeted gene editing, a 100% showed the wild-type alleles of the paralog in the target site region. Nevertheless, from the rapeseed hypocotyl transformations accomplished in this laboratory before this study, hardly any editing of the target site region was observed on tissue culture stage, while the regenerated plants in the greenhouse showed to be edited. Therefore is still possible to find a gene editing of the shoots that are currently in tissue culture after rooting and transfer to the greenhouse.

In the end, the efficiency of the CRISPR/Cas9 system to knock-out *BnIPK2β* paralogs will be assessed in a later experimental stage.

6.7 Prospects

The aim of this study was to knock-out the four *BnIPK2β* paralogs by the CRISPR/Cas9 technology to be able to decide whether a frameshift mutation would lead to an $InsP_6$ reduction in *B. napus* seeds, or if the plant carrying a functional *ipk2β* mutation can overcome the shutdown of the lipid-dependent pathway, by deviating the metabolites through the lipid-independent pathway, which as we know, also leads to $InsP_6$ accumulation in seeds.

Different phenotypes in *A. thaliana* were observed while knocking-out genes of the inositol phosphate synthesis, from poor plant performance (seed germination, development, and fertility) to lethal phenotypes, and in the case of this study, it is necessary to allow the regeneration of complete plants to take samples, and test if they are transgenic, and if the paralogs were edited, and it will be necessary to get the T_2 and establish the phenotype of the mutants.

Two *BnIPK2α* and four *BnIPK2β* paralogs were successfully identified in the *in silico* analysis, but further experiments like carrying out an expression analysis of all the paralogs will give a clue which ones are functional and focus on targeting those paralogs, might help to elucidate their role in $InsP_6$ synthesis.

Regarding the CRISPR/Cas9 construct development, it remains to be clarified why an unspecific band was observed in the PCR reaction if two sgRNAs are cloned in tandem in the pCas9-TPC vector, and if recombination is the reason for the occurrence of that band. Therefore is advisable to clone the band and send it to sequencing. Nevertheless, the construct development was succesfully carried out and confirmed by Sanger sequencing.

From the tissue culture performance, it was possible to state that the genotype Haydn has higher regeneration rates than Express 617, but it will not be possible to calculate the final regeneration and transformation rates or the efficiency of the CRISPR/Cas9 technology to knock-out the *BnIPK2* paralogs until all the explants are dead or regenerated.

7 Summary

Oilseed rape is a significant crop due to its seed oil and protein content, being used as an oil source for human consumption and the remaining meal after oil extraction (high in protein) as animal feed. Another compound present in high amounts (2 to 4% for the whole dry seed) is phytic acid (*myo*-inositol-1,2,3,4,5,6-hexa*kis*phosphate; $InsP_6$), which is ubiquitous in eukaryotic cells and the most abundant form of phosphorus in plants, serving as an energy reservoir for seeds during germination. $InsP_6$ is considered as an antinutritional compound due to its great ability to chelate mineral nutrients in non-ruminant animals because their lack of phytases to digest it.

$InsP_6$ synthesis starts with the *de novo* production of *myo*-inositol (Ins) which can be further phosphorylated through the lipid-dependent or lipid-independent branch of the pathway until it reaches a higher state of phosphorylation, producing $InsP_6$.

The aim of this study was to knock-out the *BnIPK2β* gene, from the phytic acid lipid-dependent pathway by CRISPR/Cas9 to reduce the $InsP_6$ content in seeds.

Four *BnIPK2β* orthologues from *A. thaliana* were found in the *in silico* analysis, and functional domains and structural domains were annotated, identifying the characteristic PxxxDxKxG motif of the inositol polyphosphate multikinase superfamily (IP binding) and an ATP/Mg^{+2} binding domain as well. Two target sites upstream these domains were identified to knock-out the four *BnIPK2β* paralogs, and off-target analysis was performed by blastn.

The sgRNAs were successfully cloned in the pChimera and subcloned in the pCas9-TPC as described by Fauser et al., 2014, and a strategy to generate a pCas9-TPC carrying two sgRNAs was developed, by cloning a second sgRNA in a BbsI restriction site adjacent to AvrII.

In total 5,054 rapeseed hypocotyls were transformed using *A. tumefaciens*, from those 1,165 were from the Express 617 winter genotype and 3,889 were from the Haydn spring genotype, where Haydn showed to respond better than Express 617 to the regeneration protocol followed.

At the time there are only Haydn explants, in total 1,360 regenerated explants that were transferred to selection medium (with phosphinothricin), and from those, 947 shoots where surviving, but at the time no positive selection for transgenics was observed. Nevertheless, leaf samples from 19 putative transgenic shoots were taken during subculture, and DNA was isolated to screening the plants for transgenes and gene editing, but no edition was found.

The putative transgenic shoots will be further grown and regenerated to T_1 plants, genotyped and selfed to measure the $InsP_6$ content in seeds of the T_2 generation.

8 References

Bhalla, P. L., & Singh, M. B. (2008). Agrobacterium-mediated transformation of Brassica napus and Brassica oleracea. *Nature Protocols*, *3*(2), 181–189. http://doi.org/10.1038/nprot.2007.527

Bočkor, L., & Malenica, N. (2013). Comparison of Intraplasmid Rearrangements in Agrobacterium tumefaciens and Esherichia coli. *Food Technology and Biotechnology*, *51*(4), 441–445. Retrieved from http://hrcak.srce.hr/index.php?show=clanak&id_clanak_jezik=169393

Bortesi, L., & Fischer, R. (2014). The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnology Advances*, *33*(1), 41–52. http://doi.org/10.1016/j.biotechadv.2014.12.006

Boszoradova, E., Libantova, J., Matusikova, I., Poloniova, Z., Jopcik, M., Berenyi, M., & Moravcikova, J. (2011). Agrobacterium-mediated genetic transformation of economically important oilseed rape cultivars. *Plant Cell, Tissue and Organ Culture*, *107*(2), 317–323. http://doi.org/10.1007/s11240-011-9982-y

Braatz, J., Harloff, H.-J., Mascher, M., Stein, N., Himmelbach, A., & Jung, C. (2017). CRISPR-Cas9 targeted mutagenesis leads to simultaneous modification of different homoeologous gene copies in polyploid oilseed rape (Brassica napus L.). *Plant Physiology*, pp.00426.2017. http://doi.org/10.1104/pp.17.00426

Bregitzer, P., & Raboy, V. (2006). Effects of four independent low-phytate mutations on barley agronomic performance. *Crop Science*, *46*(3), 1318–1322. http://doi.org/10.2135/cropsci2005.09-0301

Bzymek, M., & Lovett, S. T. (2001). Instability of repetitive DNA sequences: The role of replication in multiple mechanisms. *Proceedings of the National Academy of Sciences*, *98*(15), 8319–8325. http://doi.org/10.1073/pnas.111008398

Chalhoub B, Denoeud F, Liu SY, Parkin IAP, Tang HB, Wang XY, Chiquet J, Belcram H, Tong CB, Samans B, Correa M, Da Silva C, Just J, Falentin C, Koh CS, Le Clainche I, Bernard M, Bento P, Noel B, Labadie K, Alberti A, Charles M, Arnaud D, Guo H, Daviaud C, Alamery S, Jabbari K, Zhao MX, Edger PP, Chelaifa H, Tack D, Lassalle G, Mestiri I,

Schnel N, Le Paslier MC, Fan GY, Renault V, Bayer PE, Golicz AA, Manoli S, Lee TH, Thi VHD, Chalabi S, Hu Q, Fan CC, Tollenaere R, Lu YH, Battail C, Shen JX, Sidebottom CHD, Wang XF, Canaguier A, Chauveau A, Berard A, Deniot G, Guan M, Liu ZS, Sun FM, Lim YP, Lyons E, Town CD, Bancroft I, Wang XW, Meng JL, Ma JX, Pires JC, King GJ, Brunel D, Delourme R, Renard M, Aury JM, Adams KL, Batley J, Snowdon RJ, Tost J, Edwards D, Zhou YM, Hua W, Sharpe AG, Paterson AH, Guan CY, Wincker P (2014) Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. *Science 345* (6199):950-953. doi:10.1126/science.1253435

