Genome editing and establishment of efficient gene targeting approaches in Arabidopsis using the CRISPR/Cas9 system.

Inaugural-Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Florian Hahn

aus Haan

Wuppertal, Januar 2018

aus dem Institut für Biochemie der Pflanzen
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Berichterstatter:

1. Prof. Dr. Andreas P. M. Weber

2. Prof. Dr. Peter Westhoff

Tag der mündlichen Prüfung: 29.03.2018

EIDESSTATTLICHE ERKLÄRUNG

Ich versichere an Eides Statt, dass die Dissertation von mir selbständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist.

Die Dissertation habe ich in der vorgelegten oder in ähnlicher Form noch bei keiner anderen Institution eingereicht.

Ich habe bisher keine erfolglosen Promotionsversuche unternommen.

Wuppertal, der 27.01.2018

Für meine Familie – Danke für alles.

TABLE OF CONTENTS

PREFACE

This PhD thesis starts with a summary in English and German language, followed by in introduction into photorespiration. The central role of the GLYCINE DECARBOXYLASE system within this essential metabolic pathway is highlighted as well as its role in the evolution of the carbon concentrating mechanisms of C_2 and C_4 plants. In the following, the mechanisms of genome editing using the CRISPR/Cas9 system are explained with an insight into double-strand repair mechanisms in plant cells and a detailed explanation of the origin of CRISPR/Cas9 as a bacterial immune system. The significance of visual markers for exploring this new genome editing technique is indicated. The introduction ends with a description of the aim of this thesis.

The main part of the thesis comprises four manuscripts:
Manuscript I, entitled "An Efficient Visual Screen for CRISPR/Cas9 Activity in *Arabidopsis thaliana*" was published in Frontiers in Plant Science and explores two alternative marker systems for tracking Cas9-mediated gene-editing *in vivo*: *BIALAPHOS RESISTANCE* and *GLABROUS1*. GLABROUS1 is established as visual marker for assessing and optimizing Cas9-mediated gene-editing in *Arabidopsis* (Hahn et al., 2017b).

Manuscript II, entitled "Generation of Targeted Knockout Mutants in *Arabidopsis thaliana* Using CRISPR/Cas9" was published in Bio-Protocol. It presents a detailed cloning strategy for generation of knockout constructs and describes methods to detect Cas9-induced mutations (Hahn et al., 2017a).

Manuscript III, entitled "Homology-directed repair of a defective *glabrous* gene in *Arabidopsis* with Cas9-based gene targeting" describes the comparison of two gene targeting approaches by using the *GLABROUS1* gene for visual detection of mutations. This manuscript was submitted to Frontiers in Plant Science in January 2018 and is deposited at bioRxiv (Hahn et al., 2018).

Manuscript IV, entitled "Towards establishment of a photorespiratory CO_2 pump in *Arabidopsis thaliana* using CRISPR/Cas9-mediated modifications of the *GLYCINE DECARBOXYLASE P-PROTEIN* promoter" describes the application of the CRISPR/Cas9 system for targeted modification of the promoter of the *GLYCINE DECARBOXYLASE P-PROTEIN1* gene in order to selectively exclude this protein from mesophyll cells.

The thesis ends with concluding remarks to summarize the manuscripts.

The following manuscripts derived from the work performed for this thesis:

Hahn, F., Eisenhut, M., Mantegazza, O., and Weber, A. P. M. (2017a). Generation of Targeted Knockout Mutants in *Arabidopsis thaliana* Using CRISPR/Cas9. *BIO-Protoc.* 7. doi:10.21769/BioProtoc.2384.

Hahn, F., Eisenhut, M., Mantegazza, O., and Weber, A. P. M. (2018). Homology-directed repair of a defective *glabrous* gene in *Arabidopsis* with Cas9-based gene targeting. *bioRxiv*, 243675. doi:10.1101/243675.

Hahn, F., Mantegazza, O., Greiner, A., Hegemann, P., Eisenhut, M., and Weber, A. P. M. (2017b). An Efficient Visual Screen for CRISPR/Cas9 Activity in *Arabidopsis thaliana*. *Front. Plant Sci.* 8. doi:10.3389/fpls.2017.00039.

SUMMARY

The CRISPR/Cas9 system has emerged recently as a novel tool for genome editing in plants and other organisms. It relies on the targeted introduction of DNA double-strand breaks by the programmable Cas9 nuclease and subsequent exploitation of the DNA repair machinery of the host cell. Double-strand break repair can be conducted by the error-prone non-homologous end joining pathway, which frequently results in small indel mutations at the site of the double-strand break, thus leading to gene knockouts. Alternatively, a homologous DNA molecule can be used as repair template for a conservative repair by homologous recombination. This approach can be employed for targeted introduction of foreign DNA by adding an artificial repair template. This process is called gene targeting and allows highly specific manipulation of genomes. However, gene targeting frequencies are intrinsically low and require optimization of the repair template delivery.

This thesis established the gene for the trichome master regulator GLABROUS1 (GL1) as visual marker for successful CRISPR/Cas9-based genome editing. Trichome formation can be used as direct visual read out for Cas9-induced mutations. To target the *GL1* gene, a generic vector system was established, which allows easy cloning of any target sequence. Finally, the *GL1* gene was used as marker for comparison of two gene targeting methods. Here, stable gene targeting events could be generated in *Arabidopsis* using an *in planta* gene targeting approach.

Genome editing techniques, such as the CRISPR/Cas9 system, are crucial to develop highly efficient crop varieties in times of a growing world population and loss of arable land. Plants employ photosynthesis to convert atmospheric CO_2 into biomass. This process still harbors potential for optimization. More precisely, photorespiration, which is a salvage pathway for unavoidable O_2 fixation during photosynthesis, is an attractive target for optimization as it releases previously fixed CO_2 and consumes ATP. C_2 plants exploit the photorespiratory pathway as an efficient CO_2 pump by bundle sheath cell specific expression of the *GLYCINE DECARBOXYLASE P-PROTEIN*, thereby minimizing the negative effects of photorespiration. Introduction of this pump into C_3 plants, which represent the majority of today's crop plants, is therefore of interest.

In this thesis, the promoter of the *Arabidopsis GLYCINE DECARBOXYLASE P-PROTEIN1* was modified by genome editing in order to express this protein specifically in bundle sheath cells. The generated mutants accumulated glycine and were slightly delayed in growth. The presented data provide a starting point for further characterization of these mutants and explore a promising path for engineering photorespiration.

ZUSAMMENFASSUNG

Das CRISPR/Cas9-System dient als molekularbiologisches Werkzeug zur gezielten Genveränderung und wird seit seiner Entdeckung vor einigen Jahren zunehmend in Pflanzen und anderen Organismen angewandt. Die Veränderung der genetischen Sequenz basiert auf der gezielten Einführung von DNS-Doppelstrangbrüchen durch die Aktivität der Cas9 Nuklease und der anschließenden Reparatur durch die DNS-Reparaturmaschinerie der Zelle. Doppelstrangbrüche können durch den fehleranfälligen „non-homologous end joining"-Mechanismus repariert werden, der jedoch oft kleine InDel-Mutationen an dem Ort des Doppelstrangbruchs einfügt. Diese Mutationen können zur Genausschaltung führen. Alternativ kann bei der homologen Rekombination ein homologes DNS Molekül als Reparaturmatrize für eine konservative Reparatur genutzt werden. Dieser Mechanismus kann durch die Hinzugabe einer künstlichen Reparaturmatrize zur Einfügung fremder DNS in einen Organismus ausgenutzt werden. Diesen Vorgang, der sehr präzise Genommodifikationen erlaubt, nennt man „Gene targeting". Jedoch sind die „Gene targeting"-Frequenzen sehr niedrig und eine Optimierung der Reparaturmatrizenbereitstellung ist nötig.

In dieser Doktorarbeit wurde das Gen für den Hauptregulator der Trichomentstehung, GLABROUS1 (GL1), als visueller Marker für erfolgreiche Genveränderungen durch das CRISPR/Cas9-System etabliert. Die Bildung von Trichomen kann so zum direkten Auslesen von Cas9-induzierten Mutationen genutzt werden. Um das *GL1* Gen zu mutieren, wurde ein generisch nutzbares Vektorsystem etabliert, welches die einfache Klonierung jeder Zielsequenz ermöglicht. Außerdem wurde das *GL1* Gen als Marker genutzt, um zwei „Gene targeting"-Methoden zu vergleichen. Hierbei konnten mit Hilfe der „*in planta* gene targeting"-Methode stabile „Gene targeting"-Ereignisse in *Arabidopsis* generiert werden.

Manipulationen von Pflanzengenomen durch Methoden wie das CRISPR/Cas9-System sind in Zeiten einer wachsenden Weltbevölkerung wichtig, um hocheffiziente Nutzpflanzen zu erzeugen. Pflanzen nutzen die Photosynthese um CO_2 in Biomasse umzusetzen. Dieser Prozess kann jedoch noch optimiert werden: Um die unvermeidbare O_2-Fixierung während der Photosynthese auszugleichen, nutzen Pflanzen die Photorespiration. Diese ist ein attraktives Ziel zur Optimierung, da hierbei CO_2 freigesetzt und ATP verbraucht wird. C_2-Pflanzen nutzen die Photorespiration als effiziente CO_2-Pumpe durch bündelscheidenzellspezifische Expression des *GLYCINE DECARBOXYLASE P-PROTEIN*, wodurch die negativen Effekte der Photorespiration minimiert werden. Deshalb ist das

Einführen dieser Pumpe in C_3-Pflanzen, die die Mehrzahl der heutigen Nutzpflanzen bilden, von großem Interesse.

In dieser Arbeit wurde der Promoter des *GLYCINE DECARBOXYLASE P-PROTEIN1*-Gens von *Arabidopsis* durch das CRISPR/Cas9-System modifiziert, um dieses Gen spezifisch in Bündelscheidenzellen zu exprimieren. Die erzeugten Mutanten akkumulierten Glycin und waren leicht wachstumsverzögert. Die in dieser Arbeit präsentierten Daten können als Ausgangspunkt für eine weiterführende Charakterisierung der Mutanten genutzt werden. Außerdem zeigen sie einen vielversprechenden Weg auf, die Photorespiration zu verbessern.

INTRODUCTION

A growing world population benefits from improvements in crop yield potential

A growing world population requires a constant increase in crop production. By 2050, today's crop yields have to double to meet the demands for calories and biofuel production, which translate into an annual yield increase of 2.4 %. However, yield increases in major crop plants are far below this threshold in many countries (Ray et al., 2013). Increasing the amount of arable land is one possible solution but involves high environmental costs. Additionally, climate change might actually lower the areas of arable land (Hertel, 2011). A better use of available calories by focusing on plant-based diets is desirable but meat consumption is projected to increase (OECD/FAO, 2017). Therefore, increasing the yield potential, meaning the possible yields per hectare farmland of a given crop, is necessary. Yield potential is based on several factors (Zhu et al., 2010). First, the amount of solar radiation, which reaches the plants, second, the interception capacities of the plant for this radiation, third, the efficiency of the plant to convert the radiation energy into biomass, and fourth, the partitioning of this biomass into harvested plant parts (harvest index). While breeders and plant scientists have improved interception capacities and harvest index close to a theoretical maximum, large improvements are still possible for a more efficient conversion of sunlight into biomass (Long et al., 2015). The major determinant for conversion efficiencies is photosynthesis, which allows the production of carbohydrates from CO_2 and water using solar energy. Understanding and optimizing this process is therefore of high importance.

Photosynthetic efficiencies are lowered by photorespiration, an essential, but wasteful metabolic pathway

Photosynthesis is a two-step process. In the light reactions, light energy is used to transfer electrons between the two photosystems, which ultimately leads to reduction of $NADP^+$ and generation of ATP. ATP and NADPH then power the assimilation of CO_2 in the following Calvin-Benson cycle (CBC) to generate carbohydrates (Johnson, 2016). The hallmark enzyme of this process, RIBULOSE-1,5-BISPHOSPATE CARBOXYLASE/OXYGENASE (RuBisCO), carboxylates ribulose-1,5-bisphosphate (RuBP) yielding two molecules of 3-phosphoglycerate (3-

PGA). However, in about 2 of 7 reactions (Sharkey, 1988; Walker et al., 2016), RuBisCO also catalyzes the addition of atmospheric O_2 to RuBP, leading to generation of one molecule 3-PGA and one molecule of the dead-end metabolite 2-phosphoglycolate (2-PG). 2-PG and its derivatives glyoxylate and glycerate are known inhibitors of enzymes of the CBC and other central carbon pathways, such as PHOSPHOFRUCTOKINASE, FRUCTOSE-1,6-BISPHOSPHATASE, TRIOSE-PHOSPHATE-ISOMERASE and the RuBisCO activation machinery (Kelly and Latzko, 1976; Campbell and Ogren, 1990; Schimkat et al., 1990; Norman and Colman, 1991). Therefore, 2-PG is recycled to 3-PGA via a complex salvage pathway including nine core enzymes and spanning chloroplasts, mitochondria, peroxisomes and the cytosol (Peterhansel et al., 2010). This pathway is called photorespiration (PR) or oxidative C_2 cycle. PR allows photosynthesis in an oxygen-containing environment and is therefore essential for all oxygenic photosynthetic organisms (Eisenhut et al., 2008; Bauwe, 2010; Rademacher et al., 2016). PR is also assumed to play a role in photoprotection under high light and drought stress, as it reoxidizes electron acceptors of the photosynthetic electron transport chain (Kozaki and Takeba, 1996; Guan et al., 2004; Bai et al., 2008; Lima Neto et al., 2017). Additionally, a positive correlation between photorespiration and nitrogen uptake was detected in wheat and *Arabidopsis* (Rachmilevitch et al., 2004; Bloom et al., 2010). However, PR is also a highly wasteful process, as it releases previously fixed CO_2 and NH_3, which must be reassimilated. This leads to a high investment of reducing equivalents and ATP (12.25 ATP per oxygenation reaction, Peterhansel et al., 2010). PR represents the second highest flux of metabolites in plants directly after photosynthesis. The wasteful nature of this process imposes therefore economical damage to farmers as shown by calculation of US wheat and soybean productions. Here, a decrease in production by 20% and 36%, respectively, can be accounted by PR, translating in a loss of 148 trillion calories. Even slight reductions in photorespiratory losses of 5% by enhancing the PR efficiency would therefore lead to an increase in wheat and soybean yield worth 540 million international dollars in the US alone (Walker et al., 2016). This underlines that PR is a major target for crop improvement.

The glycine decarboxylase system is the central enzyme of photorespiration

The photorespiratory pathway in plants (Fig. 1) starts with the hydrolyzation of 2-PG to glycolate by 2-PHOSPHOGLYCOLATE PHOSPHATASE in the chloroplasts (Somerville and Ogren, 1979; Schwarte and Bauwe, 2007). Glycolate is shuttled to the peroxisomes, where it is converted

by GLYCOLATE OXIDASE to glyoxylate. This reaction releases H_2O_2 that is detoxified by peroxisomal CATALASE2 (Queval et al., 2007; Dellero et al., 2016). GLUTAMATE:GLYOXYLATE AMINOTRANSFERASE transaminates glyoxylate to glycine, which is transported into the mitochondria (Igarashi et al., 2003). Here, the combined activity of the multienzyme system GLYCINE DECARBOXYLASE (GDC) combined with SERINE HYDROXYMETHYLTRANSFERASE (SHMT) converts two molecules of glycine into one molecule of serine, thereby releasing NH_3 and CO_2 (Voll et al., 2006; Engel et al., 2007, 2011). NH_3 is reassimilated to glutamate in the chloroplast by the concerted action of GLUTAMINE SYNTHETASE and FERREDOXIN-DEPENDENT GLUTAMINE:OXOGLUTARATE AMINOTRANSFERASE (Keys, 2006). Serine is shuttled back to the peroxisomes, where it is deaminated by SERINE:GLYOXYLATE AMINOTRANSFERASE, yielding one molecule 3-hydroxypyruvate and one molecule of glycine. Glycine is shuttled back to the mitochondria to fuel the GDC/SHMT reaction (Liepman and Olsen, 2001). 3-Hydroxypyruvate is reduced to glycerate by a cytosolic or peroxisomal 3-HYDROXYPYRUVATE REDUCTASE (Givan and Kleczkowski, 1992; Timm et al., 2008). The last step of PR takes place in the chloroplast, where glycerate is phosphorylated by D-GLYCERATE 3-KINASE to 3-phosphoglycerate, which can reenter the CBC (Boldt et al., 2005). Next to these core enzymes, PR also requires many transport proteins of which only a few are characterized so far (Eisenhut et al., 2013; Pick et al., 2013).

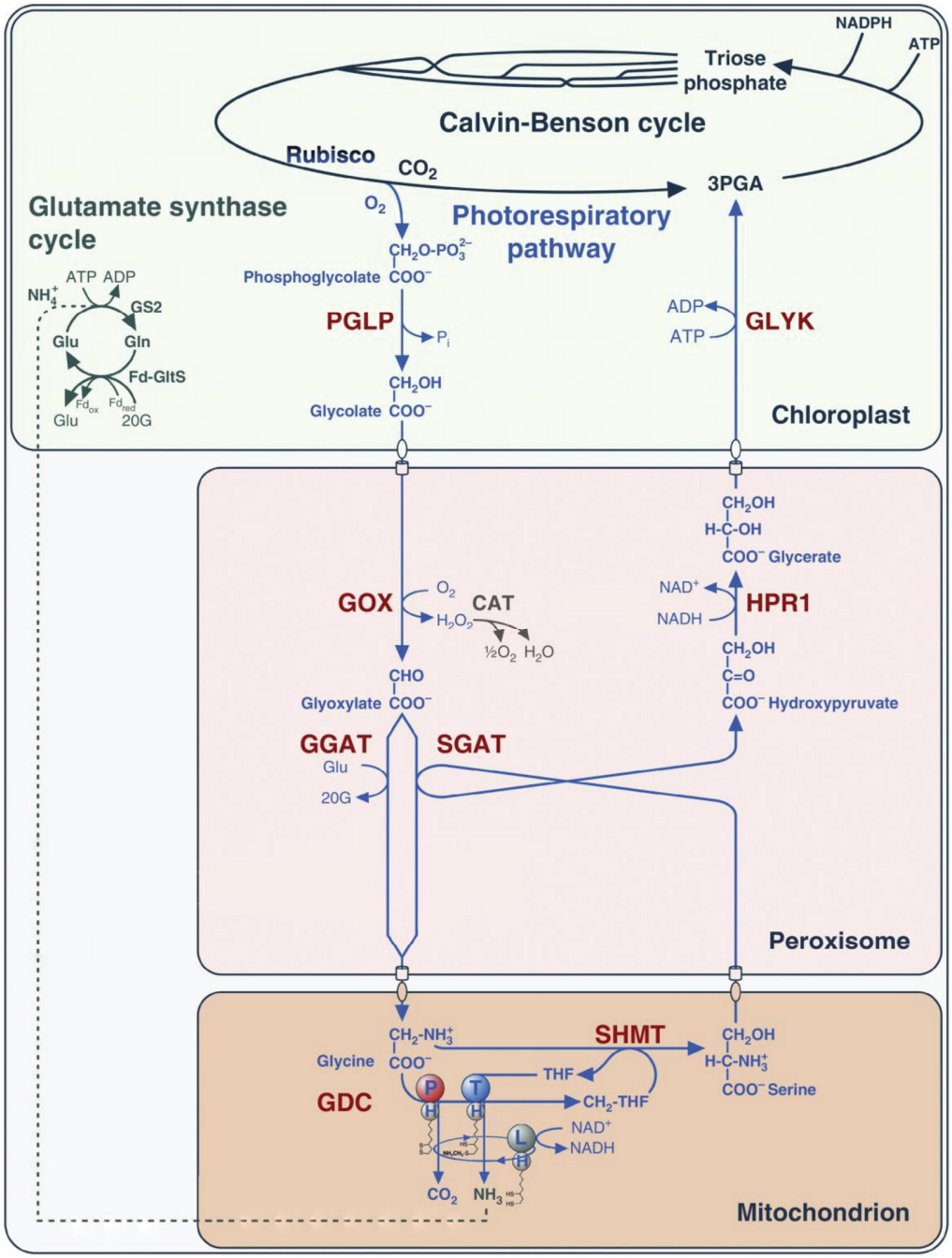


Figure 1: The photorespiratory pathway in higher plants. 2OG: 2-oxoglutarate; 3PGA: 3-phosphoglycerate; Asn: asparagine; CAT: CATALASE; FD-GltS: FERREDOXIN-DEPENDENT GLUTAMINE:OXOGLUTARATE AMINOTRANSFERASE; GDC: GLYCINE DECARBOXYLASE; GGAT: GLUTAMATE:GLYOXYLATE AMINOTRANSFERASE; Gln: glutamine; Glu: glutamate; GLYK: GLYCERATE 3-KINASE; GOX: GLYCOLATE OXIDASE; GS2: GLUTAMINE SYNTHETASE; HPR1: HYDROXYPYRUVATE REDUCTASE; PGLP: PHOSPHOGLYCOLATE PHOSPHATASE; Rubisco: RIBULOSE 1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE; SGAT: SERINE:GLYOXYLATE AMINOTRANSFERASE;

SHMT, SERINE HYDROXYMETHYLTRANSFERASE. Cytosolic HPR2 is not depicted. Figure modified from Hagemann and Bauwe (2016).

The GDC has a central role in photorespiration, as it is responsible for CO_2 loss and the release of NH_3. However, the GDC/SHMT system must be involved in further metabolic pathways as knockout mutants of the GDC/SHMT system are the only photorespiratory mutants that are lethal even under non-photorespiratory conditions (Engel et al., 2007, 2011; Timm and Bauwe, 2013; Timm et al., 2017). This is mainly due to the provision of important metabolites by GDC, such as glycine and serine, which are necessary for protein biosynthesis, phospholipid generation and purine generation (Mouillon et al., 1999). Furthermore, the two proteins are key player of the C_1 metabolism, as they provide one-carbon units in form of 5,10-methylene-tetrahydrofolate, which are necessary for basic cellular processes such as protein synthesis and nucleic acid synthesis (Hanson and Roje, 2001). This additional role next to photorespiration is the major reason that the GDC is not only found in phototropic organisms but essentially in all eukaryotes and also bacteria (Schulze et al., 2016). The GDC consists of four different subunits, namely the H-protein, the L-Protein, the T-protein, and the P-protein in a subunit stoichiometry of 27:2:9:4, respectively. Interaction studies suggested that the H-protein forms the central core of the system, with the other subunits solely binding to it. This complex is very fragile and only stable due to the very high concentrations of the subunits as found in the mitochondrial matrix of leaves, where it comprises about 1/3 of the soluble mitochondrial matrix proteins (Oliver et al., 1990). This high abundance of the protein allows the plant to react immediately to higher photorespiratory fluxes, e.g. caused by stomata closure during drought stress, and avoids therefore toxic glycine accumulation (Heineke et al., 2001; Eisenhut et al., 2007). The reaction scheme of the GDC is depictured in Fig. 2. Glycine binds to the homodimeric P-protein (GLDP) of GDC by forming a Schiff base with pyridoxal phosphate, the prosthetic group of GLDP. Decarboxylation by GLDP leads to a release of CO_2. The H-Protein (GLDH) contains a covalently bound lipoate group and acts as a shuttle between the other subunits of GDC. The methylamine, which is generated by decarboxylation of the glycine can bind to the distal sulfur residue of the GLDH lipoate. GLDH stabilizes the NH_3^+ -residue of the methylamine by interaction with various amino acid residues to avoid non-enzymatic release of NH_3 and formaldehyde (Pares et al., 1994). The protective interaction is disturbed by the T-Protein (GLDT). This leads to a release of the methylamine and is accompanied by a nucleophilic attack by tetrahydrofolate (THF), the cofactor of GLDT, to the methyl residue (Guilhaudis et al., 2000; Douce et al., 2001). This step releases NH_3, which is

refixated in the chloroplasts, and provides methylene-THF for the formation of serine by SHMT. The homodimeric L-protein (GLDL) is a dihydrolipoamide dehydrogenase and part of various multienzyme complexes, such as pyruvate dehydrogenase. It contains a tightly bound FAD cofactor as well as a redox disulfide. This enables the transfer of electrons from the lipoate group of GLDH to NAD^+, thereby reoxidizing the lipoate group for a new round of glycine decarboxylation (Bourguignon et al., 1992, 1996; Douce et al., 2001). Reoxidation of NADH might then be mediated by the mitochondrial electron chain or by the HPR proteins after export from the mitochondria (Douce et al., 2001; Bykova et al., 2005).

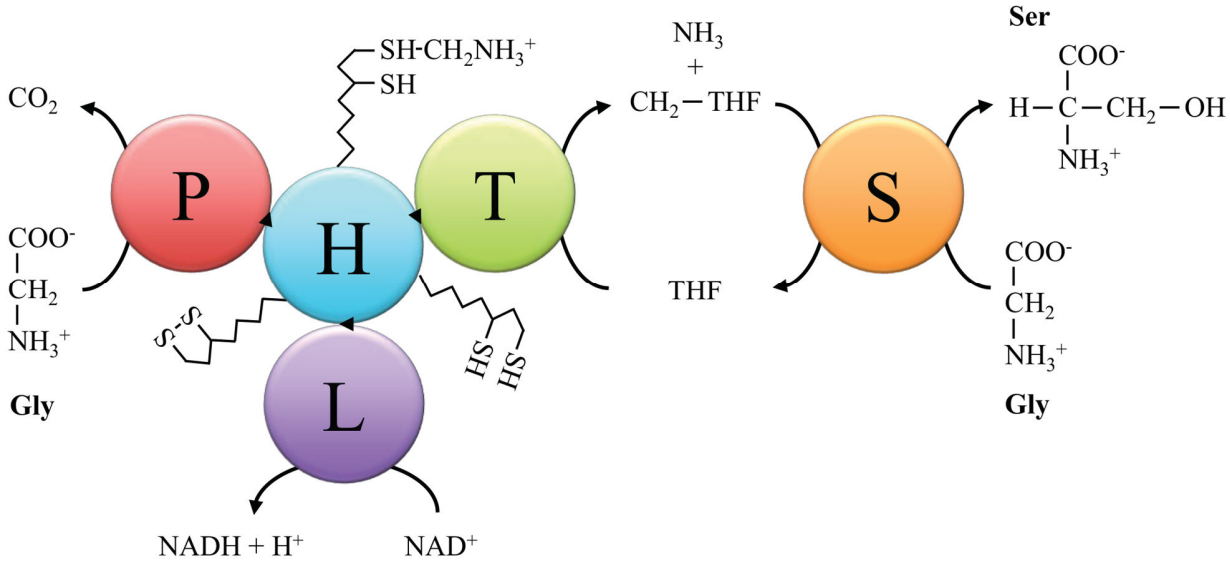


Figure 2: Reaction scheme of GDC/SHMT. Gly: Glycine; H: H-protein; L: L-Protein; P: P-Protein; S: SERINE HYDROXYMETHYLTRANSFERASE; Ser: Serine; T: T-Protein; THF: Tetrahydrofolate. Figure adapted from Engel (2010) and Hasse et al. (2013).

Carbon losses from photorespiration can be reduced by carbon concentrating mechanisms

The wasteful photorespiratory pathway limits the photosynthetic capacity of plants. Therefore, plants have evolved mechanisms to concentrate CO_2 within their cells to suppress the initial oxygenation reaction of RuBisCO. Some C_3 plants, such as rice, show specific subcellular localizations of their organelles with their chloroplasts forming a continuous layer around the outer periphery of the mesophyll cells. Additional chloroplastidic extrusions allow trapping of CO_2, which is released by GDC activity in mitochondria in the inner regions of the cell (Sage and Sage, 2009; Sage and Khoshravesh, 2016; Stata et al., 2016). In some tissues, such as woody stems or green fruits, respiratory CO_2 is trapped to high concentrations (>1,000 ppm) due to lack

of stomata and diffusional barriers and can be efficiently fixed by chloroplasts (Sage and Khoshravesh, 2016).

Next to these passive CO_2 concentrating mechanisms (CCM), evolution of the active CCM C_4 photosynthesis 25-30 Mya allowed adaption to a drastic decrease in atmospheric CO_2 levels and climate condition changes favoring dry environments (Sage et al., 2012). The term C_4 photosynthesis describes a number of different pathways, which share a primary CO_2 fixation step by hydration of CO_2 to HCO_3^- and subsequent fixation by PHOSPHOENOLPYRUVATE CARBOXYLASE (PEPC) into the four-carbon compound oxaloacetate (Sage et al., 2012). In most C_4 plants, primary CO_2 fixation by PEPC takes place in the mesophyll cells (M cells). Oxaloacetate is then converted to malate or aspartate and transferred to the bundle sheath cells (BS cells). In the BS cells, a decarboxylation step releases CO_2 and the resulting three-carbon compound is shuttled back to the M cells (Fig. 3). The concerted decarboxylation within the BS cells leads to a 10- to 30-fold enrichment of CO_2 in these cells (Sage, 2013). As RuBisCO activity is restricted to BS cells in C_4 plants, its oxygenation reaction is repressed and PR highly reduced compared to C_4 plants. Under hot and dry conditions, the lowered costs of PR compensate for the enhanced ATP consumption of the C_4 pathway due to PEP regeneration (Sage, 2004; Slack and Hatch, 1967). This translates into 40-50 % higher growth rates of C_4 crops compared to C_3 crops (Monteith, 1978). Engineering the C_4 pathway into C_3 crops is therefore a possible way for generating yield increases (Hibberd et al., 2008). Understanding the distinct steps that led to the evolution of C_4 photosynthesis is therefore crucial.

Studying C_3-C_4 intermediate plants as found in the genus *Flaveria* and *Cleome* amongst others allowed the assignment of distinct steps, which are necessary to integrate a C_4 cycle (Sage, 2004; Sage et al., 2012, 2014). Next to anatomical changes to increase vein density and BS cells size, the initial biochemical change towards C_4 photosynthesis was the establishment of a photorespiratory glycine shuttle (Fig. 3; Sage, 2004; Heckmann et al., 2013; Mallmann et al., 2014). By a relative simple relocation of a single protein, plants could turn the wasteful photorespiratory process into an efficient carbon pump: Plants employing this preliminary CCM lost the GDC activity in the M cells and show BS cells specific expression of *GLDP* and partly also the other components of the glycine cleavage system (Schulze et al., 2016). During photorespiration, glycine, produced in the M cells, needs to diffuse to BS cells to complete the photorespiratory cycle (Hylton et al., 1988; Rawsthorne et al., 1988). Restricted decarboxylation of glycine in the BS cells leads to a 3-fold enhancement of CO_2 concentration in plastids of leaf

cells of C_2 plants compared to C_3 plants (Keerberg et al., 2014). This suppresses the oxygenation reaction of RuBisCO in these cells, leading to more effective photosynthesis under photorespiration-promoting conditions as shown by lowered CO_2 compensation points compared to C_3 plants (Vogan and Sage, 2012).

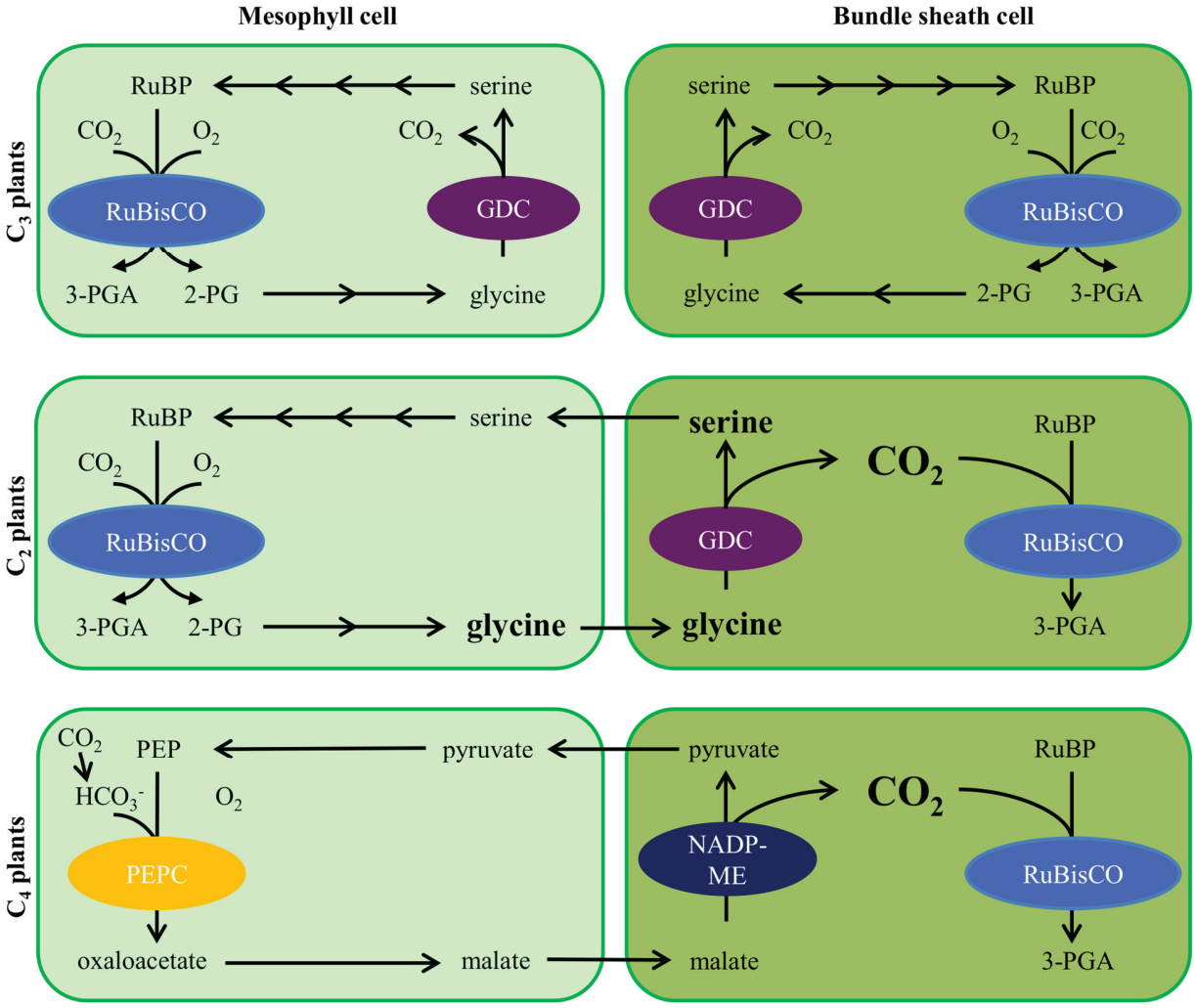


Figure 3: Comparison of carbon fixation mechanisms in C_3, C_2, and C_4 plants. Top: C_3 plants have a fully functional CBC and photorespiratory cycle in M cells and BS cells. Middle: C_2 plants have lost GDC activity in M cells, photorespiratory glycine accumulates and diffuses to BS cells where decarboxylation leads to CO_2 enrichment. This suppresses the oxygenation reaction of RuBisCO in BS cells. Bottom: C_4 plants (here NADP-ME subtype) fix CO_2 in the M cells via PEPC, the resulting C_4 acid is shuttled to BS cells where decarboxylation leads to an enrichment of CO_2. Only the BS cells harbor a fully functional CBC. 2-PG: 2-phosphoglycolate; 3-PGA: 3-phosphoglycerate; GDC: GLYCINE DECARBOXYLASE; NADP-ME: NADP-MALIC ENZYME; PEPC: PHOSPHOENOLPYRUVATE-CARBOXYLASE; RuBisCO: RIBULOSE-1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE; RuBP: Ribulose-1,5-bisphosphate.

In the GDC reaction, two molecules of M cell-derived glycine are converted into one molecule of serine, which is shuttled back to the BS cells. This leads to a nitrogen imbalance between M cells and BS cells that needs to be rebalanced. Modelling results indicate that the energetically most favorable N-rebalancing pathways highly resemble basic C_4 cycles. This implies a high evolutionary pressure towards C_4 PS in these intermediate plants (Heckmann et al., 2013; Mallmann et al., 2014; Schlüter and Weber, 2016). Still, numerous species (e.g. in the genera *Moricandia* and *Diplotaxis*) have been detected, which employ the glycine shuttle but are not related to C_4 species, indicating that the glycine shuttle is not always an intermediate step towards C_4 photosynthesis but can already be advantageous on its own. As the two-carbon compound 2-PG is the primary product of the photorespiratory cycle, the name C_2 photosynthesis has been established recently to describe this CCM (Vogan et al., 2007). To date, 40 species in 21 lineages have been described as C_2 plants, which is far below the number of known C_4 species (~8100). This can partially be explained by the reason that detection of C_2 photosynthesis is more difficult than C_4 photosynthesis. Rapid carbon isotope screens, which are typically used to identify C_4 plants, allow large-scale screens of photosynthetic properties, but are not amenable for detection of C_2 plants. Additionally, the evolutionary pressure towards C_4 photosynthesis might explain the low abundance of C_2 plants (Vogan and Sage, 2012; Sage, 2017). Next to the compartmentalization of GDC activity, most C_2 species share common physiological features, such as decreased M cell layers between veins and enlarged BS cells with larger and/or more mitochondria. Additionally, the mitochondria are in close association with chloroplasts. C_2 photosynthesis can thereby refix up to 90% of the photorespiratory CO_2 (Sage et al., 2012).This allows the plants to move to warmer areas than their C_3 relatives as well as to areas with elongated periods of drought (Lundgren and Christin, 2017).

C_2 and C_4 photosynthesis recruit genes, which are already present in C_3 plants albeit with different expression patterns or strength (Aubry et al., 2011). A full C_4 cycle requires differential regulation of up to several thousands of genes (Bräutigam et al., 2011; Gowik et al., 2011; Lauterbach et al., 2017). Compared to a full C_4 cycle, the C_2 photorespiratory pump requires less transcriptional changes and might be easier to implement into C_3 plants like *Arabidopsis* or major crop plants. Deciphering the detailed mechanisms underlying the path to C_2 photosynthesis is therefore of high biotechnological interest (Winzer et al., 2001; Bauwe and Kolukisaoglu, 2003; Engel, 2010; Lin et al., 2016; Schulze et al., 2016; Schlüter et al., 2017).

Variations in *cis*-regulatory elements in the *GLDP* promoter explain evolution of C_2 photosynthesis

Understanding the regulatory elements that lead to a BS cell specific accumulation of the GDC is crucial as this is the driving force of C_2 photosynthesis. Cell specific accumulation of proteins can be mediated on transcript level by *cis*-elements and *trans*-factors but also via posttranscriptional processes or changes in translation efficiencies (Hibberd and Covshoff, 2010). Detailed promoter studies from differentially expressed C_4 genes revealed diverse regulatory elements within all segments of genes, namely in the promoter regions (Gowik et al., 2004; Akyildiz et al., 2007; Engelmann et al., 2008; Wiludda et al., 2012; Adwy et al., 2015; Gowik et al., 2017), the untranslated regions (Kajala et al., 2012; Williams et al., 2016), or the exons (Brown et al., 2011). Detailed studies have assessed the influence of promoter elements for BS cell specific expression of *GLDP*, the decarboxylating enzyme of GDC. The *GLDPA* promoter of the C_4 plant *Flaveria trinervia* shows a complex interplay of transcriptional and posttranscriptional regulation. It consists of a tandem promoter with a strong general leaf promoter element at the distal region of the promoter and a weak BS cells/vasculature specific promoter element proximal to the coding region. Expression strength of the BS cell specific promoter element is however increased by transcriptional enhancer modules in other regions of the promoter. In contrast, the activity of the unspecific distal promoter is mostly repressed post-transcriptionally by triggering of nonsense-mediated RNA decay due to inefficient splicing sites and upstream open reading frames within the transcript. An additional layer of repression might be present due to sterical hindrance of the transcription machinery by polymerases binding to the proximal promoter. This leads to a dominant occurrence of *FtGLDPA* transcripts in BS cells with leaky expression in M cells to fuel C_1 metabolism (Engelmann et al., 2008; Wiludda et al., 2012). Interestingly, the complex regulatory system of the *FtGLDPA* promoter is also recognized in C_3 plants as shown in transgenic *Arabidopsis* plants (Engelmann et al., 2008). It was demonstrated that the C_3 species *Flaveria robusta* and *Flaveria pringlei* already contain the BS cell specific expressed *GLDPA* next to a general chlorenchyma specific *GLDPB*. Gradual upregulation of *GLDPA* and simultaneous downregulation of *GLDPB* led finally to the establishment of a glycine shuttle (Schulze et al., 2013).

The C_3 plant *Arabidopsis thaliana* contains two copies of *GLDP*, *AtGLDP1* and *AtTGLDP2*. In contrast to the *GLDPA* gene from *Flaveria*, both are ubiquitously expressed in all plant tissues

(Adwy et al., 2015). Detailed promoter analyses revealed two distinct motives in the *AtGLDP* promoters, which are highly conserved within the *Brassicaceae*. The V-Box confers expression in vasculature and bundle sheath cells, the M-Box confers expression in mesophyll cells. Interestingly, the C_2 *Brassicaceae Moricandia nitens*, which shows a BS cell specific expression of *GLDP*, lacks the M-Box motif in the *GLDP* promoter (Rawsthorne et al., 1998; Adwy et al., 2015). Additionally, phylogenetic analyses revealed that all *Moricandia* species lost the *GLDP2* copy during evolution (Schlüter et al., 2017). Therefore, establishment of C_2 photosynthesis in *Brassicaceae* probably was preceded by loss of one ubiquitously expressed *GLDP* copy and then enforced by deletion of the M-Box motif in the promoter of the other *GLDP* copy (Schlüter et al., 2017).

Both examples for cell-type specific expression of *GLDP* provide simple mechanisms towards first steps of establishing a photorespiratory pump in C_3 plants, either by transferring a BS cell specific promoter from a C_2 plant into a C_3 plant or by deleting a *cis*-element in the endogenous promoter. However, traditional breeding methods cannot introduce such precise changes. Therefore, more specific methods are necessary to achieve these genomic modifications.

Establishment of new methods is necessary for targeted genomic modifications

Increasing genetic variation in plant genomes by induction of mutations is a key concept for introduction of beneficial traits that cannot be introgressed by breeding. Early methods to manipulate plant genomes included physical mutagenesis by ionizing radiation, chemical mutagenesis by alkylating agents, such as ethyl methanesulfonate, but also introduction of mutations and new genes via biological methods such as T-DNAs or transposons. However, the mutagenic outcome of all these techniques is random and often induces manifold mutations that have to be removed by backcrossing (Arora and Narula, 2017). One major source for mutations in genomes are double-strand breaks (DSB). DSB are highly problematic for cells as large chromosomal fragments can be lost during the division process (Waterworth et al., 2011). All living organisms have therefore evolved sophisticated repair mechanisms for DSB. Two major repair mechanisms compete in plant cells for repair of DSB: the non-homologous end joining pathway (NHEJ) induces direct relegation of the DNA ends while homologous recombination (HR) depends on the presence of homologous sequences, which are used as template for repair of the DSB. However, these repair pathways can also be exploited to introduce targeted changes at the DSB site (Pacher and Puchta, 2017). To allow efficient and specific modifications of

genomes, two concepts need to be mastered: Understanding the mechanisms underlying DSB repair pathways and targeted introduction of DSB in the DNA.

DNA double strand break are repaired by two distinct repair pathways, which can be exploited for mutagenesis

DSB are detected by the MRE11-RAD50-NBS1 (MRN) complex. This complex then recruits the protein kinase ATAXIA TELANGIECTASIA MUTATED, which activates downstream proteins involved in cell cycle arrest and DNA repair (Mannuss et al., 2012; Manova and Gruszka, 2015). The most common repair pathway in somatic cells of higher eukaryotes is NHEJ (Sargent et al., 1997; Puchta, 2005). Understanding the repair pathways in plant somatic cells is crucial, as plant germline cells arise from somatic cells in the meristem, so mutations in these somatic cells can also be passed on to the progeny (Walbot, 1996; Puchta, 2005). In the classical NHEJ (cNHEJ) pathway, the DSB ends are recognized by the ring-like KU70/KU80 heterodimer which protects the DNA against degradation and results in tethering of the free ends (Cary et al., 1997; Walker et al., 2001). KU70/KU80 recruit further proteins, which restore hydroxyl- and phosphogroups. Religation of the DNA ends is then mediated by DNA LIGASE 4 (LIG4) and associated proteins such as X-RAY REPAIR CROSS-COMPLEMENTING PROTEIN 4 (XRCC4, West et al., 2000; Manova and Gruszka, 2015). Principally, NHEJ can join any given DNA ends irrespective of their origin. Additionally, minimal processing of the DNA ends prior to ligation, which can include removal and integration of nucleotides makes NHEJ an error-prone process (Fig. 4A), which can result in random small insertion or deletion (indel) mutations at the DSB site (Gorbunova and Levy, 1997; Salomon and Puchta, 1998; Dueva and Iliakis, 2013; Bétermier et al., 2014; Fauser et al., 2014). These indel mutations can interrupt open reading frames and knock-out gene functions (Pacher and Puchta, 2017). Next to the cNHEJ pathway, alternative NHEJ (aNHEJ) pathways have been described, which are independent of the KU-heterodimer and LIG4 but require other factors such as POLY (ADP-RIBOSE) POLYMERASE I (PARP1) and XRCC1 (Dueva and Iliakis, 2013). A prominent example of these pathways employs small microhomologies of 1-16 nucleotides for repair (Fig. 4B) and is therefore also called microhomology-mediated end joining (MMEJ). Here, the DSB ends are recognized by PARP1, which might hinder processing of the DSB via cNHEJ (Jia et al., 2013). In the following, the 5'-ends of the DSB ends are trimmed by the MRN complex and other nucleases leading to ssDNA overhangs. Microhomologies within the ssDNA ends anneal and the remaining ssDNA regions

are filled up by polymerases. The 3'-flaps are removed by exonucleases. LIG1 or LIG3 ligates the break ends in mammals (the corresponding plant ligase is still uncharacterized) with the help of the scaffolding protein XRCC1 (Charbonnel et al., 2010; Sfeir and Symington, 2015; Wang and Xu, 2017). MMEJ creates large deletions and is therefore highly mutagenic (Puchta and Fauser, 2014; Shen et al., 2016; Collonnier et al., 2017).

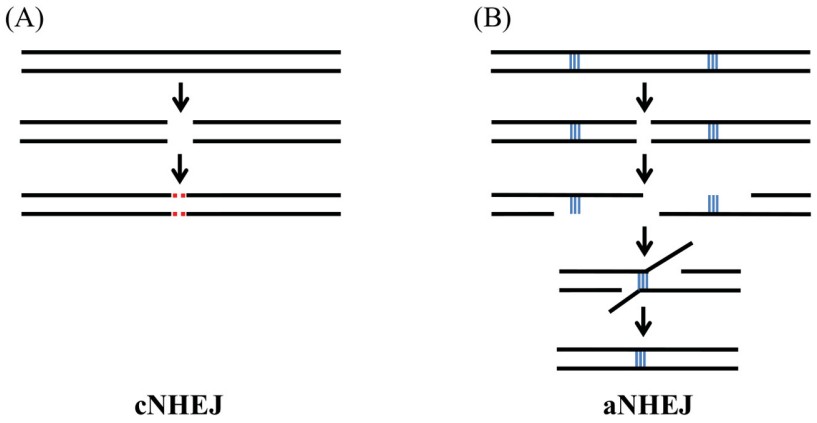


Figure 4: NHEJ-mediated repair pathways in cells. (A) The classical NHEJ pathway induces direct religation of DSB ends. Minimal processing prior to ligation results in small indel mutations (red). (B) Microhomologies (blue) in the DSB ends induce alternative NHEJ pathways. After resection of DSB ends, microhomologies can anneal. Trimming of free 3'-ends results in loss of genetic information.

In contrast to NHEJ, HR occurs only rarely in plant somatic cells with only 1:10,000 DSB repaired by HR using ectopic repair templates (Puchta, 1999). The two most prominent HR pathways in somatic plant cells are the single-strand annealing pathway (SSA) and the synthesis-dependent strand annealing pathway (SDSA, Puchta, 2005). All HR pathways start with resection of the 5'-DNA ends, which is mediated by the MNR complex amongst others. The free ssDNA ends are then coated by REPLICATION PROTEIN A (RPA), which prevents degradation and hinders formation of secondary structures. RPA also prevents MMEJ mediated repair of the DSB. RADIATION SENSITIVE 52 (RAD52), one of the hallmark enzymes for HR is then additionally recruited to the nucleoprotein complex (Mortensen et al., 2009; Samach et al., 2011; Sfeir and Symington, 2015; Manova and Gruszka, 2015). If the DSB occurs between large homologous regions (> 30 nucleotides), e.g. in regions of gene duplications, SSA is the preferred HR pathway (Fig. 5A). Homologous ssDNA ends can anneal. 3'-overhangs are trimmed by a RAD1/RAD10 heterodimer and the remaining gaps are filled by polymerases and sealed by LIG1 (Dubest et al., 2002; Siebert and Puchta, 2002; Orel et al., 2003; Manova and Gruszka, 2015; Sfeir and

Symington, 2015). The non-conservative SSA pathway resembles the MMEJ pathway as it also generates large deletions and is highly mutagenic. In case of a lack of homologous ssDNA sequences, the SDSA pathway is induced, which employs a homologous dsDNA molecule as repair template (Fig. 5B). Therefore, REP52 mediates replacement of RPA with RAD51 (Mortensen et al., 2009; Samach et al., 2011). RAD51-loaded 3'-ssDNA ends can then detect and invade a homologous DNA template by replacing the second strand of the DNA template. This generates a displacement loop (D-Loop) structure, which is stabilized by numerous enzymes (Mannuss et al., 2012; Manova and Gruszka, 2015). The 3'-end is then elongated by polymerases with the homologous DNA strand as template. If the 3'-end elongation proceeds up to the homology of the other resected 3'-end (Fig. 5B, left), microhomology-based second end capture allows annealing of both strands. This leads to a conservative repair with no sequence information loss if a highly homologous sequence (e.g. a sister chromatid) is used as repair template. SDSA also allows integration of foreign DNA in the DSB, if this DNA is flanked by sequences of high homology to the DSB ends (Puchta and Fauser, 2014). This process is called gene targeting (GT). GT is of high biotechnological interest as it allows targeted and precise changes in the genome (Paszkowski et al., 1988; Puchta and Fauser, 2013). SDSA is based on one-sided initiation of DSB repair by one of the two 3'-ends. Therefore, repair template homology to one side of the DSB can be sufficient to initiate SDSA repair. However, the elongated 3'-end will then lack homology to the other resected 3'-end and the two ends will anneal via error-prone NHEJ. This also occurs if the elongation of the 3'-end stops before the homology of the other 3'-end (Fig. 5B, middle) due to weak interaction between extended strand and repair template strand (Puchta, 1998; Voytas, 2013; Manova and Gruszka, 2015). In contrast to somatic cells, meiotic cells repair DSB mostly by the classical DSB repair model (DSBR) pathway (Szostak et al., 1983; Osman et al., 2011). Here, following establishment of the D-loop, both 3'-ends use homologous sequences of the donor template for elongation. This leads to the formation of a double Holliday junction, which can be broken up by endonucleases, such as GEN1 and SEND1 (Bauknecht and Kobbe, 2014). Depending of the orientation of these enzymes, dissolution of the double Holliday junction results in noncrossover events (Fig. 5B, right) while resolution results in crossover events, which are important for genetic material exchange between sister chromosomes during meiosis (Osman et al., 2011; Pacher and Puchta, 2017).

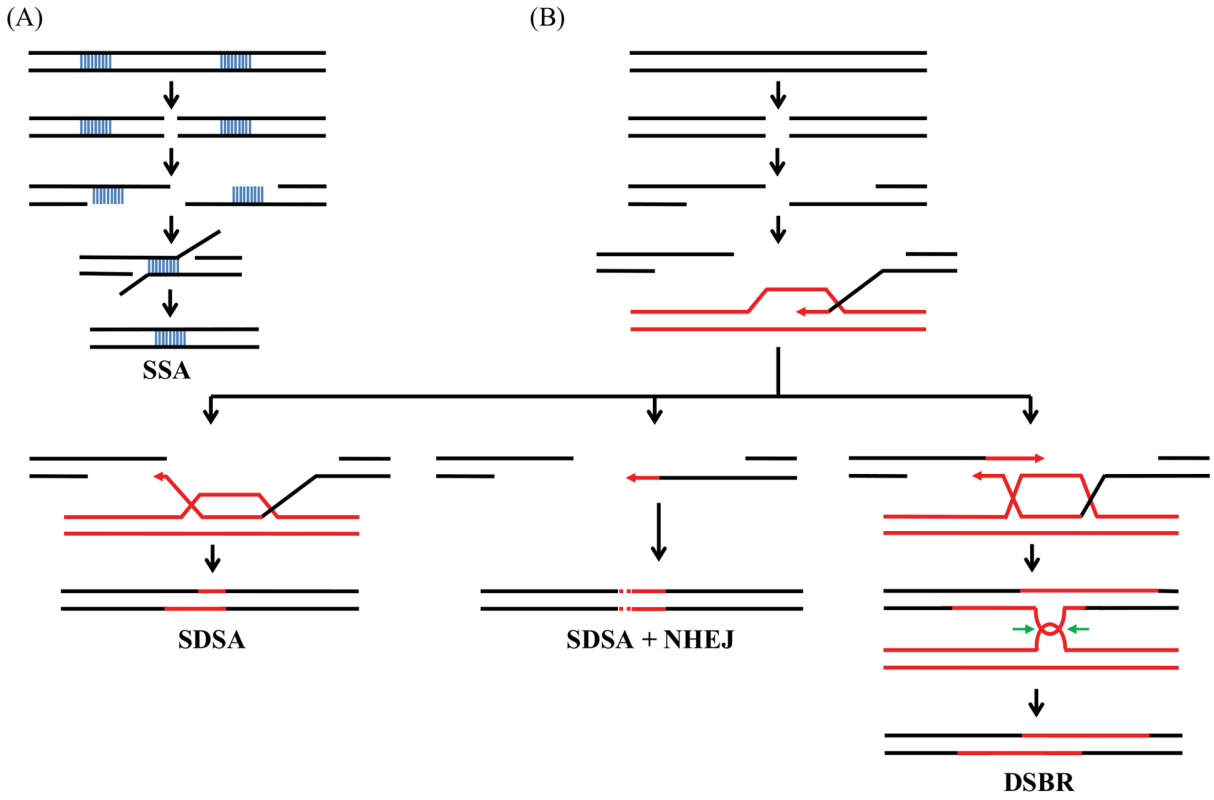


Figure 5: Homologous recombination-mediated repair pathways. (A) In the non-conservative SSA repair pathway, large microhomologies (blue) anneal after 5'-end resections. Trimming of 3'-ends results in loss of genetic information. (B) In the conservative HR pathways, following 5'-end resections, one 3'-end can invade a homologous sequence, thereby forming a D-loop structure. In the SDSA model, elongation of the 3'-end up to homologous regions of the other 3'end allows reannealing of both DSB ends via microhomology-based second end capture (left). One-sided homologies or premature stop of elongation results in one-sided repair by HR while the other end of the DSB is repaired by error-prone NHEJ (middle). In the DSBR model, both 3'ends use homologous sequences for elongation, leading to formation of a double Holliday junction. Resolution of the double Holliday junction by endonucleases (green arrows) can lead to non-crossover events (right) or cross-over events (not shown). Figure adapted from Pacher and Puchta (2017).

In summary, both repair pathways for DSB allow the introduction of mutations into the plant genome. For exploiting these pathways in genome editing, targeted introduction of DSB is required.

Targeted introduction of double strand breaks via sequence specific nucleases

DSB of the DNA can be introduced enzymatically by endonucleases. Many endonucleases, such as restriction enzymes, cut DNA sequence-specific. However, the target sites of these enzymes usually consist only of few nucleotides, which does not allow specific cleavage of single sites in

an average-sized genome. However, 35 years ago, Puchta and colleagues first used a yeast-derived I-*Sce*I meganuclease to introduce DSB in tobacco protoplasts and later plantlets (Puchta et al., 1993, 1996) to study mechanisms of DSB repair. I-*Sce*I meganucleases reliably detect a specific 18-mer sequence and can therefore be used to cleave genomic DNA at unique positions. However, meganucleases do not allow targeting sequences of choice and the 18-mer sequence has to be introduced artificially into the plant genome. Additionally, mutating meganucleases to modify binding sites is challenging as the DNA-binding domains overlap with the nuclease domain (Smith et al., 2006). In 1996, Kim et al. combined DNA binding zinc-fingers with nuclease domains to develop Zinc-finger nucleases (ZFN). ZFN employ arrays of usually 3 to 4 zinc-fingers, each of which can detect a specific nucleotide triplet. The zinc-fingers domain is coupled to a FokI-nuclease domain. If a pair of ZFNs targets opposite DNA strands in close proximity, the two FokI domains dimerize and introduce a DSB in the DNA. However, context-dependency of the recognition domains and overlapping specificities of the individual zinc-fingers hindered efficient modular assembly of ZFN (Ramirez et al., 2008), which prevented broad application of this technique. Transcription Activator-like Effector Nucleases (TALENs) are also modular proteins consisting of an array of TALE proteins to target a genomic sequence, fused to a FokI-nuclease domain (Christian et al., 2010). TALEs bind DNA by 13-28 copies of a 33-35-amino-acid repeat. Every repeat region is largely conserved but contains two variable amino acid residues at position 12 and 13, which recognize a target nucleotide (Boch et al., 2009; Moscou and Bogdanove, 2009). Like ZFN, two TALENs are combined to target opposite DNA strands to allow FokI-mediated cleavage of the DNA. While TALENs allow the targeting of any DNA sequence by combining a TALE-array of choice, the assembly of the arrays is challenging due to the highly repetitive sequence of the repeat regions (Voytas, 2013). The most recent addition to the sequence-specific nucleases (SSN) was the CRISPR/Cas9 system.

The CRISPR/Cas9 system derives from a bacterial defense system and can be adapted for targeted genome modifications

Prokaryotes have developed a range of defense systems to cope with the constant threat of bacteriophages (phages) infections. These systems involve innate immune mechanisms, such as secretion of polysaccharides to restrict phage adsorption, production of restriction enzymes that recognize unmethylated viral DNA, or induced cell death amongst others (Marraffini, 2015; Hille and Charpentier, 2016). While the innate immune system detects generic features of pathogens,

an adaptive immune system can learn to deal with newly arising threats from mutated viruses or new pathogens (Rath et al., 2015). CRISPR/Cas (clustered regulatory interspaced short palindromic repeats/CRISPR-associated proteins) represents the only known adaptive immune system in prokaryotes, where it can be found in 45 % of the so far sequenced bacterial genomes and 87 % of archaeal genomes (Grissa et al., 2007). CRISPR derived immunity is based on acquisition of small fragments of foreign DNA and subsequent integration in the host cell's CRISPR array. This locus consists of fragments of previous infections (so called spacers) of 26-47 bp, separated by short repetitive sequences of 21-48 bp (Grissa et al., 2007; Rath et al., 2015). The spacers act as a memory of previous infections, which can be used for targeted defense mechanisms by Cas effector proteins in following infections (Ishino et al., 1987; Mojica et al., 2005; Barrangou et al., 2007). The size of the *CRISPR* locus is dependent on the number of spacers, which range from very few up to several hundred depending on the species (Grissa et al., 2007). The so-far described CRISPR/Cas system can be separated in two classes, six types and 19 subtypes according to the Cas effector proteins, targeting requirements and biogenesis of effector RNAs (Hille and Charpentier, 2016; Jackson et al., 2017). In the following, the type II-A CRISPR/Cas9 system from *Streptococcus pyogenes* (*S. pyogenes*) will be described in more detail, as it is the most relevant system for genome editing.

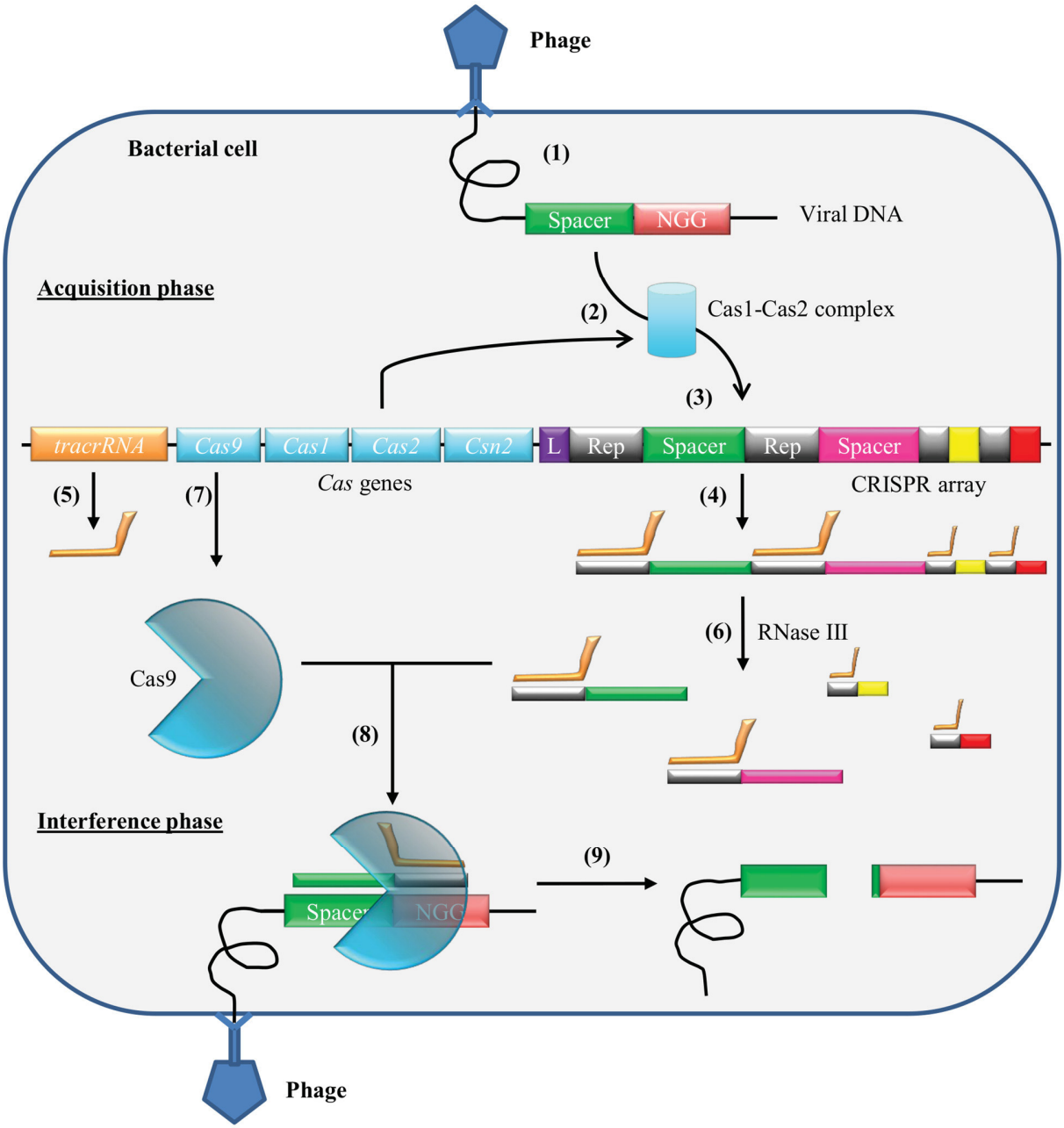


Figure 6: The CRISPR/Cas9 system from *Streptococcus pyogenes*. During the acquisition phase, the bacterial cell integrates small fragments of foreign DNA (spacers) into the CRISPR array. The spacers act as a memory of previous infections for targeted defense mechanisms by Cas effector proteins in the interference phase of following infections. L = Leader sequence, Rep = Repeat. Numbers according to main text.

During an infection, a phage inserts its DNA into the bacterial host cell (Fig. 6, 1). In the acquisition phase of the immune reaction, the bacterial host cell selects a protospacer from the genetic material of the phage and generates a spacer, which it integrates into the CRISPR array.

The genetic material used by the acquisition machinery is probably not the complete viral genome but rather DNA degradation products of the bacterial helicase/nuclease complex RecBCD. RecBCD degrades exposed linear DNA, such as phage DNA or DNA resulting from DSB in replicating vectors, another possible source of foreign DNA (Levy et al., 2015). By this, the cell preferentially takes up new protospacers from foreign DNA and can discriminate against host DNA. The selection of a protospacer sequence is not random. A short motif (NGG in *S. pyogenes*) directly adjacent to the protospacer (protospacer adjacent motif = PAM) is necessary for acquisition. This PAM motif is detected by the Cas9 nuclease (Anders et al., 2014; Heler et al., 2015). Cas9 then recruits Cas1 and Cas2, two nucleases, which are conserved among most CRISPR/Cas system, and which function in a complex (2) as key factors for processing and integration of a protospacer (Nuñez et al., 2014, 2015). Another protein, Csn2, is also part of the acquisition machinery but its role has not been fully resolved yet. Similarly, a small trans-activating CRISPR RNA (tracrRNA) also participates in spacer acquisition (Heler et al., 2015). The Cas1-Cas2 complex with the bound DNA protospacer recognizes intrinsic sequences in the CRISPR array, such as an A-T-rich leader sequence upstream of the CRISPR array and the first nucleotides of the first repeat (Nozawa et al., 2011; Wei et al., 2015; Wright and Doudna, 2016; Van Orden et al., 2017). By this, a polar addition of newly acquired spacer at the 5'-end of the CRISPR array is achieved (McGinn and Marraffini, 2016). In a reaction catalyzed by the Cas1-Cas2-complex, the two strands of the first repeat are separated and the newly acquired spacer is integrated (3). In the interference phase of an infection, promoter elements in the leader region lead to transcription of the CRISPR array (4). Anti-repeat sequences within the tracrRNA molecules (5) hybridize to the conserved repeat regions of the precursor CRISPR RNA (pre-crRNA). These double stranded RNA elements are recognized by RNaseIII, which (together with Cas9) processes the pre-crRNA:tracrRNA hybrids to small guide RNA (gRNA) molecules (6). These gRNAs consist of one repeat-spacer element, in which the spacer element is shortened to 20 bp (Deltcheva et al., 2011; Jinek et al., 2012). Cas9, the hallmark nuclease of type II CRISPR systems (7), is then activated by interaction with the repeat-antirepeat duplex of the mature gRNA molecule (Jinek et al., 2014; Nishimasu et al., 2014; Lim et al., 2016). Cas9 is guided to target DNA (8) by Watson-Crick base pairing between the gRNA molecule and the target DNA (Nishimasu et al., 2014). First, two conserved arginine residues recognize the PAM motif in the target DNA. The necessity of a PAM motif in the target site avoids self-cleavage of the CRISPR array. Detection of a PAM motif leads to repositioning of the DNA strand and local DNA strand

separation. This allows probing of the identity of the first nucleotides directly upstream of the PAM motif. Consequently, matching of the first 8-12 nucleotides upstream of the PAM (seed region) is crucial for efficient target cleavage while mismatches further distal from the PAM motif are partly tolerated (Jinek et al., 2012; Anders et al., 2014). In case of pairing between the seed region and the gRNA, the DNA strand is further opened and a gRNA-DNA-heteroduplex is formed (Anders et al., 2014). Further structural rearrangements of Cas9 allow cleavage of both DNA strands 3 bp upstream of the PAM motif (9) by its two nuclease domains RuvC and HNH (Makarova et al., 2011; Jinek et al., 2012; Nishimasu et al., 2014). The CRISPR/Cas defense system is highly effective and can lead to complete virus extinction under laboratory conditions (van Houte et al., 2016). Since the discovery of the basic mechanisms of the CRISPR/Cas9 system of *S. pyogenes*, it has gained major interest due to the possibility to target a generic nuclease with a programmable short RNA duplex to any DNA sequence for induction of DSB (Fig. 7). The 20 bp target sequence is sufficient to target unique sites in a genome even though off-target cleavage in highly similar sequences has been reported (Deltcheva et al., 2011; Fu et al., 2013; Hsu et al., 2013). For genome editing purposes, the system has furthermore been simplified by the combination of crRNA and tracrRNA to an artificial hybrid single guide RNA (sgRNA), which allows targeting of any DNA sequence containing a PAM motif (Jinek et al., 2012). This two component system, consisting of Cas9 and a custom-tailored sgRNA to induce a DSB, and employing the internal mechanisms of cells to repair DSB for mutagenesis, has since then been successfully applied for genome editing in human cells and embryos (Cong et al., 2013; Jinek et al., 2013; Mali et al., 2013; Ma et al., 2017), animals (Gratz et al., 2013; Wang et al., 2013; Niu et al., 2014), and plants (Jiang et al., 2013; Li et al., 2013; Nekrasov et al., 2013; Fauser et al., 2014; Feng et al., 2014), amongst others.

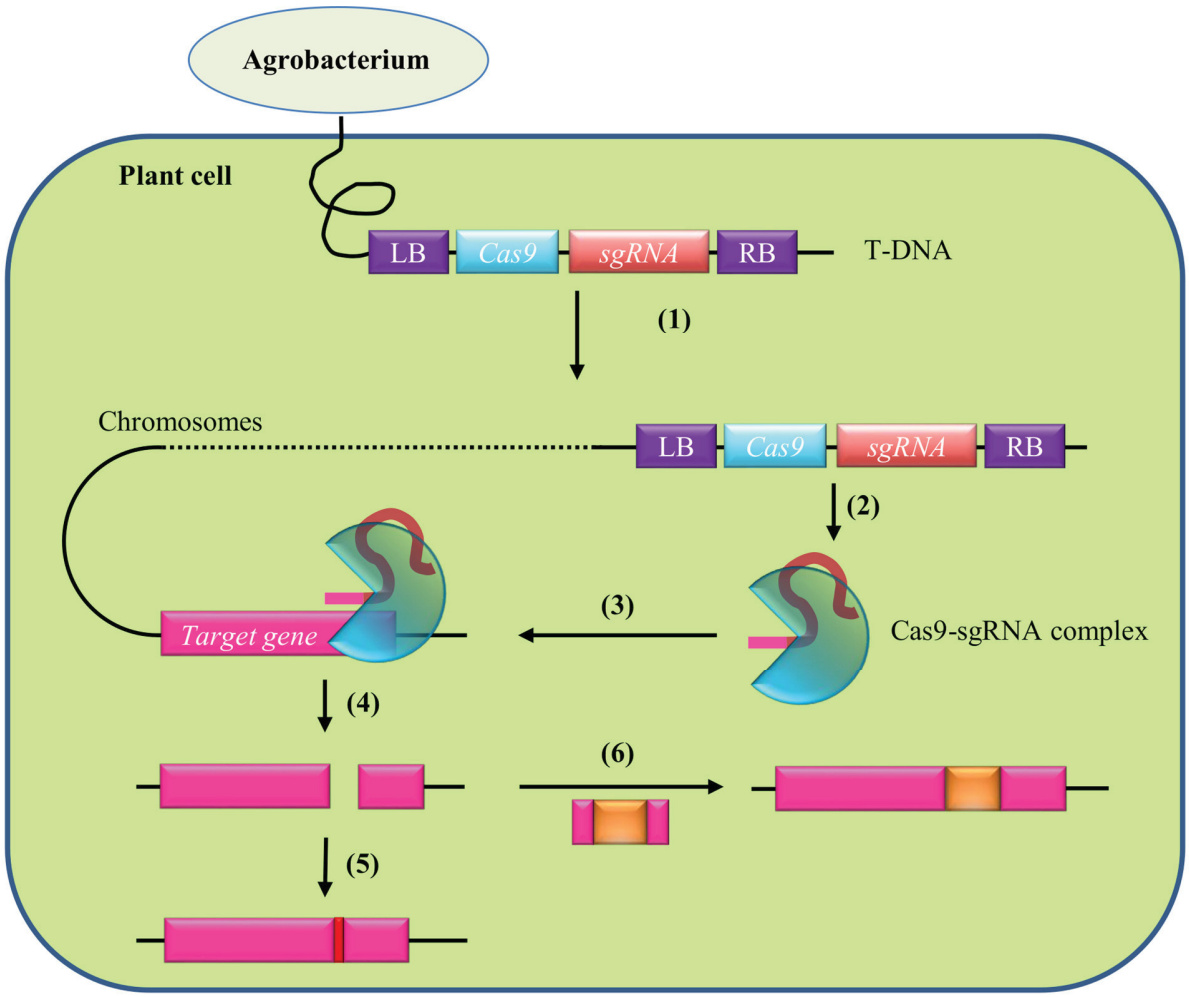


Figure 7: Application of the CRISPR/Cas9 system for targeted modifications of plant genomes. Genes for *Cas9* and a sgRNA can be stably integrated into the plant genome via *Agrobacterium*-mediated transformation (1). After expression, Cas9 and sgRNA form a ribonucleoprotein complex (2), which is guided by RNA-DNA interaction to the target gene (3). Cleavage by Cas9 induces a DSB (4). Repair by error-prone NHEJ can induce indel mutations (5). Offering an artificial homology template can induce gene knock-ins (6). LB = left T-DNA border, RB = right T-DNA border.

Two main points make the Cas9 system advantageous to previous genome editing techniques: First, target strand recognition is mediated by RNA-DNA-interactions and not protein-DNA interaction, making programing of the nuclease far easier. Second, multiplex genome engineering is possible by introduction of several sgRNAs (Cong et al., 2013).

Efficiencies of gene targeting are enhanced by homology template mobilization

SSN allow targeted introduction of DSB into plant cells and GT provides the opportunity to induce highly specific targeted changes in the plant genome if a suitable homologous repair target

is present in the plant cell. Still, GT is hindered by the intrinsically low frequency of HR repair of DSB. However, early analyses of HR repair events in plants indicated that T-DNAs provided better repair templates than ectopic sequences. It was interpreted that double stranded T-DNA, which was not yet integrated into the genome, would be more accessible for the DNA repair machinery and would also provide less steric hindrance than ectopic genomic loci (Puchta, 2005). In recent years, two techniques have been developed that allow this kind of repair template mobilization also for stably integrated repair templates. The first approach (Fig. 8A), employs a geminiviral endoreplication system to induce production of several thousand circular repair templates in the cell nucleus (Baltes et al., 2014). The SSN and the repair template are integrated in the plant genome via T-DNA transfer. The repair template is flanked by *cis*-acting sequences, namely two long intergenic regions (LIR) and the short intergenic region (SIR) from deconstructed geminiviruses. Additionally, a gene encoding the *trans*-acting viral replication initiator protein Rep is incorporated on the T-DNA. Expression of *Rep* is mediated by the bidirectional promoter activity of the LIR. The SIR functions as terminator of Rep expression and as polyadenylation signal (Chen et al., 2011). Rep nicks the DNA strand within an invariant 9 nucleotide motif in a stem-loop structure of the LIR. Rep binds the 5'-end of the DNA nick and the replication machinery of the host cell synthesizes a complementary strand using the 3'-end of the terminus as primer. Complementary strand synthesis stops at the second LIR and Terminase-activity of Rep releases the newly synthesized strand as ssDNA molecule. The ssDNA molecule is circularized by Rep and the cell machinery converts the ssDNA replicon in a dsDNA replicon using the SIR as origin for complementary strand synthesis (Stenger et al., 1991; Mor et al., 2003; Huang et al., 2009; Chen et al., 2011). Further rounds of replication using either the replicon or the stably integrated T-DNA as templates can lead to several thousand repair template copies in the plant cell nucleus, which can serve for HR mediated DSB repair without sterical hindrance of an unfavorable chromosomal location (Timmermans et al., 1992; Baltes et al., 2014). Next to the replicational effect, further pleiotropic effects of the viral system might enhance GT frequencies. Alterative splicing of *Rep* mRNA generates RepA protein, which promotes entry of the cells into the S-phase in which frequencies of HR are enhanced in higher eukaryotes (Liu et al., 1999; Saleh-Gohari and Helleday, 2004; Baltes et al., 2014; Mathiasen and Lisby, 2014). Furthermore, nicked donor molecules (as generated by Rep) are described as highly effective HR repair templates (Lozano-Durán, 2016; Chen et al., 2017). Depending on the host species, viral sequences can be derived from various geminiviruses, e.g the bean yellow dwarf

virus for *Arabidopsis*, tobacco, potato and tomato (Liu et al., 1997; Chen et al., 2011; Baltes et al., 2014; Čermák et al., 2015; Butler et al., 2016), or the wheat dwarf virus for wheat and rice (Gil-Humanes et al., 2017; Wang et al., 2017). Using this system, GT frequencies could be enhanced by one to two orders of magnitude compared to classical T-DNA delivery in tobacco, wheat, tomato, rice and potato (Baltes et al., 2014; Čermák et al., 2015; Butler et al., 2016; Gil-Humanes et al., 2017; Wang et al., 2017).

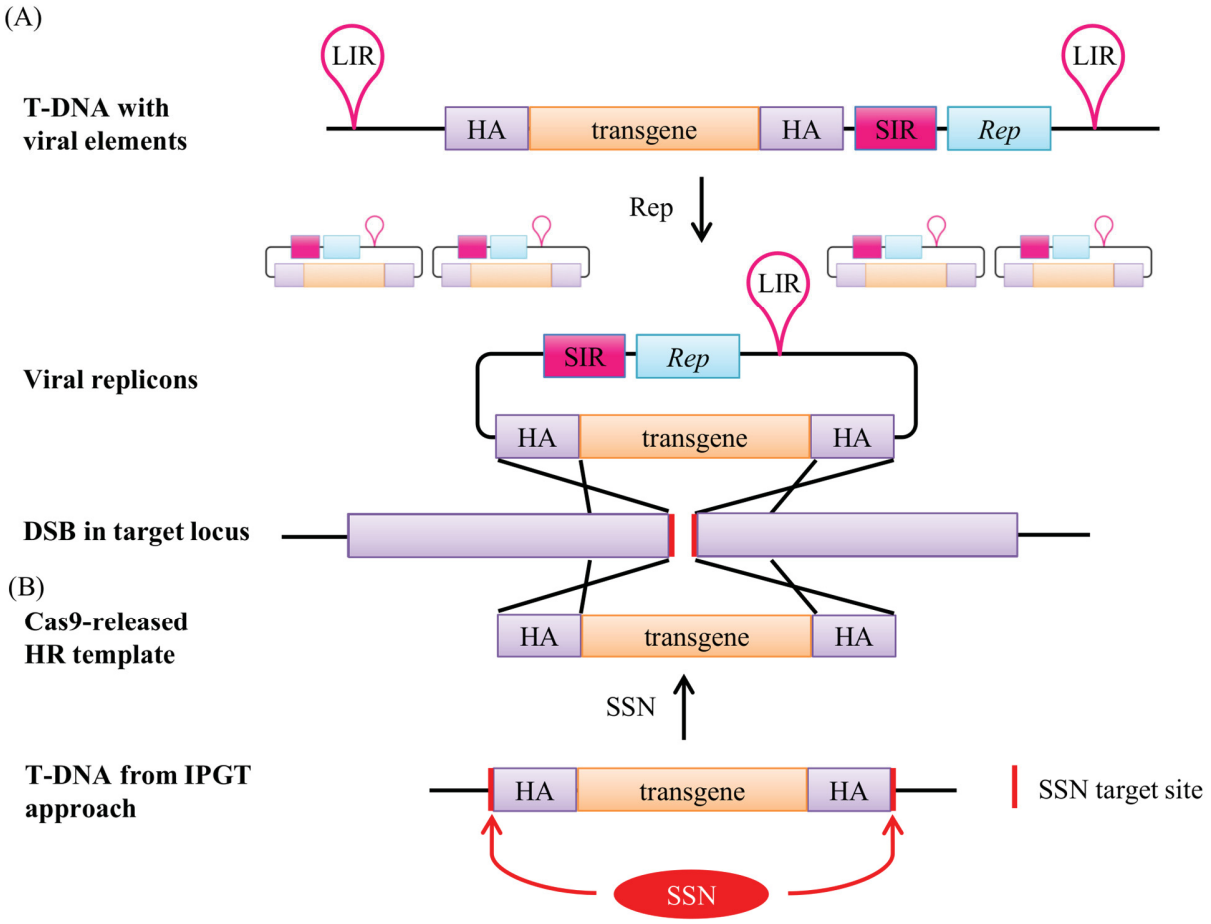


Figure 8: Mobilization of the homology repair template for gene knock-ins. (A) Flanking of the homology template (consisting of a transgene with homology arms) with viral *cis*-elements (LIR/SIR) leads to production of circular viral replicons by the viral replication initiator protein Rep. The replicons can function as repair template for DSB repair in a target locus. (B) In the *in planta* gene targeting (IPGT) approach, flanking of the repair template with target sites for sequence specific nucleases (SSN) allows simultaneous release of the repair template from the T-DNA backbone and DSB induction in the target locus. LIR = long intergenic region, SIR = short intergenic region, HA = homology arm.

An alternative *in planta* gene targeting (IPGT) approach (Fig.8B) employs the SSN for release of the template (Fauser et al., 2012; Schiml et al., 2014). Therefore, the homology template, which

can be integrated in the plant cell on the same T-DNA as the SSN, is flanked by SSN recognition sites. Two DSB by the SSN will therefore release the homology template, while a third DSB primes the target locus for HR mediated repair (Schiml et al., 2017). If the T-DNA was inserted only in one genomic locus, two copies of homology templates per cell are released as maximum. Thus, the probability of random integration of templates into the genome is low (Fauser et al., 2012). Using the IPGT approach, stable GT events were recovered in *Arabidopsis* at frequencies of 0.14-0.83 % (Fauser et al., 2012; Schiml et al., 2014; Zhao et al., 2016).

Both approaches allow the release of the homology template throughout the whole life cycle of transformable plants, which allows GT in tissues that might be more prone to HR (Fauser et al., 2012; Schiml et al., 2014).

Variations of the CRISPR/Cas9 system broaden the range of this technique

Only a few years have passed since the establishment of the Cas9 system for targeted genome modifications. Still, manifold improvements and modifications have been made to extend the CRISPR toolbox. Mutations of one of the two cleavage domains by a single amino acid exchange (D10A in RuvC cleavage domain or H840A in HNH cleavage domain) resulted in a Cas9 nickase (Ran et al., 2013). DNA nicks are repaired with high fidelity and are therefore not susceptible to error-prone NHEJ. This makes the nickase an attractive option for HR experiments, since the target sites cannot be destroyed by random mutagenesis (Fauser et al., 2014). Pairing of two nickases with adjacent target sites on complementary DNA strands can induce DSBs. This approach can reduce off-targeting, as the target sequence is enhanced to 40 nucleotides (Ran et al., 2013). Another approach of reducing off-targeting is the use of high-fidelity Cas9 versions, which have been improved by rational mutagenesis. Several variants of Cas9 are nowadays available that show less conformational activation of the HNH cleavage domain by DNA substrates with mismatches to the sgRNA sequence (Kleinstiver et al., 2016; Slaymaker et al., 2016; Chen et al., 2017). Further reduction in off-targeting was achieved by usage of preassembled Cas9-sgRNA-ribonucleoproteins for genome engineering due to rapid degradation of the genome editing tools (Kim et al., 2014). Mutation of both cleavage domains of Cas9 results in a dead Cas9 (dCas9). This dCas9 can be used as cargo protein to target bound effectors to genomic loci of choice. Fusion of fluorescent proteins to dCas9 allows visualization of genomic regions and was used in human cells and plants to track telomere dynamics (Chen et al., 2013; Dreissig et al., 2017). Other applications include determination of chromatin ultrastructures

(Anton et al., 2014) or studies of protein-DNA interactions by simultaneous labeling of the DNA locus and the protein of interest (Dreissig et al., 2017). The Cas9 system can also be used for gene expression modifications, which is of interest for C_4 photosynthesis research, since many genes are differentially expressed between C_3 and C_4 sister species (Bräutigam et al., 2011; Gowik et al., 2011; Lauterbach et al., 2017). Targeting dCas9 to promoter sites lowers gene expression by hindering the assembly of the transcription complex (Bikard et al., 2013; Qi et al., 2013; Wang et al., 2016). Fusion of dCas9 to repressor domains such as SRDX can enhance theses repressor capacities (Lowder et al., 2015; Piatek et al., 2015; Tang et al., 2017). Similarly, fusion to activator domains, such as VP64, EDLL or TAD can strongly increase gene expression (Lowder et al., 2015; Piatek et al., 2015). By fusing dCas9 to catalytic domains of epigenetic modifiers, such as the CpG demethylase TET1, the methyl transferase DNMT3A or the histone acetyltransferase p300, targeted changes in the epigenome can be induced (Hilton et al., 2015; Choudhury et al., 2016; Vojta et al., 2016). These changes also influence the transcription pattern of genes but the effects of epigenetic changes can be inherited to daughter cells and might therefore persist longer (Vojta et al., 2016). The outcome of NHEJ-mediated repair of Cas9-induced DSB is relatively random and HR-mediated approaches still lack efficiency. Alternatively, targeted nucleotide exchanges can be specifically introduced by base changer Cas9 variants in which dCas9 or Cas9 nickase are fused to cytidine deaminase (Komor et al., 2016; Zong et al., 2017) or artificial adenosine deaminases (Gaudelli et al., 2017). By this, C>T or A>G substitutions can be introduced. This might be also of interest for C_4 photosynthesis research, as some differences between C_3 and C_4 plants, such as differences of feedback inhibition of phosphoenolpyruvate carboxylase can be explained by single amino acid exchanges (Paulus et al., 2013). Finally, manifold Cas9 variants have been isolated by now from other bacteria or have been engineered. These variants have different properties, such as different PAM motives, which enhances the amount of target sites. Additionally, some of these orthologues harbor various advantages compared to Cas9 from *S. pyogenes*. As an example, Cas9 from *Staphylococcus aureu*s induces stable mutations in *Arabidopsis* at very high rates of over 90 % within two generations and is considerably smaller than *S. pyogenes* Cas9 (Steinert et al., 2015). The nuclease CpfI (CRISPR from *Prevotella* and *Francisella*) generates staggered DSB distal from the PAM site which can be advantageous for GT (Zetsche et al., 2015). Finally, Cas13 from *Leptotrichia wadei* targets RNA instead of DNA and can be used for RNA modifications and as a catalytically inactive form analogous to dCas9 for a broad array of RNA research (Abudayyeh et

al., 2017; Cox et al., 2017).

Visual markers allow testing of mutagenesis approaches in plants

Cas9-mediated genome editing is a highly dynamic research area with new tools being described frequently and validation of these tools is indispensable. Mutations on genomic level can be detected by sequencing or by various detection methods such as restriction fragment length polymorphism assays or T7 endonuclease assays (Belhaj et al., 2013). However, visual markers are advantageous as they allow tracking of mutational events by eye. Several genes have been employed in plants in early experiments to quantitate mutational efficiencies. Stably expressed fluorescent proteins, such as GREEN FLUORESCENT PROTEIN (GFP), YELLOW FP, and BLUE FP, have been targeted in numerous studies and are especially well suited for studies in protoplasts, since they allow high throughput phenotypical analysis via flow cytometry (Feng et al., 2013; Jiang et al., 2013; Mao et al., 2013; Sauer et al., 2016). However, the corresponding genes first need to be introduced into the plant cell. Another exogenous marker is the ß-GLUCURONIDASE (GUS) gene. Here, a repair of a dysfunctional *GUS* gene by CRISPR/Cas9 can be detected by eye after a staining procedure (Miao et al., 2013; Fauser et al., 2014; Butler et al., 2016; Čermák et al., 2017). The disadvantage of this technique is that the staining procedure is lethal for the plant and does not allow propagation of verified mutants. Induction of resistance to an otherwise lethal treatment by mutation in endogenous genes has been used manifold. Mutations by CRISPR/Cas9 in the endogenous *ADH1* gene of *Arabidopsis* conferred resistance to allyl alcohol treatment (Fauser et al., 2014; Steinert et al., 2015). Cas9-induced amino acid changes in the 5'-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATESYNTHASE conferred resistance to glyphosate in flax (Sauer et al., 2016). Similarly, herbicide-inhibiting point mutations were introduced in the *ACETOLACTATE SYNTHASE 1* gene of potato (Butler et al., 2016). While these approaches allow regeneration of stably mutated plants, they do not allow tracking of mutational progress during development as chimeric plants do not survive the treatments. Various other endogenous marker genes have been that provide distinct phenotypical changes in case of mutations. PHYTOENE DESATURASE (PDS) is involved in carotenoid biosynthesis and a knockout of the *PDS* gene leads to albino phenotypes. *PDS* was already targeted in *Arabidopsis*, tobacco, rice, cassava, watermelon, and banana amongst others (Li et al., 2013; Shan et al., 2013; Odipio et al., 2017; Tian et al., 2017; Kaur et al., 2018). Similar leaf color phenotypes could be obtained by targeting the *Arabidopsis MAGNESIUM-CHELATASE*

SUBUNITS 1 and *2* genes (Mao et al., 2013) or the rice *CHLOROPHYLL A OXYGENASE 1* gene (Miao et al., 2013). In tomato, the promoter of *ANTHOCYANIN MUTANT 1* was modified by CRISPR/Cas9. Purple spots indicated anthocyanin accumulation and were thus used as readout for successful gene editing (Čermák et al., 2015). Feng et al. (2013) analyzed a range of endogenous marker, such as *Arabidopsis* JASMONATE-ZIM-DOMAIN PROTEIN 1, GIBBERELLIC ACID INSENSITIVE, or rice YOUNG SEEDLING ALBINO, which all showed drastic growth phenotypes. Morphology changes were used as markers in rice, with mutations of the *LAZY1* gene to induce a tiller-spreading phenotype (Miao et al., 2013), or in the *DROOPING LEAF* gene for changed leaf morphology (Ikeda et al., 2016). Leaf morphology was also changed by targeting the tomato *ARGONAUTE 7* gene (Brooks et al., 2014).