Cheek, S., Ginalski, K., Zhang, H., & Grishin, N. V. (2005). A comprehensive update of the sequence and structure classification of kinases. *BMC Structural Biology*, *5*(6), 1–19. http://doi.org/10.1186/1472-6807-5-6

Chen, Y., Wei, Z., Yang, Q., Sang, S., & Wang, P. (2017). Rice inositol polyphosphate kinase gene (OsIPK2), a putative new player of gibberellic acid signaling, involves in modulation of shoot elongation and fertility. *Plant Cell, Tissue and Organ Culture*, *0*(0), 1–12. http://doi.org/10.1007/s11240-017-1298-0

Chen, Y., Yang, Q., Sang, S., Wei, Z., & Wang, P. (2017). Rice Inositol Polyphosphate Kinase (OsIPK2) Directly Interacts with OsIAA11 to Regulate Lateral Root Formation. *Plant & Cell Physiology*, *58*(11), 1891–1900. http://doi.org/10.1093/pcp/pcx125

Cong, L., Ran, F. A., Cox, D., Lin, S., Barreto, R., Habib, N., … Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*, *339*(6121), 819–823. http://doi.org/10.1126/science.1231143

De Block, M., De Brouwer, D., & Tenning, P. (1989). Transformation of Brassica napus and Brassica oleracea Using Agrobacterium tumefaciens and the Expression of the bar and neo Genes in the Transgenic Plants. *Plant Physiology*, *91*(2), 694–701. http://doi.org/10.1104/pp.91.2.694

El-Batal, A. I., & Abdel Karem, H. (2001). Phytase production and phytic acid reduction in rapeseed meal by Aspergillus niger during solid state fermentation. *Food Research International*, *34*(8), 715–720. http://doi.org/10.1016/S0963-9969(01)00093-X

Endo-Streeter, S., Tsui, M.-K. M., Odom, A. R., Block, J., & York, J. D. (2012). Structural studies and protein engineering of inositol phosphate multikinase. *The Journal of Biological Chemistry*, *287*(42), 35360–35369. http://doi.org/10.1074/jbc.M112.365031

Fauser, F., Schiml, S., & Puchta, H. (2014). Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in Arabidopsis thaliana. *The Plant Journal*, *79*(2), 348–359. http://doi.org/10.1111/tpj.12554

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., … Bateman, A. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research*, *44*(D1), D279–D285. http://doi.org/10.1093/nar/gkv1344

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook* (pp. 571–607). http://doi.org/10.1385/1592598900

Hatch, A. J., & York, J. D. (2010). SnapShot: Inositol phosphates. *Cell*, *143*(6), 1030–1031. http://doi.org/10.1016/j.cell.2010.11.045

Hofgen, R., & Willmitzer, L. (1988). Storage of competent cells for Agrobacterium transformation. *Nucleic Acids Research*, *16*(20), 9877.

Hüsken, A., Baumert, A., Strack, D., Becker, H. C., Möllers, C., & Milkowski, C. (2005). Reduction of sinapate ester content in transgenic oilseed rape (Brassica napus) by dsRNAi-based suppression of BnSGT1 gene expression. *Molecular Breeding*, *16*(2), 127–138. http://doi.org/10.1007/s11032-005-6825-8

Innan, H., & Kondrashov, F. (2010). The evolution of gene duplications: classifying and distinguishing between models. *Nature Reviews Genetics*, *11*(4), 97–108. http://doi.org/10.1038/nrg2689

Jackson, J. F., Jones, G., & Linskens, H. F. (1982). Phytic acid in pollen. *Phytochemistry*, *21*(6), 1255–1258. http://doi.org/10.1016/0031-9422(82)80121-0

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptative Bacterial Immunity. *Science*, *337*, 816–821. http://doi.org/10.1126/science.1225829

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., & Madden, T. L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Research*, *36*(Web Server), W5–W9. http://doi.org/10.1093/nar/gkn201

Josefsen, L., Bohn, L., Sørensen, M. B., & Rasmussen, S. K. (2007). Characterization of a multifunctional inositol phosphate kinase from rice and barley belonging to the ATP-grasp superfamily. *Gene*, *397*(1–2), 114–125. http://doi.org/10.1016/j.gene.2007.04.018

Kent, W. J. (2002). BLAT — The BLAST -Like Alignment Tool. *Genome Research*, *12*, 656–664. http://doi.org/10.1101/gr.229202.

Kim, S. I., & Tai, T. H. (2011). Identification of genes necessary for wild-type levels of seed phytic acid in Arabidopsis thaliana using a reverse genetics approach. *Molecular Genetics and Genomics*, *286*(2), 119–133. http://doi.org/10.1007/s00438-011-0631-2

Kumar, V., Sinha, A. K., Makkar, H. P. S., & Becker, K. (2010). Dietary roles of phytate and phytase in human nutrition: A review. *Food Chemistry*, *120*(4), 945–959. http://doi.org/10.1016/j.foodchem.2009.11.052

Lawrenson, T., Shorinola, O., Stacey, N., Li, C., Østergaard, L., Patron, N., … Harwood, W. (2015). Induction of targeted, heritable mutations in barley and Brassica oleracea using RNA-guided Cas9 nuclease. *Genome Biology*, *16*(1), 1–13. http://doi.org/10.1186/s13059-015-0826-7

Lemtiri-Chlieh, F., MacRobbie, E. A. C., & Brearley, C. A. (2000). Inositol hexakisphosphate is a physiological signal regulating the K+-inward rectifying conductance in guard cells. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(15), 8687–8692. http://doi.org/10.1073/pnas.140217497

Liu, H., Ding, Y., Zhou, Y., Jin, W., Xie, K., & Chen, L. L. (2017). CRISPR-P 2.0: An Improved CRISPR-Cas9 Tool for Genome Editing in Plants. *Molecular Plant*, *10*(3), 530–532. http://doi.org/10.1016/j.molp.2017.01.003

Malabanan, M. M., & Bilnd, R. D. (2016). Inositol polyphosphate multikinase (IPMK) in transcriptional regulation and nuclear inositide metabolism. *Biochemical Society Transactions*, *1*(44), 279–285. http://doi.org/10.1042/BST20150225

Paul, J. W., & Qi, Y. (2016). CRISPR/Cas9 for plant genome editing: accomplishments, problems and prospects. *Plant Cell Reports*, *35*(7), 1417–1427. http://doi.org/10.1007/s00299-016-1985-z

Puchta, H. (2017). Applying CRISPR/Cas for genome engineering in plants: the best is yet to come. *Current Opinion in Plant Biology*, *36*, 1–8. http://doi.org/10.1016/j.pbi.2016.11.011

Raboy, V. (2002). Symposium : Plant Breeding : A New Tool for Fighting Micronutrient Malnutrition. *Journal of Nutrition*, *132*, 503–505.

Raboy, V. (2003). myo-Inositol-1,2,3,4,5,6-hexakisphosphate. *Phytochemistry*, *64*(6), 1033–1043. http://doi.org/10.1016/S0031-9422(03)00446-1

Raboy, V. (2009). Approaches and challenges to engineering seed phytate and total phosphorus. *Plant Science*, *177*(4), 281–296. http://doi.org/10.1016/j.plantsci.2009.06.012

Raboy, V., Gerbasi, P. F., Young, K. A., Stoneberg, S. D., Pickett, S. G., Bauman, A. T., … Ertl, D. S. (2000). Origin and Seed Phenotype of Maize low phytic 1-1 and low phytic acid 2-1. *Plant Physiology*, *124*(1), 355–368. http://doi.org/10.1104/pp.124.1.355

Raboy, V., Young, K. a, Dorsch, J. a, & Cook, A. (2001). Genetics and breeding of seed phosphorus and phytic acid. *Journal of Plant Physiology*, *158*(4), 489–497. http://doi.org/10.1078/0176-1617-00361

Schiml, S., Fauser, F., & Puchta, H. (2014). The CRISPR/Cas system can be used as nuclease for in planta gene targeting and as paired nickases for directed mutagenesis in Arabidopsis resulting in heritable progeny. *The Plant Journal*, *80*(6), 1139–1150. http://doi.org/10.1111/tpj.12704

Seeds, A. M., Frederick, J. P., Tsui, M. M. K., & York, J. D. (2007). Roles for inositol polyphosphate kinases in the regulation of nuclear processes and developmental biology. *Advances in Enzyme Regulation*, *47*, 10–25. http://doi.org/10.1016/j.advenzreg.2006.12.019

Shears, S. B. (2001). Assessing the omnipotence of inositol hexakisphosphate. *Cellular Signalling*, *13*(3), 151–158. http://doi.org/10.1016/S0898-6568(01)00129-2

Shears, S. B. (2004). How versatile are inositol phosphate kinases? *The Biochemical Journal*, *377*(2), 265–280. http://doi.org/10.1042/bj20031428

Sparvoli, F., & Cominelli, E. (2015). Seed Biofortification and Phytic Acid Reduction: A Conflict of Interest for the Plant? *Plants*, *4*(4), 728–755. http://doi.org/10.3390/plants4040728

Stevenson-Paulik, J., Bastidas, R. J., Chiou, S.-T., Frye, R. A., & York, J. D. (2005). Generation of phytate-free seeds in Arabidopsis through disruption of inositol polyphosphate kinases. *PNAS*, *102*(35), 12612–12617. http://doi.org/10.1073?pnas.0504172102

Stevenson-Paulik, J., Odom, A. R., & York, J. D. (2002). Molecular and Biochemical Characterization of Two Plant Inositol Polyphosphate 6-/3-/5-Kinases *. *The Journal of Biological Chemistry*, *277*(45), 42711–42718. http://doi.org/10.1074/jbc.M209112200

Uppström, B., & Svensson, R. (1980). Determination of phytic acid in rapeseed meal. *Journal of the Science of Food and Agriculture*, *31*(7), 651–656. http://doi.org/10.1002/jsfa.2740310706

Wang, P., Yang, Q., Sang, S., Chen, Y., Zhong, Y., & Wei, Z. (2017). Arabidopsis inositol polyphosphate kinase AtIpk2β is phosphorylated by CPK4 and positively modulates ABA signaling. *Biochemical and Biophysical Research Communications*, *490*(2), 441–446. http://doi.org/10.1016/j.bbrc.2017.06.060

Wolt, J. D., Wang, K., Sashital, D., & Lawrence-Dill, C. J. (2016). Achieving Plant CRISPR Targeting that Limits Off-Target Effects. *The Plant Genome*, *9*(3), 1–8. http://doi.org/10.3835/plantgenome2016.05.0047

Xia, H.-J., Brearley, C., Elge, S., Kaplan, B., Fromm, H., & Meuller-Roeber, B. (2003). Arabidopsis Inositol Polyphosphate 6-/3-Kinase Is a Nuclear Protein That Complements a Yeast Mutant Lacking a Functional ArgR-Mcm1 Transcription Complex. *The Plant Cell*, *15*(2), 449–463. http://doi.org/10.1105/tpc.006676

Xia, H.-J., & Yang, G. (2005). Inositol 1,4,5-trisphosphate 3-kinases: functions and regulations. *Cell Research*, *15*(2), 83–91. http://doi.org/10.1038/sj.cr.7290270

Xu, X., Duan, D., & Chen, S.-J. (2017). CRISPR-Cas9 cleavage efficiency correlates strongly with target-sgRNA folding stability: from physical mechanism to off-target assessment. *Scientific Reports*, *7*(1), 1–9. http://doi.org/10.1038/s41598-017-00180-1

Yamaji, N., Takemoto, Y., Miyaji, T., Mitani-Ueno, N., Yoshida, K. T., & Ma, J. F. (2016). Reducing phosphorus accumulation in rice grains with an impaired transporter in the node. *Nature*, *541*(7635), 92–95. http://doi.org/10.1038/nature20610

Yang, L., Tang, R., Zhu, J., Liu, H., Mueller-Roeber, B., Xia, H.-J., & Zhang, H. (2008). Enhancement of stress tolerance in transgenic tobacco plants constitutively expressing AtIpk2β, an inositol polyphosphate 6-/3-kinase from Arabidopsis thaliana. *Plant Molecular Biology*, *66*(4), 329–343. http://doi.org/10.1007/s11103-007-9267-3

Yang, H., Wu, J. J., Tang, T., Liu, K. De, & Dai, C. (2017). CRISPR/Cas9-mediated genome editing efficiently creates specific mutations at multiple loci using one sgRNA in Brassica napus. *Scientific Reports*, *7*(1), 1–13. http://doi.org/10.1038/s41598-017-07871-9

Yang, Y., Zhu, K., Li, H., Han, S., Meng, Q., Ullah Khan, S., Zhou, Y. (2017a). *Precise editing of CLAVATA genes in Brassica napus L. regulates multilocular silique development. ARPN Journal of Engineering and Applied Sciences* (Vol. 12). http://doi.org/10.1111/ijlh.12426

Yin, K., Gao, C., & Qiu, J. (2017). Progress and prospects in plant genome editing. *Nature*

Publishing Group, *3*(July), 1–6. http://doi.org/10.1038/nplants.2017.107

Zarhloul, M. K., Stoll, C., Lühs, W., Syring-Ehemann, A., Hausmann, L., Töpfer, R., & Friedt, W. (2006). Breeding high-stearic oilseed rape (Brassica napus) with high- and low-erucic background using optimised promoter-gene constructs. *Molecular Breeding*, *18*(3), 241–251. http://doi.org/10.1007/s11032-006-9032-3

Zhan, H., Zhong, Y., Yang, Z., & Xia, H.-J. (2015). Enzyme activities of Arabidopsis inositol polyphosphate kinases AtIPK2a and AtIPK2b are involved in pollen development, pollen tube guidance and embryogenesis. *Plant Journal*, *82*(5), 758–771. http://doi.org/10.1111/tpj.12846

Zhang, K., Raboanatahiry, N., Zhu, B., & Li, M. (2017). Progress in Genome Editing Technology and Its Application in Plants. *Frontiers in Plant Science*, *8*(177), 1–17. http://doi.org/10.3389/fpls.2017.00177

Zhang, X.-H., Tee, L. Y., Wang, X.-G., Huang, Q.-S., & Yang, S.-H. (2015). Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Molecular Therapy. Nucleic Acids*, *4*(11), 1–8. http://doi.org/10.1038/mtna.2015.37

Zhang, Z.-B., Yang, G., Arana, F., Chen, Z., Li, Y., & Xia, H.-J. (2007). Arabidopsis Inositol Polyphosphate 6-/3-Kinase (AtIpk2beta) Is Involved in Axillary Shoot Branching via Auxin Signaling. *Plant Physiology*, *144*(2), 942–951. http://doi.org/10.1104/pp.106.092163

Zhu, J.-Q., Zhang, J.-T., Tang, R.-J., Lv, Q.-D., Wang, Q.-Q., Yang, L., & Zhang, H.-X. (2009). Molecular characterization of ThIPK2, an inositol polyphosphate kinase gene homolog from Thellungiella halophila, and its heterologous expression to improve abiotic stress tolerance in brassica napus. *Physiologia Plantarum*, *136*, 407–425. http://doi.org/10.1111/j.1399-3054.2009.01235.x

9 Supplementary data

Supplementary Table S. 1: Primers used in this study.