The phenotypic effect of mutations in most endogenous markers used so far was drastic with huge pleiotropic effects up to lethality. Trichome formation might provide an alternative endogenous visual marker system for analysis of mutational efficiency by eye. Trichomes are specialized epidermal cells, which are present on aerial parts of nearly all terrestrial plants (Yang and Ye, 2013). While they provide protection against herbivores and pathogens, UV light, water loss, and extreme temperatures, they are not essential for plant growth under laboratory conditions (Oppenheimer et al., 1991; Pattanaik et al., 2014). In *Arabidopsis*, trichome development is initiated in early stages of leaf development in protodermal cells by an increase in cell and nuclear size due to several rounds of endoreplication. The trichome cell then grows and branches twice until a mature trichome is formed (Hülskamp et al., 1994). Trichome development and patterning is regulated by a complex network of transcription factors and also influenced by phytohormones and microRNAs (Pattanaik et al., 2014). The master regulator of trichome formation, GLABROUS1 (GL1), is a R2R3-MYB transcription factor, which contains two MYB-DNA-binding domain repeats at the N-terminus and an acidic C-terminus (Oppenheimer et al., 1991; Larkin et al., 1994). GL1 interacts with the bHLH transcription factors GL3, or ENHANCER OF GLABROUS 3 (EGL), and with the WD40 repeat protein TRANSPARENT TESTA GLABRA 1 (TTG1). The resulting GL1-GL3/EGL3-TTG1 complex induces trichome formation by transcriptional activation of the trichome activators GL2 and TTG2, which are key determinators for trichome expansion and shape (Rerie et al., 1994; Johnson et al., 2002; Morohashi et al., 2007; Zhao et al., 2008). Additionally, expression of R3 MYB repressors, such as *ENHANCER OF TRIPTYCHON 1* (*ET*C1) and CAPRICE (*CPC*), which lack an activation domain, is induced. ETC1 and CPC can move into neighboring cells where they form an inactive

complex with TTG1 and EGL3 in competition to GL1. This competitive inhibition act as negative regulator of trichome formation as a certain threshold of active GL1-GL3/EGL3-TTG1 complex is required to produce enough GL2 and TTG2 for induction of trichome fate in a cell (Schellmann et al., 2002; Wada et al., 2002; Zhao et al., 2008). Knockout of any of the components of the activation complex leads to loss of trichomes. However, only knockout of *GL1* has no pleiotropic effects, while disruption of *GL3*/*EGL3* or *TTG* leads to anthocyanin accumulation (Koornneef et al., 1982; Oppenheimer et al., 1991; Walker et al., 1999; Zhang et al., 2003). This makes the *GL1* gene an attractive endogenous target for CRISPR/Cas9 experiments.

References

Abudayyeh, O. O., Gootenberg, J. S., Essletzbichler, P., Han, S., Joung, J., Belanto, J. J., et al. (2017). RNA targeting with CRISPR–Cas13. *Nature* 550, 280. doi:10.1038/nature24049.

Adwy, W., Laxa, M., and Peterhansel, C. (2015). A simple mechanism for the establishment of C2-specific gene expression in Brassicaceae. *Plant J.* 84, 1231–1238. doi:10.1111/tpj.13084.

Akyildiz, M., Gowik, U., Engelmann, S., Koczor, M., Streubel, M., and Westhoff, P. (2007). Evolution and Function of a cis-Regulatory Module for Mesophyll-Specific Gene Expression in the C4 Dicot Flaveria trinervia. *Plant Cell* 19, 3391–3402. doi:10.1105/tpc.107.053322.

Anders, C., Niewoehner, O., Duerst, A., and Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* 513, 569. doi:10.1038/nature13579.

Anton, T., Bultmann, S., Leonhardt, H., and Markaki, Y. (2014). Visualization of specific DNA sequences in living mouse embryonic stem cells with a programmable fluorescent CRISPR/Cas system. *Nucleus* 5, 163–172. doi:10.4161/nucl.28488.

Arora, L., and Narula, A. (2017). Gene Editing and Crop Improvement Using CRISPR-Cas9 System. *Front. Plant Sci.* 8, 1932. doi:10.3389/fpls.2017.01932.

Aubry, S., Brown, N. J., and Hibberd, J. M. (2011). The role of proteins in C3 plants prior to their recruitment into the C4 pathway. *J. Exp. Bot.* 62, 3049–3059. doi:10.1093/jxb/err012.

Bai, J., Xu, D. H., Kang, H. M., Chen, K., and Wang, G. (2008). Photoprotective function of photorespiration in Reaumuria soongorica during different levels of drought stress in natural high irradiance. *Photosynthetica* 46, 232. doi:10.1007/s11099-008-0037-5.

Baltes, N. J., Gil-Humanes, J., Cermak, T., Atkins, P. A., and Voytas, D. F. (2014). DNA replicons for plant genome engineering. *Plant Cell* 26, 151–163. doi:10.1105/tpc.113.119792.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., et al. (2007). CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science* 315, 1709–1712. doi:10.1126/science.1138140.

Bauknecht, M., and Kobbe, D. (2014). AtGEN1 and AtSEND1, Two Paralogs in Arabidopsis, Possess Holliday Junction Resolvase Activity. *Plant Physiol.* 166, 202–216. doi:10.1104/pp.114.237834.

Bauwe, H. (2010). "Photorespiration: The Bridge to C4 Photosynthesis," in *C4 Photosynthesis and Related CO2 Concentrating Mechanisms* Advances in Photosynthesis and Respiration. (Springer, Dordrecht), 81–108. doi:10.1007/978-90-481-9407-0_6.

Bauwe, H., and Kolukisaoglu, Ü. (2003). Genetic manipulation of glycine decarboxylation. *J. Exp. Bot.* 54, 1523–1535. doi:10.1093/jxb/erg171.

Belhaj, K., Chaparro-Garcia, A., Kamoun, S., and Nekrasov, V. (2013). Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. *Plant Methods* 9, 39. doi:10.1186/1746-4811-9-39.

Bétermier, M., Bertrand, P., and Lopez, B. S. (2014). Is Non-Homologous End-Joining Really an Inherently Error-Prone Process? *PLOS Genet.* 10, e1004086. doi:10.1371/journal.pgen.1004086.

Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., and Marraffini, L. A. (2013). Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res.* 41, 7429–7437. doi:10.1093/nar/gkt520.

Bloom, A. J., Burger, M., Rubio Asensio, J. S., and Cousins, A. B. (2010). Carbon dioxide enrichment inhibits nitrate assimilation in wheat and Arabidopsis. *Science* 328, 899–903. doi:10.1126/science.1186440.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., et al. (2009). Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science* 326, 1509–1512. doi:10.1126/science.1178811.

Boldt, R., Edner, C., Kolukisaoglu, Ü., Hagemann, M., Weckwerth, W., Wienkoop, S., et al. (2005). d-GLYCERATE 3-KINASE, the Last Unknown Enzyme in the Photorespiratory Cycle in Arabidopsis, Belongs to a Novel Kinase Family. *Plant Cell* 17, 2413–2420. doi:10.1105/tpc.105.033993.

Bourguignon, J., Macherel, D., Neuburger, M., and Douce, R. (1992). Isolation, characterization, and sequence analysis of a cDNA clone encoding L-protein, the dihydrolipoamide dehydrogenase component of the glycine cleavage system from pea-leaf mitochondria. *Eur. J. Biochem.* 204, 865–873. doi:10.1111/j.1432-1033.1992.tb16706.x.

Bourguignon, J., Merand, V., Rawsthorne, S., Forest, E., and Douce, R. (1996). Glycine decarboxylase and pyruvate dehydrogenase complexes share the same dihydrolipoamide dehydrogenase in pea leaf mitochondria: evidence from mass spectrometry and primary-structure analysis. *Biochem. J.* 313, 229–234.

Bräutigam, A., Kajala, K., Wullenweber, J., Sommer, M., Gagneul, D., Weber, K. L., et al. (2011). An mRNA Blueprint for C4 Photosynthesis Derived from Comparative Transcriptomics of Closely Related C3 and C4 Species. *Plant Physiol.* 155, 142–156. doi:10.1104/pp.110.159442.

Brooks, C., Nekrasov, V., Lippman, Z. B., and Eck, J. V. (2014). Efficient Gene Editing in Tomato in the First Generation Using the Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-Associated9 System. *Plant Physiol.* 166, 1292–1297. doi:10.1104/pp.114.247577.

Brown, N. J., Newell, C. A., Stanley, S., Chen, J. E., Perrin, A. J., Kajala, K., et al. (2011). Independent and Parallel Recruitment of Preexisting Mechanisms Underlying C4 Photosynthesis. *Science* 331, 1436–1439. doi:10.1126/science.1201248.

Butler, N. M., Baltes, N. J., Voytas, D. F., and Douches, D. S. (2016). Geminivirus-Mediated Genome Editing in Potato (Solanum tuberosum L.) Using Sequence-Specific Nucleases. *Front. Plant Sci.* 7, 1045. doi:10.3389/fpls.2016.01045.

Bykova, N. V., Keerberg, O., Pärnik, T., Bauwe, H., and Gardeström, P. (2005). Interaction between photorespiration and respiration in transgenic potato plants with antisense reduction in glycine decarboxylase. *Planta* 222, 130–140. doi:10.1007/s00425-005-1505-9.

Campbell, W. J., and Ogren, W. L. (1990). Glyoxylate inhibition of ribulosebisphosphate carboxylase/oxygenase activation in intact, lysed, and reconstituted chloroplasts. *Photosynth. Res.* 23, 257–268. doi:10.1007/BF00034856.

Cary, R. B., Peterson, S. R., Wang, J., Bear, D. G., Bradbury, E. M., and Chen, D. J. (1997). DNA looping by Ku and the DNA-dependent protein kinase. *Proc. Natl. Acad. Sci.* 94, 4267–4272.

Čermák, T., Baltes, N. J., Čegan, R., Zhang, Y., and Voytas, D. F. (2015). High-frequency, precise modification of the tomato genome. *Genome Biol.* 16, 232. doi:10.1186/s13059-015-0796-9.

Čermák, T., Curtin, S. J., Gil-Humanes, J., Čegan, R., Kono, T. J. Y., Konečná, E., et al. (2017). A Multipurpose Toolkit to Enable Advanced Genome Engineering in Plants. *Plant Cell* 29, 1196–1217. doi:10.1105/tpc.16.00922.

Charbonnel, C., Gallego, M. E., and White, C. I. (2010). Xrcc1-dependent and Ku-dependent DNA double-strand break repair kinetics in Arabidopsis plants. *Plant J.* 64, 280–290. doi:10.1111/j.1365-313X.2010.04331.x.

Chen, B., Gilbert, L. A., Cimini, B. A., Schnitzbauer, J., Zhang, W., Li, G.-W., et al. (2013). Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System. *Cell* 155, 1479–1491. doi:10.1016/j.cell.2013.12.001.

Chen, J. S., Dagdas, Y. S., Kleinstiver, B. P., Welch, M. M., Sousa, A. A., Harrington, L. B., et al. (2017). Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. *Nature* 550, 407. doi:10.1038/nature24268.

Chen, Q., He, J., Phoolcharoen, W., and Mason, H. S. (2011). Geminiviral vectors based on bean yellow dwarf virus for production of vaccine antigens and monoclonal antibodies in plants. *Hum. Vaccin.* 7, 331–338.

Choudhury, S. R., Cui, Y., Lubecka, K., Stefanska, B., and Irudayaraj, J. (2016). CRISPR-dCas9 mediated TET1 targeting for selective DNA demethylation at BRCA1 promoter. *Oncotarget* 7, 46545–46556. doi:10.18632/oncotarget.10234.

Christian, M., Cermak, T., Doyle, E. L., Schmidt, C., Zhang, F., Hummel, A., et al. (2010). Targeting DNA Double-Strand Breaks with TAL Effector Nucleases. *Genetics* 186, 757–761. doi:10.1534/genetics.110.120717.

Collonnier, C., Epert, A., Mara, K., Maclot, F., Guyon-Debast, A., Charlot, F., et al. (2017). CRISPR-Cas9-mediated efficient directed mutagenesis and RAD51-dependent and RAD51-independent gene targeting in the moss Physcomitrella patens. *Plant Biotechnol. J.* 15, 122–131. doi:10.1111/pbi.12596.

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

Cox, D. B. T., Gootenberg, J. S., Abudayyeh, O. O., Franklin, B., Kellner, M. J., Joung, J., et al. (2017). RNA editing with CRISPR-Cas13. *Science* 358, 1019–1027. doi:10.1126/science.aaq0180.

Dellero, Y., Jossier, M., Glab, N., Oury, C., Tcherkez, G., and Hodges, M. (2016). Decreased glycolate oxidase activity leads to altered carbon allocation and leaf senescence after a transfer from high CO2 to ambient air in Arabidopsis thaliana. *J. Exp. Bot.* 67, 3149–3163. doi:10.1093/jxb/erw054.

Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., et al. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 471, 602–607. doi:10.1038/nature09886.

Douce, R., Bourguignon, J., Neuburger, M., and Rébeillé, F. (2001). The glycine decarboxylase system: a fascinating complex. *Trends Plant Sci.* 6, 167–176. doi:10.1016/S1360-1385(01)01892-1.

Dreissig, S., Schiml, S., Schindele, P., Weiss, O., Rutten, T., Schubert, V., et al. (2017). Live-cell CRISPR imaging in plants reveals dynamic telomere movements. *Plant J.* 91, 565–573. doi:10.1111/tpj.13601.

Dubest, S., Gallego, M. E., and White, C. I. (2002). Role of the AtRad1p endonuclease in homologous recombination in plants. *EMBO Rep.* 3, 1049–1054. doi:10.1093/embo-reports/kvf211.

Dueva, R., and Iliakis, G. (2013). Alternative pathways of non-homologous end joining (NHEJ) in genomic instability and cancer. *Transl. Cancer Res.* 2, 163–177. doi:10.21037/1152.

Eisenhut, M., Bauwe, H., and Hagemann, M. (2007). Glycine accumulation is toxic for the cyanobacterium Synechocystis sp. strain PCC 6803, but can be compensated by supplementation with magnesium ions. *FEMS Microbiol. Lett.* 277, 232–237. doi:10.1111/j.1574-6968.2007.00960.x.

Eisenhut, M., Pick, T. R., Bordych, C., and Weber, A. P. M. (2013). Towards closing the remaining gaps in photorespiration - the essential but unexplored role of transport proteins. *Plant Biol. Stuttg. Ger.* 15, 676–685. doi:10.1111/j.1438-8677.2012.00690.x.

Eisenhut, M., Ruth, W., Haimovich, M., Bauwe, H., Kaplan, A., and Hagemann, M. (2008). The photorespiratory glycolate metabolism is essential for cyanobacteria and might have been conveyed endosymbiontically to plants. *Proc. Natl. Acad. Sci.* 105, 17199–17204. doi:10.1073/pnas.0807043105.

Engel, N. (2010). Funktionelle Charakterisierung der mitochondrialen GDC- und SHM-Genfamilien in Arabidopsis thaliana durch Komplementation mit heterologen Genen. Available at: http://rosdok.uni-rostock.de/metadata/rosdok_disshab_0000000531.

Engel, N., Daele, K. van den, Kolukisaoglu, Ü., Morgenthal, K., Weckwerth, W., Pärnik, T., et al. (2007). Deletion of Glycine Decarboxylase in Arabidopsis Is Lethal under Nonphotorespiratory Conditions. *Plant Physiol.* 144, 1328–1335. doi:10.1104/pp.107.099317.

Engel, N., Ewald, R., Gupta, K. J., Zrenner, R., Hagemann, M., and Bauwe, H. (2011). The Presequence of Arabidopsis Serine Hydroxymethyltransferase SHM2 Selectively Prevents Import into Mesophyll Mitochondria. *Plant Physiol.* 157, 1711–1720. doi:10.1104/pp.111.184564.

Engelmann, S., Wiludda, C., Burscheidt, J., Gowik, U., Schlue, U., Koczor, M., et al. (2008). The Gene for the P-Subunit of Glycine Decarboxylase from the C4 Species Flaveria trinervia: Analysis of Transcriptional Control in Transgenic Flaveria bidentis (C4) and Arabidopsis (C3). *Plant Physiol.* 146, 1773–1785. doi:10.1104/pp.107.114462.

Fauser, F., Roth, N., Pacher, M., Ilg, G., Sánchez-Fernández, R., Biesgen, C., et al. (2012). In planta gene targeting. *Proc. Natl. Acad. Sci. U. S. A.* 109, 7535–7540. doi:10.1073/pnas.1202191109.

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

Feng, Z., Mao, Y., Xu, N., Zhang, B., Wei, P., Yang, D.-L., et al. (2014). Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. *Proc. Natl. Acad. Sci.* 111, 4632–4637. doi:10.1073/pnas.1400822111.

Feng, Z., Zhang, B., Ding, W., Liu, X., Yang, D.-L., Wei, P., et al. (2013). Efficient genome editing in plants using a CRISPR/Cas system. *Cell Res.* 23, 1229–1232. doi:10.1038/cr.2013.114.

Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K., et al. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* 31, 822. doi:10.1038/nbt.2623.

Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., et al. (2017). Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* 551, 464. doi:10.1038/nature24644.

Gil-Humanes, J., Wang, Y., Liang, Z., Shan, Q., Ozuna, C. V., Sánchez-León, S., et al. (2017). High-efficiency gene targeting in hexaploid wheat using DNA replicons and CRISPR/Cas9. *Plant J.* 89, 1251–1262. doi:10.1111/tpj.13446.

Givan, C. V., and Kleczkowski, L. A. (1992). The enzymic reduction of glyoxylate and hydroxypyruvate in leaves of higher plants. *Plant Physiol.* 100, 552–556.

Gorbunova, V., and Levy, A. A. (1997). Non-homologous DNA end joining in plant cells is associated with deletions and filler DNA insertions. *Nucleic Acids Res.* 25, 4650–4657.

Gowik, U., Bräutigam, A., Weber, K. L., Weber, A. P. M., and Westhoff, P. (2011). Evolution of C4 Photosynthesis in the Genus Flaveria: How Many and Which Genes Does It Take to Make C4?[W]. *Plant Cell* 23, 2087–2105. doi:10.1105/tpc.111.086264.

Gowik, U., Burscheidt, J., Akyildiz, M., Schlue, U., Koczor, M., Streubel, M., et al. (2004). cis-Regulatory Elements for Mesophyll-Specific Gene Expression in the C4 Plant Flaveria trinervia, the Promoter of the C4 Phosphoenolpyruvate Carboxylase Gene. *Plant Cell* 16, 1077–1090. doi:10.1105/tpc.019729.

Gowik, U., Schulze, S., Saladié, M., Rolland, V., Tanz, S. K., Westhoff, P., et al. (2017). A MEM1-like motif directs mesophyll cell-specific expression of the gene encoding the C4 carbonic anhydrase in Flaveria. *J. Exp. Bot.* 68, 311–320. doi:10.1093/jxb/erw475.

Gratz, S. J., Cummings, A. M., Nguyen, J. N., Hamm, D. C., Donohue, L. K., Harrison, M. M., et al. (2013). Genome Engineering of Drosophila with the CRISPR RNA-Guided Cas9 Nuclease. *Genetics* 194, 1029–1035. doi:10.1534/genetics.113.152710.

Grissa, I., Vergnaud, G., and Pourcel, C. (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* 35, W52–W57. doi:10.1093/nar/gkm360.

Guan, X. Q., Zhao, S. J., Li, D. Q., and Shu, H. R. (2004). Photoprotective Function of Photorespiration in Several Grapevine Cultivars Under Drought Stress. *Photosynthetica* 42, 31–36. doi:10.1023/B:PHOT.0000040566.55149.52.

Guilhaudis, L., Simorre, J.-P., Blackledge, M., Marion, D., Gans, P., Neuburger, M., et al. (2000). Combined Structural and Biochemical Analysis of the H−T Complex in the Glycine Decarboxylase Cycle: Evidence for a Destabilization Mechanism of the H-Protein. *Biochemistry (Mosc.)* 39, 4259–4266. doi:10.1021/bi992674w.

Hagemann, M., and Bauwe, H. (2016). Photorespiration and the potential to improve photosynthesis. *Curr. Opin. Chem. Biol.* 35, 109–116. doi:10.1016/j.cbpa.2016.09.014.

Hanson, A. D., and Roje, and S. (2001). One-Carbon Metabolism in Higher Plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 52, 119–137. doi:10.1146/annurev.arplant.52.1.119.

Hasse, D., Andersson, E., Carlsson, G., Masloboy, A., Hagemann, M., Bauwe, H., et al. (2013). Structure of the Homodimeric Glycine Decarboxylase P-protein from Synechocystis sp. PCC 6803 Suggests a Mechanism for Redox Regulation. *J. Biol. Chem.* 288, 35333–35345. doi:10.1074/jbc.M113.509976.

Heckmann, D., Schulze, S., Denton, A., Gowik, U., Westhoff, P., Weber, A. P. M., et al. (2013). Predicting C4 Photosynthesis Evolution: Modular, Individually Adaptive Steps on a Mount Fuji Fitness Landscape. *Cell* 153, 1579–1588. doi:10.1016/j.cell.2013.04.058.

Heineke, D., Bykova, N., Gardeström, P., and Bauwe, H. (2001). Metabolic response of potato plants to an antisense reduction of the P-protein of glycine decarboxylase. *Planta* 212, 880–887. doi:10.1007/s004250000460.

Heler, R., Samai, P., Modell, J. W., Weiner, C., Goldberg, G. W., Bikard, D., et al. (2015). Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature* 519, 199–202. doi:10.1038/nature14245.

Hertel, T. W. (2011). The Global Supply and Demand for Agricultural Land in 2050: A Perfect Storm in the Making? *Am. J. Agric. Econ.* 93, 259–275. doi:10.1093/ajae/aaq189.

Hibberd, J. M., and Covshoff, S. (2010). The regulation of gene expression required for C4 photosynthesis. *Annu. Rev. Plant Biol.* 61, 181–207. doi:10.1146/annurev-arplant-042809-112238.

Hibberd, J. M., Sheehy, J. E., and Langdale, J. A. (2008). Using C4 photosynthesis to increase the yield of rice - rationale and feasibility. *Curr. Opin. Plant Biol.* 11, 228–231. doi:10.1016/j.pbi.2007.11.002.

Hille, F., and Charpentier, E. (2016). CRISPR-Cas: biology, mechanisms and relevance. *Phil Trans R Soc B* 371, 20150496. doi:10.1098/rstb.2015.0496.

Hilton, I. B., D'Ippolito, A. M., Vockley, C. M., Thakore, P. I., Crawford, G. E., Reddy, T. E., et al. (2015). Epigenome editing by a CRISPR/Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat. Biotechnol.* 33, 510–517. doi:10.1038/nbt.3199.

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* 31, 827. doi:10.1038/nbt.2647.

Huang, Z., Chen, Q., Hjelm, B., Arntzen, C., and Mason, H. (2009). A DNA replicon system for rapid high-level production of virus-like particles in plants. *Biotechnol. Bioeng.* 103, 706–714. doi:10.1002/bit.22299.

Hülskamp, M., Miséra, S., and Jürgens, G. (1994). Genetic dissection of trichome cell development in Arabidopsis. *Cell* 76, 555–566. doi:10.1016/0092-8674(94)90118-X.

Hylton, C. M., Rawsthorne, S., Smith, A. M., Jones, D. A., and Woolhouse, H. W. (1988). Glycine decarboxylase is confined to the bundle-sheath cells of leaves of C3/C4 intermediate species. *Planta* 175, 452–459. doi:10.1007/BF00393064.

Igarashi, D., Miwa, T., Seki, M., Kobayashi, M., Kato, T., Tabata, S., et al. (2003). Identification of photorespiratory glutamate:glyoxylate aminotransferase (GGAT) gene in Arabidopsis. *Plant J.* 33, 975–987. doi:10.1046/j.1365-313X.2003.01688.x.

Ikeda, T., Tanaka, W., Mikami, M., Endo, M., and Hirano, H.-Y. (2016). Generation of artificial drooping leaf mutants by CRISPR-Cas9 technology in rice. *Genes Genet. Syst.* 90, 231–235. doi:10.1266/ggs.15-00030.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., and Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J. Bacteriol.* 169, 5429–5433.

Jackson, S. A., McKenzie, R. E., Fagerlund, R. D., Kieper, S. N., Fineran, P. C., and Brouns, S. J. J. (2017). CRISPR-Cas: Adapting to change. *Science* 356. doi:10.1126/science.aal5056.

Jia, Q., Dulk-Ras, A. den, Shen, H., Hooykaas, P. J. J., and Pater, S. de (2013). Poly(ADP-ribose)polymerases are involved in microhomology mediated back-up non-homologous end joining in Arabidopsis thaliana. *Plant Mol. Biol.* 82, 339–351. doi:10.1007/s11103-013-0065-9.

Jiang, W., Zhou, H., Bi, H., Fromm, M., Yang, B., and Weeks, D. P. (2013). Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. *Nucleic Acids Res.* 41, e188. doi:10.1093/nar/gkt780.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821. doi:10.1126/science.1225829.

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013). RNA-programmed genome editing in human cells. *eLife* 2. doi:10.7554/eLife.00471.

Jinek, M., Jiang, F., Taylor, D. W., Sternberg, S. H., Kaya, E., Ma, E., et al. (2014). Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation. *Science* 343, 1247997. doi:10.1126/science.1247997.

Johnson, C. S., Kolevski, B., and Smyth, D. R. (2002). TRANSPARENT TESTA GLABRA2, a Trichome and Seed Coat Development Gene of Arabidopsis, Encodes a WRKY Transcription Factor. *Plant Cell* 14, 1359–1375. doi:10.1105/tpc.001404.

Johnson, M. P. (2016). Photosynthesis. *Essays Biochem.* 60, 255–273. doi:10.1042/EBC20160016.

Kajala, K., Brown, N. J., Williams, B. P., Borrill, P., Taylor, L. E., and Hibberd, J. M. (2012). Multiple Arabidopsis genes primed for recruitment into C4 photosynthesis. *Plant J.* 69, 47–56. doi:10.1111/j.1365-313X.2011.04769.x.

Kaur, N., Alok, A., Shivani, null, Kaur, N., Pandey, P., Awasthi, P., et al. (2018). CRISPR/Cas9-mediated efficient editing in phytoene desaturase (PDS) demonstrates precise manipulation in banana cv. Rasthali genome. *Funct. Integr. Genomics* 18, 89–99. doi:10.1007/s10142-017-0577-5.

Keerberg, O., Pärnik, T., Ivanova, H., Bassüner, B., and Bauwe, H. (2014). C2 photosynthesis generates about 3-fold elevated leaf CO2 levels in the C3–C4 intermediate species Flaveria pubescens. *J. Exp. Bot.* 65, 3649–3656. doi:10.1093/jxb/eru239.

Kelly, G. j., and Latzko, E. (1976). Inhibition of spinach-leaf phosphofructokinase by 2-phosphoglycollate. *FEBS Lett.* 68, 55–58. doi:10.1016/0014-5793(76)80403-6.

Keys, A. J. (2006). The re-assimilation of ammonia produced by photorespiration and the nitrogen economy of C3 higher plants. *Photosynth. Res.* 87, 165–175. doi:10.1007/s11120-005-9024-x.

Kim, S., Kim, D., Cho, S. W., Kim, J., and Kim, J.-S. (2014). Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* 24, 1012–1019. doi:10.1101/gr.171322.113.

Kim, Y. G., Cha, J., and Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci.* 93, 1156–1160.

Kleinstiver, B. P., Pattanayak, V., Prew, M. S., Tsai, S. Q., Nguyen, N., Zheng, Z., et al. (2016). High-fidelity CRISPR-Cas9 variants with undetectable genome-wide off-targets. *Nature* 529, 490–495. doi:10.1038/nature16526.

Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., and Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533, 420–424. doi:10.1038/nature17946.

Koornneef, M., Dellaert, L. W., and van der Veen, J. H. (1982). EMS- and radiation-induced mutation frequencies at individual loci in Arabidopsis thaliana (L.) Heynh. *Mutat. Res.* 93, 109–123.

Kozaki, A., and Takeba, G. (1996). Photorespiration protects C3 plants from photooxidation. *Nature* 384, 557. doi:10.1038/384557a0.

Larkin, J. C., Oppenheimer, D. G., Lloyd, A. M., Paparozzi, E. T., and Marks, M. D. (1994). Roles of the GLABROUS1 and TRANSPARENT TESTA GLABRA Genes in Arabidopsis Trichome Development. *Plant Cell* 6, 1065–1076. doi:10.1105/tpc.6.8.1065.

Lauterbach, M., Schmidt, H., Billakurthi, K., Hankeln, T., Westhoff, P., Gowik, U., et al. (2017). De novo Transcriptome Assembly and Comparison of C3, C3-C4, and C4 Species of Tribe Salsoleae (Chenopodiaceae). *Front. Plant Sci.* 8. doi:10.3389/fpls.2017.01939.

Levy, A., Goren, M. G., Yosef, I., Auster, O., Manor, M., Amitai, G., et al. (2015). CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature* 520, 505–510. doi:10.1038/nature14302.

Li, J.-F., Aach, J., Norville, J. E., McCormack, M., Zhang, D., Bush, J., et al. (2013). Multiplex and homologous recombination-mediated plant genome editing via guide RNA/Cas9. *Nat. Biotechnol.* 31, 688–691. doi:10.1038/nbt.2654.

Liepman, A. H., and Olsen, L. J. (2001). Peroxisomal alanine : glyoxylate aminotransferase (AGT1) is a photorespiratory enzyme with multiple substrates in Arabidopsis thaliana. *Plant J.* 25, 487–498. doi:10.1046/j.1365-313x.2001.00961.x.

Lim, Y., Bak, S. Y., Sung, K., Jeong, E., Lee, S. H., Kim, J.-S., et al. (2016). Structural roles of guide RNAs in the nuclease activity of Cas9 endonuclease. *Nat. Commun.* 7, 13350. doi:10.1038/ncomms13350.

Lima Neto, M. C., Cerqueira, J. V. A., da Cunha, J. R., Ribeiro, R. V., and Silveira, J. a. G. (2017). Cyclic electron flow, NPQ and photorespiration are crucial for the establishment of young plants of Ricinus communis and Jatropha curcas exposed to drought. *Plant Biol. Stuttg. Ger.* 19, 650–659. doi:10.1111/plb.12573.

Lin, H., Karki, S., Coe, R. A., Bagha, S., Khoshravesh, R., Balahadia, C. P., et al. (2016). Targeted Knockdown of GDCH in Rice Leads to a Photorespiratory-Deficient Phenotype Useful as a Building Block for C4 Rice. *Plant Cell Physiol.* 57, 919–932. doi:10.1093/pcp/pcw033.

Liu, L., Saunders, K., Thomas, C. L., Davies, J. W., and Stanley, J. (1999). Bean Yellow Dwarf Virus RepA, but Not Rep, Binds to Maize Retinoblastoma Protein, and the Virus Tolerates Mutations in the Consensus Binding Motif. *Virology* 256, 270–279. doi:10.1006/viro.1999.9616.

Liu, L., van Tonder, T., Pietersen, G., Davies, J. W., and Stanley, J. (1997). Molecular characterization of a subgroup I geminivirus from a legume in South Africa. *J. Gen. Virol.* 78 (Pt 8), 2113–2117. doi:10.1099/0022-1317-78-8-2113.

Long, S. P., Marshall-Colon, A., and Zhu, X.-G. (2015). Meeting the Global Food Demand of the Future by Engineering Crop Photosynthesis and Yield Potential. *Cell* 161, 56–66. doi:10.1016/j.cell.2015.03.019.

Lowder, L. G., Zhang, D., Baltes, N. J., Paul, J. W., Tang, X., Zheng, X., et al. (2015). A CRISPR/Cas9 Toolbox for Multiplexed Plant Genome Editing and Transcriptional Regulation. *Plant Physiol.* 169, 971–985. doi:10.1104/pp.15.00636.

Lozano-Durán, R. (2016). Geminiviruses for biotechnology: the art of parasite taming. *New Phytol.* 210, 58–64. doi:10.1111/nph.13564.

Lundgren, M. R., and Christin, P.-A. (2017). Despite phylogenetic effects, C3–C4 lineages bridge the ecological gap to C4 photosynthesis. *J. Exp. Bot.* 68, 241–254. doi:10.1093/jxb/erw451.

Ma, H., Marti-Gutierrez, N., Park, S.-W., Wu, J., Lee, Y., Suzuki, K., et al. (2017). Correction of a pathogenic gene mutation in human embryos. *Nature* 548, 413. doi:10.1038/nature23305.

Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J. J., Charpentier, E., Horvath, P., et al. (2011). Evolution and classification of the CRISPR–Cas systems. *Nat. Rev. Microbiol.* 9, 467. doi:10.1038/nrmicro2577.

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., et al. (2013). RNA-Guided Human Genome Engineering via Cas9. *Science* 339, 823–826. doi:10.1126/science.1232033.

Mallmann, J., Heckmann, D., Bräutigam, A., Lercher, M. J., Weber, A. P., Westhoff, P., et al. (2014). The role of photorespiration during the evolution of C4 photosynthesis in the genus Flaveria. *eLife* 3, e02478. doi:10.7554/eLife.02478.

Mannuss, A., Trapp, O., and Puchta, H. (2012). Gene regulation in response to DNA damage. *Biochim. Biophys. Acta* 1819, 154–165. doi:10.1016/j.bbagrm.2011.08.003.

Manova, V., and Gruszka, D. (2015). DNA damage and repair in plants – from models to crops. *Front. Plant Sci.* 6. doi:10.3389/fpls.2015.00885.

Mao, Y., Zhang, H., Xu, N., Zhang, B., Gou, F., and Zhu, J.-K. (2013). Application of the CRISPR–Cas System for Efficient Genome Engineering in Plants. *Mol. Plant* 6, 2008–2011. doi:10.1093/mp/sst121.

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

Mathiasen, D. P., and Lisby, M. (2014). Cell cycle regulation of homologous recombination in Saccharomyces cerevisiae. *FEMS Microbiol. Rev.* 38, 172–184. doi:10.1111/1574-6976.12066.

McGinn, J., and Marraffini, L. A. (2016). CRISPR-Cas Systems Optimize Their Immune Response by Specifying the Site of Spacer Integration. *Mol. Cell* 64, 616–623. doi:10.1016/j.molcel.2016.08.038.

Miao, J., Guo, D., Zhang, J., Huang, Q., Qin, G., Zhang, X., et al. (2013). Targeted mutagenesis in rice using CRISPR-Cas system. *Cell Res.* 23, 1233–1236. doi:10.1038/cr.2013.123.

Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J., and Soria, E. (2005). Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *J. Mol. Evol.* 60, 174–182. doi:10.1007/s00239-004-0046-3.

Monteith, J. L. (1978). Reassessment of Maximum Growth Rates for C3 and C4 Crops. *Exp. Agric.* 14, 1–5. doi:10.1017/S0014479700008255.

Mor, T. S., Moon, Y.-S., Palmer, K. E., and Mason, H. S. (2003). Geminivirus vectors for high-level expression of foreign proteins in plant cells. *Biotechnol. Bioeng.* 81, 430–437. doi:10.1002/bit.10483.

Morohashi, K., Zhao, M., Yang, M., Read, B., Lloyd, A., Lamb, R., et al. (2007). Participation of the Arabidopsis bHLH Factor GL3 in Trichome Initiation Regulatory Events. *Plant Physiol.* 145, 736–746. doi:10.1104/pp.107.104521.

Mortensen, U. H., Lisby, M., and Rothstein, R. (2009). Rad52. *Curr. Biol.* 19, R676–R677. doi:10.1016/j.cub.2009.06.001.

Moscou, M. J., and Bogdanove, A. J. (2009). A Simple Cipher Governs DNA Recognition by TAL Effectors. *Science* 326, 1501–1501. doi:10.1126/science.1178817.

Mouillon, J.-M., Aubert, S., Bourguignon, J., Gout, E., Douce, R., and Rébeillé, F. (1999). Glycine and serine catabolism in non-photosynthetic higher plant cells: their role in C1 metabolism. *Plant J.* 20, 197–205. doi:10.1046/j.1365-313x.1999.00591.x.

Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J. D. G., and Kamoun, S. (2013). Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* doi:10.1038/nbt.2655.

Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S., Dohmae, N., et al. (2014). Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell* 156, 935–949. doi:10.1016/j.cell.2014.02.001.

Niu, Y., Shen, B., Cui, Y., Chen, Y., Wang, J., Wang, L., et al. (2014). Generation of Gene-Modified Cynomolgus Monkey via Cas9/RNA-Mediated Gene Targeting in One-Cell Embryos. *Cell* 156, 836–843. doi:10.1016/j.cell.2014.01.027.

Norman, E. G., and Colman, B. (1991). Purification and Characterization of Phosphoglycolate Phosphatase from the Cyanobacterium Coccochloris peniocystis. *Plant Physiol.* 95, 693–698. doi:10.1104/pp.95.3.693.

Nozawa, T., Furukawa, N., Aikawa, C., Watanabe, T., Haobam, B., Kurokawa, K., et al. (2011). CRISPR Inhibition of Prophage Acquisition in Streptococcus pyogenes. *PLoS ONE* 6. doi:10.1371/journal.pone.0019543.

Nuñez, J. K., Kranzusch, P. J., Noeske, J., Wright, A. V., Davies, C. W., and Doudna, J. A. (2014). Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. *Nat. Struct. Mol. Biol.* 21, 528–534. doi:10.1038/nsmb.2820.

Nuñez, J. K., Lee, A. S. Y., Engelman, A., and Doudna, J. A. (2015). Integrase-mediated spacer acquisition during CRISPR–Cas adaptive immunity. *Nature* 519, 193–198. doi:10.1038/nature14237.

Odipio, J., Alicai, T., Ingelbrecht, I., Nusinow, D. A., Bart, R., and Taylor, N. J. (2017). Efficient CRISPR/Cas9 Genome Editing of Phytoene desaturase in Cassava. *Front. Plant Sci.* 8, 1780. doi:10.3389/fpls.2017.01780.

OECD/FAO (2017). OECD-FAO Agricultural Outlook 2017-2026. doi:http://dx.doi.org/10.1787/agr_outlook-2017-en.

Oliver, D. J., Neuburger, M., Bourguignon, J., and Douce, R. (1990). Interaction between the Component Enzymes of the Glycine Decarboxylase Multienzyme Complex. *Plant Physiol.* 94, 833–839. doi:10.1104/pp.94.2.833.

Oppenheimer, D. G., Herman, P. L., Sivakumaran, S., Esch, J., and Marks, M. D. (1991). A myb gene required for leaf trichome differentiation in Arabidopsis is expressed in stipules. *Cell* 67, 483–493. doi:10.1016/0092-8674(91)90523-2.

Orel, N., Kyryk, A., and Puchta, H. (2003). Different pathways of homologous recombination are used for the repair of double-strand breaks within tandemly arranged sequences in the plant genome. *Plant J.* 35, 604–612. doi:10.1046/j.1365-313X.2003.01832.x.

Osman, K., Higgins, J. D., Sanchez-Moran, E., Armstrong, S. J., and Franklin, F. C. H. (2011). Pathways to meiotic recombination in Arabidopsis thaliana. *New Phytol.* 190, 523–544. doi:10.1111/j.1469-8137.2011.03665.x.

Pacher, M., and Puchta, H. (2017). From classical mutagenesis to nuclease-based breeding - directing natural DNA repair for a natural end-product. *Plant J. Cell Mol. Biol.* 90, 819–833. doi:10.1111/tpj.13469.

Pares, S., Cohen-Addad, C., Sieker, L., Neuburger, M., and Douce, R. (1994). X-ray structure determination at 2.6-A resolution of a lipoate-containing protein: the H-protein of the glycine decarboxylase complex from pea leaves. *Proc. Natl. Acad. Sci. U. S. A.* 91, 4850–4853.

Paszkowski, J., Baur, M., Bogucki, A., and Potrykus, I. (1988). Gene targeting in plants. *EMBO J.* 7, 4021–4026.

Pattanaik, S., Patra, B., Singh, S. K., and Yuan, L. (2014). An overview of the gene regulatory network controlling trichome development in the model plant, Arabidopsis. *Front. Plant Sci.* 5. doi:10.3389/fpls.2014.00259.

Paulus, J. K., Schlieper, D., and Groth, G. (2013). Greater efficiency of photosynthetic carbon fixation due to single amino-acid substitution. *Nat. Commun.* 4, 1518. doi:10.1038/ncomms2504.

Peterhansel, C., Horst, I., Niessen, M., Blume, C., Kebeish, R., Kürkcüoglu, S., et al. (2010). Photorespiration. *Arab. Book Am. Soc. Plant Biol.* 8. doi:10.1199/tab.0130.

Piatek, A., Ali, Z., Baazim, H., Li, L., Abulfaraj, A., Al-Shareef, S., et al. (2015). RNA-guided transcriptional regulation in planta via synthetic dCas9-based transcription factors. *Plant Biotechnol. J.* 13, 578–589. doi:10.1111/pbi.12284.

Pick, T. R., Bräutigam, A., Schulz, M. A., Obata, T., Fernie, A. R., and Weber, A. P. M. (2013). PLGG1, a plastidic glycolate glycerate transporter, is required for photorespiration and defines a unique class of metabolite transporters. *Proc. Natl. Acad. Sci. U. S. A.* 110, 3185–3190. doi:10.1073/pnas.1215142110.

Puchta, H. (1998). Repair of genomic double-strand breaks in somatic plant cells by one-sided invasion of homologous sequences. *Plant J.* 13, 331–339. doi:10.1046/j.1365-313X.1998.00035.x.

Puchta, H. (1999). Double-Strand Break-Induced Recombination Between Ectopic Homologous Sequences in Somatic Plant Cells. *Genetics* 152, 1173–1181.

Puchta, H. (2005). The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. *J. Exp. Bot.* 56, 1–14. doi:10.1093/jxb/eri025.

Puchta, H., Dujon, B., and Hohn, B. (1993). Homologous recombination in plant cells is enhanced by in vivo induction of double strand breaks into DNA by a site-specific endonuclease. *Nucleic Acids Res.* 21, 5034–5040.

Puchta, H., Dujon, B., and Hohn, B. (1996). Two different but related mechanisms are used in plants for the repair of genomic double-strand breaks by homologous recombination. *Proc. Natl. Acad. Sci. U. S. A.* 93, 5055–5060.

Puchta, H., and Fauser, F. (2013). Gene targeting in plants: 25 years later. *Int. J. Dev. Biol.* 57, 629–637. doi:10.1387/ijdb.130194hp.

Puchta, H., and Fauser, F. (2014). Synthetic nucleases for genome engineering in plants: prospects for a bright future. *Plant J.* 78, 727–741. doi:10.1111/tpj.12338.

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., et al. (2013). Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* 152, 1173–1183. doi:10.1016/j.cell.2013.02.022.

Queval, G., Issakidis-Bourguet, E., Hoeberichts, F. A., Vandorpe, M., Gakière, B., Vanacker, H., et al. (2007). Conditional oxidative stress responses in the Arabidopsis photorespiratory mutant cat2 demonstrate that redox state is a key modulator of daylength-dependent gene expression, and define photoperiod as a crucial factor in the regulation of H2O2-induced cell death. *Plant J. Cell Mol. Biol.* 52, 640–657. doi:10.1111/j.1365-313X.2007.03263.x.

Rachmilevitch, S., Cousins, A. B., and Bloom, A. J. (2004). Nitrate assimilation in plant shoots depends on photorespiration. *Proc. Natl. Acad. Sci. U. S. A.* 101, 11506–11510. doi:10.1073/pnas.0404388101.

Rademacher, N., Kern, R., Fujiwara, T., Mettler-Altmann, T., Miyagishima, S., Hagemann, M., et al. (2016). Photorespiratory glycolate oxidase is essential for the survival of the red alga Cyanidioschyzon merolae under ambient CO2 conditions. *J. Exp. Bot.* 67, 3165–3175. doi:10.1093/jxb/erw118.

Ramirez, C. L., Foley, J. E., Wright, D. A., Müller-Lerch, F., Rahman, S. H., Cornu, T. I., et al. (2008). Unexpected failure rates for modular assembly of engineered zinc fingers. *Nat. Methods* 5, 374. doi:10.1038/nmeth0508-374.

Ran, F. A., Hsu, P. D., Lin, C.-Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., et al. (2013). Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell* 154, 1380–1389. doi:10.1016/j.cell.2013.08.021.

Rath, D., Amlinger, L., Rath, A., and Lundgren, M. (2015). The CRISPR-Cas immune system: Biology, mechanisms and applications. *Biochimie* 117, 119–128. doi:10.1016/j.biochi.2015.03.025.

Rawsthorne, S., Hylton, C. M., Smith, A. M., and Woolhouse, H. W. (1988). Photorespiratory metabolism and immunogold localization of photorespiratory enzymes in leaves of C3 and C3-C4 intermediate species of Moricandia. *Planta* 173, 298–308. doi:10.1007/BF00401016.

Rawsthorne, S., Morgan, C. L., O'Neill, C. M., Hylton, C. M., Jones, D. A., and Frean, M. L. (1998). Cellular expression pattern of the glycine decarboxylase P protein in leaves of an intergeneric hybrid between the C3-C4 intermediate species Moricandia nitens and the C3 species Brassica napus. *Theor. Appl. Genet.* 96, 922–927. doi:10.1007/s001220050821.

Ray, D. K., Mueller, N. D., West, P. C., and Foley, J. A. (2013). Yield Trends Are Insufficient to Double Global Crop Production by 2050. *PLOS ONE* 8, e66428. doi:10.1371/journal.pone.0066428.

Rerie, W. G., Feldmann, K. A., and Marks, M. D. (1994). The GLABRA2 gene encodes a homeo domain protein required for normal trichome development in Arabidopsis. *Genes Dev.* 8, 1388–1399.

Sage, R. F. (2004). The evolution of C4 photosynthesis. *New Phytol.* 161, 341–370. doi:10.1111/j.1469-8137.2004.00974.x.

Sage, R. F. (2013). Photorespiratory compensation: a driver for biological diversity. *Plant Biol.* 15, 624–638. doi:10.1111/plb.12024.

Sage, R. F. (2017). A portrait of the C4 photosynthetic family on the 50th anniversary of its discovery: species number, evolutionary lineages, and Hall of Fame. *J. Exp. Bot.* 68, e11–e28. doi:10.1093/jxb/erx005.