Aim	Primer name	Sequence	Tm (°C)	Amplicon size (bp)
BnA06_IPK2β target region genotyping	NS_P164 NS_P30	GGCTTGGTCCGCTCGTGAAT GCCTAAACCCCAAGGAAACA	60	354
BnCnn_IPK2β target region genotyping	NS_P170 NS_P32	GGCTTGGTCCGCTCGTGTAC ACTAGAACTGCCATAAACTTCCGAG	60	536
BnA09_IPK2β target region genotyping	NS_P318 NS_P26	ATGAACTCGAGGAACTCTGTTACAT GCCTCCTCTCCTCATCACCC	60	815
BnC09_IPK2β target region genotyping	BnaC09_f NS_P87	CCGAAACCCTGAACCCCAAAT ACATCGTATCTGGACTCTGCG	60	518
pChimera colony PCR	Target1_f pChimera_r	CGTGTATCACGGCACTCAAG GTTAATGCAGCTGGCACGAC	59	228
	Target2_f pChimera_r	TGTGTACCACGGCGTTCAGC GTTAATGCAGCTGGCACGAC	59	228
pCas9-TPC colony PCR single target	pCas9_f pCas9_r	CAGTCTTTCACCTCTCTTTGG CCATCTTTGGGACCACTGTC	59	483 without insert / 1,545 with insert
pCas9-TPC colony PCR multiplex target	pCas9_f pCas9_r	CAGTCTTTCACCTCTCTTTGG CCATCTTTGGGACCACTGTC	59	1,019 without insert / 1,545 with insert
Spectinomycin detection	Spec_for Spec_rev	TGCCGACTACCTTGGTGATC GTACAGTCTATGCCTCGGGC	61	910

Supplementary Table S. 2: Digestion of pChimera

Reagent	Volume (µl)
pChimera (~500 ng)	1
10x Buffer G (Thermo Scientific)	2
BbsI (10U/µl) (Thermo Scientific)	1
ddH_2O	To 20

Supplementary Table S. 3: Oligo annealing

Reagent	Volume (µl)
Oligo forward (100µM)	1
Oligo reverse (100µM)	1
ddH_2O	48

Supplementary Table S. 4: Oligo ligation into pChimera

Reagent	Volume (µl)
Digested pChimera (5 ng/µl)	2
Annealed Oligos	3
10 x T4 DNA Ligase buffer	2
T4 DNA ligase (Thermo Scientific)	1

Supplementary Table S. 5: Bacterial medium and antibiotics

***E.coli* LB medium**		**Selection Antibiotic**	
Reagent	**Weight (g)**	**Bacteria/Vector**	**Antibiotic (mg l^{-1})**
Yeast extract	5	pChimera	Ampicillin (100)
Trypton	10	pCas9-TPC	Spectinomycin (100)
NaCl	5		
Agar (optional)	15		
dH_2O	To 1 liter		
***A.tumefaciens* YEB medium**			
Reagent	**Weight (g)**		
Beef extract	5	*A.tumefaciens*	Rifampicin (25)
Yeast extract	1	pMP90RK	Gentamicin (50)
Trypton	5	pCas9-TPC	Spectinomycin (100)
Sucrose	5		
NaCl	0.5		
Agar (optional)	15		
dH_2O	To 1 liter		

Supplementary Table S. 6: PCR and Colony PCR, reagents and conditions.

Reagent	Concentration	Volume (µl)
H_2O	-	15.9
PCR Buffer ($MgCl_2$ included)	10 x	2
dNTP's	10 mM	0.4
Forward primer	10 pMol/µl	0.3
Reverse primer	10 pMol/µl	0.3
Taq DNA Polymerase (Biozym)	5 U/µl	0.1
Template	~5 ng/µl	1

Temperature (°C)	Time	Cycles
94	5 min*	1
94	30 sec	
60**	30 sec	35
72	30 sec***	
72	10 min	1
12	∞	-

* 8 to 10 minutes for colony PCR

** depending on primer combination

*** depending on product size

Supplementary Table S. 7: sgRNA restriction digestion from pChimera.

Reagent	Volume (µl)
pChimera:TS (200 ng/µl)	17
Buffer Tango (Thermo Scientific)	2
XmaJI (10 U/µl) (Thermo Scientific)	1

Supplementary Table S. 8: Digestion of pCas9-TPC

Reagent	Volume (µl)
pCas9-TPC (200 ng/µl)	17
Buffer Tango (Thermo Scientific)	2
XmaJI (10 U/µl) (Thermo Scientific)	1

Supplementary Table S. 9: pCas9-TPC dephosphorylation.

Reagent	Volume (µl)
Digested pCas9-TPC (200 ng/µl)	17
Buffer AP (Thermo Scientific)	2
Fast AP alkaline Phosphatase	1

Supplementary Table S. 10: sgRNA:TS_1 ligation into pCas9-TPC

Reagent	Volume (µl)
H_2O	2
sgRNA:TS_1 (26.3 ng/µl)*	6
Digested and dephosphorylated pCas9-TPC (17.2 ng/µl)*	9
Buffer ligase (10 x) (Thermo scientific)	2
T4 ligase (10 U/µl) (Thermo scientific)	1

*vector:insert in a molar 3:1 ratio

Supplementary Table S. 11: Digestion of pCas9-TPC:TS_1

Reagent	Volume (µl)
pCas9-TPC:TS_1 (200 ng/µl)	17
Buffer Tango (Thermo Scientific)	2
BcuI (Thermo scientific)	1

Supplementary Table S. 12: sgRNA:TS_2 ligation into pCas9-TPC:TS_1

Reagent	Volume (µl)
H_2O	4
sgRNA:TS_2 (20.02 ng/µl)	11
Digested and dephosphorylated pCas9-TPC (32.7 ng/µl)	2
Buffer ligase (10 x) (Thermo scientific)	2
T4 ligase (10 U/µl) (Thermo scientific)	1

*vector:insert in a molar 3:1 ratio

Supplementary Table S. 13: DNA isolation by CTAB method

2x CTAB Buffer 1L	
H_2O	460 mL
Tris pH 7,5	200 mL
0,5M EDTA pH 8	40 mL
5M NaCl	280 mL
CTAB	20g

Wash Buffer I 100 mL	
4M Sodium acetate	5 mL
Ethanol	76 mL
H_2O	19 mL

Wash Buffer II 100 mL	
4M Ammonium acetate	0.25 mL
Ethanol	76 mL
H_2O	23,75 mL

Chloroform/isoamyl alcohol (24:1) 250 mL	
Trichlormethane/Chloroform	240 mL
Isoamyl alcohol	10 mL

Supplementary Table S. 14: putative off-target genes found by blast. PAM site of target sequences are underlined, and mismatches between the target sequences and putative off-target sequences are in bold.

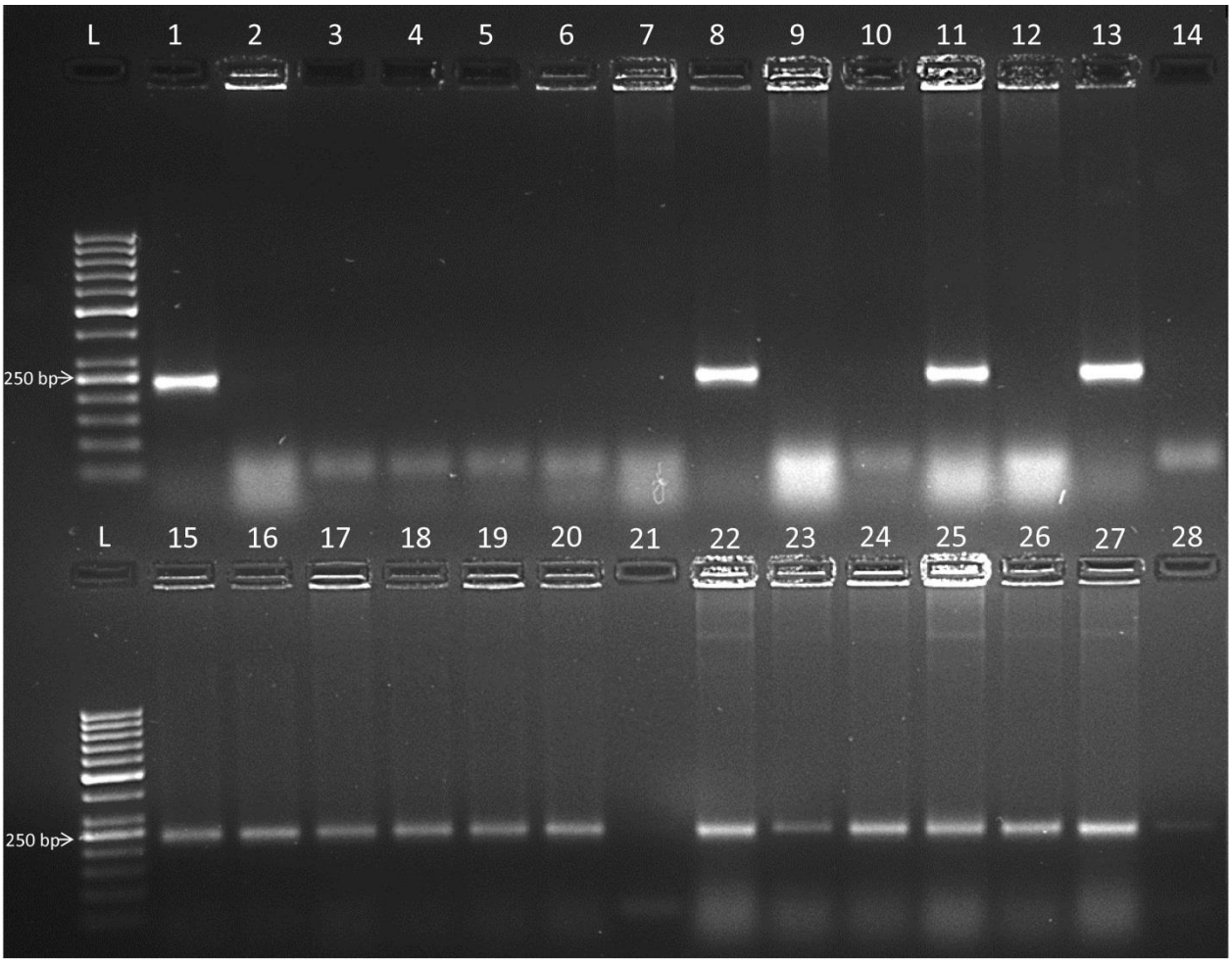
	Target sequence 1 CTTGAGTGCCGTGATACACGGGG	**ID**	**Name**	**Type**
BLAST Match	**GCAT**AA**T**C**TA**GTGATACA**C**GGG**T**	XM_013867872	PREDICTED: Brassica napus putative glycosyltransferase 5	mRNA
	AGAGAGTGC**A**GTGATACAC**CG**C**A**	XM_013860474.1	PREDICTED: Brassica napus NEDD8-specific protease 1-like	mRNA
	GCACTAGAGAGTGATACACGGG**T**	XM_013849607.1	PREDICTED: Brassica napus enoyl-[acyl-carrier-protein] reductase[NADH], chloroplastic-like	mRNA
	TGT**AGA**T**CT**CGTGATACACGG**CC**	XM_013829183.1	PREDICTED: Brassica napus transmembrane	mRNA

		emp24 domain p24beta2	
Target sequence 2 GCTGAACGCCGTGGTACACAGGG	**ID**	**Name**	**Type**
CTTTCTTGC**AT**TGGTACACAGGG	XR_001284971.1	PREDICTED: Brassica napus tubulin beta-2 chain-like (LOC106424884)	misc_RNA (ncRNA)
CAAACTC**T**CC**ATT**GGTACACAGGG	XM_013861135.1	PREDICTED: Brassica napus sucrose transport protein SUC4-like (LOC106420277)	mRNA
CGT**TCG**C**TGTT**TGGTACACAGGG	XM_013850960.1	PREDICTED: Brassica napus auxin response factor 9-like (LOC106410461), transcript variant X3	mRNA
CGT**TCG**C**TGTT**TGGTACACAGGG	XM_013850959.1	PREDICTED: Brassica napus auxin response factor 9-like (LOC106410461), transcript variant X2	mRNA
GGT**ATTAA**C**TA**TGGTACACAGGG	XM_013835765.1	PREDICTED: Brassica napus glucan endo-1,3-beta-glucosidase-like	mRNA
CTTTCTTGCATTGGTACACAGGG	XM_013829838.1	PREDICTED: Brassica napus tubulin beta-7 chain-like	mRNA
CTT**TCCTT**C**AT**TGGTACACAGGG	XM_013815661.1	PREDICTED: Brassica napus tubulin beta-9 chain	mRNA

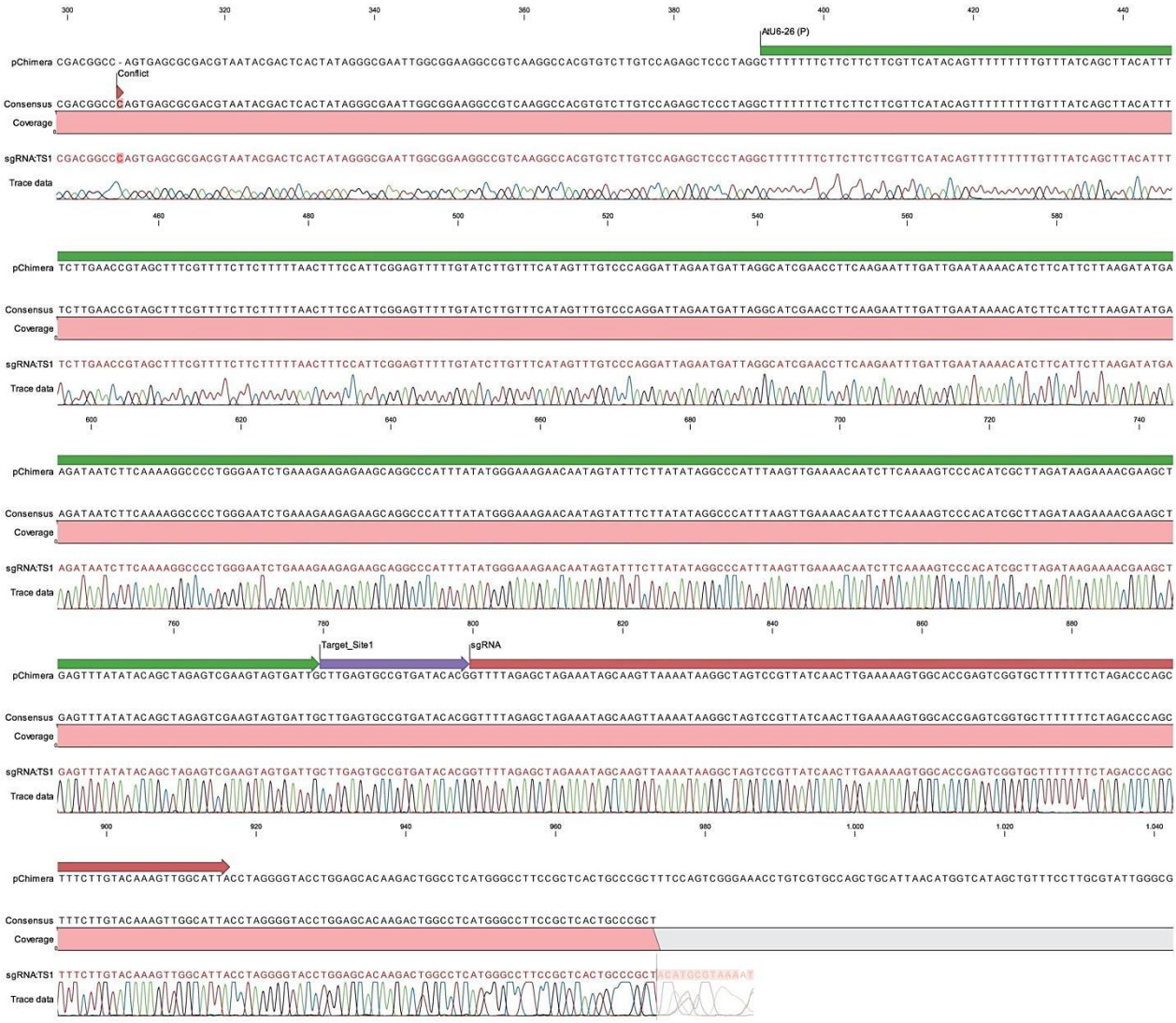
Supplementary Table S. 15: Identity matrix of protein sequences analyzed in this study.