Sage, R. F., and Khoshravesh, R. (2016). Passive CO2 concentration in higher plants. *Curr. Opin. Plant Biol.* 31, 58–65. doi:10.1016/j.pbi.2016.03.016.

Sage, R. F., Khoshravesh, R., and Sage, T. L. (2014). From proto-Kranz to C4 Kranz: building the bridge to C4 photosynthesis. *J. Exp. Bot.* 65, 3341–3356. doi:10.1093/jxb/eru180.

Sage, R. F., Sage, T. L., and Kocacinar, F. (2012). Photorespiration and the evolution of C4 photosynthesis. *Annu. Rev. Plant Biol.* 63, 19–47. doi:10.1146/annurev-arplant-042811-105511.

Sage, T. L., and Sage, R. F. (2009). The functional anatomy of rice leaves: implications for refixation of photorespiratory CO2 and efforts to engineer C4 photosynthesis into rice. *Plant Cell Physiol.* 50, 756–772. doi:10.1093/pcp/pcp033.

Saleh-Gohari, N., and Helleday, T. (2004). Conservative homologous recombination preferentially repairs DNA double-strand breaks in the S phase of the cell cycle in human cells. *Nucleic Acids Res.* 32, 3683–3688. doi:10.1093/nar/gkh703.

Salomon, S., and Puchta, H. (1998). Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. *EMBO J.* 17, 6086–6095. doi:10.1093/emboj/17.20.6086.

Samach, A., Melamed-Bessudo, C., Avivi-Ragolski, N., Pietrokovski, S., and Levy, A. A. (2011). Identification of Plant RAD52 Homologs and Characterization of the Arabidopsis thaliana RAD52-Like Genes. *Plant Cell* 23, 4266–4279. doi:10.1105/tpc.111.091744.

Sargent, R. G., Brenneman, M. A., and Wilson, J. H. (1997). Repair of site-specific double-strand breaks in a mammalian chromosome by homologous and illegitimate recombination. *Mol. Cell. Biol.* 17, 267–277.

Sauer, N. J., Narváez-Vásquez, J., Mozoruk, J., Miller, R. B., Warburg, Z. J., Woodward, M. J., et al. (2016). Oligonucleotide-Mediated Genome Editing Provides Precision and Function to Engineered Nucleases and Antibiotics in Plants. *Plant Physiol.* 170, 1917–1928. doi:10.1104/pp.15.01696.

Schellmann, S., Schnittger, A., Kirik, V., Wada, T., Okada, K., Beermann, A., et al. (2002). TRIPTYCHON and CAPRICE mediate lateral inhibition during trichome and root hair patterning in Arabidopsis. *EMBO J.* 21, 5036–5046. doi:10.1093/emboj/cdf524.

Schimkat, D., Heineke, D., and Heldt, H. W. (1990). Regulation of sedoheptulose-1,7-bisphosphatase by sedoheptulose-7-phosphate and glycerate, and of fructose-1,6-bisphosphatase by glycerate in spinach chloroplasts. *Planta* 181, 97–103. doi:10.1007/BF00202330.

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

Schiml, S., Fauser, F., and Puchta, H. (2017). CRISPR/Cas-Mediated In Planta Gene Targeting. *Methods Mol. Biol. Clifton NJ* 1610, 3–11. doi:10.1007/978-1-4939-7003-2_1.

Schlüter, U., Bräutigam, A., Gowik, U., Melzer, M., Christin, P.-A., Kurz, S., et al. (2017). Photosynthesis in C3–C4 intermediate Moricandia species. *J. Exp. Bot.* 68, 191–206. doi:10.1093/jxb/erw391.

Schlüter, U., and Weber, A. P. M. (2016). The Road to C4 Photosynthesis: Evolution of a Complex Trait via Intermediary States. *Plant Cell Physiol.* 57, 881–889. doi:10.1093/pcp/pcw009.

Schulze, S., Mallmann, J., Burscheidt, J., Koczor, M., Streubel, M., Bauwe, H., et al. (2013). Evolution of C4 Photosynthesis in the Genus Flaveria: Establishment of a Photorespiratory CO2 Pump. *Plant Cell* 25, 2522–2535. doi:10.1105/tpc.113.114520.

Schulze, S., Westhoff, P., and Gowik, U. (2016). Glycine decarboxylase in C3, C4 and C3–C4 intermediate species. *Curr. Opin. Plant Biol.* 31, 29–35. doi:10.1016/j.pbi.2016.03.011.

Schwarte, S., and Bauwe, H. (2007). Identification of the Photorespiratory 2-Phosphoglycolate Phosphatase, PGLP1, in Arabidopsis. *Plant Physiol.* 144, 1580–1586. doi:10.1104/pp.107.099192.

Sfeir, A., and Symington, L. S. (2015). Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway? *Trends Biochem. Sci.* 40, 701–714. doi:10.1016/j.tibs.2015.08.006.

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., et al. (2013). Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat. Biotechnol.* doi:10.1038/nbt.2650.

Sharkey, T. D. (1988). Estimating the rate of photorespiration in leaves. *Physiol. Plant.* 73, 147–152. doi:10.1111/j.1399-3054.1988.tb09205.x.

Shen, H., Strunks, G. D., Klemann, B. J. P. M., Hooykaas, P. J. J., and de Pater, S. (2016). CRISPR/Cas9-Induced Double-Strand Break Repair in Arabidopsis Nonhomologous End-Joining Mutants. *G3 GenesGenomesGenetics* 7, 193–202. doi:10.1534/g3.116.035204.

Siebert, R., and Puchta, H. (2002). Efficient Repair of Genomic Double-Strand Breaks by Homologous Recombination between Directly Repeated Sequences in the Plant Genome. *Plant Cell* 14, 1121–1131. doi:10.1105/tpc.001727.

Slack, C. R., and Hatch, M. D. (1967). Comparative studies on the activity of carboxylases and other enzymes in relation to the new pathway of photosynthetic carbon dioxide fixation in tropical grasses. *Biochem. J.* 103, 660–665.

Slaymaker, I. M., Gao, L., Zetsche, B., Scott, D. A., Yan, W. X., and Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. *Science* 351, 84–88. doi:10.1126/science.aad5227.

Smith, J., Grizot, S., Arnould, S., Duclert, A., Epinat, J.-C., Chames, P., et al. (2006). A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. *Nucleic Acids Res.* 34, e149–e149. doi:10.1093/nar/gkl720.

Somerville, C. R., and Ogren, W. L. (1979). A phosphoglycolate phosphatase-deficient mutant of Arabidopsis. *Nature* 280, 833. doi:10.1038/280833a0.

Stata, M., Sage, T. L., Hoffmann, N., Covshoff, S., Ka-Shu Wong, G., and Sage, R. F. (2016). Mesophyll Chloroplast Investment in C3, C4 and C2 Species of the Genus Flaveria. *Plant Cell Physiol.* 57, 904–918. doi:10.1093/pcp/pcw015.

Steinert, J., Schiml, S., Fauser, F., and Puchta, H. (2015). Highly efficient heritable plant genome engineering using Cas9 orthologues from Streptococcus thermophilus and Staphylococcus aureus. *Plant J.* 84, 1295–1305. doi:10.1111/tpj.13078.

Stenger, D. C., Revington, G. N., Stevenson, M. C., and Bisaro, D. M. (1991). Replicational release of geminivirus genomes from tandemly repeated copies: evidence for rolling-circle replication of a plant viral DNA. *Proc. Natl. Acad. Sci. U. S. A.* 88, 8029–8033.

Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J., and Stahl, F. W. (1983). The double-strand-break repair model for recombination. *Cell* 33, 25–35.

Tang, X., Lowder, L. G., Zhang, T., Malzahn, A. A., Zheng, X., Voytas, D. F., et al. (2017). A CRISPR–Cpf1 system for efficient genome editing and transcriptional repression in plants. *Nat. Plants* 3, 17018. doi:10.1038/nplants.2017.18.

Tian, S., Jiang, L., Gao, Q., Zhang, J., Zong, M., Zhang, H., et al. (2017). Efficient CRISPR/Cas9-based gene knockout in watermelon. *Plant Cell Rep.* 36, 399–406. doi:10.1007/s00299-016-2089-5.

Timm, S., and Bauwe, H. (2013). The variety of photorespiratory phenotypes – employing the current status for future research directions on photorespiration. *Plant Biol.* 15, 737–747. doi:10.1111/j.1438-8677.2012.00691.x.

Timm, S., Giese, J., Engel, N., Wittmiß, M., Florian, A., Fernie, A. R., et al. (2017). T-protein is present in large excess over the other proteins of the glycine cleavage system in leaves of Arabidopsis. *Planta*, 1–11. doi:10.1007/s00425-017-2767-8.

Timm, S., Nunes-Nesi, A., Pärnik, T., Morgenthal, K., Wienkoop, S., Keerberg, O., et al. (2008). A Cytosolic Pathway for the Conversion of Hydroxypyruvate to Glycerate during Photorespiration in Arabidopsis. *Plant Cell* 20, 2848–2859. doi:10.1105/tpc.108.062265.

Timmermans, M. C., Das, O. P., and Messing, J. (1992). Trans replication and high copy numbers of wheat dwarf virus vectors in maize cells. *Nucleic Acids Res.* 20, 4047–4054.

van Houte, S., Ekroth, A. K. E., Broniewski, J. M., Chabas, H., Ashby, B., Bondy-Denomy, J., et al. (2016). The diversity-generating benefits of a prokaryotic adaptive immune system. *Nature* 532, 385–388. doi:10.1038/nature17436.

Van Orden, M. J., Klein, P., Babu, K., Najar, F. Z., and Rajan, R. (2017). Conserved DNA motifs in the type II-A CRISPR leader region. *PeerJ* 5. doi:10.7717/peerj.3161.

Vogan, P. J., Frohlich, M. W., and Sage, R. F. (2007). The functional significance of C3–C4 intermediate traits in Heliotropium L. (Boraginaceae): gas exchange perspectives. *Plant Cell Environ.* 30, 1337–1345. doi:10.1111/j.1365-3040.2007.01706.x.

Vogan, P. J., and Sage, R. F. (2012). Effects of low atmospheric CO2 and elevated temperature during growth on the gas exchange responses of C3, C3–C4 intermediate, and C4 species from three evolutionary lineages of C4 photosynthesis. *Oecologia* 169, 341–352. doi:10.1007/s00442-011-2201-z.

Vojta, A., Dobrinić, P., Tadić, V., Bočkor, L., Korać, P., Julg, B., et al. (2016). Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Res.* 44, 5615–5628. doi:10.1093/nar/gkw159.

Voll, L. M., Jamai, A., Renné, P., Voll, H., McClung, C. R., and Weber, A. P. M. (2006). The Photorespiratory Arabidopsis shm1 Mutant Is Deficient in SHM1. *Plant Physiol.* 140, 59–66. doi:10.1104/pp.105.071399.

Voytas, D. F. (2013). Plant genome engineering with sequence-specific nucleases. *Annu. Rev. Plant Biol.* 64, 327–350. doi:10.1146/annurev-arplant-042811-105552.

Wada, T., Kurata, T., Tominaga, R., Koshino-Kimura, Y., Tachibana, T., Goto, K., et al. (2002). Role of a positive regulator of root hair development, CAPRICE, in Arabidopsis root epidermal cell differentiation. *Development* 129, 5409–5419. doi:10.1242/dev.00111.

Walbot, V. (1996). Sources and consequences of phenotypic and genotypic plasticity in flowering plants. *Trends Plant Sci.* 1, 27–32. doi:10.1016/S1360-1385(96)80020-3.

Walker, A. R., Davison, P. A., Bolognesi-Winfield, A. C., James, C. M., Srinivasan, N., Blundell, T. L., et al. (1999). The TRANSPARENT TESTA GLABRA1 locus, which regulates trichome differentiation and anthocyanin biosynthesis in Arabidopsis, encodes a WD40 repeat protein. *Plant Cell* 11, 1337–1350.

Walker, B. J., VanLoocke, A., Bernacchi, C. J., and Ort, D. R. (2016). The Costs of Photorespiration to Food Production Now and in the Future. *Annu. Rev. Plant Biol.* 67, 107–129. doi:10.1146/annurev-arplant-043015-111709.

Walker, J. R., Corpina, R. A., and Goldberg, J. (2001). Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature* 412, 607. doi:10.1038/35088000.

Wang, H., and Xu, X. (2017). Microhomology-mediated end joining: new players join the team. *Cell Biosci.* 7, 6. doi:10.1186/s13578-017-0136-8.

Wang, H., Yang, H., Shivalila, C. S., Dawlaty, M. M., Cheng, A. W., Zhang, F., et al. (2013). One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. *Cell* 153, 910–918. doi:10.1016/j.cell.2013.04.025.

Wang, M., Lu, Y., Botella, J. R., Mao, Y., Hua, K., and Zhu, J.-K. (2017). Gene Targeting by Homology-Directed Repair in Rice Using a Geminivirus-Based CRISPR/Cas9 System. *Mol. Plant* 10, 1007–1010. doi:10.1016/j.molp.2017.03.002.

Wang, Y., Zhang, Z.-T., Seo, S.-O., Lynn, P., Lu, T., Jin, Y.-S., et al. (2016). Gene transcription repression in Clostridium beijerinckii using CRISPR-dCas9. *Biotechnol. Bioeng.* 113, 2739–2743. doi:10.1002/bit.26020.

Waterworth, W. M., Drury, G. E., Bray, C. M., and West, C. E. (2011). Repairing breaks in the plant genome: the importance of keeping it together. *New Phytol.* 192, 805–822. doi:10.1111/j.1469-8137.2011.03926.x.

Wei, Y., Chesne, M. T., Terns, R. M., and Terns, M. P. (2015). Sequences spanning the leader-repeat junction mediate CRISPR adaptation to phage in Streptococcus thermophilus. *Nucleic Acids Res.* 43, 1749–1758. doi:10.1093/nar/gku1407.

West, C. E., Waterworth, W. M., Jiang, Q., and Bray, C. M. (2000). Arabidopsis DNA ligase IV is induced by γ-irradiation and interacts with an Arabidopsis homologue of the double strand break repair protein XRCC4. *Plant J.* 24, 67–78. doi:10.1046/j.1365-313x.2000.00856.x.

Williams, B. P., Burgess, S. J., Reyna-Llorens, I., Knerova, J., Aubry, S., Stanley, S., et al. (2016). An Untranslated cis-Element Regulates the Accumulation of Multiple C4 Enzymes in Gynandropsis gynandra Mesophyll Cells. *Plant Cell* 28, 454–465. doi:10.1105/tpc.15.00570.

Wiludda, C., Schulze, S., Gowik, U., Engelmann, S., Koczor, M., Streubel, M., et al. (2012). Regulation of the Photorespiratory GLDPA Gene in C4 Flaveria: An Intricate Interplay of Transcriptional and Posttranscriptional Processes. *Plant Cell* 24, 137–151. doi:10.1105/tpc.111.093872.

Winzer, T., Heineke, D., and Bauwe, H. (2001). Growth and phenotype of potato plants expressing an antisense gene of P-protein of glycine decarboxylase under control of a promoter with preference for the mesophyll. *Ann. Appl. Biol.* 138, 9–15. doi:10.1111/j.1744-7348.2001.tb00080.x.

Wright, A. V., and Doudna, J. A. (2016). Protecting genome integrity during CRISPR immune adaptation. *Nat. Struct. Mol. Biol.* 23, 876. doi:10.1038/nsmb.3289.

Yang, C., and Ye, Z. (2013). Trichomes as models for studying plant cell differentiation. *Cell. Mol. Life Sci.* 70, 1937–1948. doi:10.1007/s00018-012-1147-6.

Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., et al. (2015). Cpf1 is a single RNA-guided endonuclease of a Class 2 CRISPR-Cas system. *Cell* 163, 759–771. doi:10.1016/j.cell.2015.09.038.

Zhang, F., Gonzalez, A., Zhao, M., Payne, C. T., and Lloyd, A. (2003). A network of redundant bHLH proteins functions in all TTG1-dependent pathways of Arabidopsis. *Development* 130, 4859–4869. doi:10.1242/dev.00681.

Zhao, M., Morohashi, K., Hatlestad, G., Grotewold, E., and Lloyd, A. (2008). The TTG1-bHLH-MYB complex controls trichome cell fate and patterning through direct targeting of regulatory loci. *Development* 135, 1991–1999. doi:10.1242/dev.016873.

Zhao, Y., Zhang, C., Liu, W., Gao, W., Liu, C., Song, G., et al. (2016). An alternative strategy for targeted gene replacement in plants using a dual-sgRNA/Cas9 design. *Sci. Rep.* 6, 23890. doi:10.1038/srep23890.

Zhu, X.-G., Long, S. P., and Ort, D. R. (2010). Improving Photosynthetic Efficiency for Greater Yield. *Annu. Rev. Plant Biol.* 61, 235–261. doi:10.1146/annurev-arplant-042809-112206.

Zong, Y., Wang, Y., Li, C., Zhang, R., Chen, K., Ran, Y., et al. (2017). Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nat. Biotechnol.* 35, 438. doi:10.1038/nbt.3811.

AIM OF THE THESIS

Targeted genome modifications are helpful to design high yielding crop varieties but also valuable for basic research. The CRISPR/Cas9 system has revolutionized the field of genome editing and opened manifold new possibilities for plant scientists to manipulate the genome in a targeted manner. A small, programmable guide RNA is used to target the bacterial Cas9 nuclease to any genomic region of choice for introduction of double strand breaks (Jinek et al., 2012). The internal double strand break repair machinery of the plant cell is exploited to introduce mutations at the target locus. These changes can range from small indel mutations due to error-prone non-homologous end joining repair to integration of new genomic sequences by gene targeting (Pacher and Puchta, 2017). Optimization of the CRISPR/Cas9 system for biotechnological use is ongoing and new applications occur frequently. However, non-destructive endogenous visual markers to establish these new techniques in plants are scarce.

The aim of this thesis was to establish the trichome master regulator GLABROUS1 as an efficient visual marker for tracking of Cas9-induced mutations in *Arabidopsis thaliana*. Therefore, a generic vector system for targeted gene knockouts in *Arabidopsis* using the CRISPR/Cas9 system from *S. pyogenes* had to be developed (Manuscript I and II). In a second step (Manuscript III), the GLABROUS1 marker was used for evaluation of two previously described gene targeting systems (Baltes et al., 2014; Schiml et al., 2014). The final project (Manuscript IV) was the transfer of the knowledge gained from these experiments to a biological problem, namely studying the effects of Cas9-induced modifications in the *Arabidopsis thaliana GLYCINE DECARBOXYLASE P-PROTEIN1* promoter.

References

Baltes, N. J., Gil-Humanes, J., Cermak, T., Atkins, P. A., and Voytas, D. F. (2014). DNA replicons for plant genome engineering. *Plant Cell* 26, 151–163. doi:10.1105/tpc.113.119792.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821. doi:10.1126/science.1225829.

Pacher, M., and Puchta, H. (2017). From classical mutagenesis to nuclease-based breeding - directing natural DNA repair for a natural end-product. *Plant J. Cell Mol. Biol.* 90, 819–833. doi:10.1111/tpj.13469.

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

MANUSCRIPT I

An Efficient Visual Screen for CRISPR/Cas9 Activity in *Arabidopsis thaliana*

frontiers in Plant Science

ORIGINAL RESEARCH
published: 24 January 2017
doi: 10.3389/fpls.2017.00039

An Efficient Visual Screen for CRISPR/Cas9 Activity in *Arabidopsis thaliana*

*Florian Hahn[1], Otho Mantegazza[1], André Greiner[2], Peter Hegemann[2], Marion Eisenhut[1] and Andreas P. M. Weber[1]**

[1] Institute of Plant Biochemistry, Cluster of Excellence on Plant Science, Center for Synthetic Life Sciences, Heinrich Heine University, Düsseldorf, Germany, [2] Institute of Biology, Experimental Biophysics, Humboldt-Universität zu Berlin, Berlin, Germany

The CRISPR/Cas9 system enables precision editing of the genome of the model plant *Arabidopsis thaliana* and likely of any other organism. Tools and methods for further developing and optimizing this widespread and versatile system in *Arabidopsis* would hence be welcomed. Here, we designed a generic vector system that can be used to clone any sgRNA sequence in a plant T-DNA vector containing an ubiquitously expressed *Cas9* gene. With this vector, we explored two alternative marker systems for tracking Cas9-mediated gene-editing *in vivo*: *BIALAPHOS RESISTANCE* (*BAR*) and *GLABROUS1* (*GL1*). *BAR* confers resistance to glufosinate and is widely used as a positive selection marker; *GL1* is required for the formation of trichomes. Reversion of a frameshift null *BAR* allele to a functional one by Cas9-mediated gene editing yielded a higher than expected number of plants that are resistant to glufosinate. Surprisingly, many of those plants did not display reversion of the *BAR* gene through the germline. We hypothesize that few *BAR* revertant cells in a highly chimeric plant likely provide system-wide resistance to glufosinate and thus we suggest that BAR is not suitable as marker for tracking Cas9-mediated gene-editing. Targeting the *GL1* gene for disruption with Cas9 provided clearly visible phenotypes of partially and completely glabrous plants. 50% of the analyzed T1 plants produced descendants with a chimeric phenotype and we could recover fully homozygous plants in the T3 generation with high efficiency. We propose that targeting of *GL1* is suitable for assessing and optimizing Cas9-mediated gene-editing in *Arabidopsis*.

Keywords: CRISPR/Cas9, trichome, Glabrous1, Glufosinate, BASTA, gene editing, marker

OPEN ACCESS

Edited by:
Soren K. Rasmussen, University of Copenhagen, Denmark

Reviewed by:
Tejinder Kumar Mall, Dow Chemical Company, USA
Eva Stoger, University of Natural Resources and Life Sciences, Vienna, Austria

****Correspondence:***
Andreas P. M. Weber
andreas.weber@uni-duesseldorf.de

Specialty section:
This article was submitted to Plant Biotechnology, a section of the journal Frontiers in Plant Science

Received: *08 November 2016*
Accepted: *09 January 2017*
Published: *24 January 2017*

Citation:
Hahn F, Mantegazza O, Greiner A, Hegemann P, Eisenhut M and Weber APM (2017) An Efficient Visual Screen for CRISPR/Cas9 Activity in Arabidopsis thaliana. Front. Plant Sci. 8:39. doi: 10.3389/fpls.2017.00039

INTRODUCTION

A growing world population facing increasingly scarce arable land would benefit from applying synthetic biology approaches to generate rationally designed plants with low effort and within short time frames (Baltes and Voytas, 2015). Recently, it became possible to modify plant genomes by introduction of double-strand breaks (DSBs) at a locus of interest, using site-specific custom nucleases. DSBs are lethal for the cell and must be repaired by either homologous recombination (HR) or non-homologous end joining (NHEJ), which are carried out by plant internal DNA break repair mechanisms (Puchta and Fauser, 2014). Both mechanisms can be exploited for targeted

genome editing. While HR enables the targeted integration of sequences of choice into the DSB mediated by homologous regions between a DNA donor template and the target region, the error-prone NHEJ can be exploited to destroy open reading frames of genes and thus to disrupt gene function (Salomon and Puchta, 1998). The development of the CRISPR/Cas9 (hereafter Cas9) system, which relies on the bacterial Cas9 nuclease from *Streptococcus pyogenes* and a single guide RNA (sgRNA) molecule that directs the nuclease to its specific target (Jinek et al., 2012) has greatly reduced the effort to create DSBs at specific loci in the genome of living cells (Doudna and Charpentier, 2014). The Cas9 system allows the creation of DSBs in virtually any gene and even in several genes simultaneously. It has already been successfully applied in many model and crop plants species (Bortesi and Fischer, 2015). Cas9 has quickly established itself as a powerful and versatile technology and strong efforts are made on developing and optimizing Cas9 related tools and methods. However, markers as indicators for the successful mutagenesis are still scarce.

A commonly used marker for positive selection of transgenic plants is resistance to the broad-spectrum herbicide glufosinate also known as phosphinothricin or BASTA® (Deblock et al., 1987). Glufosinate is a glutamate analog that inhibits glutamine synthetase *in planta* and thus hinders the conversion of glutamate and ammonia to glutamine (Bayer et al., 1972; Lea et al., 1984). The accumulation of ammonia in the plant inhibits photosynthetic reactions and uncouples photophosphorylation, finally resulting in cell death (Tachibana et al., 1986; Sauer et al., 1987; Wild et al., 1987). Resistance to glufosinate is conferred by the bialaphos resistance gene (*BAR*) from *Streptomyces hygroscopicus*, which encodes the enzyme phosphinotricin acetyl transferase (PAT). PAT acetylates the amino group of glufosinate and thereby inactivates it (Thompson et al., 1987; Mullner et al., 1993). Since the *BAR* gene has been highly successfully applied as positive selection marker for screening of T-DNA insertion lines, a *BAR*/glufosinate based screening could be implemented also for assessing Cas9 based gene-editing in plants.

The genes that control trichome development could be used as alternative non-invasive markers for visually tracking mutations in the model plant *Arabidopsis thaliana*. Trichomes are specialized epidermal cells, which are found on most aerial parts of the plant. They function in protection against UV light, herbivores and pathogen attack, extreme temperatures, and water loss (reviewed in Yang and Ye, 2013; Pattanaik et al., 2014). Importantly, they are not essential for plant growth and development under laboratory conditions (Oppenheimer et al., 1991). The development and distribution of trichomes on the leaf surface are highly regulated processes. In *Arabidopsis*, trichome initiation starts very early during leaf development with an increase in cell and nuclear size by at least four rounds of endoreplication. The trichome cell then elongates and branches. While the first trichomes form at the distal region of the leaf primordium, subsequent trichomes develop at the more basal region of the leaf primordium and in between the mature trichomes on the distal parts (Larkin et al., 1994, 1996; Marks, 1994). Trichome development is strongly reduced by knocking out the gene for the R2R3-MYB transcriptional master regulator GLABROUS1 (GL1) (Koornneef et al., 1982). Like other plant MYB proteins, GL1 contains two MYB DNA-binding domain repeats at the N-terminus and an acidic C-terminus (Larkin et al., 1993). GL1 interacts with the WD40-repeat factor TTG1 and the bHLH factor GL3/EGL3 to form a MYB/bHLH/WD-repeat complex. This complex activates the expression of numerous downstream activators, such as TTG2 and GL2, which allow epidermal cells to differentiate into trichomes (Yang and Ye, 2013; Pattanaik et al., 2014). Homozygous *gl1* mutants have been described to be completely or partly glabrous or showing at least a reduced number of trichomes (Koornneef et al., 1982), depending on the strength of the mutational effect. Mutations in the MYB DNA-binding domains seem to produce the most drastic phenotypes (Hauser et al., 2001).

In this report, we tested and compared the positive selection marker BAR, which confers resistance to glufosinate, and the non-pleiotropic visual marker GL1 as markers for Cas9 activity in *Arabidopsis*. Therefore, we first created a vector system that can be used to clone any sgRNA into a plant T-DNA vector that contains the *Cas9* gene under the control of the *UBIQUITIN10* promoter. Using this vector system, we were able to repair an inactive *bar* allele that carries a premature stop codon in *Arabidopsis* using Cas9 and to positively select plants by glufosinate treatment. However, low incidence of the repaired allele in the leaves of the plants resistant to BASTA suggests that few cells producing a functional BAR protein are sufficient to rescue the whole plant by conferring some degree of system-wide resistance. This low stringency of the selection system did not allow detection of homozygous mutants in the following generations. We were also able to knock out the *GL1* gene using the Cas9 technique. With a *GL1* specific sgRNA we generated several independent glabrous *Arabidopsis* lines. Genotyping showed that small InDels at the *GL1* target site caused shifts in its reading frame and premature stop codons resulting in a disruption of the GL1 protein function. With the *gl1* phenotype, we provide an efficient visual screen that can be used for testing mutagenesis efficiency under various experimental conditions in further studies.

MATERIALS AND METHODS

Plant Growth Conditions

Arabidopsis Col-0 seeds were surface-sterilized and stratified for 3 days at 4°C. Seeds were germinated on 0.8% (w/v) agar-solidified half-strength MS medium (Duchefa, http://www.duchefa.com) supplemented with 1% (w/v) sucrose under long day conditions (16-h-light/8-h-dark cycle, 22/18°C) in growth chambers with a light intensity of 100 μmol photons m^{-2} s^{-1}. 14-day-old seedlings were transferred to soil and grown under the same conditions.

Isolation of a Homozygous *gl1* T-DNA Insertion Line

T-DNA insertion lines are a common way to analyze the impact of gene knockouts in *Arabidopsis*. The T-DNA insertion line *gl1* (SAIL_1149_D03) (Sessions et al., 2002) that contains a T-DNA

integration in between the first and second exon of the *GL1* gene was obtained from the Arabidopsis Biological Resource Centre. Homozygous T-DNA insertion plants were identified using the gene specific primer pair FH214 and FH215 (Supplementary Table S1) as well as the T-DNA specific primer pair FH215 and P49 (Supplementary Table S1). Homozygosity was additionally confirmed by visual control of trichome loss. Homozygous *gl1* plants were further propagated.

Generation of a Binary T-DNA Vector with the Null *bar-1* Allele

To construct a binary T-DNA vector with a null *bar-1* allele, we first amplified the *BAR* gene from the vector pUB-Dest (Grefen et al., 2010) with the primer pair FH5/FH6 and cloned the PCR product into pDONR207 via BP reaction (Invitrogen) resulting in the vector pFH24. We then created an internal stop codon in the *BAR* gene via overlapping PCR. Therefore, we first amplified the 5′-end of the *BAR* gene with the primer pair FH25/FH26 and the 3′-end with primer pair FH27/FH28 from pFH24 and then used both PCR products as PCR template with the primer pair FH25/FH28. The PCR product was subcloned into pJET1.2 (ThermoFisher Scientific) resulting in the vector pFH10. Finally, we digested the plant binary T-DNA vector pUT-Kan (Fendrych et al., 2014) with PstI and BamHI, PCR-amplified the *BAR* gene with the internal stop codon from pFH10 with the primer pair FH39/FH40 and assembled the final vector pFH17 via Gibson cloning (NewEngland Biolabs). All primer sequences are given in Supplementary Table S1. The sequence of pFH17 can be found on GenBank (accession number KY080690).

Cas9 and sgRNA Plasmid Construction

We aimed at generating a single T-DNA vector for expression of the Cas9 protein and the sgRNA cassette. Therefore, a *Chlamydomonas reinhardtii* codon-optimized *Sp*Cas9 was amplified with the primers NH117/NH119 (Supplementary Table S1) from the vector CasYFP and cloned under the control of a *UBIQUTIN10* promoter (*UB10p*) from *Arabidopsis* into the vector pKB65 (Bernhardt, 2012) via Gateway® Cloning (Invitrogen) resulting in the vector pUB-Cas9. For sgRNA expression, the *Arabidopsis U6-26* promoter (*U6-26p*) was PCR-amplified with an additional NheI restriction site from *Arabidopsis* genomic DNA using the primers FH14 and FH15 (Supplementary Table S1). The PCR fragment was subcloned into the pGEM T-easy™ vector (Promega) resulting in the vector pFH13. The sgRNA scaffold was amplified from the vector pKS Chlamy dual with the primers FH16 and FH18 (Supplementary Table S1) and subcloned into pJET1.2 (ThermoFisher Scientific) resulting in the vector pFH4. To create a sgRNA expression cassette with the sgRNA scaffold and the *U6-26p*, the *U6-26p* was PCR-amplified from pFH13 with FH21 and FH22 (Supplementary Table S1) and the sgRNA scaffold was PCR-amplified from pFH4 with FH23 and FH24 (Supplementary Table S1). The two PCR products were combined with NheI/SpeI-digested pFH13 to the construct pFH6 via Gibson assembly® (NEB). To introduce the 20 bp target sequence designed against the *GL1* gene, pFH6 was digested with BbsI (NEB) and the vector backbone was ligated with two annealed primers (FH35/FH36, Supplementary Table S1) containing the target sequence resulting in the vector pFH26. The final T-DNA transformation vector pUB-Cas9-@GL1 (with the @ representing the targeted gene) was constructed by Gibson Assembly© (New England Biolabs) of KpnI and HindIII-digested pUB-Cas9 and the sgRNA cassette that was amplified from pFH26 with FH41 and FH42 (Supplementary Table S1). Similarly, a target sequence designed against the disrupted *bar* gene was prepared. Since, the U6 RNA polymerase III promoter takes guanine as transcription start nucleotide, we could not use the usual 20 bp protospacer sequence. We therefore added an extra G at the 5′-end of the sgRNA (5′-GCAGGAACCGCAGGAGTAGGA-3′, **Figure 1B**) that should not interfere with Cas9 activity (Belhaj et al., 2013; Ran et al., 2013). The target sequence was created by annealing the two primers FH145/FH146 (Supplementary Table S1) and ligation into BbsI-digested pFH6 resulting in the vector pFH25. Gibson Assembly© of KpnI and HindIII-digested pUB-Cas9 and the sgRNA cassette that was amplified from pFH25 with the primers FH41 and FH42 (Supplementary Table S1) resulted in the final T-DNA transformation vector pUB-Cas9-@BAR. The sequence of the vectors CasYFP (accession number KY080688), pKS Chlamy dual (accession number KY080687), pFH6 (accession number KY080689), pUB-Cas9 (accession number KY080691), pUB-Cas9-@GL1 (accession number KY080693) and pUB-Cas9-@BAR (accession number KY080692) can be found on GenBank.

Plant Transformation

The T-DNA vectors pFH17, pUB-Cas9-@BAR and pUB-Cas9-@GL1 were used to transform *Agrobacterium tumefaciens*, strain GV3101::pMP90. *Arabidopsis* Columbia-0 (Col-0) wild type plants (WT) were then transformed by floral dipping as described in Clough and Bent (1998) with pFH17 and pUB-Cas9-@GL1, respectively. Primary transformants were selected using kanamycin (pFH17) or hygromycin (pUB-Cas9-@GL1). Additionally, genomic DNA was isolated from leaves of surviving plants via isopropanol precipitation (Weigel and Glazebrook, 2009) and the presence of T-DNA insertion was verified via PCR using plant genomic DNA as template and the primer pairs P16/P98 (Supplementary Table S1; pFH17) or FH158/FH159 (Supplementary Table S1; pUB-Cas9-@GL1).

Similarly, the vector pUB-Cas9-@BAR was transformed by floral dipping in *Arabidopsis* Col-0 plants homozygous for the insertion of the T-DNA vector pFH17 containing the *bar-1* null allele of the *BAR* gene. Primary transformants were selected on hygromycin and kanamycin. T-DNA insertion was verified via PCR with the primer pair FH61/FH201 (Supplementary Table S1) using plant genomic DNA as template.

Molecular Determination of Cas9-Mediated Gene Editing

For detection of mutations in the pUB-Cas9-@BAR lines, plants were grown on soil under long day conditions. After 3 weeks, the plants were sprayed for 2 weeks with glufosinate ammonium (120 mg/L) every 3 days to isolate mutants that

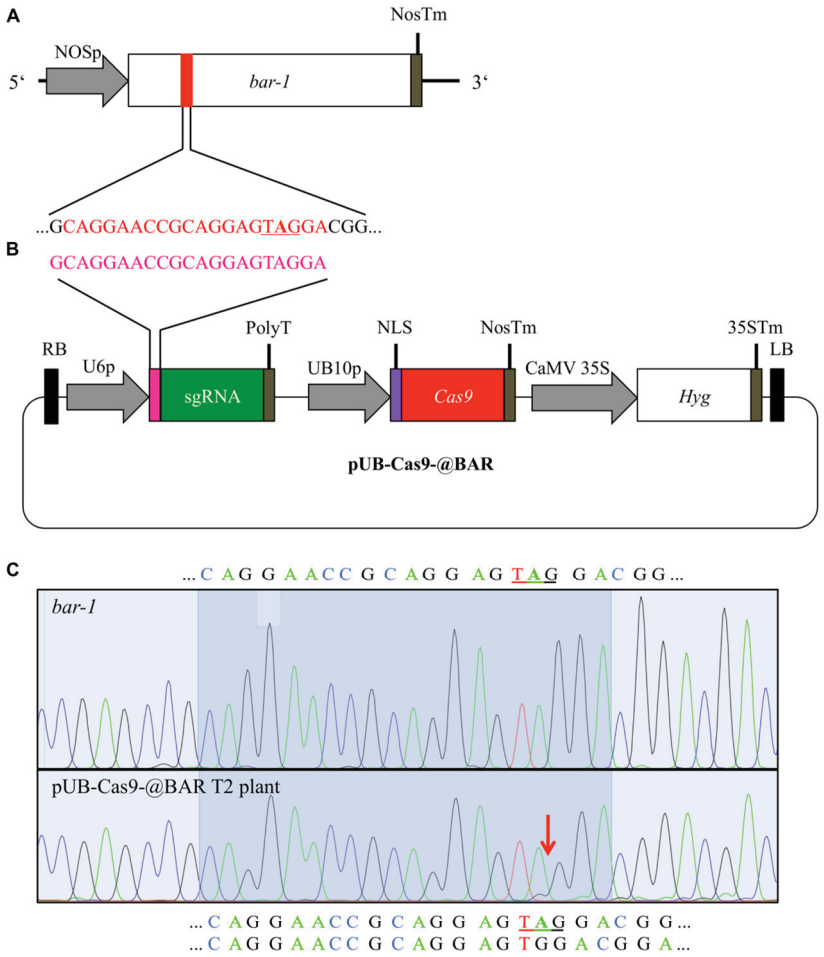


FIGURE 1 | Cas9-mediated repair of a disrupted *bar* gene. We generated an *Arabidopsis* line with a null allele of the *BAR* gene (*bar-1*) that contains an extra adenine (**A** in bold) leading to a premature STOP codon **(A)**. In order to induce mutations, *Arabidopsis* plants homozygous for the *bar-1* insertion were transformed with the single T-DNA construct pUB-Cas9-@BAR **(B)** which contains the sgRNA (green) with the 20 bp target motive (pink) and an additional guanine at the beginning (necessary for transcription) under the control of the *U6-26* promoter (U6p), the *Cas9* (red) controlled by the *UBIQUITIN10* promoter (UB10p) and the selection marker hygromycin (*Hyg*). Gene sizes are not to scale. T2 transformants were then treated with glufosinate and resistant plants were analyzed for rescuing mutations in the *bar* gene in leaf genomic DNA. The direct comparison of sequencing chromatograms **(C)** of a glufosinate resistant plant (bottom) and one of a control plant containing only the *bar-1* allele (top) shows small additional peaks appearing at the expected site of Cas9 induced mutagenesis (red arrow) at 3 bp in front of the PAM site in the glufosinate resistant plant. The 20 bp target site is highlighted and the validated sequence is given for the null allele plant (top) and the resistant plant (bottom; first line: main peaks, second line: minor peaks). The premature stop codon is underlined. NOSp, nopaline synthase promoter; NosTm, nopaline synthase terminator; RB, right T-DNA border; PolyT, poly thymidine transcription terminator; NLS, nuclear localization signal; CaMV 35S, *Cauliflower Mosaic Virus* 35S promoter; 35STm, 35S terminator; LB, left T-DNA border.

repaired the inactivated reading frame of the *BAR* gene via Cas9-mediated gene editing. Genomic DNA was isolated from leaves of surviving plants. The *BAR* gene was PCR-amplified with the primer pair FH154/FH155 (Supplementary Table S1) using the proofreading Phusion© polymerase (NewEngland Biolabs) and the obtained PCR amplicon was sent for Sanger sequencing (Macrogen). Sequencing histograms were compared using 4Peaks software.

For mutation detection in the pUB-Cas9-@GL1 lines, plant genomic DNA was isolated and the targeted region in the *GL1* gene was amplified with the primers FH189/FH190 (Supplementary Table S1) using the proofreading Phusion© polymerase (NewEngland Biolabs). The PCR product was then either digested with DdeI (NewEngland Biolabs) for detection of restriction digest length polymorphisms or Sanger sequenced (Macrogen). We used the CasOFF-Finder program

(Bae et al., 2014) to detect putative off-target sites. For detection of mutations in the off-target *AT1G08810*, the putatively mutated region was amplified with primers FH258/FH259 (Supplementary Table S1) and then processed as described before.

Detection of pUB-Cas9-@GL1 T-DNA in Glabrous Plants

Genomic DNA was isolated from leaves. Presence of the T-DNA was verified with primers amplifying specifically the *Cas9* gene (FH61/FH201, Supplementary Table S1) or the sgRNA gene (FH14/FH36, Supplementary Table S1). The T-DNA vector pUB-Cas9-@GL1 was used as positive control. High quality of the genomic DNA was verified via PCR on the housekeeping gene *Actin7* (*At5g09810*) using the primers P67/P68 (Supplementary Table S1).

RESULTS

We aimed at testing and evaluating markers for easy detection of Cas9 induced mutagenesis. Therefore, we explored the two alternative markers, resistance to glufosinate and development of glabrous leaves.

Plants Carrying a Mutated Allele of the *BAR* Gene Are Not Resistant to Glufosinate

Glufosinate resistance is a well-established marker for positive selection of plants (Miki and McHugh, 2004) and allows screening of large plant populations by simple spraying of glufosinate. For a positive selection screening system, we first produced a mutant allele of the *BAR* gene (from here on called *bar-1*) and stably transformed *Arabidopsis* Col-0 plants with it. The *bar-1* allele contains an adenine insertion, which leads to a premature STOP codon (**Figure 1A**). We verified that this allele is indeed non-functional and does not confer resistance to glufosinate by growing homozygous transformants on soil for 3 weeks and treating them several times with glufosinate spraying. Out of 77 homozygous plants, none survived the treatment (Supplementary Figure S1).

Design of a Single Generic Cas9 Vector for *Arabidopsis*

We wanted to design a vector system that can be used for targeted knockout of any gene in *Arabidopsis*. Therefore, we designed first the binary vector pUB-Cas9 encoding a Cas9 controlled by the *UBIQUITIN10* (*UB10)* promoter and containing insertion sites for the integration of the sgRNA cassette as well as a hygromycin resistance cassette for selection of transformants. Second, we designed a subcloning vector containing the *U6-26p*, followed by two BbsI-sites for protospacer integration and the sgRNA scaffold. After integration of the target sequence, any complete sgRNA cassette can basically be amplified via PCR using the same primer pair and be introduced in the binary vector pUB-Cas9 via Gibson cloning.

Design of CRISPR/Cas9 System for Repairing the Mutant *BAR* Gene

In order to repair the mutated *bar-1* allele previously generated in *Arabidopsis*, we used the binary T-DNA vector pUB-Cas9-@BAR (**Figure 1B**). We designed our target sequence so that the Cas9 nuclease was expected to cut next to the adenine insertion in the *bar-1* gene, which causes the premature STOP codon, 3 bp upstream of the PAM motive. As small InDels were expected to occur there by incorrect NHEJ, we expected almost certainly a reading frame restoration of the *BAR* gene.

T2 Plants Show High Survival Rate after Glufosinate Treatment

We transformed *Arabidopsis* plants homozygous for the *bar-1* allele with the pUB-Cas9-@BAR construct. We selected two positive transformants in the T1 generation. From both independent lines, we grew 77 T2 plants on soil and treated them after 3 weeks with repetitive glufosinate spraying. In both lines, the majority of plants (99 and 96%, respectively) survived the glufosinate treatment displaying different degrees of resistance to glufosinate (Supplementary Figure S1).

Mutation Analysis Shows Only Minor Mutational Events

We amplified the *BAR* gene from leaves of 12 T2 glufosinate resistant plants from each line and sequenced the PCR product. As a control, we also amplified the *BAR* gene from a homozygous *bar-1* knockout plant that was not transformed with the pUB-Cas9-@BAR construct. All sequencing chromatograms displayed the sequence of the defective *bar-1* allele in the main peaks. However, starting at the Cas9 target sequence, minor chromatogram peaks were often visible, confirming that indeed Cas9 was actively targeting and inducing mutations in the *bar-1* reading frame (**Figure 1C**; Supplementary Figure S2A). In two of the 24 analyzed plants, we could conclude from the sequencing chromatograms that the inserted adenine was deleted in a part of the cells and the reading frame of the *BAR* gene was restored (**Figure 1C**). In the other plants, either additional small InDel mutations occurred that did not restore the reading frame or no change compared to the control DNA was visible (Supplementary Figure S2). This result indicated that even though most cells were not able to produce a functional PAT protein, resistance to glufosinate was still achieved.

Generation of Homozygous Mutants Was Not Successful

We analyzed 8 T3 plants from the progeny of two T2 pUB-Cas9-@BAR plants that were resistant to glufosinate. We searched for mutations in the *bar* gene as described above. As with the T2 generation, we obtained sequencing chromatograms that mainly displayed the sequence of the *bar-1* allele with additional peaks starting at 3 bp in front of the PAM motive (Supplementary Figure S2A). We could not obtain non-chimeric plants with a fully restored *bar* gene.

The *gl1* T-DNA Knockout Line Displays the Glabrous Phenotype

As the positive selection system using glufosinate did not yield completely mutated plants, we tested alternatively the non-invasive marker gene *GL1*. The *GL1* gene was chosen as target of Cas9 induced mutagenesis since *gl1* knockout plants are not able to form leaf and stem trichomes (Koornneef et al., 1982) and provide therefore a suitable visual marker to track mutation progress *in planta*. To establish a visual reference of the expected glabrous phenotype, an *Arabidopsis gl1* T-DNA insertion line was isolated (Supplementary Figure S3) and grown in parallel with *Arabidopsis* WT plants, ecotype Col-O. As expected, homozygous insertion lines did not produce trichomes on leaves and stems (Supplementary Figure S3). Additional differences in growth or development could not be detected as expected from previous reports (Oppenheimer et al., 1991).