	AtIPK2α	BnC09_ IPK2α	BnA10_ IPK2α	At_IPK2β	BnA09_IPK2β	BnC09_IPK2β	BnA06_IPK2β	BnCnn_IPK2β
AtIPK2α		86.93	87.28	72.98	71.53	70.71	66.79	66.79
BnC09_ IPK2α	86.93		98.94	70.21	69.04	69.29	66.06	65.33
BnA10_ IPK2α	87.28	98.94		70.21	69.04	69.29	66.06	65.69
At_IPK2β	72.98	70.21	70.21		81.12	80.35	78.85	77.78
BnA09_IPK2β	71.53	69.04	69.04	81.12		94.04	78.62	77.90
BnC09_IPK2β	70.71	69.29	69.29	80.35	94.04		77.17	76.45
BnA06_IPK2β	66.79	66.06	66.06	78.85	78.62	77.17		97.51
BnCnn_IPK2β	66.79	65.33	65.69	77.78	77.90	76.45	97.51	

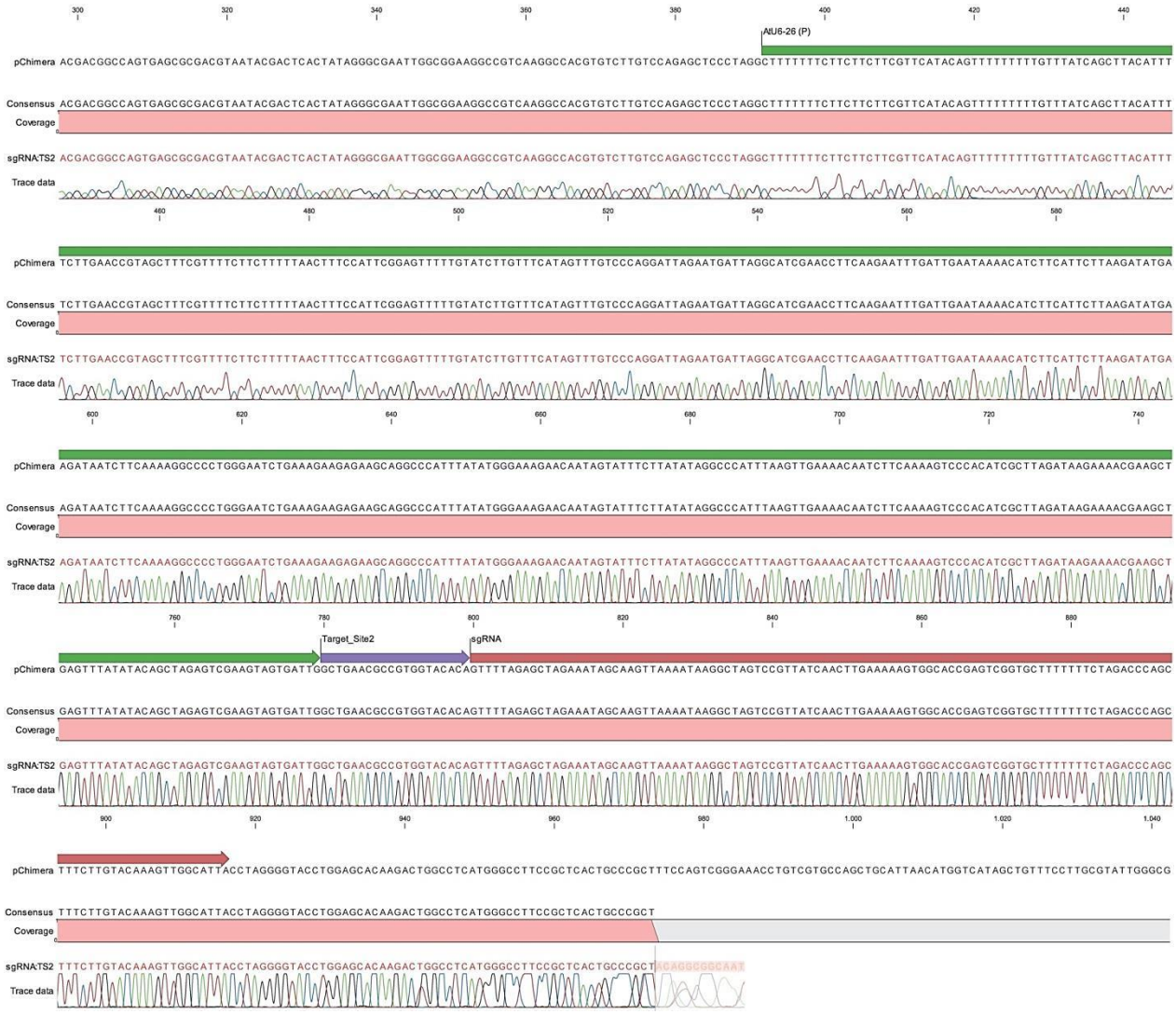
Supplementary Figure S. 1: pChimera colony PCR, where L corresponds to 50 bp ladder, 1 to 13 colonies samples from pChimera:TS_1, 15 to 28 colonies samples from pChimera:TS_2, 14 and 28 are negative controls.



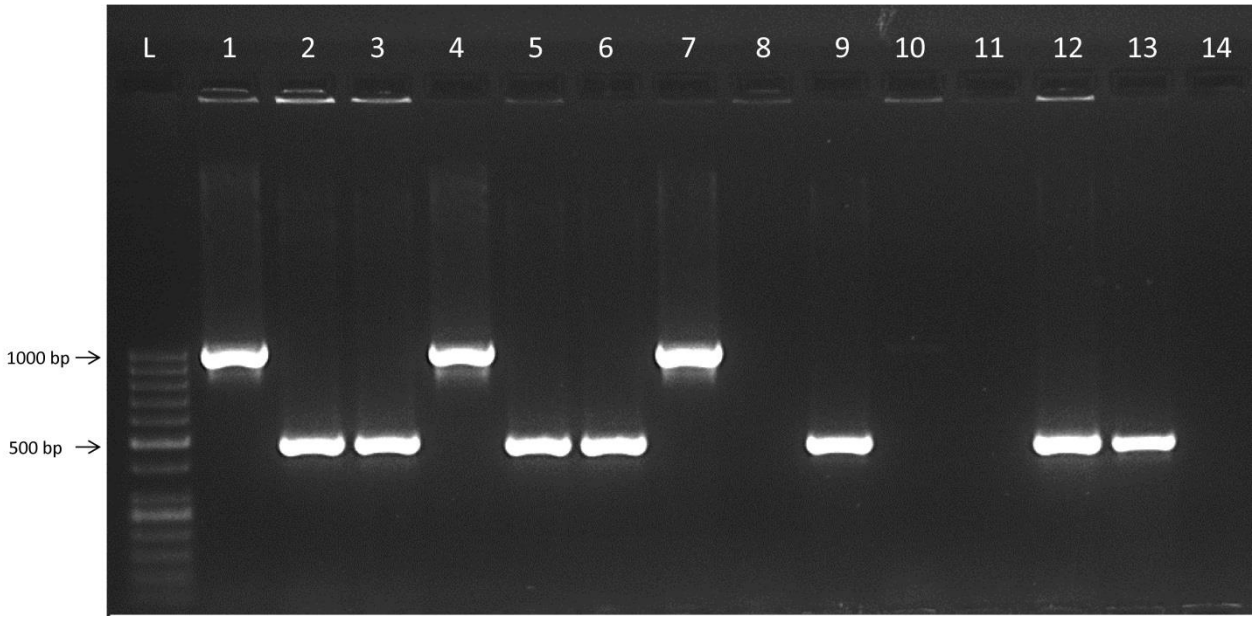
Supplementary Figure S. 2: Sequence alignment of PCR product from pChimera with the sgRNA (TS_1) annealed between the promoter and the sgRNA.



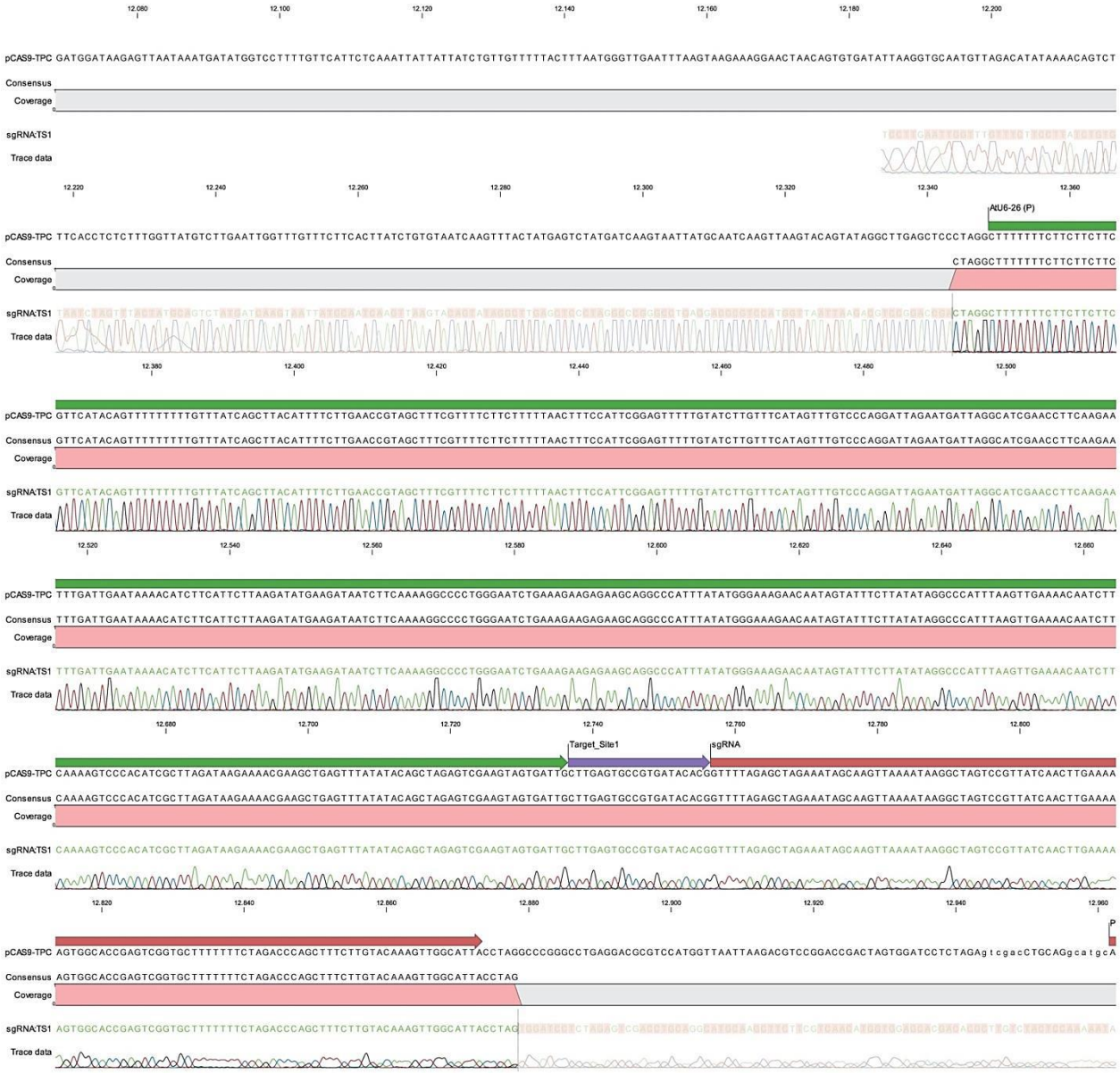
Supplementary Figure S. 3: Sequence alignment of PCR product from pChimera with the sgRNA(TS_2) annealed between the promoter and the sgRNA.



Supplementary Figure S. 4: Colony PCR after cloning sgRNA:TS_1 into pCAS9-TPC, yielding pCAS9-TPC:TS_1. Where L corresponds to 50 bp ladder, 1 to 13 colonies used as a template, and 14 to the negative control.

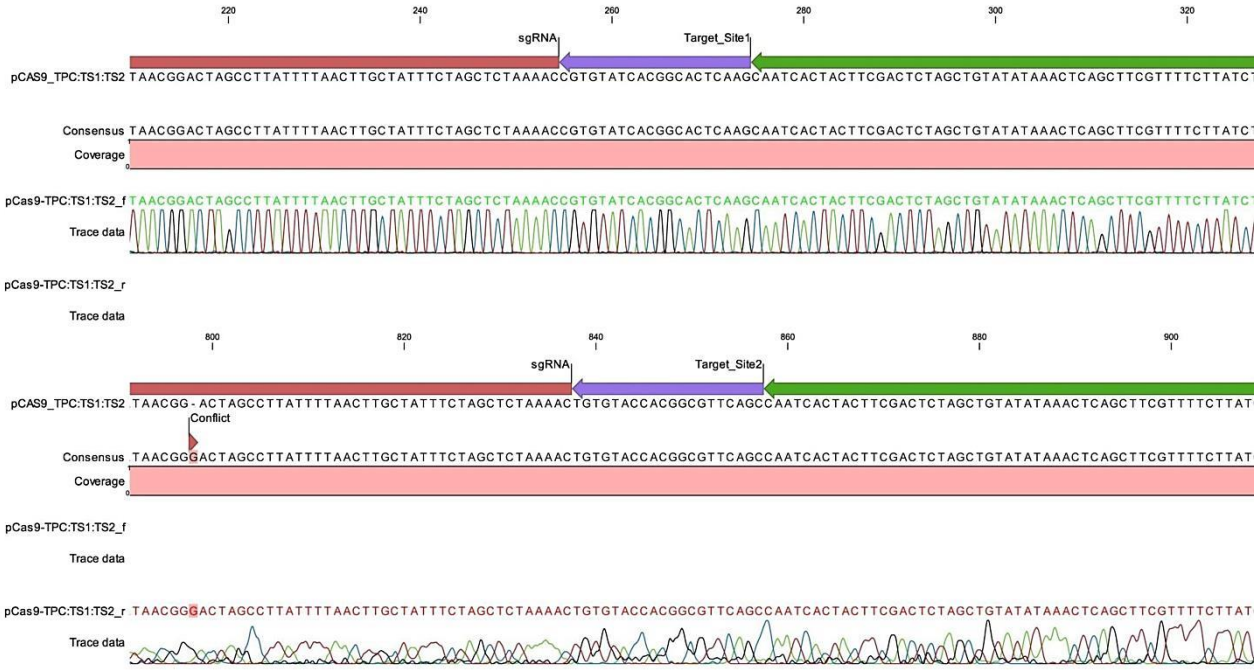


Supplementary Figure S. 5: Sequence alignment of PCR products from pCas9-TPC:TS_1 and of pCas9-TPC:TS_2.