Design of Cas9 System for Editing *GL1*

Using the vector system described above, we generated the binary T-DNA vector pUB-Cas9-@GL1 (**Figure 2B**) to introduce a targeted gene knockout in the *GL1* gene of *Arabidopsis* by the Cas9 system. pUB-Cas9-@GL1 contained all the features of pUB-Cas9 and additionally the sgRNA cassette with the *U6-26* promoter, a 20 bp protospacer sequence targeted against *GL1* and the sgRNA scaffold. We selected a targeting site that is located at the beginning of the second exon of the gene (5′-GGAAAAGTTGTAGACTGAGA-3′; **Figure 2A**).

Chimeric Trichome Patterns Appear in the T2 Generation

We visually analyzed trichome appearance over five generations in different independent transformant lines (**Figure 3**). In the following, all examined plants will be named CasGL1 followed by an unique number containing information about their progenitor and referring to the scheme in **Figure 3**. As an example, the T2 plant marked by the red circle in **Figure 3** is named CasGL1_2.2 because it derives from the T1 plant number 2 (first number; CasGL1_2) and is the second out of four analyzed plants from that progenitor in that generation. The seven plants in its progeny are then called CasGL1_2.2.1-7.

After transformation of WT plants with the construct pUB-Cas9-@GL1, we visually analyzed four T1 generation plants but could not detect any completely or partly glabrous plants. Four descendants of each T1 plant were analyzed in the T2 generation. While most plants still displayed WT-like trichome patterns, two plants from two independent T1 lines displayed a chimeric phenotype (CasGL1_2.2 and CasGL1_4.3). That is, these plants had few either completely or partly glabrous leaves, which hints at a disruption of both alleles of the *GL1* gene in the glabrous tissues (**Figures 4A,B**).

Restriction Fragment Length Polymorphism Assay Indicate Mutations in Glabrous Leaves

To verify that the genomic DNA from glabrous leaves indeed contained InDels in the *GL1* gene in the targeted sequence, we used a restriction fragment length polymorphism assay. We made use of the DdeI restriction site within the target site of the Cas9 nuclease that was expected to be destroyed by InDel mutations (**Figure 4C**). We amplified the genomic region that spans the target site from genomic DNA of a WT leaf and a glabrous leaf of plant CasGL1_2.2, respectively, and digested the PCR product with DdeI (**Figure 4D**). The PCR product from the glabrous leaf was resistant to DdeI digestion, which provided evidence for a mutation in the DdeI target site (**Figure 4E**). The PCR product was also subcloned into pJET1.2 and seven clones were sequenced (**Figure 4F**). Three different InDels were found 3 bp upstream of the PAM motive, four clones out of seven contained an adenine insertion, two a thymine insertion and one clone had a guanine deletion. As both alleles of *GL1* have to be disrupted to generate a glabrous phenotype, the high proportion of clones with the adenine insertion suggests that this mutation happened at early developmental stages or in the previous generation, while the thymine insertion and guanine deletion happened later in independent progenitor cells on the other allele disrupting trichome formation in that specific leaf.

Completely Glabrous Plants Emerge in the T3 Generation

Four T2 plants (chimeric plant CasGL1_2.2, three WT trichome pattern plants CasGL1_2.4, CasGL1_4.2 and CasGL1_4.4) were carried to the next generation. While two T3 lines displayed normal trichome patterns, all descendants of the chimeric line CasGL1_2.2 and of the WT-like line CasGL1_4.2 were completely glabrous (**Figures 3** and **5A**). In the glabrous lines, trichomes were neither visible on leaves nor on stems. We did not observe any difference in growth or development.

Loss of Function Is Caused by Small InDels

To analyze the nature of the generated mutations, we amplified the *GL1* gene from the glabrous plants and sequenced the PCR products (**Figure 5B**). As control, we also sequenced the *GL1* gene from the non-glabrous T3 plants CasGL1_4.4.1 and CasGL1_2.4.1 (Supplementary Figure S4). While the control plants did not show any changes in the *GL1* sequence, the three analyzed descendants of line CasGL1_4.2 showed a 10 bp deletion. The other descendants from the chimeric T2 line CasGL1_2.2 showed either a homozygous adenine insertion (three out of seven analyzed plants, 3/7 plants), a homozygous guanine deletion (3/7 plants) or a biallelic mutation (+A/-G; 1/7 plants). All mutations generated a premature stop codon. The high amount of homozygous mutant plants in the T3 generation cannot be explained by Mendelian segregation only and might hint at allelic conversion events.

Presence of *Cas9* Gene Does Not Influence the Stability of the Mutation

To test whether the mutations were stably inherited, we brought the glabrous plant CasGL1_4.2.1 with the 10 bp deletion in the *gl1* gene into the T4 and T5 generation and visually analyzed the plants for reappearance of trichomes. All progeny maintained

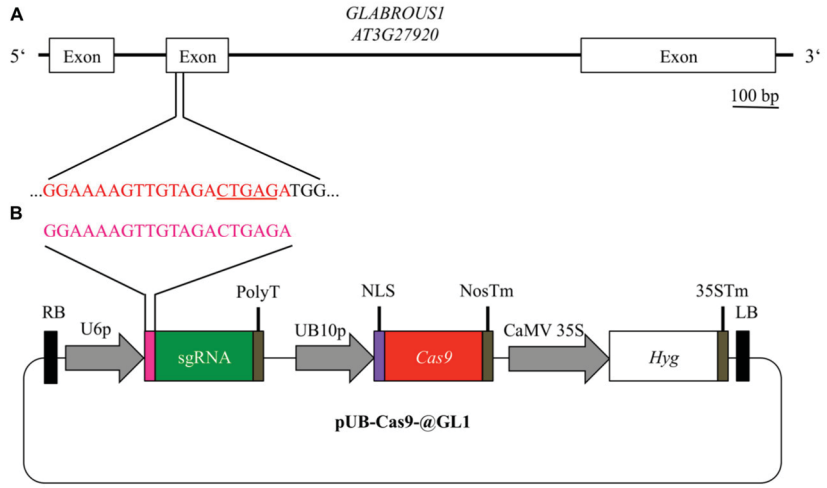


FIGURE 2 | Experimental design for disruption of the *GL1* gene. For targeting the *GLABROUS1* gene with the Cas9 system, we selected a 20 bp target site (red) in the second exon of the *GL1* gene **(A)**. The target site contains a DdeI restriction site (underlined) spanning the area where Cas9 cleaves the dsDNA. In order to induce mutations, *Arabidopsis* WT plants were transformed with a single T-DNA construct pUB-Cas9-@GL1 **(B)** containing the sgRNA (green) with the target motive (pink) under the control of the *U6-26* promoter (U6p), the *Cas9* gene (red) controlled by the *UBIQUITIN10* promoter (UB10p) and the selection marker hygromycin (Hyg). Gene sizes in the vector map are not to scale. RB, right T-DNA border; PolyT, poly thymidine transcription terminator; NLS, nuclear localization signal; NosTm, nopaline synthase terminator; CaMV 35S, *Cauliflower Mosaic Virus* 35S promoter; 35STm, 35S terminator; LB, left T-DNA border.

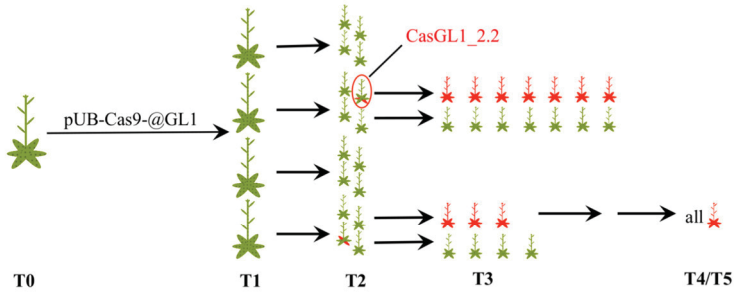


FIGURE 3 | Schematic overview of analyzed plants. We analyzed several independent lines transformed with pUB-Cas9-@GL1 over five generations. This scheme shows the number of analyzed plants up to the T3 generation and the appearance of chimeric and glabrous plants over time (red areas = glabrous areas). The analyzed T4 and T5 plants were all glabrous. The plant CasGL1_2.2 (red circle) is the chimeric plant that was analyzed with a restriction fragment length polymorphism assay (see **Figure 4**).

their glabrous phenotype. We also investigated whether it was possible to obtain stably mutated plants that do not carry the Cas9 cassette in the genomic DNA anymore. Therefore, we grew the T4 generation plants on non-selective medium to recover both plants with the T-DNA cassette (including Cas9 and the sgRNA) and plants that lost the cassette by segregation. The genomic DNA of seven plants was analyzed by PCR for the presence of the Cas9 and the sgRNA gene (**Figure 6**). Three out of seven plants had lost the T-DNA with Cas9 and sgRNA but maintained their glabrous phenotype. This shows that the T-DNA is independently inherited from the generated mutation and the presence or absence of Cas9 does not influence a Cas9-generated mutation.

No Off-Targets were Detected in Closest Homologous Sequence

A common issue for Cas9-based generation of mutants is the possibility of off-target cleavage of similar regions in the genome (Cho et al., 2014). We analyzed six glabrous descendants of the plant CasGL1_4.2.1 (T4 generation) for mutations in the closest homologous site to our target sequence. This site contains only one mismatch near the 5′ end of the sgRNA target sequence and lies in the gene *AT1G08810* (5′-GCAAAAGTTGTAGACTGAGA-3′, mismatch is underlined), which encodes for the putative Myb-type transcription factor MYB60. Sequencing of genomic DNA did not reveal mutations in this gene (Supplementary Figure S5).

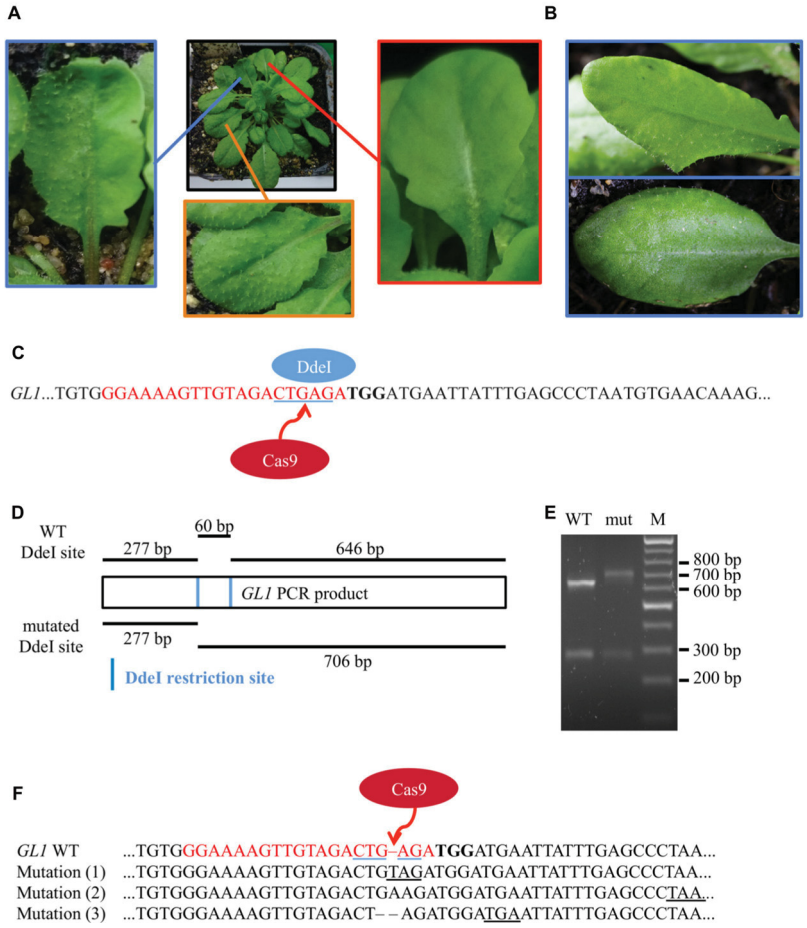


FIGURE 4 | Mutation analysis in chimeric T2 generation plants. (A) Several T2 plants displayed a combination of partly glabrous (blue), to completely glabrous (red), or WT-like leaves (orange). The trichome pattern on the partly glabrous leaves was different on individual chimeric leaves in different plants **(B)**. The restriction fragment length polymorphism assay was used to verify mutations on gDNA level. **(C)** As mutations induced by Cas9 nuclease disrupt a DdeI restriction site (blue underlined) in the *GL1* target site (red), digestion of *GL1* PCR amplicons of mutant and WT alleles yield a different restriction pattern. **(D)** We amplified the *GL1* gene from genomic DNA from the glabrous leaf shown in **(A)** (mut) and a WT leaf and digested it with DdeI. The resulting DNA fragments **(E)** match the length expected for WT (277 and 646 bp) and mutant (277 and 706 bp) alleles, the 60 bp band in the WT allele is not retained by the gel (M, Marker). Sequencing of the *GL1* gene revealed three different InDel mutations in the glabrous leaf **(F)** 3 bp upstream of the PAM (bold) within the sgRNA target sequence (red), all causing frameshifts and the formation of premature STOP codons (underlined).

DISCUSSION

Positive Selection by Regaining Glufosinate Resistance Is Not a Suitable Marker for Cas9-Induced Mutagenesis

In this study, we compared two visual markers for monitoring the efficiency of Cas9-induced mutations. First, we aimed at repairing a non-functional *BAR* gene *in planta* by removing a premature STOP codon using a sgRNA directed against the mutated site (**Figure 1**). While we were able to show Cas9-induced mutations in that site in the T2 generation (**Figure 1C**), we could not explain the discrepancy between the high percentage of plants resistant to glufosinate (over 95%, Supplementary Figure S1) and the low occurrence of edited sequences in the genome of resistant plants (Supplementary Figure S2). A likely explanation is that few cells carrying an active allele of the *BAR* gene are sufficient for conferring systemic glufosinate resistance to the plant. It has been shown recently in cotton that small amounts of the PAT protein are sufficient to provide sufficient glufosinate resistance in regard to visual symptoms (Carbonari et al., 2016). Additionally, the PAT protein consists of only 183 amino acids (app. 20 kDa) and is therefore small enough to be transported through plasmodesmata between cells (Crawford and Zambryski, 2001)

FIGURE 5 | Glabrous plants can be generated within three generations. Completely glabrous plants **(A)** were obtained in the T3 generation. Plant CasGL1_4.2.1 is shown as a representative example. Sequencing of the glabrous plants revealed three different frameshift mutations **(B)**, which all lead to premature stop codons (underlined). The PAM is marked in bold letters, the sgRNA target site in red letters.

and perhaps even through the phloem (Stadler et al., 2005). This means that few cells producing functional PAT might be able to supplement many cells with functional PAT protein and thus providing systemic resistance. Consequently, we did not isolate plants with a fully repaired *BAR* gene with the same genotype in all cells. The observation that Cas9 induced mutations occurred in the plants at the somatic level, suggests that likely those genetic changes also occur in germline cells. Germline mutations would be passed to the next generation and sequencing a higher amount of plants might have therefore ultimately led to detection of completely mutated plants. However, as the stringency of glufosinate selection is low, at least in the standard concentration used in our work, a possible usage of this system to compare the efficiency of different Cas9 variants or expression vectors is not recommended, as differences in the efficiency of Cas9 repair would not be detectable.

The Cas9 System Is a Highly Efficient Method to Generate Independent *gl1* Knockout Lines

A non-invasive alternative to visually track Cas9 nuclease activity in *Arabidopsis* is the disruption of the *GL1* gene. Therefore, we designed a sgRNA targeting the second exon of the gene. This site was favored for several reasons. First, no completely redundant sequence could be found in the genome of *Arabidopsis*. Second, a frameshift mutation due to small insertion or deletions (InDels) caused by the error-prone NHEJ was likely to have more drastic effects at the beginning of the gene. Third, the target site is located in one of the two MYB-type DNA-binding domains and mutations in these domains lead to more drastic phenotypes compared to mutations elsewhere in the gene (Hauser et al., 2001). Fourth, a DdeI restriction enzyme site overlapped with the cutting site. Thus, we decided to detect mutations by screening for the loss of the restriction enzyme recognition site.

In this report, we could show that the Cas9 system was highly efficient for generating knockouts in the *GL1* gene. Using a single sgRNA and a *Chlamydomonas reinhardtii* codon optimized *Cas9* gene (**Figure 2B**), we found that 50% of our analyzed T1 lines contained glabrous progenitors (**Figure 3**). In total, we were able to generate four different homozygous/biallelic glabrous lines. This was achieved in the T3 generation by screening a very limited amount of plants. Screening a higher amount of plants would have probably yielded knockouts even in the earlier generation, as shown by Cas9 experiments of other groups (Fauser et al., 2014; Feng et al., 2014). This puts the application of Cas9 technology within the reach of every laboratory equipped with basic molecular biology equipment.

The high frequency of homozygous mutants might be caused by allelic gene conversion (Li et al., 2014; Yoshimi et al., 2014), meaning that one allele is mutated first and then serves as repair template for the DSB in the second allele. Therefore, producing homozygous mutants with Cas9 can be more efficient than expected by chance.

All observed mutations were small InDels that caused a shift in the reading frame and thus the generation of premature STOP codons (**Figures 4F** and **5B**). These small InDels are typical for error-prone repair of the induced DSB by the NHEJ pathway (Salomon and Puchta, 1998; Puchta and Fauser, 2014).

Moreover, targeting the *GL1* gene allowed us to visualize both homozygous plants and chimeric leaves (**Figure 4A**). We suggest

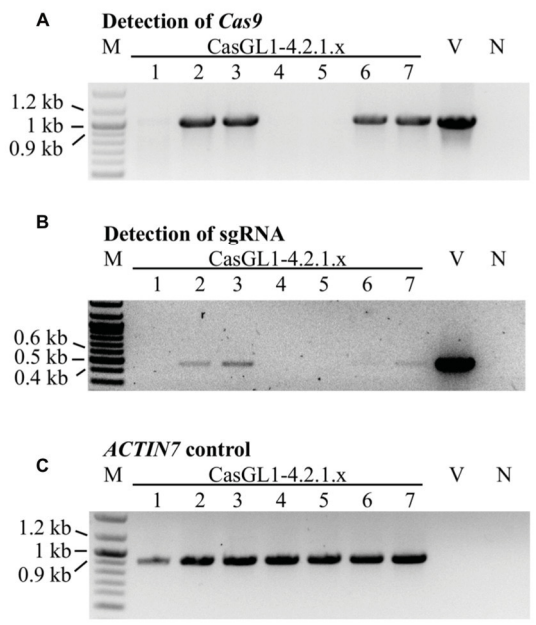


FIGURE 6 | Segregation analysis of the Cas9 T-DNA cassette in glabrous T4 generation plants. Seven Cas9-generated glabrous plants of the T4 generation (CasGL1_4.2.1.1-7) were analyzed by PCR for the presence of the *Cas9* gene (expected size: 1 kb, **A**) and the sgRNA cassette (expected size: 0.49 kb, **B**) using the *Arabidopsis ACTIN7* gene as DNA quality control (expected size: 0.92 kb, **C**). Images were color inverted for better visibility. M, marker; V, vector control (pUB-Cas9-@GL1); N, negative control with water.

that the partially glabrous phenotype of the chimeric leaves could be used in order to track cell lineages during leaf development, which might ultimately lead to a model of cell migration through space and time (Kalve et al., 2014). Another advantage of chimeric mutations is the possibility to study the function of essential genes *in vivo* by direct comparison of mutated and non-mutated parts within the same plant and thus overcome the problem of lethality of complete knock-out plants. However, it has to be taken into account that the CRISPR/Cas9 system is constitutively active in plants under the control of the *UBIQUITIN10* promoter and new mutations can thus occur in every cell at any time. A precise delivery of Cas9 only to certain parts of a plant (e.g., by transient *Agrobacterium* leaf infiltration or tissue-specific, inducible promoters) might therefore be the favorable approach for such intraplant-comparisons.

Transgene-Free Plants Can Be Generated with the Cas9 System

A major issue for the acceptance of genetically modified crops is the introduction of transgenes into crop plants by random integration of a T-DNA (Ishii and Araki, 2016). The Cas9 system is also classically introduced as T-DNA cassette. However, since allelic point mutations are established on different genomic loci, the T-DNA containing the Cas9 components can be removed by backcrossing. We could show via PCR analysis that some of the glabrous plants in the T4 generation had actually lost the T-DNA cassette over time while still keeping the mutation in the *GL1* gene (**Figure 6**). This is in accordance with results from other groups that generated transgene-free mutated rice (Xu et al., 2015) or *Arabidopsis* (Feng et al., 2014) plants. This underlines the opportunities for the introduction of new traits in transgene-free crop plants that are given by the Cas9 system and other targeted nucleases, such as TALENs (Wang et al., 2014). As long as these new traits are only created by allelic point mutations and not by integration of foreign transgenes that cannot be introgressed by genetic crosses, site-directed mutations are not distinguishable from mutations generated by natural processes or breeding and are hence likely to be not subject to regulatory assessments in many countries (Wolt et al., 2016). If applicable, transient delivery systems for Cas9 RNA and corresponding sgRNA (Zhang et al., 2016), or preassembled protein–RNA complexes (Woo et al., 2015) facilitate the genomic editing, without DNA integration into the target genome at any time.

Off-Targeting Effects were Not Detectable in Plants

One potential pitfall of the Cas9 system is the possibility for off-target mutations to occur in sequences highly similar to the target site (Fu et al., 2013). In plant research, off-target mutations are highly problematic if the generated mutants are used to characterize gene functions. We targeted the DNA binding helix 3 of the MYB transcription factor GL1. Since this domain is highly conserved on the DNA sequence level in the R2R3-MYB gene family (Jia et al., 2004), off-targeting was possible. We used the CasOFF-Finder program (Bae et al., 2014) to detect putative off-target sites. While in *Arabidopsis* there is no other DNA sequence perfectly matching to the target site, three gene sequences have only one mismatch to the sgRNA target site against *GL1* used in this study. Two of them have a mismatch in the 3′-end of the protospacer motive. This region is called seed sequence. A mismatch in the seed sequence strongly reduces the affinity of the Cas9-sgRNA complex toward its off-target (Semenova et al., 2011; Jinek et al., 2012). Therefore, we focused on the third putative off-target site, where the mismatch is near the 5′-end of the protospacer. However, sequencing of this gene in six glabrous T4 generation plants did not show off-target mutations (Supplementary Figure S5). This result is in line with previous observations that off-targeting does not occur at high rates in *Arabidopsis* as shown by in-depth whole genome analysis of *Arabidopsis* Cas9-edited lines where no off-targets in highly homologous sequences could be detected (Feng et al., 2014). Additionally, it has to be taken into account that off-target mutations can be removed in plants by backcrossing to the parental line.

GL1 Is a Suitable Marker for Visual Detection of Mutation Efficiency

Even though successful gene knockout using the Cas9 system has already been reported for *Arabidopsis* and other plant species

(Bortesi and Fischer, 2015), our study provides a direct readout of the mutagenic process in living plants throughout the plant development by directing Cas9 to *GL1*. Previous visual markers mostly relied on fluorescence proteins (Jiang et al., 2014) or on the GUS reporter system (Mao et al., 2013), which had to be transformed and expressed in the plant. Instead, the disruption of trichome formation allows following the progress of mutagenesis without having pleiotropic effects on the development of the plant (Oppenheimer et al., 1991). The glabrous lines generated in this work now provide the opportunity to detect heterozygous mutations using a reverse approach by repairing the *GL1* gene via Cas9. Especially the lines CasGL1_4.2.1/2/3 with the 10 bp deletion is of interest for HR based repair approaches. We propose that trichome loss can be exploited in all trichome-containing crop species as a marker for mutagenesis efficiency if the genes for trichome initiation are described [e.g., the *Glabrous Rice 1* gene in rice (Li et al., 2012) or the *Wooly* gene in tomato (Yang et al., 2011)].

CONCLUSION

In this study, we confirmed and showed in detail that the Cas9 system can be used to introduce small mutations with precision in the genome of *Arabidopsis*. The mutations are stably maintained in the genome even when the *Cas9* gene is lost. The efficiency of obtaining plants carrying homozygous mutation was higher than expected by mutation frequency alone. Thus, Cas9 mediated genome editing is a powerful and easy to use system for genome editing in *Arabidopsis*. GL1 is a reliable marker for observing mutation efficiency in living plants. Our marker approach might be of use for comparing different mutagenesis approaches, to enhance the mutagenesis efficiency, and finally for optimization of the Cas9 toolbox.

AUTHOR CONTRIBUTIONS

FH, OM, ME, and AW designed the study. FH carried out all laboratory experiments except the design of the codon-optimized Cas9 and cloning of the vectors CasYFP and pKS Chlamy dual (done by AG and PH). FH, OM, ME, and AW interpreted the data. FH, OM, ME, and AW wrote the manuscript with contributions of AG and PH.

FUNDING

This work was funded by the Cluster of Excellence on Plant Science (CEPLAS, EXC 1028), the HHU Center for Synthetic Life Sciences (CSL), and by the German Research Foundation (DFG), FOR 1261 (P.H.).

ACKNOWLEDGMENTS

We thank Prof. Dr. Holger Puchta, Prof. Dr. Peter Westhoff, Dr. Shizue Matsubara, and Dr. Nadine Rademacher for helpful discussions. We thank Steffen Köhler for photography support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpls.2017.00039/full#supplementary-material

REFERENCES

Bae, S., Park, J., and Kim, J. S. (2014). Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* 30, 1473–1475. doi: 10.1093/bioinformatics/btu048

Baltes, N. J., and Voytas, D. F. (2015). Enabling plant synthetic biology through genome engineering. *Trends Biotechnol.* 33, 120–131. doi: 10.1016/j.tibtech.2014.11.008

Bayer, E., Gugel, K. H., Hagele, K., Hagenmaier, H., Jessipow, S., Konig, W. A., et al. (1972). Metabolic products of microorganisms. 98. Phosphinothricin and phosphinothricyl-alanyl-analine. *Helv. Chim. Acta* 55, 224–239. doi: 10.1002/hlca.19720550126

Belhaj, K., Chaparro-Garcia, A., Kamoun, S., and Nekrasov, V. (2013). Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. *Plant Methods* 9:39. doi: 10.1186/1746-4811-9-39

Bernhardt, K. (2012). *The Role of a Peroxisomal NAD+ Carrier in Plants*. Düsseldorf: Heinrich-Heine Universität Düsseldorf.

Bortesi, L., and Fischer, R. (2015). The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnol. Adv.* 33, 41–52. doi: 10.1016/j.biotechadv.2014.12.006

Carbonari, C. A., Latorre, D. O., Gomes, G. L. G. C., Velini, E. D., Owens, D. K., Pan, Z. Q., et al. (2016). Resistance to glufosinate is proportional to phosphinothricin acetyltransferase expression and activity in LibertyLink (R) and WideStrike (R) cotton. *Planta* 243, 925–933. doi: 10.1007/s00425-015-2457-3

Cho, S. W., Kim, S., Kim, Y., Kweon, J., Kim, H. S., Bae, S., et al. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.* 24, 132–141. doi: 10.1101/gr.162339.113

Clough, S. J., and Bent, A. F. (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* 16, 735–743. doi: 10.1046/j.1365-313x.1998.00343.x

Crawford, K. M., and Zambryski, P. C. (2001). Non-targeted and targeted protein movement through plasmodesmata in leaves in different developmental and physiological states. *Plant Physiol.* 125, 1802–1812. doi: 10.1104/pp.125.4.1802

Deblock, M., Botterman, J., Vandewiele, M., Dockx, J., Thoen, C., Gossele, V., et al. (1987). Engineering herbicide resistance in plants by expression of a detoxifying enzyme. *Embo J.* 6, 2513–2518.

Doudna, J. A., and Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* 346:1258096. doi: 10.1126/science.1258096

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

Fendrych, M., Van Hautegem, T., Van Durme, M., Olvera-Carrillo, Y., Huysmans, M., Karimi, M., et al. (2014). Programmed cell death controlled by ANAC033/SOMBRERO determines root cap organ size in *Arabidopsis*. *Curr. Biol.* 24, 931–940. doi: 10.1016/j.cub.2014.03.025

Feng, Z., Mao, Y., Xu, N., Zhang, B., Wei, P., Yang, D. L., et al. (2014). Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 111, 4632–4637. doi: 10.1073/pnas.1400822111

Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K., et al. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* 31, 822–826. doi: 10.1038/nbt.2623

Grefen, C., Donald, N., Hashimoto, K., Kudla, J., Schumacher, K., and Blatt, M. R. (2010). A ubiquitin-10 promoter-based vector set for fluorescent protein tagging facilitates temporal stability and native protein distribution in transient and stable expression studies. *Plant J.* 64, 355–365. doi: 10.1111/j.1365-313X.2010.04322.x

Hauser, M. T., Harr, B., and Schlotterer, C. (2001). Trichome distribution in *Arabidopsis thaliana* and its close relative *Arabidopsis lyrata*: molecular analysis of the candidate gene GLABROUS1. *Mol. Biol. Evol.* 18, 1754–1763. doi: 10.1093/oxfordjournals.molbev.a003963

Ishii, T., and Araki, M. (2016). Consumer acceptance of food crops developed by genome editing. *Plant Cell Rep.* 35, 1507–1518. doi: 10.1007/s00299-016-1974-2

Jia, L., Clegg, M. T., and Jiang, T. (2004). Evolutionary dynamics of the DNA-binding domains in putative R2R3-MYB genes identified from rice subspecies indica and japonica genomes. *Plant Physiol.* 134, 575–585. doi: 10.1104/pp.103.027201

Jiang, W., Yang, B., and Weeks, D. P. (2014). Efficient CRISPR/Cas9-mediated gene editing in *Arabidopsis thaliana* and inheritance of modified genes in the T2 and T3 generations. *PLoS ONE* 9:e99225. doi: 10.1371/journal.pone.0099225

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821. doi: 10.1126/science.1225829

Kalve, S., De Vos, D., and Beemster, G. T. (2014). Leaf development: a cellular perspective. *Front. Plant Sci.* 5:362. doi: 10.3389/fpls.2014.00362

Koornneef, M., Dellaert, L. W., and Van Der Veen, J. H. (1982). EMS- and radiation-induced mutation frequencies at individual loci in *Arabidopsis thaliana* (L.) Heynh. *Mutat. Res.* 93, 109–123. doi: 10.1016/0027-5107(82)90129-4

Larkin, J. C., Oppenheimer, D. G., Lloyd, A. M., Paparozzi, E. T., and Marks, M. D. (1994). Roles of the GLABROUS1 and TRANSPARENT TESTA GLABRA Genes in *Arabidopsis* trichome development. *Plant Cell* 6, 1065–1076. doi: 10.2307/3869885

Larkin, J. C., Oppenheimer, D. G., Pollock, S., and Marks, M. D. (1993). *Arabidopsis* GLABROUS1 gene requires downstream sequences for function. *Plant Cell* 5, 1739–1748. doi: 10.1105/tpc.5.12.1739

Larkin, J. C., Young, N., Prigge, M., and Marks, M. D. (1996). The control of trichome spacing and number in *Arabidopsis*. *Development* 122, 997–1005.

Lea, P. J., Joy, K. W., Ramos, J. L., and Guerrero, M. G. (1984). The action of 2-amino-4-(methylphosphinyl)-butanoic acid (phosphinothricin) and its 2-oxo-derivative on the metabolism of cyanobacteria and higher-plants. *Phytochemistry* 23, 1–6. doi: 10.1016/0031-9422(84)83066-6

Li, J., Yuan, Y., Lu, Z., Yang, L., Gao, R., Lu, J., et al. (2012). Glabrous rice 1, encoding a homeodomain protein, regulates trichome development in rice. *Rice (N Y)* 5, 32. doi: 10.1186/1939-8433-5-32

Li, K., Wang, G., Andersen, T., Zhou, P., and Pu, W. T. (2014). Optimization of genome engineering approaches with the CRISPR/Cas9 system. *PLoS ONE* 9:e105779. doi: 10.1371/journal.pone.0105779

Mao, Y., Zhang, H., Xu, N., Zhang, B., Gou, F., and Zhu, J. K. (2013). Application of the CRISPR-Cas system for efficient genome engineering in plants. *Mol Plant* 6, 2008–2011. doi: 10.1093/mp/sst121

Marks, M. D. (1994). Plant development. The making of a plant hair. *Curr. Biol.* 4, 621–623. doi: 10.1016/S0960-9822(00)00136-6

Miki, B., and McHugh, S. (2004). Selectable marker genes in transgenic plants: applications, alternatives and biosafety. *J. Biotechnol.* 107, 193–232. doi: 10.1016/j.jbiotec.2003.10.011

Mullner, H., Eckes, P., and Donn, G. (1993). Engineering crop resistance to the naturally-occurring glutamine-synthetase inhibitor phosphinothricin. *ACS Symp. Ser.* 524, 38–47. doi: 10.1021/bk-1993-0524.ch003

Oppenheimer, D. G., Herman, P. L., Sivakumaran, S., Esch, J., and Marks, M. D. (1991). A myb gene required for leaf trichome differentiation in *Arabidopsis* is expressed in stipules. *Cell* 67, 483–493. doi: 10.1016/0092-8674(91)90523-2

Pattanaik, S., Patra, B., Singh, S. K., and Yuan, L. (2014). An overview of the gene regulatory network controlling trichome development in the model plant, *Arabidopsis*. *Front. Plant Sci.* 5:259. doi: 10.3389/fpls.2014.00259

Puchta, H., and Fauser, F. (2014). Synthetic nucleases for genome engineering in plants: prospects for a bright future. *Plant J.* 78, 727–741. doi: 10.1111/tpj.12338

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* 8, 2281–2308. doi: 10.1038/nprot.2013.143

Salomon, S., and Puchta, H. (1998). Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. *EMBO J.* 17, 6086–6095. doi: 10.1093/emboj/17.20.6086

Sauer, H., Wild, A., and Ruhle, W. (1987). The effect of phosphinothricin (glufosinate) on photosynthesis.2. The causes of inhibition of photosynthesis. *Z. Naturforsch.* 42, 270–278. doi: 10.1007/BF00032644

Semenova, E., Jore, M. M., Datsenko, K. A., Semenova, A., Westra, E. R., Wanner, B., et al. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. U.S.A.* 108, 10098–10103. doi: 10.1073/pnas.1104144108

Sessions, A., Burke, E., Presting, G., Aux, G., Mcelver, J., Patton, D., et al. (2002). A high-throughput *Arabidopsis* reverse genetics system. *Plant Cell* 14, 2985–2994. doi: 10.1105/tpc.004630

Stadler, R., Wright, K. M., Lauterbach, C., Amon, G., Gahrtz, M., Feuerstein, A., et al. (2005). Expression of GFP-fusions in *Arabidopsis* companion cells reveals non-specific protein trafficking into sieve elements and identifies a novel post-phloem domain in roots. *Plant J.* 41, 319–331. doi: 10.1111/j.1365-313X.2004.02298.x

Tachibana, K., Watanabe, T., Sekizawa, Y., and Takematsu, T. (1986). Action mechanism of bialaphos.2. Accumulation of ammonia in plants treated with bialaphos. *J. Pestic. Sci.* 11, 33–37. doi: 10.1584/jpestics.11.33

Thompson, C. J., Movva, N. R., Tizard, R., Crameri, R., Davies, J. E., Lauwereys, M., et al. (1987). Characterization of the herbicide-resistance gene bar from *Streptomyces-hygroscopicus*. *Embo J.* 6, 2519–2523.

Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C., et al. (2014). Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat. Biotechnol.* 32, 947–951. doi: 10.1038/nbt.2969

Weigel, D., and Glazebrook, J. (2009). Quick miniprep for plant DNA isolation. *Cold Spring Harb. Protoc.* 2009:dbrot5179. doi: 10.1101/pdb.prot5179

Wild, A., Sauer, H., and Ruhle, W. (1987). The effect of phosphinothricin (glufosinate) on photosynthesis.1. Inhibition of photosynthesis and accumulation of ammonia. *Z. Naturforsch.* 42, 263–269. doi: 10.1007/BF00032644

Wolt, J. D., Wang, K., and Yang, B. (2016). The regulatory status of genome-edited crops. *Plant Biotechnol. J.* 14, 510–518. doi: 10.1111/pbi.12444

Woo, J. W., Kim, J., Kwon, S. I., Corvalan, C., Cho, S. W., Kim, H., et al. (2015). DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. *Nat. Biotechnol.* 33, 1162–1164. doi: 10.1038/nbt.3389

Xu, R. F., Li, H., Qin, R. Y., Li, J., Qiu, C. H., Yang, Y. C., et al. (2015). Generation of inheritable and "transgene clean" targeted genome-modified rice in later generations using the CRISPR/Cas9 system. *Sci. Rep.* 5:11491. doi: 10.1038/srep11491

Yang, C., Li, H., Zhang, J., Luo, Z., Gong, P., Zhang, C., et al. (2011). A regulatory gene induces trichome formation and embryo lethality in tomato. *Proc. Natl. Acad. Sci. U.S.A.* 108, 11836–11841. doi: 10.1073/pnas.1100532108

Yang, C., and Ye, Z. (2013). Trichomes as models for studying plant cell differentiation. *Cell Mol. Life. Sci.* 70, 1937–1948. doi: 10.1007/s00018-012-1147-6

Yoshimi, K., Kaneko, T., Voigt, B., and Mashimo, T. (2014). Allele-specific genome editing and correction of disease-associated phenotypes in rats using the CRISPR-Cas platform. *Nat. Commun.* 5:4240. doi: 10.1038/ncomms5240

Zhang, Y., Liang, Z., Zong, Y., Wang, Y., Liu, J., Chen, K., et al. (2016). Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA. *Nat. Commun.* 7:12617. doi: 10.1038/ncomms12617

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Supplementary Material
An Efficient Visual Screen for CRISPR/Cas9 Activity in *Arabidopsis thaliana*

Florian Hahn, Otho Mantegazza, André Greiner, Peter Hegemann, Marion Eisenhut, Andreas P.M. Weber*

*** Correspondence:** Corresponding Author: Andreas.Weber@uni-duesseldorf.de

1 SUPPLEMENTARY FIGURES AND TABLES

1.1 Supplementary Figures

Figure S1

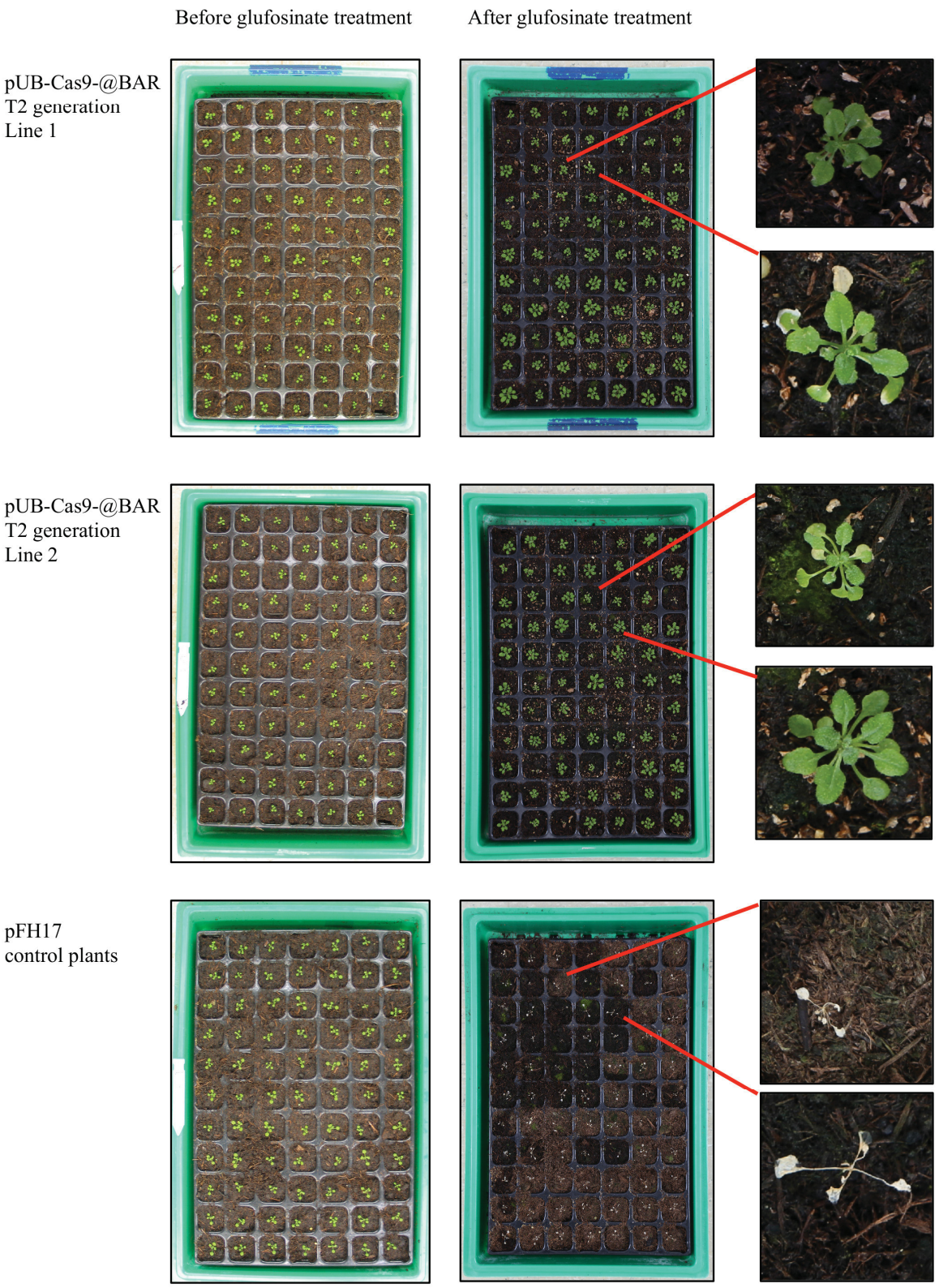


Fig. S1: Glufosinate treatment of T2 generation plants transformed with the vector pUB-Cas9-@BAR. We grew two independent lines of *Arabidopsis* T2 transformants with the vector pUB-Cas9-@BAR on soil. Repetitive glufosinate spraying after three weeks of growth yielded in over 95% surviving plants in both transformant lines with green glufosinate-resistant tissue. Still, some plants showed partly yellowish leaf tissue in variable amounts (compare detail

pictures on the right). In contrast, the background line carrying the null *bar*-1 allele in homozygosis and no *Cas9* or sgRNA did not survive the glufosinate treatment (pFH17, bottom).

Figure S2

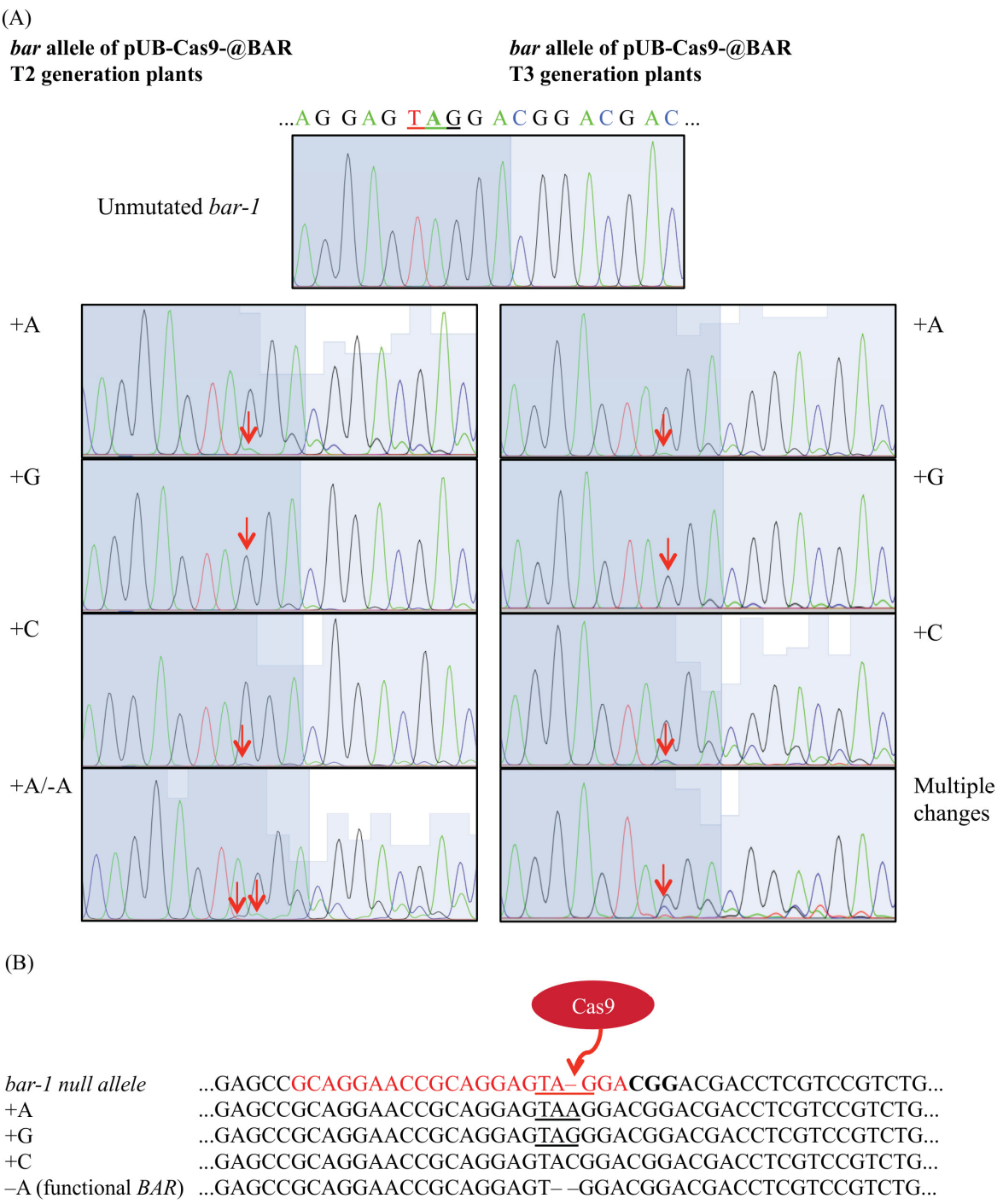


Fig. S2: Representative sequencing results of glufosinate resistant plants. We sequenced the *bar* gene in plants of the T2 and the T3 generation (A) that survived the glufosinate treatment. Next to plants showing an adenine deletion, which restores the reading frame (compare Figure 1C), most of the plants showed either no change (top), 1 bp insertions, multiple changes or no change, all of them not restoring a functional reading frame (B) but still resulting in premature STOP codons (underlined). A = adenine, C = cytosine, G = guanine.

Figure S3

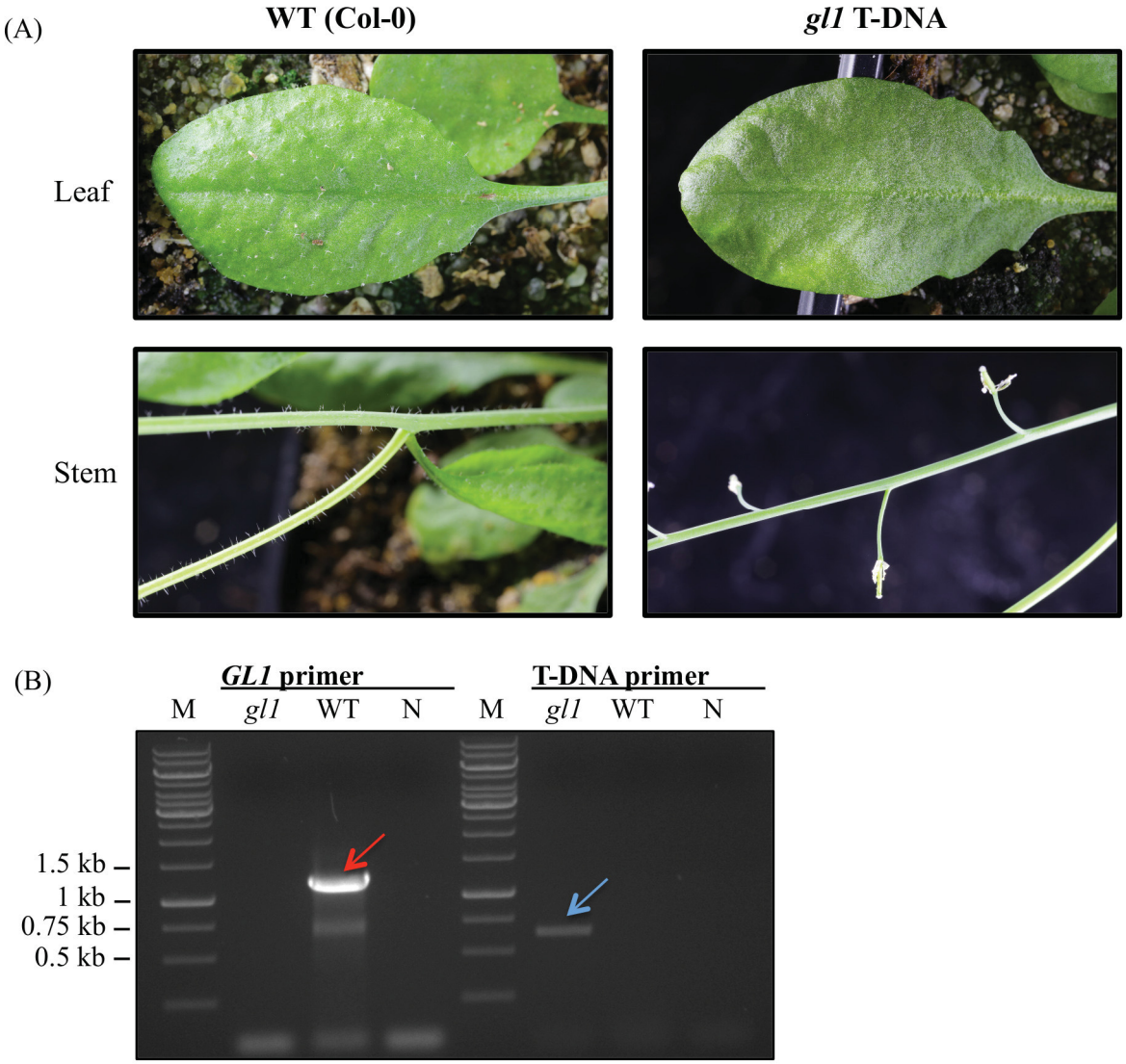
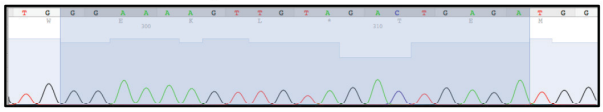


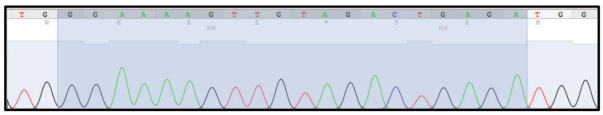
Figure S3: Visual phenotype and homozygosity proof of *gl1* T-DNA mutant plants. (A) We compared trichome formation of an *Arabidopsis* WT Col-0 plant and a *gl1* T-DNA knockout line. While WT plants (left) show dense trichome growth on leaves and stems, the *gl1* plants (right) display a glabrous phenotype. (B) Homozygous T-DNA insertion in the knockout line (*gl1*) was additionally confirmed with PCR using gene specific primers (left, expected product size 1267 bp) and T-DNA specific primers (right, expected product size 589-889 bp) on leaf gDNA. WT Col-0 gDNA (WT) was used as comparison; no DNA was used as negative control (N). The gene specific PCR product can only be found in the WT sample (red arrow), the T-DNA product only in the T-DNA gDNA (blue arrow). M = Marker

Figure S4

WT Col-0 *GL1*sequence



CasGL1_2.4.1 *GL1* sequence



CasGL1_4.4.1 *GL1* sequence

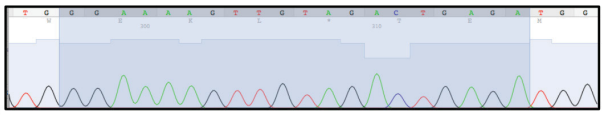
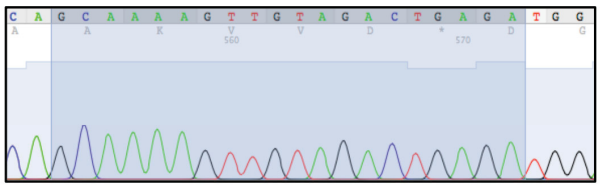


Fig. S4: *GL1* sequence analysis of a non-glabrous T3 plants. Two T3 plants (CasGL1_2.4.1 and CasGL1_4.4.1) transformed with the vector pUB-Cas9-@GL1 were analyzed for mutations in the *GL1* gene. Therefore, the gene was amplified via PCR and sequenced. WT *Arabidopsis* Col-0 DNA was used as a control. No mutations could be detected on sequence level.

Figure S5

(A) WT Col-0 *MYB60* sequence



(B) CasGL1 *MYB60* sequence

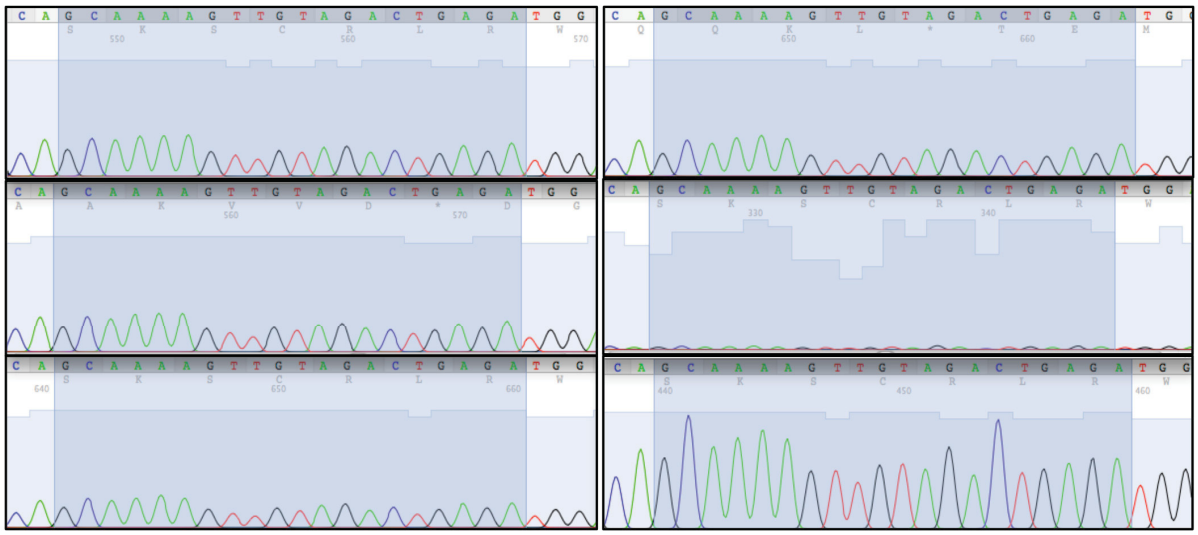


Fig. S5: Sequence analysis of a putative pUB-Cas9-@GL1 off-target site in T4 glabrous plants. Six Cas9-generated glabrous plants of the T4 generation (CasGL1_4.2.1.1/4/5 and CasGL1_4.2.2.1/3/4) were analyzed for mutations in a highly homologous off-target sequence in the gene *AT1G08810* (*MYB60*). Therefore, the gene was amplified via PCR and sequenced. WT *Arabidopsis* Col-0 DNA was used as a control (A). No mutations could be detected on the putative off target site in the six glabrous plants (B).

1.2 Supplementary Tables

Table S1: Primer sequences

Primer name	Sequence (5'→3')
FH5	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCACCTATCGATCATGAGCGGAGAAT
FH6	GGGGACCACTTTGTACAAGAAAGCTGGGTGATCAGCTTGCATGCCGGT
FH14	CACCAAGCTTCGTTGAACAACGGAAA
FH15	CACCGCTAGCGAAGGGACAATCACTACTTC
FH16	CACCGCTAGCGGGTCTTCGAGAAGACCTGTTT
FH18	CACCACTAGTTTAAAAAAATTTGCACCGACT
FH21	CGGGAATTCGATTCACCGAAGCTTCGTTGAACAACG
FH22	CGAAGACCCCGAAGGGACAATCACTACTTC
FH23	GTCCCTTCGGGGTCTTCGAGAAGACCTG
FH24	CAGGCGGCCGCGAATTCAAGATCTTCCGGATGGCTC
FH25	CACCTATCGATCATGAGCGGAGAA
FH26	CGGACGAGGTCGTCCGTCCTACTCCTGCGGTTCCTGC
FH27	GCCGCAGGAACCGCAGGAGTAGGACGGACGACCTCGTC
FH28	CACCATCAGCTTGCATGCCGG
FH35	TTCGGGAAAAGTTGTAGACTGAGA
FH36	AAACTCTCAGTCTACAACTTTTCC
FH39	TTAACAGCTCGAGTGCGGTATCGATCATGAGCGGAGAATTAAG
FH40	ACGAACGAAAGCTCTGCAATCAGCTTGCATGCCGGT
FH41	AAACGACGGCCAGTGCCAGAATTGGGCCCGACGTCG
FH42	TACTGACTCGTCGGGTACCAAGCTATGCATCCAACGCG
FH61	CTGCCCAACGAGAAGGTGC
FH145	TTCGGCAGGAACCGCAGGAGTAGGA
FH146	AAACTCCTACTCCTGCGGTTCCTGC
FH154	ATGAGCCCAGAACGACGC
FH155	AGATTTCGGTGACGGGCA

FH158	ATGAAAAAGCCTGAACTCACC
FH159	GGTTTCCACTATCGGCGAG
FH189	CTCTCACACACACACACAGACA
FH190	GTTTGCACGACTAATACACTTATG
FH201	AGTCGGACAGGCGGTTGATG
FH214	AAAGACCATCTGAATTTGGGG
FH215	TAAGCACGTGTCACGAAAACC
FH258	GATCAGTTCCCACCAACACT
FH259	TGTCGAAGAAGTTTGCAGC
NH117	GGGGACAAGTTTGTACAAAAAAGCAGGCTACCATATGGCCAGCAGCCCCCC
NH119	GGGGACCACTTTGTACAAGAAAGCTGGGTATTAGTCGCCGCCCAGCTGCGA
P16	CACAGTTCGATAGCGAAAACC
P49	TAGCATCTGAATTTCATAACCAATCTCGATACAC
P67	TTCAATGTCCCTGCCATGTA
P68	TGAACAATCGATGGACCTGA
P98	CGATTTTCTGGGTTTGATCG

MANUSCRIPT II

Generation of Targeted Knockout Mutants in *Arabidopsis thaliana* Using CRISPR/Cas9

bio-protocol www.bio-protocol.org/e2384 Vol 7, Iss 13, Jul 05, 2017
DOI:10.21769/BioProtoc.2384

Generation of Targeted Knockout Mutants in *Arabidopsis thaliana* Using CRISPR/Cas9

Florian Hahn, Marion Eisenhut, Otho Mantegazza and Andreas P. M. Weber*

Institute of Plant Biochemistry, Cluster of Excellence on Plant Science (CEPLAS), Center for Synthetic Life Sciences (CSL), Heinrich Heine University, Düsseldorf, Germany
*For correspondence: Andreas.Weber@uni-duesseldorf.de

[Abstract] The CRISPR/Cas9 system has emerged as a powerful tool for gene editing in plants and beyond. We have developed a plant vector system for targeted Cas9-dependent mutagenesis of genes in up to two different target sites in *Arabidopsis thaliana*. This protocol describes a simple 1-week cloning procedure for a single T-DNA vector containing the genes for Cas9 and sgRNAs, as well as the detection of induced mutations *in planta*. The procedure can likely be adapted for other transformable plant species.

Keywords: CRISPR/Cas9, Genome editing, *Arabidopsis thaliana*, Plants, Knockout

[Background] The CRISPR/Cas9 system (Cas9) provides a simple and widely applicable approach to modify genomic regions of interest and has therefore become the tool of choice for genome editing in plants and other organisms (Schiml and Puchta, 2016). The system relies on the bacterial Cas9 nuclease from *Streptococcus pyogenes* (Cas9), which can be directed by a short artificial single guide RNA molecule (sgRNA) towards a genomic DNA sequence (Jinek *et al.*, 2012), where it creates a double strand break (DSB). These DSBs are then repaired by the innate DNA repair mechanism of the plant cell. Here, two main pathways can be distinguished (Salomon and Puchta, 1998). (i) DNA molecules with high homology to the DSB site can be used as repair template. This homology directed repair (HDR) approach can be exploited to introduce specific sequences at the site of the DSB (Schiml *et al.*, 2014; Baltes and Voytas, 2015). However, due to low integration rates of these sequences, HDR mediated gene editing in plants remains challenging. (ii) An easier and more efficient approach is the use of the non-homologous end joining (NHEJ) repair pathway of the plant, which is the dominant repair pathway in most plants, such as *Arabidopsis thaliana* (*Arabidopsis*). Since NHEJ is error-prone, small insertions or deletions (indels) of a few base pairs (bp) occur often at the DSB site, leading to frameshift mutations and gene knockouts (Pacher and Puchta, 2016). Here, we provide a detailed protocol for targeted gene knockout in the model plant *Arabidopsis* including a simple 1-week cloning protocol for a plant vector system containing the Cas9 and sgRNA, and then *Arabidopsis* transformation and detection of mutations.

Materials and Reagents

1. 1.5 ml microcentrifuge tubes (SARSTEDT, catalog number: 72.690.001)
2. 200 µl PCR tubes (Labomedic, catalog number: 2081644AA)

bio-protocol www.bio-protocol.org/e2384 Vol 7, Iss 13, Jul 05, 2017
DOI:10.21769/BioProtoc.2384

3. Petri dishes (SARSTEDT, catalog number: 82.1472)
4. 2 ml microcentrifuge tubes (SARSTEDT, catalog number: 72.691)
5. 20 µl pipette tips (SARSTEDT, catalog number: 70.1116)
6. 200 µl pipette tips (SARSTEDT, catalog number: 70.760.012)
7. 1,000 µl pipette tips (SARSTEDT, catalog number: 70.762.010)
8. *Agrobacterium tumefaciens (A. tumefaciens)* strain GV3101::pMP90
9. *Arabidopsis thaliana* seeds (Col-0)
10. Vectors (see Figure 1)

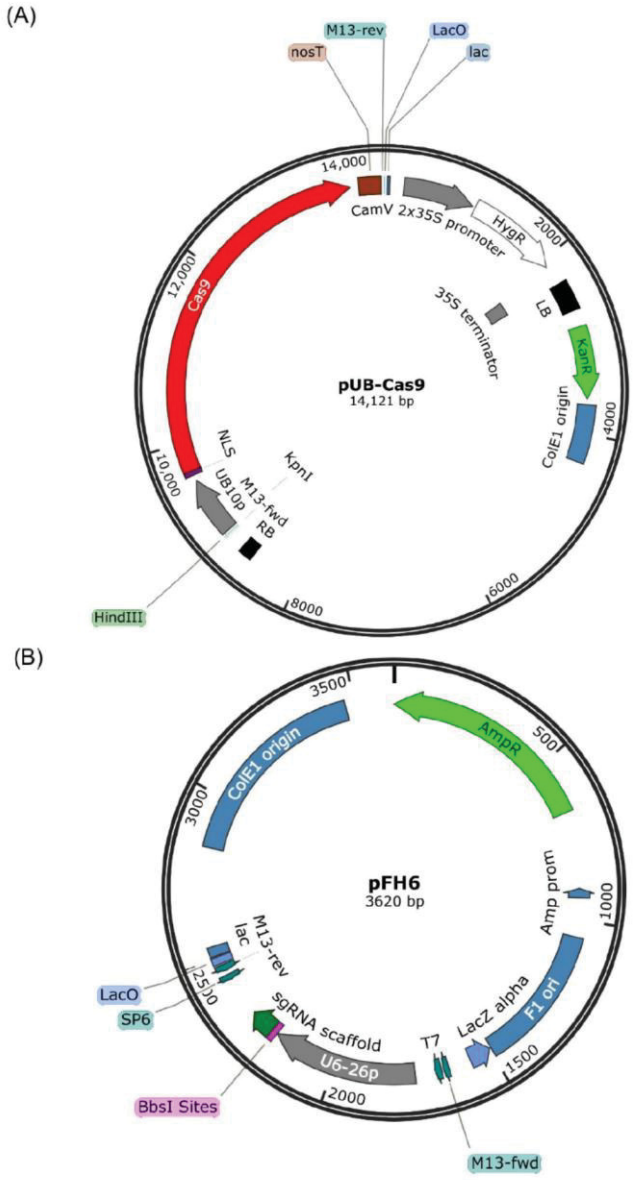


Figure 1. Vector maps of pUB-Cas9 (A) and pFH6 (B). pFH6 is used to integrate the 20 bp target sequence upstream of the sgRNA scaffold and under the control of the *Arabidopsis U6-26* RNA polymerase III promoter. The whole sgRNA cassette is then transferred via Gibson

bio-protocol www.bio-protocol.org/e2384 Vol 7, Iss 13, Jul 05, 2017
DOI:10.21769/BioProtoc.2384

cloning into the binary T-DNA vector pUB-Cas9 (contains the *Cas9* gene under the control of the *Ubiquitin10* promoter) for plant transformation. Maps were generated using SnapGene Viewer (http://www.snapgene.com/products/snapgene_viewer/).

a. Plant T-DNA Cas9 vector pUB-Cas9 (GenBank accession number KY080691), containing a *Chlamydomonas reinhardtii* codon-optimized ubiquitously (*UBIQUITIN10* promoter) expressed *Cas9* gene, a kanamycin resistance cassette for bacterial selection and a hygromycin resistance cassette as plant selection marker (Hahn *et al.*, [2017]; available at Addgene, catalog number: 86556)
b. sgRNA subcloning vector pFH6 (GenBank accession number KY080689) containing the *Arabidopsis U6-26* promoter, the integration site for the 20 bp protospacer sequence, the sgRNA scaffold and an ampicillin resistance cassette (Hahn *et al.* [2017]; available at Addgene, catalog number: 86555)
   Note: pFH6 contains an additional 9 bp fragment (GTCCCTTCG) between the 3' end of the U6-26 promoter and the protospacer integration site. In several experiments, we could show that this does not affect gene editing activity. However, we have also cloned a new version of the subcloning vector without the additional fragment (pFH6_new), which shows high cleavage activity in preliminary experiments and can be obtained from us. This version only contains an additional guanine (G) between the 3' end of the U6-26 promoter and the protospacer integration site, which allows the integration of any 20 bp protospacer without restriction of G as first bp (compare e.g., Fauser et al. [2014]). The cloning strategy for pFH6_new is analogous to the one described in this protocol, the only difference is that the forward primer for cloning your 20 bp protospacer sequence contains a different overlap and lacks the need for an initial G at the beginning (ATTG-N_{20}, compare procedure section).

11. Competent *Escherichia coli* (*E. coli*) cells (*e.g.*, Mach1™ competent cells, Thermo Fisher Scientific, Invitrogen™, catalog number: C862003)
12. *Bbs*I-HF + CutSmart buffer (New England Biolabs, catalog number: R3539S)
13. Distilled H_2O
14. Plasmid mini prep kit and agarose gel extraction kit (*e.g.*, GeneMATRIX 3 in 1–Basic DNA Purification Kit, Roboklon, catalog number: E3545)
15. T4 DNA ligase with 10x ligation buffer (New England Biolabs, catalog number: M0202S)
16. Ampicillin (Amp) (Carl Roth, catalog number: K029.2)
17. Primers (see Table 1)

bio-protocol www.bio-protocol.org/e2384 Vol 7, Iss 13, Jul 05, 2017
DOI:10.21769/BioProtoc.2384

Table 1. List of oligonucleotides

Primer name	Sequence (5'→3')
FH41	aaacgacggccagtgccaGAATTGGGCCCGACGTCG
FH42	tactgactcgtcgggtacCAAGCTATGCATCCAACGCG
FH61	CTGCCCAACGAGAAGGTGC
FH179	TATTACTGACTCGTCGGGTA
FH201	AGTCGGACAGGCGGTTGATG
FH254	gcccaattcCAAGCTATGCATCCAACGCG
FH255	catagcttgGAATTGGGCCCGACGTCG
M13F	GTAAAACGACGGCCAG
M13R	CAGGAAACAGCTATGAC

18. 5x Green GoTaq Reaction buffer (Promega, catalog number: M791A)
19. dNTPs (10 mM each) (Thermo Fisher Scientific, Thermo Scientific™, catalog number: R0192)
20. GoTaq G2 polymerase (Promega, catalog number: M7841) or other standard PCR polymerase
21. *Hind*III-HF + CutSmart buffer (New England Biolabs, catalog number: R3104S)
22. *Kpn*I-HF + CutSmart buffer (New England Biolabs, catalog number: R3142S)
23. Phusion High-Fidelity polymerase (New England Biolabs, catalog number: M0530S) or other proofreading polymerases
24. Gibson Assembly Cloning Kit (New England Biolabs, catalog number: E5510S)
25. Kanamycin sulfate (Kan) (Carl Roth, catalog number: T832.4)
26. Rifampicin (Rif) (Molekula, catalog number: 32609202)
27. Gentamycin sulfate (Gent) (Carl Roth, catalog number: 0233.3)
28. Hygromycin B (Hyg) (Carl Roth, catalog number: CP12.2)
29. T7 Endonuclease I (optional; New England Biolabs, catalog number: M0302S)
30. LB medium (agar plates and liquid), supplemented with 200 µg/ml ampicillin (see Recipes)
31. LB medium (agar plates and liquid), supplemented with 30 µg/ml kanamycin sulfate (see Recipes)
32. YEP medium (agar plates and liquid), supplemented with 150 µg/ml rifampicin, 50 µg/ml gentamycin sulfate, 50 µg/ml kanamycin (see Recipes)
33. ½ MS medium agar plates, supplemented with 33.3 µg/ml hygromycin B (see Recipes)

Equipment

1. Agarose gel electrophoresis equipment (*e.g.*, VWR, Peqlab, model: PerfectBlue™ Gel System Mini M, catalog number: 700-0434)
2. Bacteria plate incubators (28 °C, 37 °C, *e.g.*, Memmert, model: IN55) and shaker (28 °C, 37 °C; *e.g.*, Eppendorf, New Brunswick™, model: Innova® 44, catalog number: M1282-0002)
3. Heating blocks (*e.g.*, Eppendorf, model: Thermomixer Compact, catalog number: T1317-1EA)
4. PCR cycler (*e.g.*, Thermo Fisher Scientific, Applied Biosytems™, model: Veriti™ 96-well Thermal Cycler, catalog number: 4375786)
5. 10 µl pipette (*e.g.*, Pipetman Neo P10N, Gilson, catalog number: F144562)

bio-protocol www.bio-protocol.org/e2384 Vol 7, Iss 13, Jul 05, 2017
DOI:10.21769/BioProtoc.2384

6. 20 µl pipette (*e.g.*, P20N, Gilson, catalog number: F144563)
7. 200 µl pipette (*e.g.*, P200N, Gilson, catalog number: F144565)
8. 1,000 µl pipette (*e.g.*, P1000N, Gilson, catalog number: F144566)
9. Plant growth chamber (*e.g.*, CLF Plant Climatics, Percival Scientifici, model: AR-66L)

Software

1. SerialCloner
2. Cas-OFFinder
3. 4Peaks software
4. MUSCLE (Optional)

Procedure

A. Design of sgRNA sequence for gene knockout

The design of the sgRNA is possibly the most critical step of Cas9-mediated gene editing. Thus, the Cas9 target site should be carefully selected. Several online tools (Heigwer *et al.*, 2014; Lei *et al.*, 2014; Xie *et al.*, 2014) assist with finding suitable target sites, however, we still choose most target sites manually in standard molecular biology software (*e.g.*, SerialCloner, http://serialbasics.free.fr/Serial_Cloner.html). When choosing the target site, the following aspects should be taken into account:

1. Search your gene of interest (*GOI*) for a $GN_{19}NGG$ sequence on both strands to find putative target sites. The NGG is the *Streptococcus pyogenes* Cas9 protospacer adjacent motive (PAM), which has to be present in the target locus for efficient binding of the Cas9 protein (Jinek *et al.*, 2012). The GN_{19} represents the 20 bp long protospacer motif, defining the target site where you want to create mutations. This 20 bp sequence, which marks the 5' end of your sgRNA molecule, should start with a G, since this is a prerequisite for transcription of the sgRNA under the control of the *U6-26* promoter of *Arabidopsis*.

   Note: If there is no suitable target sequence, you might just add the G as 21st bp, even if it does not have a complementary base in the target locus. These sgRNAs might still lead to cleavage at the target site (e.g., Hahn et al., 2017). However, we recommend using the canonical 20 bp protospacer sequences if possible.
2. Choose a target site in an exon of your *GOI*, which is preferably close to the 5' end of the coding sequence, since mutations in this region are more likely to disrupt the protein's function. If you know highly important domains of the protein, such as nucleic acid binding sites, you might want to create mutations in these regions.
3. Consider choosing several different target sites, since not all sgRNAs work equally well (Liang *et al.*, 2016). Additionally, you might want to use two sgRNAs at once (see cloning strategy) to create large deletions, which might cause more drastic effects.

bio-protocol www.bio-protocol.org/e2384 Vol 7, Iss 13, Jul 05, 2017
DOI:10.21769/BioProtoc.2384

4. Search for putative off-target sites to avoid knocking out genes that share the 20 bp target site or highly similar sites. For this purpose, we use the Cas-OFFinder program (Bae *et al*., 2014), which is available at http://www.rgenome.net/cas-offinder/. Here, we use the following settings:
   PAM-Type: SpCas9 from *Streptococcus pyogenes*: 5'-NRG-3' (R = A or G)
   Target genome: *Arabidopsis thaliana* (TAIR 10)
   Mismatch number: 3
   After searching for your 20 bp target site in the reference genome, the program displays all similar sites with up to 3 mismatches. If you have more than one hit with zero mismatches (one is always your *GOI*), you should choose a different target site. If you have 1-3 mismatches, you should take into account that Cas9 might also mutagenize these sites, especially if these mismatches are at the 5' end of the 20 bp sequence (Semenova *et al*., 2011). Even though the affinity for Cas9 will be lower than to your *GOI* and plants seem to be in general less affected by off targeting (Feng *et al*., 2014), we recommend careful sequencing of these off target site in the knockout plants to ensure the absence of mutations in these genes.
5. Consider using a 20 bp target site that also contains a recognition site of a commercially available restriction enzyme. The target site of the restriction enzyme should span the site of Cas9 cleavage (3 bp in front of the PAM). This will allow you to use the restriction enzyme length polymorphism assay to detect mutations in your *GOI*.

B. Subcloning of 20 bp protospacer sequence into pFH6
   1. An overview how to perform the subcloning of the protospacer sequence into pFH6 is given in Figure 2.
   2. Order forward and reverse primers for your candidate target sequence N_{20} with overlaps for cloning and dilute them to a final concentration (f.c.) of 10 µM.
      Forward primer: TTCG-GN_{19}
      Reverse primer: AAAC-$N_{19\ (reverse\ complement)}$ C (see Figure 2)

bio-protocol www.bio-protocol.org/e2384 Vol 7, Iss 13, Jul 05, 2017
DOI:10.21769/BioProtoc.2384

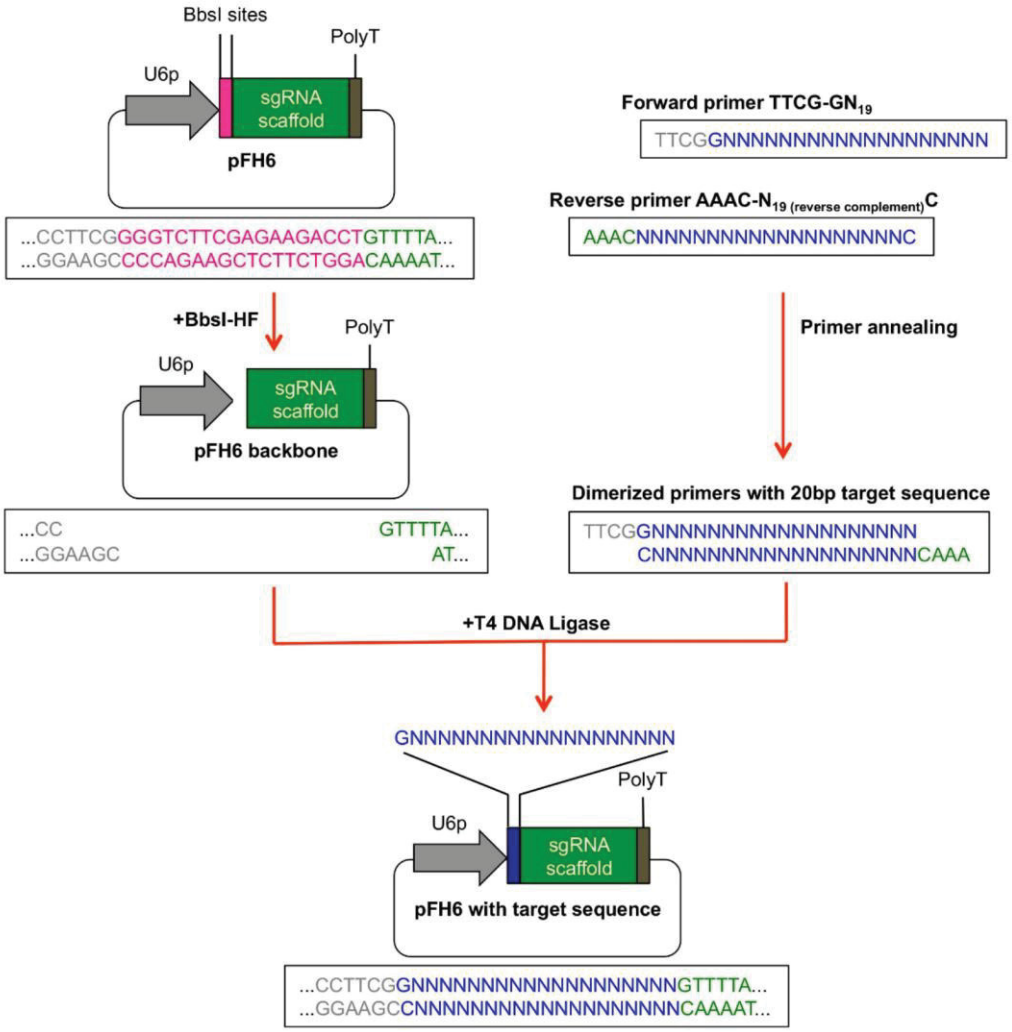


Figure 2. Subcloning of protospacer sequence in pFH6. The 20 bp protospacer sequence can be integrated between the sgRNA scaffold and the *U6-26* promoter (U6p) by ligation of dimerized primers into *Bbs*I-digested pFH6 vector.

3. Pipette 10 µl of each primer dilution into a 1.5 ml microcentrifuge tube and mix.
4. Heat primer mix to 98 °C, 10 min, then 55 °C, 10 min, afterwards slowly cool down to RT (*e.g.*, by switching off the heating block). By doing so, a double stranded DNA including your protospacer sequence (= annealed primers) will be generated.
5. Digest pFH6 with *Bbs*I-HF for 3 h at 37 °C.

pFH6	3 µg
10x CutSmart buffer	5 µl
ddH_2O	to 49 µl
*Bbs*I-HF (20 U/µl)	1 µl

bio-protocol www.bio-protocol.org/e2384 Vol 7, Iss 13, Jul 05, 2017
DOI:10.21769/BioProtoc.2384

6. Run the restriction digest on an agarose gel and extract the vector backbone (3,612 bp) from the gel using the GeneMATRIX 3 in 1–Basic DNA Purification Kit.
7. Ligate annealed primers with digested pFH6 at 4 °C overnight.

pFH6 backbone	300 ng
Annealed primers	1 µl
10x ligation buffer	2 µl
ddH_2O	to 19 µl
T4 DNA ligase	1 µl

8. Transform competent MachI *E. coli* cells with the ligation reaction and spread the transformed cells on LB agar plates supplemented with Amp.
   Note: Usually, plating 100 µl of the transformation mixture is sufficient. Protocols for transformation of E. coli cells are provided by the authors upon request.
9. Verify eight *E. coli* colonies by colony PCR using primer M13R and your protospacer primer TTCG-N_{20} with the following setup:

5x Green GoTaq Reaction buffer	4 µl
M13R (10 µM)	1 µl
TTCG-N_{20} (10 µM)	1 µl
dNTPs (10 mM each)	0.4 µl
ddH_2O	13.4 µl
GoTaq G2 polymerase (5 U/µl)	0.2 µl
Pick half of a colony with a pipette tip and suspend it in the PCR reaction.	
Use the other half of the colony for inoculation.	

PCR program:

Initial denaturation	95 °C	2 min	
Denaturation	95 °C	30 sec	35 cycles
Annealing	51 °C	30 sec	
Elongation	72 °C	15 sec	
Final elongation	72 °C	5 min	
Storage	20 °C	∞	

Note: We recommend using pFH6 as negative control. We have sometimes observed fragments at the expected band size (281 bp) in the negative control since the protospacer primer might partly bind to the BbsI-sites. However, positive colonies usually have a much stronger PCR product signal and can therefore be distinguished from negative clones.

10. Inoculate positive colony in LB medium supplemented with Amp and grow overnight.
11. Purify the plasmid using the GeneMATRIX 3 in 1–Basic DNA Purification Kit.

bio-protocol www.bio-protocol.org/e2384 Vol 7, Iss 13, Jul 05, 2017
DOI:10.21769/BioProtoc.2384

12. Sequence the plasmid and verify the sequence of the protospacer insertion using primer M13R (Table 1).

C. Cloning of final plant T-DNA Cas9 vector using Gibson Assembly

1. An overview how to clone the final plant T-DNA Cas9 vector is given in Figure 3.

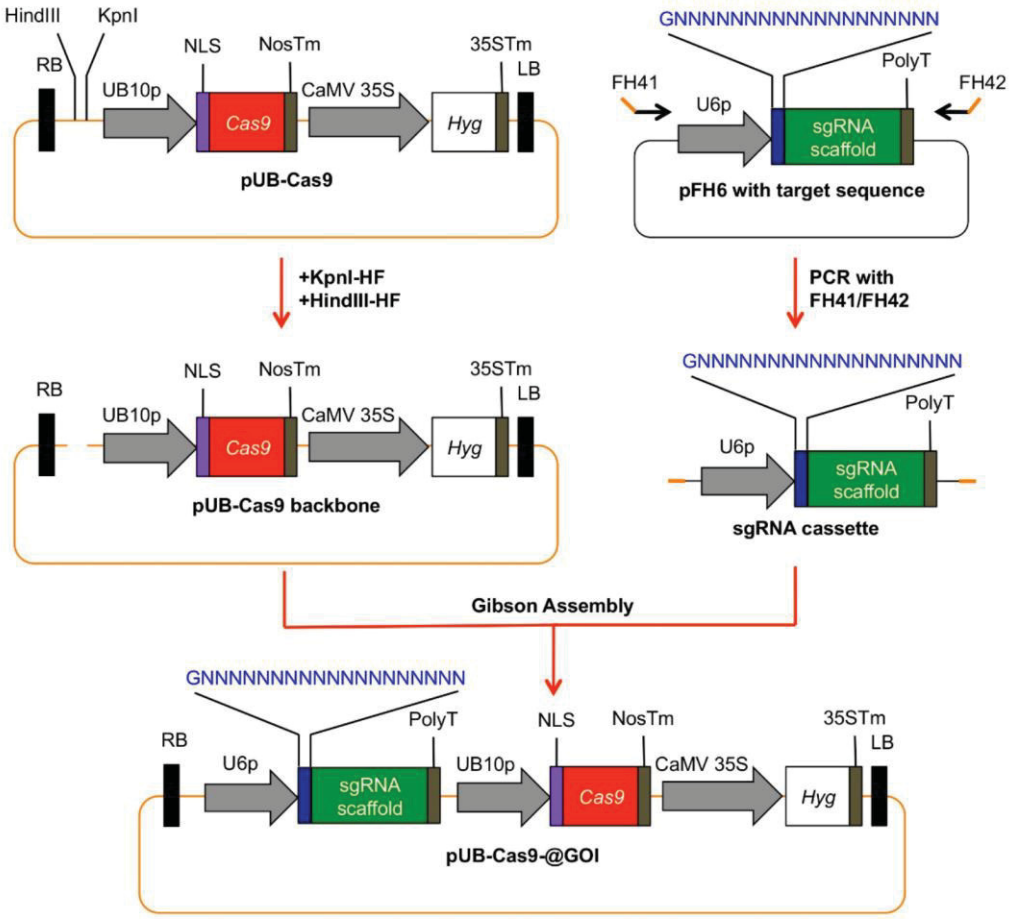


Figure 3. Generation of a plant T-DNA vector with all necessary components for Cas9 induced mutagenesis. The sgRNA cassette including the 20 bp target site is amplified from pFH6 and integrated into *Kpn*I/*Hind*III-digested pUB-Cas9 via Gibson Assembly. The resulting vector contains the sgRNA cassette as well as the *Cas9* gene under the control of the *UBIQUITIN10* promoter (UB10p), and the hygromycin (*Hyg*) resistance gene as selection marker. RB, right T-DNA border; PolyT, polythymidine transcription terminator; NLS, nuclear localization signal; NosTm, nopaline synthase terminator; CaMV35S, Cauliflower Mosaic Virus 2x35S promoter; 35STm, 35S terminator; LB, left T-DNA border.

bio-protocol www.bio-protocol.org/e2384 Vol 7, Iss 13, Jul 05, 2017
DOI:10.21769/BioProtoc.2384

2. Digest vector pUB-Cas9 with *Hind*III-HF and *Kpn*I-HF for 3 h at 37 °C.

pUB-Cas9	3 µg
10x CutSmart buffer	5 µl
ddH_2O	to 48 µl
*Kpn*I-HF (20 U/µl)	1 µl
*Hind*III-HF (20 U/µl)	1 µl

Note: Efficient digestion of the pUB-Cas9 plasmid is critical for further cloning steps. In case you have problems in the subsequent cloning steps, try to digest pUB-Cas9 with a fresh batch of enzymes and longer incubation times.

3. Run out the restriction digest reaction on an agarose gel and extract the vector backbone (14,097 bp) from the gel using the GeneMATRIX 3 in 1–Basic DNA Purification Kit.
4. Amplify the sgRNA cassette (expected size: 771 bp) from pFH6 with target sequence using primer FH41 and FH42 (Table 1) and Phusion High-Fidelity polymerase and the following setup:

5x Phusion HF-buffer	10 µl
FH41 (10 µM)	2.5 µl
FH42 (10 µM)	2.5 µl
dNTPs (10 mM each)	1 µl
ddH_2O	33 µl
pFH6 with target sequence (100 ng/µl)	0.5 µl
Phusion DNA polymerase (2 U/µl)	0.5 µl

PCR program:

Initial denaturation	98 °C	30 sec	
Denaturation	98 °C	15 sec	35 cycles
Annealing	69 °C	30 sec	
Elongation	72 °C	20 sec	
Final elongation	72 °C	5 min	
Storage	20 °C	∞	

5. Run out the PCR reaction on an agarose gel and extract the sgRNA cassette (771 bp) from the gel using the GeneMATRIX 3 in 1–Basic DNA Purification Kit.
6. Assemble the pUB-Cas9 backbone and the sgRNA cassette to generate the final vector pUB-Cas9-@GOI using Gibson Assembly Master Mix according to the manufacturer's instruction with an incubation time of 1 h.

   Note: We propose a nomenclature for the final constructs, where the target gene is marked by the @-sign. A vector designed against the GL1 gene of Arabidopsis would therefore be called pUB-Cas9-@GL1 (Hahn et al., 2017).

bio-protocol www.bio-protocol.org/e2384 Vol 7, Iss 13, Jul 05, 2017
DOI:10.21769/BioProtoc.2384

7. Transform competent *E. coli* cells with the Gibson reaction mixture and spread the transformed cells on LB agar plates supplemented with Kan.
   Note: New England Biolabs provides highly competent E. coli cells in the Gibson Assembly Cloning Kit. However, we also used competent E. coli cells produced in our lab. Since these are usually less competent, we recommend plating all transformed cells on the LB agar plate. A protocol for generation of competent MachI E. coli can be found in Sambrook and Russell (2006).
8. Verify eight positive colonies by colony PCR using primer M13F and FH179 (Table 1) with the following setup:

5x Green GoTaq Reaction buffer	4 µl
M13F (10 µM)	1 µl
FH179 (10 µM)	1 µl
dNTPs (10 mM each)	0.4 µl
ddH_2O	13.4 µl
GoTaq G2 polymerase (5 U/µl)	0.2 µl
Pick half of a colony with a pipette tip and suspend it in the PCR reaction. Use the other half of the colony for inoculation.	

   PCR program:

Initial denaturation	95 °C	2 min	
Denaturation	95 °C	30 sec	35 cycles
Annealing	52 °C	30 sec	
Elongation	72 °C	45 sec	
Final elongation	72 °C	5 min	
Storage	20 °C	∞	

   Note: Expected product size is 779 bp, pUB-Cas9 vector can be used as negative control.
9. Inoculate positive colony into LB medium supplemented with Kan and let grow overnight.
10. Purify the plasmid using the GeneMATRIX 3 in 1–Basic DNA Purification Kit.
11. Sequence the plasmid for verification of protospacer sequence insertion using primer FH179 and M13F (Table 1).
    Note: The vector system also allows cloning of two different sgRNAs into the pUB-Cas9 vector backbone to induce large deletions in the target genome or to destroy two genes at once. Therefore, both protospacer sequences have to be cloned separately into pFH6 as described. Then, the first sgRNA cassette has to be amplified from its subcloning pFH6-construct using the primers FH41 and FH254 (Table 1). The second sgRNA cassette has to be amplified with FH42 and FH255 (Table 1), using the PCR conditions described above. Using the Gibson Assembly Master Mix, both cassettes can be integrated into KpnI-HF/HindIII-HF-digested pUB-Cas9 in a single step. As two fragments are integrated into the digested vector, the amount of colonies is

bio-protocol www.bio-protocol.org/e2384 Vol 7, Iss 13, Jul 05, 2017
DOI:10.21769/BioProtoc.2384

usually lower compared to the integration of one fragment. However, the proportion of correctly assembled vectors is comparable. The colony PCR can be performed with the same primers but using a longer elongation time (1.5 min), since the PCR reaction amplifies both sgRNA cassettes (expected product size: 1,512 bp). Sequencing should be performed as described above.