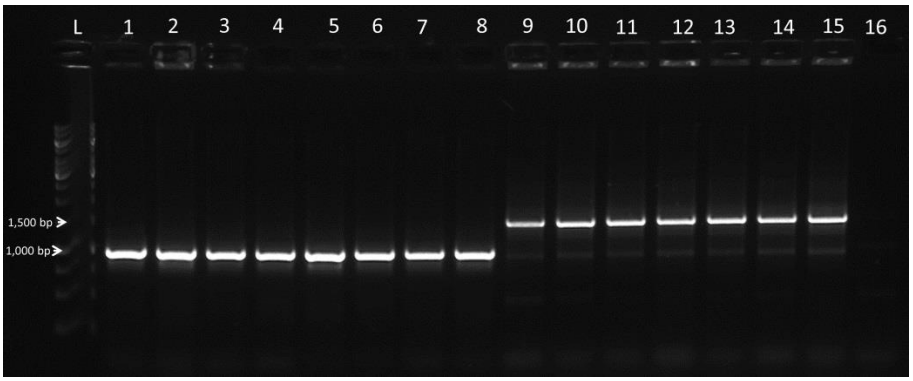

AtU6-26 (P)
pCAS9-TPC:TS2
Consensus
Coverage
sgRNA:TS2
Trace data
Target_Site2
sgRNA
PPT Cassette

Supplementary Figure S. 6: Sequence alignment of vector pCas9-TPC:TS_1:TS_2, where the sgRNAs are shown separately for the ease of understanding.

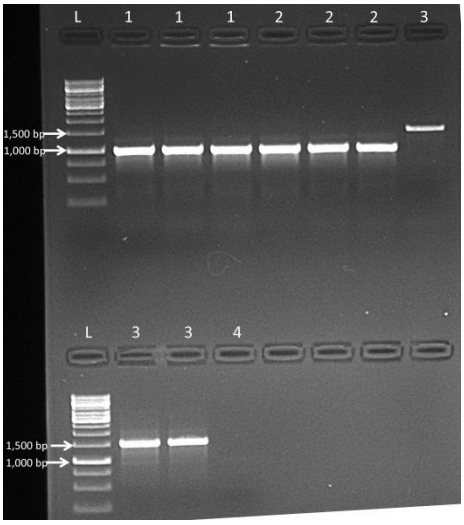


Supplementary Figure S. 7: Sequence alignment of PCR product of pCas9-TPC:TS_1:TS_2 against the reference.

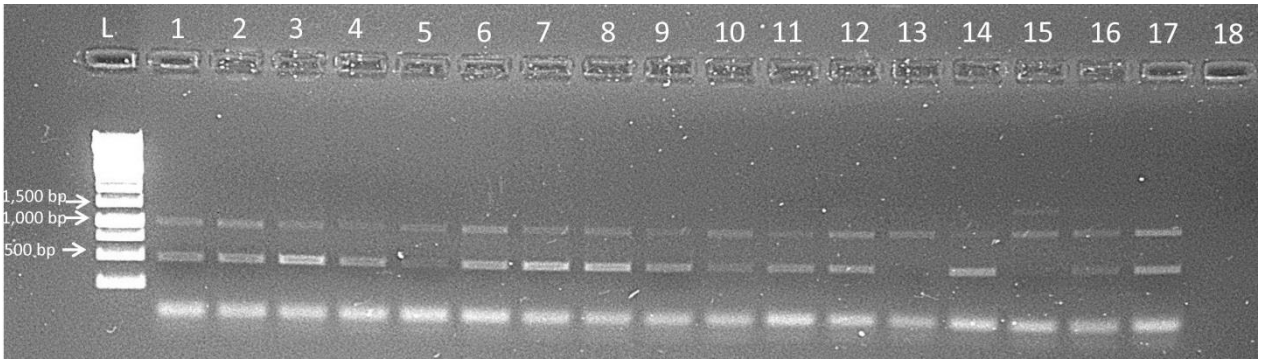
Supplementary Figure S. 8: Cas9 PCR from bacterial stock, where L corresponds to 1 kb ladder, 1 to 3 corresponds to pCas9-TPC:TS_1 samples, 4 to 8 to pCas9-TPC:TS_2 samples, 9 to 15 to pCas9-TPC:TS_1:TS_2, and 16 to the negative control.



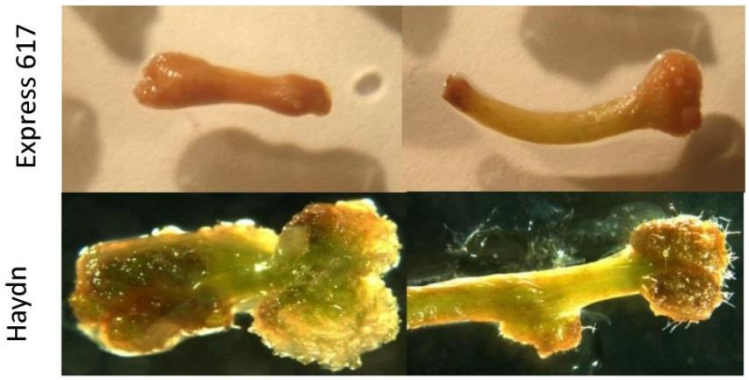
Supplementary Figure S. 9: Colony Cas9 PCR from *A. tumefaciens*, where L corresponds to 1 kb ladder, sample 1 to pCas9-TPC:TS_1, sample 2 to pCas9-TPC:TS_2, and sample 3 to pCas9-TPC:TS_1:TS_2, and sample 4 to the negative control.



Supplementary Figure S. 10: Cas9 PCR using the surrounding medium from tissue culture shoots as a template. Where L corresponds to 1 kb ladder, 1 to 17 corresponds to tissue culture medium samples, and 18 to the negative control.

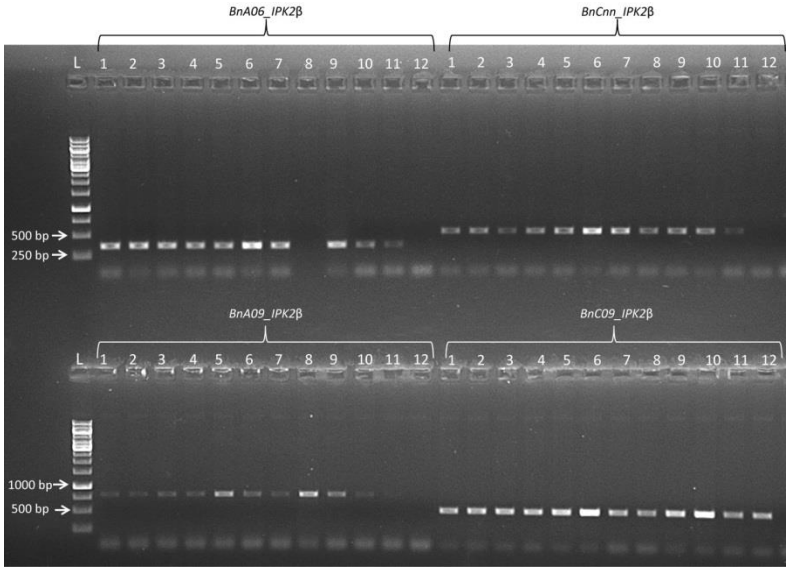


Supplementary Figure S. 11: 4 weeks old hypocotyl explants of Express 617 in regeneration medium O_4, without a selective agent.



Supplementary Figure S. 12: Pink bacteria (black arrows) growing in hypocotyls explants four weeks after *A. tumefaciens* infection.

Supplementary Figure S. 13: Gene editing PCR. Where L corresponds to 1 kb ladder, 1 to 11 to leaf samples, and 12 to the negative control.



10 Declaration

Herewith, I declare that this thesis has been completed independently and unaided and that no other sources other than the ones given here have been used. The submitted written version of this work is the same as the one electronically saved and submitted on CD. Furthermore, I declare that this work has never been submitted at any other time and anywhere else as a final thesis.

Date, Signature