All cloning steps are summarized in Table 2:

Table 2. Timetable for cloning of plant T-DNA Cas9 vector for targeted gene knockout

Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
Digest of pFH6 and gel extraction of backbone	Transformation of ligated pFH6 with target sequence in *E. coli*	Colony PCR of pFH6 with target sequence	Plasmid purification	Amplification of sgRNA cassette from pFH6 with target sequence with FH41/FH42 and gel extraction of sgRNA cassette	Colony PCR of pUB-Cas9-@GOI	Plasmid purification
Protospacer annealing		Inoculation of positive colonies	Sequencing	Gibson Assembly of digested pUB-Cas9 and sgRNA cassette resulting in pUB-Cas9-@GOI	Inoculation of positive colonies	Sequencing
Ligation of pFH6 and annealed protospacer overnight			Digest of pUB-Cas9 and gel extraction of backbone	Transformation of pUB-Cas9-@GOI in *E. coli*		

D. *Arabidopsis* transformation with pUB-Cas9-@GOI

1. Transform competent *A. tumefaciens* cells with 1 µg of pUB-Cas9-@GOI and spread the transformed cells on YEP agar plates supplemented with Rif/Gent/Kan.
   Note: Protocols for generation of competent A. tumefaciens cells and transformation of A. tumefaciens cells are described elsewhere (Hofgen and Willmitzer, 1988).
2. After two days of growth at 30 °C, verify plasmid presence in at least three colonies by colony PCR using primer FH61 and FH201 (Table 1). In order to verify colony, pick half of a colony with a pipette tip and suspend cells in a 1.5 ml microcentrifuge tube containing 100 µl ddH_2O. Heat to 98 °C for 10 min and use 2 µl as DNA template for the PCR. Streak the other half of the colony on a fresh YEP agar plate supplemented with Rif/Gent/Kan. Grow cells at 30 °C for one day and then store at 4 °C. For the colony PCR, use the following setup:

bio-protocol www.bio-protocol.org/e2384 Vol 7, Iss 13, Jul 05, 2017
DOI:10.21769/BioProtoc.2384

5x GoTaq Green Reaction buffer	4 µl
FH61 (10 µM)	1 µl
FH201 (10 µM)	1 µl
dNTPs (10 mM each)	0.4 µl
ddH_2O	11.4 µl
DNA template	2 µl
GoTaq G2 polymerase (5 U/µl)	0.2 µl

PCR program:

Initial denaturation	95 °C	2 min	
Denaturation	95 °C	30 sec	35 cycles
Annealing	60 °C	30 sec	
Elongation	72 °C	1 min	
Final elongation	72 °C	5 min	
Storage	20 °C	∞	

Note: Expected product size is 1,003 bp.

3. Transform *Arabidopsis* plants using *Agrobacterium*-mediated T-DNA transfer with the floral dipping method (Clough and Bent, 1998).
   Note: Per construct, we usually transform four pots with five Arabidopsis plants, each. We recommend removing of all present siliques before transformation to avoid excessive screening for transformation events afterwards.
4. Select the T1 generation plants for successful T-DNA insertion events on ½ strength MS agar plates supplemented with Hyg. Verify the integration of the T-DNA by isolating leaf genomic DNA (gDNA) and subsequent PCR using primers FH61/FH201with the parameters described before.
   Note: A detailed protocol for plant gDNA extraction by isopropanol precipitation can be found elsewhere (Weigel and Glazebrook, 2009).

E. Detection of Cas9 induced mutations

The Cas9 protein in our vector system is transcribed under the control of the *UBIQUITIN10* promoter of *Arabidopsis*, which allows Cas9 expression in nearly all tissues of the plant (Norris *et al*., 1993). Since the editing mechanisms of the Cas9 system work independently in each plant cell, T1 generation plants are usually chimeras with distinct genotypes in different cells. Only if mutations occur in germline cells, can progeny plants arise that have a uniform genotype. This is usually the prerequisite for further analyses. Therefore (and due to the limited amount of plants available in the T1 generation), screening approaches for detection of mutations should be applied to T2 or following generations. We recommend screening descendants of three to four independent transformed lines, since mutation efficiencies might vary between different lines due to the integration site of the T-

bio-protocol www.bio-protocol.org/e2384 Vol 7, Iss 13, Jul 05, 2017
DOI:10.21769/BioProtoc.2384

DNA cassette. To choose promising T1 lines for further screening, we also searched for mutational events in the T1 generation. We suggest to screen the plants for mutation in this way:

1. Amplify the genomic region around the Cas9 cutting site in your *GOI* (± 300-500 bp upstream and downstream) with gene specific primers from leaf gDNA using Phusion polymerase.

5x Phusion HF- buffer	10 µl
Forward primer (10 µM)	2.5 µl
Reverse primer (10 µM)	2.5 µl
dNTPs (10 mM each)	1 µl
ddH_2O	32.5 µl
Leaf gDNA	1 µl
Phusion DNA polymerase (2 U/µl)	0.5 µl

PCR program:

Initial denaturation	98 °C	2 min	
Denaturation	98 °C	15 sec	35 cycles
Annealing	variable	30 sec	
Elongation	72 °C	30 sec/kb	
Final elongation	72 °C	5 min	
Storage	20 °C	∞	

2. Run out the PCR products on an agarose gel and extract your *GOI* from the gel using the GeneMATRIX 3 in 1–Basic DNA Purification Kit.
3. Sequence PCR product with Sanger sequencing. When Cas9 was active and mutations occurred at least in a part of the leaf cells, the sequencing histogram will show a mixture of WT sequence and mutated sequences (Figure 4).

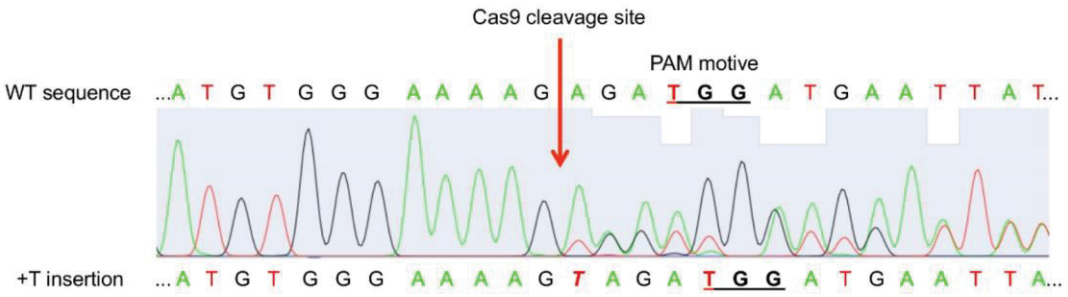


Figure 4. Exemplary sequencing result histogram of amplified target gene from chimeric leaf gDNA. Double peaks appear at the site of Cas9 cleavage (3 bp in front of the PAM motive NGG) pointing to mutation events in the *GOI* in a part of the leaf cells.

bio-protocol www.bio-protocol.org/e2384 Vol 7, Iss 13, Jul 05, 2017
DOI:10.21769/BioProtoc.2384

The efficiency of the Cas9 system varies considerably as its mutagenic effect relies on various aspects, such as: the plant species, Cas9 and sgRNA expression levels, and the sequence of the target loci (Bortesi and Fischer, 2015; Paul and Qi, 2016). Therefore, it is difficult to predict how many plants have to be screened to identify a heritable mutation. We recommend starting screening 30 plants per line. Screening for mutations in the T2 generation can be performed by amplification and Sanger sequencing of your *GOI* as described before. As this can become costly, several methods enable detecting mutations on genomic level without sequencing.

4. If you expect a specific phenotype due to the knockout of your gene, the easiest way to screen for mutations is a visual screen in the first place. These phenotypes can be changes in anatomy or survival of plants on selection medium. Plants that show a phenotype are then analyzed on genomic level as described above.
5. If you target two sites in your *GOI* at the same time, a simultaneous cleavage with two sgRNAs can result in large deletions, which can already be detected during the PCR amplification of your *GOI* (Figure 5A). Larger deletions can in theory also occur using one sgRNA, however, we usually discovered small indels in the range of a few bp, which usually cannot be detected by PCR product size difference using agarose gel electrophoresis.
   Note: If you expect that only few cells contain the expected mutation (e.g., in chimeric plants), you can improve the PCR amplification efficiency of the mutant alleles by digesting unmutated gDNA with a restriction enzyme that cuts between the two sgRNA target sites before PCR amplification of your GOI.

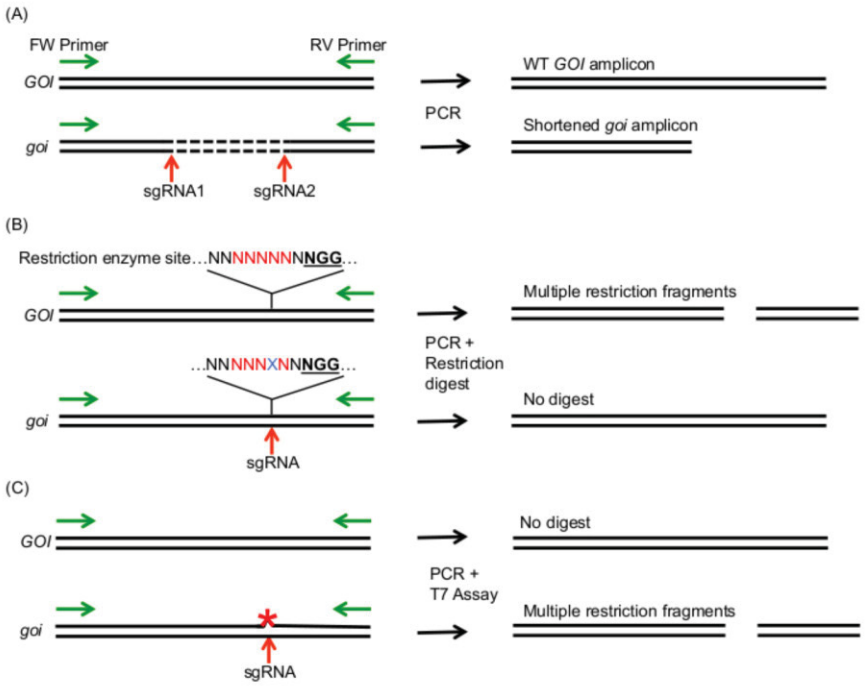


Figure 5. Summary of methods that allow the detection of induced mutations after amplification of the *GOI* with a forward (FW) and reverse (RV) primer (green arrows). A.

bio-protocol www.bio-protocol.org/e2384 Vol 7, Iss 13, Jul 05, 2017
DOI:10.21769/BioProtoc.2384

Usage of two sgRNAs can induce larger deletions in the *GOI,* which can be detected by PCR amplicon size difference between WT and mutated *GOI* (*goi*). B. If a restriction site (red letters) is present in the mutagenic region 3 bp in front of the PAM site (bold, underlined), mutations (blue X) lead to cleavage resistant PCR amplicons. C. The T7 Endonuclease assay allows cleavage of heteroduplexes, which are produced by imperfect DNA pairing due to mutations on one strand of the DNA amplicon (red asterisk).

6. The restriction fragment length polymorphism (RFLP) assay is an elegant method to detect bp exchanges or small indels in your *GOI*. The advantage of this method is that heterozygous, biallelic and homozygous mutations can be detected (Figure 5B). The RFLP assay relies on the disruption of a restriction enzyme target site by Cas9 induced mutations. Therefore, the restriction enzyme target site should span the region at 3 bp in front of the PAM motive. For detection of mutations, the *GOI* is amplified with gene specific primers as described above. The PCR amplicon is then digested with the restriction enzyme of choice according to the manufacturer's instructions. The sizes of the restriction fragments are analyzed by agarose gel electrophoresis. For the RFLP assay, amplify your *GOI* as described before. Separate the PCR products on an agarose gel and extract the PCR product from the gel using the GeneMATRIX 3 in 1–Basic DNA Purification Kit. Digest the PCR product with the suitable restriction enzyme according to the manufacturer's instruction.
   Notes:
   a. *WT gDNA should be used as negative control.*
   b. *In case you have a homozygous or a heterozygous mutation in your GOI, you either get only the undigested PCR amplicon or a mixture of undigested and digested PCR bands.*
   c. *It can be favorable if your PCR product contains more than one cleavage site of your restriction enzyme (which are not targeted by the Cas9) as it provides you with an internal control that the digest itself worked.*
   d. *If you have already detected mutations in your T1 plants by sequencing, you can use the sequenced PCR product as positive control for this assay.*
   e. *If you expect only small amounts of mutant alleles (e.g., in chimeric plants), you can enrich these amplicons by digesting the gDNA with the restriction enzyme before PCR amplification of your GOI. In this case, no other cut sites besides the one in the sgRNA target site should be present between your primer annealing sites.*
7. The T7 endonuclease assay and the Surveyor assay allow detection of small indels and bp exchanges in your target site (Figure 5C). First, the *GOI* is PCR amplified using a proofreading DNA polymerase. The PCR amplicon is then denatured and reannealed. In this step, ssDNA strands containing a mutation will anneal to DNA strands that do not contain a mutation, leading to non-perfectly matching dsDNA strands. These heteroduplexes are then recognized and cleaved by the T7 Endonuclease I or the Surveyor Nuclease. The restriction fragment pattern is analyzed by agarose gel electrophoresis. The advantage of these methods is that they do not

bio-protocol www.bio-protocol.org/e2384 Vol 7, Iss 13, Jul 05, 2017
DOI:10.21769/BioProtoc.2384

require a restriction site within your gRNA target site. For details on this method, we refer to detailed protocols of the Endonuclease suppliers (*e.g.*, https://www.neb.com/protocols/2014/08/11/determining-genome-targeting-efficiency-using-t7-endonuclease-i) and the experimental procedures of other research groups (Mao *et al.*, 2013; Xie and Yang, 2013; Qiu *et al.*, 2004).

Note: These assays are not able to detect homozygous mutations as no heteroduplexes occur.

Data analysis

1. There is no specific requirement for large-scale data analysis to generate Cas9-mediated knockout mutants. The mutagenic effect of the Cas9 system varies considerably depending on the plant species, Cas9 and sgRNA expression levels, and the sequence of the target loci (Bortesi and Fischer, 2015; Paul and Qi, 2016). Additionally, the Cas9 system produces chimeric plants due to the possibility of independent mutation events in each cell. Therefore, statistical analyses of mutation efficiencies are complicated and the results of one experiment might not be transferable to another one, especially if a different sgRNA is used.
2. Sequencing data obtained from transformed plants can be compared to sequences from WT plants using alignment tools like MUSCLE with default settings (Edgar, 2004). Sequencing histograms can be analyzed using 4Peaks software.

Notes

1. All analyses on gDNA level (sequencing, RFLP-assay, T7 assay) rely on the gDNA of a single leaf. If your knockout mutant does not display a visible phenotype and the Cas9 cassette is still integrated into the plant genome, it is difficult to tell whether you are analyzing a chimeric mutation only present in the leaf that you used for gDNA isolation or if your whole plant is already mutagenized. You should therefore consider the following aspects.
   a. Isolation of gDNA from several different leaves followed by *GOI*-amplification should result in similar sequencing histograms.
   b. The mutation should show a Mendelian inheritance pattern in the progeny generation.
   c. Loss of the Cas9 T-DNA cassette by backcrossing and/or segregation should not affect the mutation. The presence of the Cas9 T-DNA can be detected by PCR using primers FH61/FH201 and plant gDNA as template.

Recipes

1. LB-medium
   10 g/L Bacto-tryptone (BD)
   5 g/L yeast extract (BD)

bio-protocol www.bio-protocol.org/e2384 Vol 7, Iss 13, Jul 05, 2017
DOI:10.21769/BioProtoc.2384

5 g/L NaCl (Fisher Scientific)
Optional: 15 g/L Bacto agar for plates (BD)

2. YEP-medium
   10 g/L yeast extract (BD)
   10 g/L Bacto-peptone (BD)
   5 g/L NaCl
   Optional: 15 g/L Bacto agar for plates (BD)
3. ½ MS-medium
   2.2 g/L Murashige & Skoog medium (Duchefa Biochemie)
   0.5 g/L MES (Carl Roth)
   Adjust pH to 5.7 with KOH (Carl Roth)
   Optional: 8 g/L plant agar for plates (Duchefa Biochemie)

Acknowledgments

This work was funded by the Cluster of Excellence on Plant Science (CEPLAS, EXC 1028) and the HHU Center for Synthetic Life Sciences (CSL). We thank Peter Hegemann and André Greiner (Humboldt-Universität zu Berlin) for providing us with Cas9 and sgRNA genes. We thank Holger Puchta, Felix Wolter and Dr. Nadine Rademacher for helpful discussions. We thank Franziska Kuhnert for critical reading of the manuscript. The protocol is adapted from Hahn *et al*. (2017).

References

1. Bae, S., Park, J., and Kim, J. S. (2014). Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* 30(10): 1473-1475.
2. Baltes, N. J. and Voytas, D. F. (2015). Enabling plant synthetic biology through genome engineering. *Trends Biotechnol* 33(2): 120-131.
3. Bortesi, L. and Fischer, R. (2015). The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnol Adv* 33(1): 41-52.
4. Clough, S. J. and Bent, A. F. (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* 16(6): 735-743.
5. Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5), 1792-1797.
6. Fauser, F., Schiml, S., and Puchta, H. (2014). Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in *Arabidopsis thaliana*. *Plant J* 79(2): 348-359.
7. Feng, Z., Mao, Y., Xu, N., Zhang, B., Wei, P., Yang, D. L., Wang, Z., Zhang, Z., Zheng, R., Yang, L., Zeng, L., Liu, X. and Zhu, J. K. (2014). Multigeneration analysis reveals the inheritance,

bio-protocol www.bio-protocol.org/e2384 Vol 7, Iss 13, Jul 05, 2017
DOI:10.21769/BioProtoc.2384

specificity, and patterns of CRISPR/Cas-induced gene modifications in *Arabidopsis*. *Proc Natl Acad Sci U S A* 111(12): 4632-4637.

8. Hahn, F., Mantegazza, O., Greiner, A., Hegemann, P., Eisenhut, M. and Weber, A. P. (2017). An efficient visual screen for CRISPR/Cas9 activity in *Arabidopsis thaliana*. *Front Plant Sci* 8: 39.
9. Heigwer, F., Kerr, G. and Boutros, M. (2014). E-CRISP: fast CRISPR target site identification. *Nat Methods* 11(2): 122-123.
10. Hofgen, R. and Willmitzer, L. (1988). Storage of competent cells for *Agrobacterium* transformation. *Nucleic Acids Res* 16(20): 9877.
11. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337(6096): 816-821.
12. Lei, Y., Lu, L., Liu, H. Y., Li, S., Xing, F. and Chen, L. L. (2014). CRISPR-P: a web tool for synthetic single-guide RNA design of CRISPR-system in plants. *Mol Plant* 7(9): 1494-1496.
13. Liang, G., Zhang, H., Lou, D. and Yu, D. (2016). Selection of highly efficient sgRNAs for CRISPR/Cas9-based plant genome editing. *Sci Rep* 6: 21451.
14. Mao, Y., Zhang, H., Xu, N., Zhang, B., Gou, F. and Zhu, J. K. (2013). Application of the CRISPR-Cas system for efficient genome engineering in plants. *Mol Plant* 6(6): 2008-2011.
15. Norris, S. R., Meyer, S. E. and Callis, J. (1993). The intron of *Arabidopsis thaliana* polyubiquitin genes is conserved in location and is a quantitative determinant of chimeric gene expression. *Plant Mol Biol* 21(5): 895-906.
16. Pacher, M. and Puchta, H. (2016). From classical mutagenesis to nuclease-based breeding - directing natural DNA repair for a natural end-product. *Plant J*.
17. Paul, J. W., 3rd and Qi, Y. (2016). CRISPR/Cas9 for plant genome editing: accomplishments, problems and prospects. *Plant Cell Rep* 35(7): 1417-1427.
18. Qiu, P., Shandilya, H., D'Alessio, J. M., O'Connor, K., Durocher, J. and Gerard, G. F. (2004). Mutation detection using Surveyor nuclease. *Biotechniques* 36(4): 702-707.
19. Salomon, S. and Puchta, H. (1998). Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. *EMBO J* 17(20): 6086-6095.
20. Sambrook, J. and Russell, D. W. (2006). The inoue method for preparation and transformation of competent *E. coli*: "ultra-competent" cells. *CSH Protoc* 2006(1).
21. Schiml, S., Fauser, F., and Puchta, H. (2014). The CRISPR/Cas system can be used as nuclease for in planta gene targeting and as paired nickases for directed mutagenesis in *Arabidopsis* resulting in heritable progeny. *Plant J* 80: 1139-1150.
22. Schiml, S. and Puchta, H. (2016). Revolutionizing plant biology: multiple ways of genome engineering by CRISPR/Cas. *Plant Methods* 12: 8.
23. Semenova, E., Jore, M. M., Datsenko, K. A., Semenova, A., Westra, E. R., Wanner, B., van der Oost, J., Brouns, S. J. and Severinov, K. (2011). Interference by clustered regularly interspaced

bio-protocol www.bio-protocol.org/e2384 Vol 7, Iss 13, Jul 05, 2017
DOI:10.21769/BioProtoc.2384

short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc Natl Acad Sci U S A* 108(25): 10098-10103.

24. Weigel, D. and Glazebrook, J. (2009). Quick miniprep for plant DNA isolation. *Cold Spring Harb Protoc* 2009(3): pdb prot5179.
25. Xie, K. and Yang, Y. (2013). RNA-guided genome editing in plants using a CRISPR-Cas system. *Mol Plant* 6(6): 1975-1983.
26. Xie, K., Zhang, J. and Yang, Y. (2014). Genome-wide prediction of highly specific guide RNA spacers for CRISPR-Cas9-mediated genome editing in model plants and major crops. *Mol Plant* 7(5): 923-926.

MANUSCRIPT III

Homology-directed repair of a defective *glabrous* gene in *Arabidopsis* with Cas9-based gene targeting

Homology-directed repair of a defective *glabrous* gene in *Arabidopsis* with Cas9-based gene targeting

Florian Hahn[1], Marion Eisenhut[1], Otho Mantegazza[1], Andreas P.M. Weber[1]*

[1]Institute of Plant Biochemistry, Cluster of Excellence on Plant Sciences (CEPLAS), Center for Synthetic Life Sciences (CSL), Heinrich Heine University, Düsseldorf, Germany

***Correspondence:**
Andreas P.M. Weber
Andreas.Weber@uni-duesseldorf.de

Keywords: CRISPR/Cas9, homologous recombination, Glabrous1, trichomes, gene editing, marker, *in planta* gene targeting, viral replicons, gene targeting

ABSTRACT

The CRISPR/Cas9 system has emerged as a powerful tool for targeted genome editing in plants and beyond. Double-strand breaks induced by the Cas9 enzyme are repaired by the cell's own repair machinery either by the non-homologous end joining pathway or by homologous recombination. While the first repair mechanism results in random mutations at the double-strand break site, homologous recombination uses the genetic information from a highly homologous repair template as blueprint for repair of the break. By offering an artificial repair template, this pathway can be exploited to introduce specific changes at a site of choice in the genome. However, frequencies of double-strand break repair by homologous recombination are very low. In this study, we compared two methods that have been reported to enhance frequencies of homologous recombination in plants. The first method boosts the repair template availability through the formation of viral replicons, the second method makes use of an *in planta* gene targeting approach. Additionally, we comparatively applied a nickase instead of a nuclease for target strand priming. To allow easy, visual detection of homologous recombination events, we aimed at restoring trichome formation in a glabrous *Arabidopsis* mutant by repairing a defective *glabrous1* gene. Using this efficient visual marker, we were able to regenerate plants repaired by homologous recombination at frequencies of 0.12% using the *in planta* gene targeting approach, while both approaches using viral replicons did not yield any trichome-bearing plants.

INTRODUCTION

Gene targeting (GT) means integration of foreign DNA into a cell's genome by homologous recombination (HR; Paszkowski et al., 1988). GT has been a long-term goal for plant scientists since it allows modifying an endogenous gene *in planta* or integrating a transgene at a specific position in the genome (Puchta, 2002). GT is easily achieved in lower eukaryotes, such as yeast, or the moss *Physcomitrella patens,* because they include foreign DNA predominantly through a highly efficient HR pathways. However, GT is not easily achieved in higher plants, since they integrate foreign DNA via illegitimate recombination through the non-homologous end joining (NHEJ) pathway (Puchta, 2005). A simple introduction of a DNA donor molecule with homology arms (HA) only led to frequencies of 10^{-4} to 10^{-6} GT events per transformation event in higher plants, such as tobacco and *Arabidopsis* (Lee et al., 1990; Offringa et al., 1990; Miao and Lam, 1995; Risseeuw et al., 1995), making GT impractical for most applications. Nevertheless, GT frequencies in higher plants can be increased by two orders of magnitude by inducing double-strand breaks (DSB) in the target locus with sequence specific nucleases (SSN, Puchta et al., 1996). SSN are proteins that can cleave DNA at predefined sequence motives. The first generation of SSN, including meganucleases, Zinc-finger nucleases (ZFN), and Transcription Activator-like Effector Nucleases (TALENs) recognized DNA by protein-DNA interactions, which required rational protein design for every target sequence (Voytas, 2013). The most recent addition to the SSN has been the CRISPR/Cas9 system (Cas9 system).

Cas9 is an endonuclease of the bacterium *Streptococcus pyogenes* that generates DSB in DNA (Doudna and Charpentier, 2014) and that recently has become the SSN of choice for genome editing and GT. Cas9 became so popular because, in contrast to other SSNs that recognize DNA sequences through protein-DNA interactions, Cas9 is guided to its target site on DNA by complementarity matching of a single guide RNA (sgRNA, Jinek et al., 2012). The RNA-based targeting system makes Cas9 easily programmable towards a selected target. Therefore, the Cas9 system has been used extensively by numerous laboratories to edit plant genomes (Zhang et al., 2017). Additionally, mutation of one of the two endonuclease sites of Cas9 resulted into a Cas9 nickase, which introduces only single-strand nicks in the plant genome (Fauser et al., 2014).

SSN-generated DSB are lethal for the cell and must be repaired by dedicated protein machineries. Two major repair pathways can be distinguished: NHEJ and HR (Puchta, 2005). Most DSB in higher eukaryotes are repaired by NHEJ (Sargent et al., 1997), a pathway that

involves direct re-ligation of the DSB ends (Gorbunova and Levy, 1997; Dueva and Iliakis, 2013), making it an error-prone process. This process can be exploited to introduce random small indel mutations at a DSB site and has been used extensively to knock-out gene functions with SSN (Bortesi and Fischer, 2015). In contrast, HR repair pathways use a homology template (HT) to repair a DSB and are therefore of high potential for GT. In the conservative synthesis-dependent strand annealing pathway, which is the most prominent HR pathway in plant somatic tissues (Puchta, 1998), DNA ends are resected to create 3' overhangs. One 3'-end can invade into a homologous sequence forming a loop-structure and copy the sequence from the HT DNA. After release from the loop structure, the single strand can reanneal with homologous sequences on the opposite side of the DSB, leading to conservative repair without sequence information loss (Steinert et al., 2016). Artificial HT with HA have already been used for GT experiments to integrate foreign DNA sequences into a DSB generated by SSN (Puchta et al., 1996; Baltes et al., 2014; Schiml et al., 2014; Li et al., 2015).

While GT frequencies were successfully enhanced by cutting the target site with SSN, this is not sufficient for streamlined application of GT in plants. GT frequencies can be further enhanced of approximately two orders of magnitudes (compared to conventional T-DNA delivery) by mobilizing the HT after T-DNA transformation. This can be achieved with an *in planta* gene targeting (IPGT) approach and with a viral replicon approach.

The IPGT approach (Fauser et al., 2012) employs SSN not only to cut the target locus but also to excise the HT from the chromosomal T-DNA backbone. The liberation from the genomic context of the T-DNA allows the activated HT to be available at the site of the DSB where it can function as repair template. This method has been used successfully using various SSN in *Arabidopsis* (Fauser et al., 2012; Schiml et al., 2014; Zhao et al., 2016), maize (Ayar et al., 2013; Kumar et al., 2016), and rice (Sun et al., 2016).

An alternative approach has been developed by Baltes et al. in tobacco. It activates the HT by including it in a *Geminivirus* based replicon (Baltes et al., 2014). Therefore, the HT is flanked with three viral elements from the *Bean yellow dwarf virus* (BeYDV) that are required to establish the replicon. These regions are known as the *cis*-acting long intergenic region (LIR), the short interacting region (SIR) and the gene for the *trans*-acting viral replicase *Rep/RepA*. The Rep protein initiates a rolling circle endoreplication at the LIR site, leading to a replicational release of the HT in form of up to several thousand circular replicons (Chen et al., 2011). This method has been successfully applied using components from various virus strains and with different SSN for GT in wheat protoplast (Gil-Humanes et al., 2017), tomato

(Čermák et al., 2015), potato (Butler et al., 2016), and rice (Wang et al., 2017). While IPGT has already been shown to induce stable GT events in *Arabidopsis* (Fauser et al., 2012; Schiml et al., 2014), reports about successful applications of viral replicons based GT approaches in *Arabidopsis* have not been published so far.

In this report, we aimed to evaluate and compare the efficiencies of the viral replicon system and the IPGT system to enhance GT frequencies in the model plant *Arabidopsis thaliana*. Therefore, we made use of the R2R3-MYB transcriptional master regulator of trichome formation GLABROUS1 (GL1), which we established previously as non-invasive, endogenous visual marker for Cas9 activity in *Arabidopsis* (Hahn et al., 2017). Null-mutations of the *GL1* gene lead to loss of trichome formation (Koornneef et al., 1982) on stems and leaves as GL1 is part of the signaling cascade which allows epidermal cells to differentiate into trichomes (Pattanaik et al., 2014). Using the IPGT approach and the viral replicon-based approach, we aimed at repairing a non-functional *gl1* gene with a 10 bp deletion in the coding region (Hahn et al., 2017) by HR and by this, restoring trichome formation. Using IPGT, we achieved successful GT and regenerated three stable non-glabrous plants out of ~2,500 plants screened over three generations. However, we were not able to regenerate non-glabrous plants using viral replicons. The *gl1* phenotype provided a simple and efficient visual marker for tracking mutagenic activity *in planta*.

METHODS

Plant growth conditions

Surface-sterilized seeds of an *Arabidopsis gl1* mutant with a 10 bp deletion in the *GL1* gene (Hahn et al., 2017) were stratified at 4 °C for three days. Seeds were germinated on 0.8% (w/v) agar-solidified half-strength Murashige and Skoog medium (Duchefa) in growth chambers with a light intensity of 100 µmol photons m^{-2} s^{-1} under long day conditions (16-h-light/8-h-dark cycle, 22/18 °C). After 14 days, seedlings were transferred to soil and further grown under the same conditions. For large-scale screening of plants in the T2 generation, seeds were directly sowed on soil and the plants were grown in the green house after three days of stratification at 4 °C.

Generation of a homology repair template for reparation of the ten nucleotides deletion in the *gl1* gene

We designed a homology repair template (HT) containing a codon-modified version of the missing 10 bp. The 10 bp (CTGCCGTTTA) were flanked by homology arms (HA) complementary to the genomic regions adjacent to the cutting site of Cas9 in the *gl1* gene. The HA consisted of 813 bp and 830 bp, respectively. Both homology arms contained a single nucleotide exchange within the homology region to distinguish two-sided true GT events from one-sided GT events combined with NHEJ mediated repair pathways (Schiml et al., 2014). After a short linker sequence, the sequence of the sgRNA recognition site including the protospacer adjacent motif (PAM) sequence followed. The HT was synthesized at GeneART (ThermoFisher Scientific) and integrated in the vector backbone pMA-RQ (GeneART). The resulting vector was named pFH95.

Cloning of a binary T-DNA repair vector for *in planta* gene targeting repair approach

We made use of the binary T-DNA vector pDE-Cas9, containing an *Arabidopsis*-codon-optimized Cas9 controlled by the constitutive *Ubiquitin* promoter (PcUBp) from *Petroselinum crispum* (Fauser et al., 2014) and the sgRNA subcloning vector pEN-C.1.1 containing the *Arabidopsis* U6-26 promoter for sgRNA expression, the sgRNA scaffold and BbsI-sites for subcloning of the protospacer sequence (Schiml et al., 2014). After digestion of pEN-C1.1 with BbsI-HF (NewEngland Biolabs), we inserted the target sequence (AAGAGATGTGGGAAAAGAGA plus an additional guanine as first bp for transcription by the U6-26 promoter) using two annealed primers FH210/FH211 (Table S1) resulting in vector pFH89. We then digested pFH89 with MluI-HF (NEB) to extract the sgRNA cassette and ligated it into MluI-HF-digested pDE-Cas9, thereby creating the vector pFH99. We amplified the HT from pFH95 using primers FH278/FH279 (Table S1) and performed Gateway® BP-reaction with pDONR 207 (Invitrogen) resulting in vector pFH122. LR reaction (Invitrogen) of pFH99 with pFH122 led to the assembly of the IPGT repair vector pIPGT-Nuc.

Generation of binary T-DNA repair vectors for viral replicon repair approach

For the viral replicon repair approach, we amplified the viral components from BeYDV necessary for replicon formation from the vector pZLSLD.R (Baltes et al., 2014) using primers FH217/FH218 (Table S1) and cloned the PCR product into pDONR 207 (Invitrogen) via BP reaction. This vector (pFH83) was then digested with PmII (NewEngland Biolabs) and BsaXI (NewEngland Biolabs) to remove the original HT from Baltes et al. (2014). We PCR-

amplified the *GL1* HT from pFH95 using primers FH240/FH241 (Table S1) and inserted it into the digested pFH83 vector by Gibson® cloning (NewEngland Biolabs), resulting in the vector pFH84. The final viral replicon nuclease repair construct pVIR-Nuc containing Cas9 nuclease, the sgRNA cassette directed against the *gl1* gene and the HT was then cloned by Gateway® LR reaction (Invitrogen) of pFH99 with pFH84.

For the viral replicon repair approach using the Cas9 nickase, we used the MluI-HF-extracted sgRNA cassette from pFH89 and ligated it into MluI-HF-digested pDE-Cas9-D10A (Fauser et al., 2014) resulting in the vector pFH103. LR reaction of pFH103 with pFH84 led to the final viral replicon nickase repair construct pVIR-Nick.

The sequences of the vectors pIPGT-Nuc (NCBI accession number: MG760355), pVIR-Nuc (MG760356) and pVIR-Nick (MG760357) are accessible on GenBank.

Plant transformation

Agrobacterium tumefaciens, strain GV3101::pMP90 was transformed with the vectors pIPGT-Nuc, pVIR-Nuc, and pVIR-Nick. Floral dipping (Clough and Bent, 1998) was used to transform *Arabidopsis gl1* plants with the three constructs. Kanamycin (pFH136) and glufosinate (pFH114, pFH135) were used for selection of primary transformants. Furthermore, leaf genomic DNA was isolated from surviving plants via isopropanol precipitation (Weigel and Glazebrook, 2009). The presence of T-DNA insertion was verified by PCR using primer pairs FH190/FH223 (pIPGT-Nuc) or FH223/FH265 (pVIR-Nuc, pVIR-Nick, Table S1).

Detection of the *Cas9* gene in plants

Presence of the *Cas9* gene in plant genomes was verified by PCR on leaf genomic DNA using primer pairs amplifying specifically the *Cas9* gene (FH313/FH314, Table S1).

Verification of viral replicon formation

Formation of circular viral replicons was verified in T1 generation plants by PCR using leaf genomic DNA and the primers FH266/FH271 (Table S1).

Detection of Cas9-mediated gene repair

For verification of HDR events on genomic level, leaf genomic DNA was isolated via isopropanol precipitation (Weigel and Glazebrook, 2009) and the region of interest in the *gl1* gene was PCR-amplified with proofreading Q5© polymerase (NewEngland Biolabs) using

primers FH214/FH215 (Table S1). The obtained PCR amplicons were analyzed directly by Sanger sequencing (Macrogen) or subcloned into pJET1.2 (ThermoFisher Scientific) and then sequenced. 4Peaks software (Nucleobytes) and SnapGene Viewer (GSL Biotech) were used to analyze sequencing histograms.

RESULTS

Selection of a visual marker for testing gene targeting events

In a previous study, we established trichome formation as efficient visual marker to monitor Cas9 mutagenic activity in *Arabidopsis* (Hahn et al., 2017). We used a CRISPR/Cas9 approach to generate a T-DNA free glabrous *Arabidopsis* mutant that contains a deletion of 10 bp in the master regulator gene for trichome formation, *GL1*. This 10 bp deletion disrupts the reading frame of *GL1* and leads to a premature stop codon. Homozygous mutant plants show almost no leaf and stem trichomes (Fig. S1). We decided to use this mutant line as a model system to test Cas9-based GT in *Arabidopsis*. Cas9-induced cleavage in combination with homology directed repair can be used to integrate sequences of choice into a genomic region. If we successfully reintegrated the missing 10 bp into this mutant line´s genome, we would expect restoration of leaf trichome formation. Since the mutant line was created by CRISPR/Cas9, a PAM sequence was already present at the site of mutation and the predicted Cas9 cleavage site (3 bp upstream of the PAM sequence) matched the desired integration site for the missing 10 bp (Fig. S1).

Selection of a gene targeting system and design of related constructs

HR is more efficient in plants if the HT is available at the site of the DSB (Baltes et al., 2014). If the HT is delivered as T-DNA, which integrates randomly into the genome, it might not be available at the DSB site for the DNA repair machinery. Thus, we comparatively analyzed two methods to enhance the availability of the HT at the DSB site.

The first method was developed by Baltes et al (2014) in tobacco calli. This method increases the availability of the HT by including it in small circular geminiviral replicons.

For the repair of the *gl1* gene, we accordingly created a single T-DNA vector, (Fig. 1A), that contained an *Arabidopsis* codon-optimized *Cas9* gene under the control of the constitutive PcUB promoter (Fauser et al., 2014), a sgRNA cassette targeting the *gl1* gene, and the HT with ~ 800 bp of HA on both sides, flanked by LIRs, SIR, and a *Replicase* (*Rep*) gene. For an easy nomenclature of the plasmids, we decided to include the name of the HT activation

method (viral replicons = vir; *in planta* gene targeting = IPGT) and the used Cas9 variant (nuclease = Nuc; nickase = Nick) in the plasmid name and therefore named this construct pVIR-Nuc.

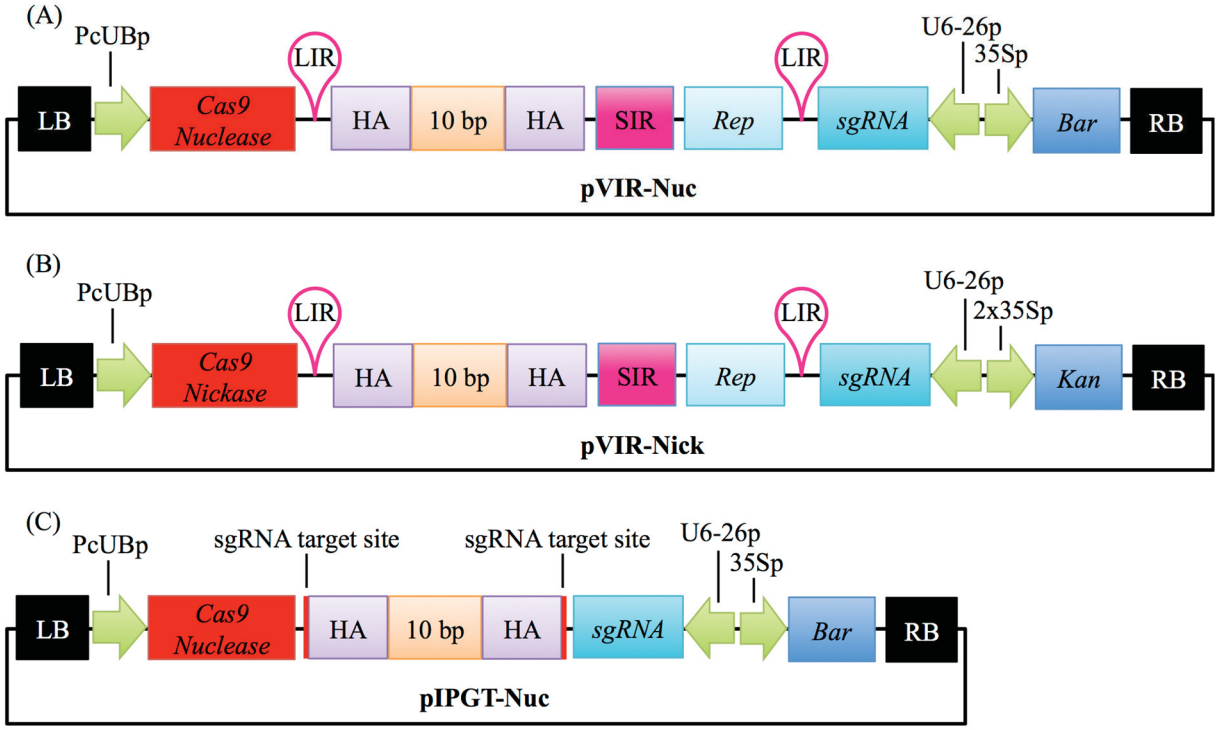


Figure 1: Vector design for repair of the *gl1* gene. Single T-DNA vectors were designed to repair the dysfunctional *gl1* gene. (A) pVIR-Nuc contains the *Cas9* Nuclease gene (red box) under a constitutive Ubiquitin6 promoter from parsley (PcUBp) and the sgRNA cassette (petrol box) controlled by the *U6-26* promoter (U6-26p). A Basta resistance cassette (*Bar*, dark blue box) is included as selection marker. The repair template itself consists of 10 nucleotides (10 bp, CTGCCGTTTA, orange box), which should restore the reading frame of the *gl1* gene flanked by homology arms (HA, purple box). Rolling circle replication of the HT is ensured by the flanking long intergenic region (LIR, pink hairpin), short intergenic region (pink box, SIR), and the gene encoding the replicase Rep (light blue box). (B) pVIR-Nick contains similar features but harbors the gene for the *Cas9* nickase instead of the nuclease and a kanamycin resistance cassette (*Kan*, dark blue box) as selection marker. (C) pIPGT-Nuc contains the gene for the *Cas9* nuclease, the sgRNA cassette, the Basta resistance cassette and the repair template with HA flanked by target sites for the Cas9 nuclease (red lines). RB = right T-DNA border, 35Sp = Cauliflower Mosaic Virus *35S* promoter, LB = left T-DNA border. Size not to scale.

After expression, Rep binds to the LIR sites and initiates rolling circle replication, leading to the formation of up to several thousand copies of the HT in the nucleus (Timmermans et al., 1992; Baltes et al., 2014), which can then be used as repair template for the DSB created by Cas9 nuclease (Fig. 2A).

To further explore the potential of this method, we additionally generated and tested a complementary vector pVIR-Nick in which the Cas9 nuclease gene was replaced by a nickase gene (Fig. 1B). DNA nicks trigger HR without triggering the competing NHEJ repair (Fauser

et al., 2014). NHEJ repair interferes with HR because it induces random indel mutations that destroy the sgRNA target site and make it unsuitable for homology directed repair. By using the Cas9 nickase, we expected to increase HR efficiency.

The second method, developed by Fauser et al. (2012), is called IPGT. This method increases the availability of the HT by releasing it from the T-DNA backbone using the same SSN that creates the DSB in the target region. Therefore, we designed the vector pIPGT-Nuc (Fig. 1C), in which the HT is flanked by the same sgRNA target site that is used to create the DSB in the *gl1* gene (Fig. 2B).

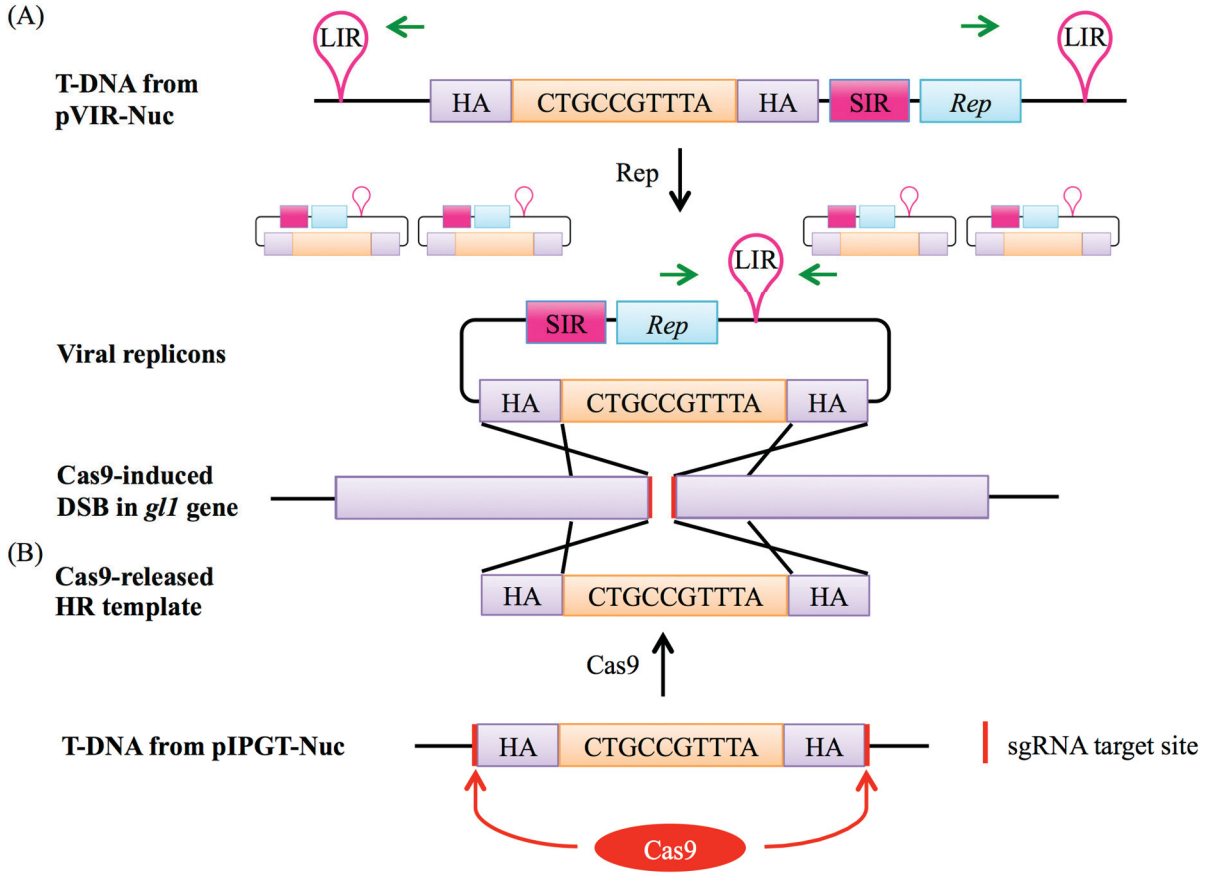


Figure 2: Two different mechanisms enhance the availability of the HT in the plant cell nucleus. (A) If the T-DNA from pVIR-Nuc is integrated into the plant genome, Rep is expressed and initiates rolling circle endoreplication of the HT leading to thousands of viral replicons. The presence of viral replicons can be detected by PCR using primers that only generate an amplicon on the circular replicons but not on the T-DNA (green arrows). The HT can then attach to the DSB site in the *gl1* gene induced by Cas9 due to homology between the HA and the genomic regions next to the DSB (purple). Repair of the DSB by HR will then lead to integration of the 10 bp, which restores the ORF. pVIR-Nick functions accordingly, except that Cas9 nickase only induces a DNA nick in the *gl1* gene. (B) The HT in the T-DNA of pIPGT-Nuc is flanked by the same sgRNA target site that is present in the *gl1* gene. Expression of Cas9 will therefore simultaneously target the *gl1* gene and release the HT, which can then attach to the DSB and function as repair template.

Plants transformed with pVIR-constructs form circular viral replicons

We transformed Arabidopsis *gl1* plants with either pVIR-Nuc, pVIR-Nick, or pIPGT-Nuc and selected three independent T1 lines per construct. Though we did not detect trichomes on the leaf surface of any T1 plant, we wanted to verify that the viral components were indeed producing circular HT in the T1 plants. Thus, we performed a PCR using primers that only yield a product in the circularized version of the HT but not in the linear T-DNA (Fig. 2A). We detected PCR products in the plants transformed with pVIR-Nuc and pVIR-Nick. PCR fragments for plants transformed with pIPGT-Nuc, which do not generate replicons (Fig. S2), were not obtained.

T2 generation plants show spots of trichomes on leaves

We analyzed several hundred plants from the progeny of each T1 plant for trichome reappearance. While we could not detect plants that fully recovered trichome production on leaves, we found several plants from the IPGT approach that harbored spots of trichomes (Tab. 1).

Table 1: Trichome appearance in T2 generation plants.

Repair construct	Transformant line	Number of analyzed T2 plants	Number of plants with trichome spots (% of T2 plants)
pIPGT-Nuc	#1	378	4 (1.1)
	#2	368	2 (0.5)
	#3	683	0 (0)
pVIR-Nuc	#1	728	0 (0)
	#2	884	1 (0.1)
	#3	514	0 (0)
pVIR-Nick	#1	925	0 (0)
	#2	528	0 (0)
	#3	826	0 (0)

These spots ranged from two trichomes on a single leaf of the plant to large spots on several leaves (Fig. 3A). In contrast, using the viral replicon method, we only detected one plant with a single trichome in plants transformed with pVIR-Nuc, which implements a full Cas9 nuclease to trigger HR. Plants transformed with pVIR-Nick (Fig. 3A), which implements the Cas9 nickase, did not form any trichomes on their leaves.

(A)

(B)

Cas9

gl1	... CTA[N_{53}]GGCTAAAGAGATGTGGGAAAAG– – – – – – – – – AGATGGATGAA...	19/42 clones
+G	... CTA[N_{53}]GGCTAAAGAGATGTGGGAAAAG– – – – – – – – GAGATGGATGAA...	2/42 clones
+T	... CTA[N_{53}]GGCTAAAGAGATGTGGGAAAAG– – – – – – – – TAGATGGATGAA...	6/42 clones
+A	... CTA[N_{53}]GGCTAAAGAGATGTGGGAAAAG– – – – – – – – AAGATGGATGAA...	4/42 clones
-AG	... CTA[N_{53}]GGCTAAAGAGATGTGGGAAAAG– – – – – – – – – – – ATGGATGAA...	4/42 clones
-14 bp	... CTA[N_{53}]GGCTAAAGAGATG – GATGAA...	1/42 clones
Indel	... CTA – – – – – – – AA – – – – – T – – – – – – AAAG– – – – – – – – – – AGATGGATGAA...	1/42 clones
+10 bp	... CTA[N_{53}]GGCTAAAGAGATGTGGGAAAAGCTGCCGTTTA AGATGGATGAA...	4/42 clones

Figure 3: T2 generation plants show spots of trichomes. (A) Representative images of T2 plants transformed with pIPGT-Nuc (red frame) or pVIR-Nuc (blue frame) showing spots of trichomes (yellow circles) ranging from one single trichome to fully covered leaves. (B) The *gl1* gene was amplified from leaves with trichomes, subcloned and sequenced. Sequence analysis revealed clones with various indel mutations next to clones containing the desired 10 bp insertion (red letters). PAM sequence is underlined, STOP codon in *gl1* line is marked in green letters, Cas9 target site is marked in blue letters.

We isolated gDNA from leaf trichome spots of three IPGT plants (the single trichome on the pVIR-Nuc plant provided not enough plant material for gDNA isolation), amplified the *gl1* gene, subcloned the PCR product and sequenced 42 clones. While most clones still contained the *gl1* sequence with the 10 bp deletion, we detected various indel mutations but also the integration of the codon-modified 10 bp HT, which restored the wildtype amino acid sequence of the *GL1* gene (Fig. 3B). This chimerism on genotype level was in line with the highly chimeric phenotype of the leaves.

Fully regenerated trichomes were detected in plants of the T3 generation

We harvested seeds from the chimeric T2 plants and analyzed ~ 200 progeny T3 plants of each T2 plant. In most cases, the progeny plants were glabrous or just contained single trichome spots. However, we detected three plants (frequency: 1.5 %) in the progeny of the

highly chimeric IPGT plant (Fig. 3A, left) with fully regenerated trichomes (Fig. 4A). We amplified the *gl1* gene from those plants and sequenced the PCR product. Two of the three plants showed clear peaks corresponding to the 10 bp insertion but also peaks corresponding to small indel mutations (Fig. 4B, Fig. S3), indicating that only one of the two alleles was repaired by HR. In the third plant, we detected the expected 10 bp insertion not in the sequencing histogram but in 5 out of 15 sequenced clones of subcloned PCR product (Fig. S3D).

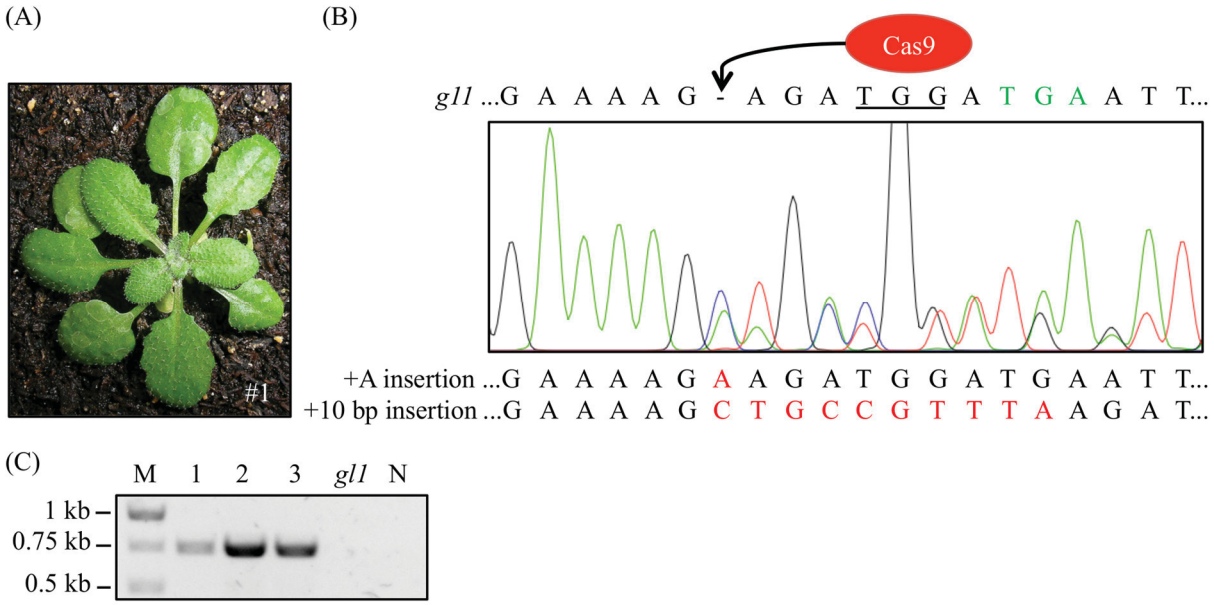


Figure 4: Wildtype-like trichome patterning was detected in three T3 generation plants. (A) Exemplary picture of one of the three T3 plants (#1) with full trichome covering. (B) We verified the repair of the *gl1* gene by amplifying the *gl1* gene and sequencing the PCR amplicon. Double peaks appeared at the site of Cas9 cleavage hinting at a chimeric or biallelic mutation, the peaks corresponded to an insertion of the 10 bp (red) from the HT and to an adenine insertion. Sequence of the dysfunctional *gl1* gene is given on top of the sequencing histogram for comparison. Peak colours: Adenine = green, cytosine = blue, guanine = black, thymine = red; PAM sequence underlined, premature STOP codon marked in green letters. (C) Presence of the *Cas9* gene in the three non-glabrous plants was detected by PCR using *Cas9*-specific primers. All three plants revealed PCR amplicons at the expected size of 749 bp. In contrast, DNA from the *gl1* background line did not yield a PCR *Cas9* signal. Image of agarose gel was colour inverted for better visibility. M = Marker, N = water control.

The repaired *GL1* allele segregates in a mendelian way

In genome edited plants, while the editing nuclease (Cas9) is active, it is hard to distinguish between somatic and heritable mutations. The three plants that fully recovered trichome production were still carrying the *Cas9* gene (Fig. 4C), which was likely still active. Thus, we confirmed in those plants that the 10 bp insertion is indeed heritable and not somatic by Mendelian segregation analysis. We analyzed the offspring of the three T3 plants displaying

trichomes and counted the amounts of trichome-bearing and glabrous plants in the T4 generation. If one allele was indeed repaired by HR in the whole plant and the other one not, we would expect a Mendelian distribution of the trichome phenotype with a ratio of non-glabrous:glabrous plants of 3:1. Indeed, the progeny of all three plants did not deviate significantly from the expected 3:1 ratio as calculated by chi-squared test (χ^2; P > 0.5, Tab. 2).

Table 2: Segregation pattern of trichome phenotype in T4 generation plants transformed with pIPGT-Nuc. Progenitor T3 plant number according to Fig. 4 and Fig. S3.

Progenitor T3 plant	Number of plants with trichomes	Number of plants without trichomes	Ratio of plants with/without trichomes	χ^2 value	P value
#1	164	55	2.98/1	0.0015	> 0.9
#2	150	52	2.88/1	0.0594	> 0.5
#3	210	75	2.80/1	0.2632	> 0.5

Sequencing of 125 progeny plants from all three T3 plants with or without trichomes revealed that the trichome phenotype was always associated with at least one repaired *GL1* allele, while the glabrous plants mostly contained various small indel mutations (Fig. S3). Among the 125 sequenced plants, we detected 30 plants that showed a homozygous insertion of the 10 nucleotides according to the sequencing histogram (Fig. 5). The small indel mutations found in glabrous plants or in heterozygous non-glabrous plants were in many cases not consistent with the indel alleles found in the progenitor plants, indicating that only the repaired *GL1* allele was already fixed in the T3 generation and that the other sequenced allele in the T3 plants was caused by somatic mutations in the leaf used for gDNA extraction. This putative chimerism in the T3 generation plants corresponds to the presence of more than two sequence peaks in some of the sequencing histograms of the T3 parental plants (Fig. S3).

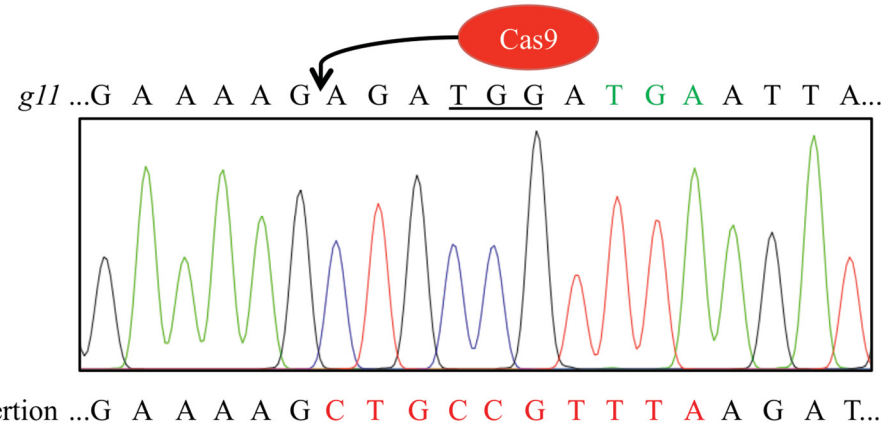


Figure 5: Sequencing histogram of T4 generation plant reveals homozygous integration of the 10 bp thereby restoring the *GL1* reading frame. Sequence of the dysfunctional *gl1* gene is given on top of the sequencing histogram for comparison. Peak colours: Adenine = green, cytosine = blue, guanine = black, thymine = red; PAM sequence underlined, premature STOP codon marked in green letters.

In summary, we screened between 2,200 to 2,500 plants containing the *gl1* repair constructs pIPGT-Nuc, pVIR-Nuc, and pVIR-Nick, respectively, over three generations. Using the viral replicon approach, we could not regenerate non-glabrous plants. Using the IPGT approach, we were able to detect three stable GT events in the T3 generation with one functional *GL1* allele.

DISCUSSION

Trichome formation is a marker for mutagenic efficiency

The Cas9 system has become a valuable and efficient tool for genome modifications in plants (Arora and Narula, 2017). However, while random mutagenesis in plants by error-prone DSB repair through the NHEJ-pathway has been successfully applied in many studies (Liu et al., 2017), introduction of targeted mutations and gene knock-ins by GT remains challenging.

To explore the efficiency of two systems that have been described to induce HR in plants, we used a Cas9-generated trichomeless *Arabidopsis* mutant (Hahn et al., 2017). This mutant contains a 10 bp deletion (Fig S1) in the *GL1* gene, which leads to a premature STOP codon. The trichome phenotype (Fig. 3) allowed easy selection of transformed plants, which were likely to produce completely repaired offspring (Fig. 4).

An important aspect of our marker system is that DSB repair has to occur via HR and not via NHEJ to induce trichome formation. Small indels of 1-2 nucleotides generated by error-prone NHEJ might restore the open reading frame (ORF) of the gene, but at least three amino acid residues are still missing in the MYB-binding domain of the GL1 protein, which are critical

for protein function (Hauser et al., 2001). As NHEJ mediated repair of DSB rarely results in insertions of more than one nucleotide (Fauser et al., 2014), it is highly unlikely that NHEJ mediated repair can restore trichome formation. Consequently, in a large-scale screening of glabrous and non-glabrous T4 plants (Fig. S4), we detected a clear correlation of trichome formation and repair via HR, while NHEJ mutations did not result in non-glabrous plants even though the ORF of the gene was partly restored in some cases. Previously used markers, such as ß-glucuronidase (GUS) or fluorescent proteins need first to be introduced into the plant and GT events can only be detected by staining or fluorescence microscopy. In contrast, our endogenous marker gene has two advantages: First, it allows to follow the mutagenesis progress by eye and second, trichome absence does not have pleiotropic effects on plant development under greenhouse conditions (Oppenheimer et al., 1991).

Activation of the homology template by viral replicons did not result in gene targeting events

Baltes et al. (2014) developed a GT approach in tobacco that employs geminivirus-based replicons to active the HT (Fig. 2). In a first approach in *Arabidopsis*, they aimed at inserting a short DNA sequence into an endogenous gene by GT using a HT, which was activated by replication in recombinant bipartite cabbage leaf curl virus genomes. While GT events could be detected on somatic level by PCR, the authors did not report inheritance of the GT events to progeny plants (Baltes et al., 2014). In further experiments, several groups used deconstructed virus replicons of BeYDV (or other viruses) depleted of the coat and movement proteins. These can be stably transformed into plants and have an increased vector cargo capacity, allowing the use of larger HT. This approach allowed efficient GT in various crop plants, including tobacco, tomato, potato, rice and wheat (Baltes et al., 2014; Butler et al., 2016; Čermák et al., 2015; Gil-Humanes et al., 2017; Wang et al., 2017).

While the initial experiments with the full virus demonstrated that GT using viral replicons can be achieved in *Arabidopsis*, no further studies reported the use of deconstructed replicons in *Arabidopsis*. We were also unable to detect GT events in form of trichome regeneration in 2,334 analyzed plants from three independent lines transformed with pVIR-Nuc except a single trichome (Tab. 1, Fig. 3). One possible explanation could be that the sgRNA that we used did not allow efficient DSB induction in the target locus. Since we generated successful GT events with the IPGT approach, using the same sgRNA, which requires efficient Cas9 cutting for HT activation, we do not consider this as likely. As we detected high frequencies of NHEJ mutations next to GT events while sequencing T4 generation plants (Fig. S4), we

conclude that our sgRNA is highly active. It could be possible that replicon formation is not efficient in *Arabidopsis* in contrast to other plant species. However, BeYDV infects *Arabidopsis* (Liu et al., 1997). Thus, replicon formation should be possible. We furthermore detected efficient replicon formation on DNA level in T1 plants (Fig. S2). Another option could be an *Arabidopsis*-specific discrimination of the repair machinery for circular HT as these do not occur naturally in the plant nucleus.

Differences between plant species with handling of foreign DNA molecules have already been observed in *Arabidopsis* and tobacco (Orel and Puchta, 2003). It has been suggested that the stimulatory effect of viral replicons on GT is not only caused by the amount of HT but also by the impact of RepA (which is derived from an alternative splice variant of *Rep*) on the plant cell cycle. RepA stimulates entry of cells into the S-phase by interaction with plant proteins that control cell cycle so that the virus can use the necessary factors for replicon replication that the cell produces for DNA synthesis (Baltes et al., 2014). Cells in S-Phase show an increased HR efficiency in eukaryotes (Saleh-Gohari and Helleday, 2004; Mathiasen and Lisby, 2014). Possibly, this effect is less pronounced in *Arabidopsis* than in other plants.

In contrast to most of the studies using viral replicons, we did not include the SSN in the replicating unit. Baltes and colleagues (2014) reported comparable GT efficiencies if the ZFN was part of the replicating unit or not, and similar results were found in a Cas9-based approach in tomato (Čermák et al., 2015). However, it might be possible that a high amount of SSN is more critical in *Arabidopsis*.

Donor architecture might also account for variable efficiencies in GT. We used HA with a length of app. 800 bp, which is in the range of the most successful GT studies so far (Fauser et al., 2012; Baltes et al., 2014; Schiml et al., 2014; Čermák et al., 2015; Butler et al., 2016; Gil-Humanes et al., 2017; Wang et al., 2017). Additionally, we used the viral components described in the original report of Baltes et al. (2014) and just integrated our HT. Even though we used a different vector backbone (Fauser et al., 2014), the replicons should therefore be comparable to the ones described in successful GT studies by Baltes et al. (2014) and Čermák et al. (2015).

While we could not detect considerable trichome formation in our experiments, it has to be taken into account that GT events might occur within cells, which are not determined to develop into trichomes. That is, using trichome formation as readout for mutagenic efficiencies, we actually underestimate the amount of successful GT events on the somatic cell level. PCR-based methods, as used for GT detection in *Arabidopsis* using cabbage leaf curl virus genomes (Baltes et al., 2014) will therefore allow detection of more GT events. In

other experiments of our group with a different target gene (unpublished data), we actually could detect GT events using PCR-based approaches on somatic level at low frequencies but we were also unable to regenerate inheritable events Taken together, these data indicate that viral replicons allow GT only at low frequencies in *Arabidopsis*, which contrasts with the considerable success in many crop species. Further experiments are needed to unravel the reasons for these differences so that this promising technique might also be applicable in this important model organism.

Cas9 nickase is not improving HR events in combination with viral replicons

We tested whether Cas9 nickase can induce GT at higher efficiencies than Cas9 nuclease as proposed by Fauser et al. (2014). We only applied the Cas9 nickase in combination with the viral replicon approach, since IPGT needs a nuclease to release the HR template. We did not detect GT events in form of trichome formation on any plant (Tab. 1). However, when we combined the viral replicon approach with a Cas9 nuclease, we also detected only a single putative GT event in ~2,300 plants screened (Tab. 1, Fig. 3). Therefore, it is likely that the viral replicon approach does not trigger GT in *Arabidopsis*, independent from its Cas9 variant.

However, besides the initial report of Fauser et al. (2014), reports on GT using nickases in plants are scarce. For their GT experiments, Fauser and colleagues used a reporter construct (Orel et al., 2003) containing a disrupted *GUS* ORF as well as a homologous sequence, which can function as HT for repair of the *GUS* ORF upon induction of a DSB by Cas9. Up to 4 times more HR events were detected compared to experiments with Cas9 nuclease. Importantly, only somatic mutations were quantified and information on the regeneration of stably mutated plants was not provided. It has been observed earlier, that GT efficiencies increase if target locus and donor construct are located on the same chromosome (Fauser et al., 2012). In the GUS experiments described above, target locus and HT are directly adjacent, which possibly explains the successful HR events. In our case, the HT integrated randomly into the genome and even after replicon formation, the replicons still had to “find” the nick site to be available for HR. Supporting this hypothesis, other studies detected lower GT efficiency when using nickases compared to nucleases (Ran et al., 2013; Čermák et al., 2017).

In a recent study, Cas9 nickase was used in combination with viral replicons to repair a dysfunctional *GUS* gene in tobacco. Using two sgRNAs to create adjacent nicks, a 2-fold increase in GT was observed (Čermák et al., 2017). However, paired nicks on different DNA strands in close proximity are considered to induce a DSB with single strand overhangs

(Schiml et al., 2014). Thus, it is questionable if nick HR repair pathways were indeed responsible for the successful GT in that study. In human cells, DNA nick repair via HR occurs 8 times more frequently on the transcribed (non-coding) DNA strand, probably due to stimulation by the transcription machinery (Davis and Maizels, 2014). The Cas9 nickase used in our study only contains a functional HNH-cleavage domain so that the nick is created on the non-coding strand. However, whether the strand preference in plants is comparable to that in human cells remains to be elucidated and might provide a possibility to enhance HR frequencies.

In general, HR repair mechanisms on DNA nicks in plants are not well studied. In humans, two competing HR mechanisms are proposed. The canonical HR pathway resembles the synthesis-dependent strand annealing repair pathway for DSB, while the alternative one preferentially employs single stranded oligonucleotides or nicked dsDNA HT for repair (Davis and Maizels, 2014). The alternative HR pathway shares features with another HR pathway, the single strand annealing pathway, which is induced by micro-homologies at the break ends. As HR via the single strand annealing pathway seems to be more efficient in plants than the synthesis-dependent strand annealing repair pathway (Orel et al., 2003; Fauser et al., 2014), it might be interesting to include nickase target site in the repair template to induce target locus repair via the alternative HR nick repair pathway. However, the detailed mechanisms underlying DNA nick repair in plants via HR needs to be elucidated to verify if the same mechanisms as in human cells apply. As the Cas9 nickase has already been shown to induce gene knock in in human embryonic stem cells using dsDNA as template (Rong et al., 2014), it remains an interesting option for future HR experiments *in planta*.

The *in planta* gene targeting approach allows regeneration of stable gene targeting events in Arabidopsis

The IPGT approach, which excises the HT from the T-DNA backbone by a SSN (Fig. 2), has been used successfully in *Arabidopsis.* By the application of the I-*Sce*I meganuclease a *GUS* marker gene was restored (Fauser et al., 2012), and by the application of the Cas9 system a resistance cassette was integrated into the *ADH* locus (Schiml et al., 2014) and a *GFP* gene introduced into the *TFL1* locus (Zhao et al., 2016). In contrast to these previous studies, we did not exploit an artificial marker gene but restored the gene function of the endogenous *GL1* gene, using trichome formation as direct read-out for GT events.

We detected spots of trichomes ranging from single trichomes to fully covered leaves in T2 generation plants (Fig. 3, Tab. 1). Other studies reported similar chimeric GT spots using

GUS staining (Fauser et al., 2012). The T2 plant, which showed the highest level of chimerism (Fig. 3) produced offspring plants with WT-like trichome patterning. We therefore conclude, that the trichome phenotype can be used to select for plants, in which GT efficiencies are enhanced and which are candidate plants for production of stably mutated progeny. This enhancement can have several reasons. The location of the T-DNA, which contains HR template and SSN gene, in the plant genome likely has influence on GT and close proximity of repair template and DSB locus seems favorable (Fauser et al., 2012). Similarly, we found differences in the frequency of chimeric T2 plants depending on the parental T1 line (Tab. 1). It is therefore highly recommended to start with as many independent transformation events as possible, in contrast to screening a lot of progeny plants from few primary transformants (Schiml et al., 2017).

T2 plants might be homozygous or heterozygous for the T-DNA insertion. It has been shown that the heterozygosity state of the HR template influences recombination frequencies (Puchta et al., 1995; Orel et al., 2003). Indeed, since all three plants, which completely restored trichome formation in the T3 generation (Fig. 4, Fig. S3), contained the Cas9-T-DNA cassette and also produced only T-DNA containing offspring, it is likely, that the parental T2 plant contained already two copies of the repair template. Crop improvement usually requires the creation of marker-free plants. Thus, more screening effort might be needed to detect T-DNA free GT plants.

It has been reported that one-sided GT events can occur during repair (Schiml et al., 2014), in which one side of the of the DSB is repaired by HR and the other strand is repaired using the stably integrated T-DNA as template. This can lead to integration of large parts of the HT, which finally results in a NHEJ-mediated ligation of the repaired strand and the loose end of the DSB. To detect such events, we had added single bp exchanges in both HA. Our sequencing results of our repaired *GL1* alleles did not contain these bp exchanges, which indicates that we achieved true GT events.

In previous studies the frequencies of GT events using IPGT in *Arabidopsis* ranged from 0.14-0.83 % (Fauser et al., 2012; Schiml et al., 2014; Zhao et al., 2016) which corresponds well to our GT frequency of 0.12 %. The single bp exchanges in the HA, which allowed us to detect true GT events, might explain the slightly lower efficiencies in our experimental approach. Additionally, efficiencies in Cas9 experiments are difficult to compare due to use of different sgRNAs (Peng et al., 2016).

Further options to enhance GT

In this report, we identified three stable GT event out of ~2,500 *Arabidopsis* plants screened across three generations using an IPGT approach. In principle, a GT event at such frequency could be detected exclusively by PCR and without a phenotype based screening. However, we achieved such a high frequency because we were able to select parental lines prone to GT by phenotype. Therefore, it might still be desirable to increase GT efficiencies if no visual marker is used for GT detection. In general, it is advisable to test sgRNA efficiencies in protoplast systems before using them in stable plant transformations as efficient induction of DSB is a prerequisite for successful GT (Puchta et al., 1993).

Higher GT efficiencies of ~5.5 % have been achieved in *Arabidopsis* protoplasts using single stranded oligonucleotides as repair template (Sauer et al., 2016) to exchange specific nucleotides in a fluorescent protein to change its emission spectrum. However, single stranded oligonucleotides are not suitable for knock-in of larger DNA fragments or whole genes due to size limitations and are not transformed via Agrobacteria. For targeted exchange of a single bp, the use of base changer Cas9 provides an attractive alternative to HR-mediated repair processes. In this approach, an inactive Cas9 is fused to cytidine deaminase, which converts cytosines to uracil without DNA cutting (Komor et al., 2016; Zong et al., 2017). Recently, an adenine base editor Cas9 has also been described (Gaudelli et al., 2017). While mutation frequencies of up to 50% can be achieved using base changer Cas9, this approach cannot be applied for gene knock-ins.

In general, protoplast transformation seems to yield higher mutation frequencies compared to Agrobacterium-mediated transformation (Li et al., 2013; Sauer et al., 2016), since more DNA donor molecules are integrated in the plant cells. However, regeneration of *Arabidopsis* plants from protoplasts is challenging and is therefore not an attractive alternative to floral dipping. Focusing research on improved regeneration protocols is therefore desirable. Several groups tried to enhance HR efficiencies by disrupting either the NHEJ pathway (Maruyama et al., 2015; Zhu et al., 2015) or introducing HR stimulating enzymes from other organisms (Reiss et al., 2000; Shaked et al., 2005; Even-Faitelson et al., 2011). These approaches can destabilize the genome as HR frequencies are enhanced between all repetitive sequences in the genome (Puchta and Fauser, 2013).

In the IPGT approach, the HT is excised from the genome. If this happens at an early stage of plant development, the HT might not be passed on during cell division. This could reduce the efficiency of this approach. Using germline-specific promoters (Wang et al., 2015; Mao et al., 2016) for the SSN might increase chances of simultaneous cleavage of target locus and

release of the HT specifically in the cells, where GT needs to occur to generate stable offspring.

A similar problem is target locus inactivation through error-prone NHEJ repair. Besides the before mentioned nickase approach, the Cas9 variant CpfI (CRISPR from *Prevotella* and *Francisella*) might provide an attractive alternative. CpfI generates DSB more distal to the PAM than Cas9 (Zetsche et al., 2015), thus making the target locus accessible to a second round of cleavage even after NHEJ mediated repair. An interesting approach would be to combine the viral replicon approach with the IPGT approach. Therefore, the HT would be flanked by SSN cutting site and the viral components. By this, large amounts of linear dsDNA molecules would be generated that can be used as repair templates. Since more target sites are present in this approach, we suggest including the gene for the SSN and the sgRNA within the replicon unit, so that sufficient amounts of all genome editing components are available.

CONCLUSION

In this report, we applied the Cas9 system to induce GT events in *Arabidopsis*. The HT needed to be activated by an IPGT approach. We were able to stably introduce a short DNA sequence into an endogenous gene and the inserted sequence was transmitted through the germline into the next generation. The high GT efficiency allows to use GT approaches also in marker-free approaches in which GT events need to be detected by PCR. Thus, we conclude that Cas9 mediated GT in combination with IPGT is a powerful tool that has high potential to become routinely applied in plant research in the future. Trichome formation as marker for editing efficiency proved to be a reliable system to compare different GT approaches. We suggest that the *GL1* gene can be used for optimization of Cas9-based mutagenesis approaches in future experiments.

AUTHOR CONTRIBUTIONS

FH, OM, ME, and AW designed the study. FH carried out all laboratory experiments. FH, OM, ME, and AW interpreted the data and wrote the manuscript.

FUNDING

This work was funded by the Cluster of Excellence on Plant Science (CEPLAS, EXC 1028) and the HHU Center for Synthetic Life Sciences (CSL).

ABBREVIATIONS

BeYDV = *bean yellow dwarf virus*; bp = base pair(s); Cas9 system = CRISPR/Cas9 system; DSB = double strand break; GT = gene targeting; GUS = ß-glucuronidase; HA = homology arm(s); HR = homologous recombination; HT = homology template; IPGT = *in planta* gene targeting; LIR = long intergenic region; NHEJ = non-homologous end joining; ORF = open reading frame; PAM = protospacer adjacent motif; sgRNA = single guide RNA; SIR = short intergenic region; SSN = sequence specific nuclease(s); TALEN = Transcription Activator-like Effector Nuclease; ZFN = Zinc finger nuclease(s)

ACKNOWLEDGMENTS

We thank Prof. Dr. Holger Puchta and Felix Wolters for providing us with the vectors pDE-Cas9, pDE-Cas9-D10A and pEN-C1.1, and for valuable feedback. We acknowledge Prof. Dr. Peter Westhoff and Dr. Shizue Matsubara for helpful discussions. We thank Steffen Köhler for photography support. We thank the gardeners for plant care and the student helper team for their support.

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

REFERENCES

Arora, L., and Narula, A. (2017). Gene Editing and Crop Improvement Using CRISPR-Cas9 System. *Front. Plant Sci.* 8, 1932. doi:10.3389/fpls.2017.01932.

Ayar, A., Wehrkamp-Richter, S., Laffaire, J.-B., Le Goff, S., Levy, J., Chaignon, S., et al. (2013). Gene targeting in maize by somatic ectopic recombination. *Plant Biotechnol. J.* 11, 305–314. doi:10.1111/pbi.12014.

Baltes, N. J., Gil-Humanes, J., Cermak, T., Atkins, P. A., and Voytas, D. F. (2014). DNA replicons for plant genome engineering. *Plant Cell* 26, 151–163. doi:10.1105/tpc.113.119792.

Bortesi, L., and Fischer, R. (2015). The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnol. Adv.* 33, 41–52. doi:10.1016/j.biotechadv.2014.12.006.

Butler, N. M., Baltes, N. J., Voytas, D. F., and Douches, D. S. (2016). Geminivirus-Mediated Genome Editing in Potato (Solanum tuberosum L.) Using Sequence-Specific Nucleases. *Front. Plant Sci.* 7, 1045. doi:10.3389/fpls.2016.01045.

Čermák, T., Baltes, N. J., Čegan, R., Zhang, Y., and Voytas, D. F. (2015). High-frequency, precise modification of the tomato genome. *Genome Biol.* 16, 232. doi:10.1186/s13059-015-0796-9.

Čermák, T., Curtin, S. J., Gil-Humanes, J., Čegan, R., Kono, T. J. Y., Konečná, E., et al. (2017). A Multipurpose Toolkit to Enable Advanced Genome Engineering in Plants. *Plant Cell* 29, 1196–1217. doi:10.1105/tpc.16.00922.

Chen, Q., He, J., Phoolcharoen, W., and Mason, H. S. (2011). Geminiviral vectors based on bean yellow dwarf virus for production of vaccine antigens and monoclonal antibodies in plants. *Hum. Vaccin.* 7, 331–338.

Clough, S. J., and Bent, A. F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J.* 16, 735–743. doi:10.1046/j.1365-313x.1998.00343.x.

Davis, L., and Maizels, N. (2014). Homology-directed repair of DNA nicks via pathways distinct from canonical double-strand break repair. *Proc. Natl. Acad. Sci. U. S. A.* 111, E924-932. doi:10.1073/pnas.1400236111.

Doudna, J. A., and Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* 346, 1258096. doi:10.1126/science.1258096.

Dueva, R., and Iliakis, G. (2013). Alternative pathways of non-homologous end joining (NHEJ) in genomic instability and cancer. *Transl. Cancer Res.* 2, 163–177. doi:10.21037/1152.

Even-Faitelson, L., Samach, A., Melamed-Bessudo, C., Avivi-Ragolsky, N., and Levy, A. A. (2011). Localized egg-cell expression of effector proteins for targeted modification of the Arabidopsis genome. *Plant J.* 68, 929–937. doi:10.1111/j.1365-313X.2011.04741.x.

Fauser, F., Roth, N., Pacher, M., Ilg, G., Sánchez-Fernández, R., Biesgen, C., et al. (2012). In planta gene targeting. *Proc. Natl. Acad. Sci. U. S. A.* 109, 7535–7540. doi:10.1073/pnas.1202191109.

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

Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., et al. (2017). Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* 551, 464. doi:10.1038/nature24644.

Gil-Humanes, J., Wang, Y., Liang, Z., Shan, Q., Ozuna, C. V., Sánchez-León, S., et al. (2017). High-efficiency gene targeting in hexaploid wheat using DNA replicons and CRISPR/Cas9. *Plant J.* 89, 1251–1262. doi:10.1111/tpj.13446.

Gorbunova, V., and Levy, A. A. (1997). Non-homologous DNA end joining in plant cells is associated with deletions and filler DNA insertions. *Nucleic Acids Res.* 25, 4650–4657.

Hahn, F., Mantegazza, O., Greiner, A., Hegemann, P., Eisenhut, M., and Weber, A. P. M. (2017). An Efficient Visual Screen for CRISPR/Cas9 Activity in Arabidopsis thaliana. *Front. Plant Sci.* 8. doi:10.3389/fpls.2017.00039.

Hauser, M. T., Harr, B., and Schlötterer, C. (2001). Trichome distribution in Arabidopsis thaliana and its close relative Arabidopsis lyrata: molecular analysis of the candidate gene GLABROUS1. *Mol. Biol. Evol.* 18, 1754–1763.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821. doi:10.1126/science.1225829.

Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., and Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533, 420–424. doi:10.1038/nature17946.

Koornneef, M., Dellaert, L. W., and van der Veen, J. H. (1982). EMS- and radiation-induced mutation frequencies at individual loci in Arabidopsis thaliana (L.) Heynh. *Mutat. Res.* 93, 109–123.

Kumar, S., Worden, A., Novak, S., Lee, R., and Petolino, J. F. (2016). A trait stacking system via intra-genomic homologous recombination. *Planta* 244, 1157–1166. doi:10.1007/s00425-016-2595-2.

Lee, K. Y., Lund, P., Lowe, K., and Dunsmuir, P. (1990). Homologous recombination in plant cells after Agrobacterium-mediated transformation. *Plant Cell* 2, 415–425. doi:10.1105/tpc.2.5.415.

Li, J.-F., Aach, J., Norville, J. E., McCormack, M., Zhang, D., Bush, J., et al. (2013). Multiplex and homologous recombination-mediated plant genome editing via guide RNA/Cas9. *Nat. Biotechnol.* 31, 688–691. doi:10.1038/nbt.2654.

Li, Z., Liu, Z.-B., Xing, A., Moon, B. P., Koellhoffer, J. P., Huang, L., et al. (2015). Cas9-Guide RNA Directed Genome Editing in Soybean. *Plant Physiol.* 169, 960–970. doi:10.1104/pp.15.00783.

Liu, L., van Tonder, T., Pietersen, G., Davies, J. W., and Stanley, J. (1997). Molecular characterization of a subgroup I geminivirus from a legume in South Africa. *J. Gen. Virol.* 78 (Pt 8), 2113–2117. doi:10.1099/0022-1317-78-8-2113.

Liu, X., Wu, S., Xu, J., Sui, C., and Wei, J. (2017). Application of CRISPR/Cas9 in plant biology. *Acta Pharm. Sin. B* 7, 292–302. doi:10.1016/j.apsb.2017.01.002.

Mao, Y., Zhang, Z., Feng, Z., Wei, P., Zhang, H., Botella, J. R., et al. (2016). Development of germ-line-specific CRISPR-Cas9 systems to improve the production of heritable gene modifications in Arabidopsis. *Plant Biotechnol. J.* 14, 519–532. doi:10.1111/pbi.12468.

Maruyama, T., Dougan, S. K., Truttmann, M. C., Bilate, A. M., Ingram, J. R., and Ploegh, H. L. (2015). Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat. Biotechnol.* 33, 538. doi:10.1038/nbt.3190.

Mathiasen, D. P., and Lisby, M. (2014). Cell cycle regulation of homologous recombination in Saccharomyces cerevisiae. *FEMS Microbiol. Rev.* 38, 172–184. doi:10.1111/1574-6976.12066.

Miao, Z. H., and Lam, E. (1995). Targeted disruption of the TGA3 locus in Arabidopsis thaliana. *Plant J. Cell Mol. Biol.* 7, 359–365.

Offringa, R., de Groot, M. J., Haagsman, H. J., Does, M. P., van den Elzen, P. J., and Hooykaas, P. J. (1990). Extrachromosomal homologous recombination and gene targeting in plant cells after Agrobacterium mediated transformation. *EMBO J.* 9, 3077–3084.

Oppenheimer, D. G., Herman, P. L., Sivakumaran, S., Esch, J., and Marks, M. D. (1991). A myb gene required for leaf trichome differentiation in Arabidopsis is expressed in stipules. *Cell* 67, 483–493. doi:10.1016/0092-8674(91)90523-2.

Orel, N., Kyryk, A., and Puchta, H. (2003). Different pathways of homologous recombination are used for the repair of double-strand breaks within tandemly arranged sequences in the plant genome. *Plant J.* 35, 604–612. doi:10.1046/j.1365-313X.2003.01832.x.

Orel, N., and Puchta, H. (2003). Differences in the processing of DNA ends in Arabidopsis thaliana and tobacco: possible implications for genome evolution. *Plant Mol. Biol.* 51, 523–531.

Paszkowski, J., Baur, M., Bogucki, A., and Potrykus, I. (1988). Gene targeting in plants. *EMBO J.* 7, 4021–4026.

Pattanaik, S., Patra, B., Singh, S. K., and Yuan, L. (2014). An overview of the gene regulatory network controlling trichome development in the model plant, Arabidopsis. *Front. Plant Sci.* 5. doi:10.3389/fpls.2014.00259.

Peng, R., Lin, G., and Li, J. (2016). Potential pitfalls of CRISPR/Cas9-mediated genome editing. *FEBS J.* 283, 1218–1231. doi:10.1111/febs.13586.

Puchta, H. (1998). Repair of genomic double-strand breaks in somatic plant cells by one-sided invasion of homologous sequences. *Plant J.* 13, 331–339. doi:10.1046/j.1365-313X.1998.00035.x.

Puchta, H. (2002). Gene replacement by homologous recombination in plants. *Plant Mol. Biol.* 48, 173–182.

Puchta, H. (2005). The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. *J. Exp. Bot.* 56, 1–14. doi:10.1093/jxb/eri025.

Puchta, H., Dujon, B., and Hohn, B. (1993). Homologous recombination in plant cells is enhanced by in vivo induction of double strand breaks into DNA by a site-specific endonuclease. *Nucleic Acids Res.* 21, 5034–5040.

Puchta, H., Dujon, B., and Hohn, B. (1996). Two different but related mechanisms are used in plants for the repair of genomic double-strand breaks by homologous recombination. *Proc. Natl. Acad. Sci. U. S. A.* 93, 5055–5060.

Puchta, H., and Fauser, F. (2013). Gene targeting in plants: 25 years later. *Int. J. Dev. Biol.* 57, 629–637. doi:10.1387/ijdb.130194hp.

Puchta, H., Swoboda, P., Gal, S., Blot, M., and Hohn, B. (1995). Somatic intrachromosomal homologous recombination events in populations of plant siblings. *Plant Mol. Biol.* 28, 281–292.

Ran, F. A., Hsu, P. D., Lin, C.-Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., et al. (2013). Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell* 154, 1380–1389. doi:10.1016/j.cell.2013.08.021.

Reiss, B., Schubert, I., Köpchen, K., Wendeler, E., Schell, J., and Puchta, H. (2000). RecA stimulates sister chromatid exchange and the fidelity of double-strand break repair, but not gene targeting, in plants transformed by Agrobacterium. *Proc. Natl. Acad. Sci. U. S. A.* 97, 3358–3363.

Risseeuw, E., Offringa, R., Franke-van Dijk, M. E., and Hooykaas, P. J. (1995). Targeted recombination in plants using Agrobacterium coincides with additional rearrangements at the target locus. *Plant J. Cell Mol. Biol.* 7, 109–119.

Rong, Z., Zhu, S., Xu, Y., and Fu, X. (2014). Homologous recombination in human embryonic stem cells using CRISPR/Cas9 nickase and a long DNA donor template. *Protein Cell* 5, 258–260. doi:10.1007/s13238-014-0032-5.

Saleh-Gohari, N., and Helleday, T. (2004). Conservative homologous recombination preferentially repairs DNA double-strand breaks in the S phase of the cell cycle in human cells. *Nucleic Acids Res.* 32, 3683–3688. doi:10.1093/nar/gkh703.

Sargent, R. G., Brenneman, M. A., and Wilson, J. H. (1997). Repair of site-specific double-strand breaks in a mammalian chromosome by homologous and illegitimate recombination. *Mol. Cell. Biol.* 17, 267–277.

Sauer, N. J., Narváez-Vásquez, J., Mozoruk, J., Miller, R. B., Warburg, Z. J., Woodward, M. J., et al. (2016). Oligonucleotide-Mediated Genome Editing Provides Precision and Function to Engineered Nucleases and Antibiotics in Plants. *Plant Physiol.* 170, 1917–1928. doi:10.1104/pp.15.01696.

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

Schiml, S., Fauser, F., and Puchta, H. (2017). CRISPR/Cas-Mediated In Planta Gene Targeting. *Methods Mol. Biol. Clifton NJ* 1610, 3–11. doi:10.1007/978-1-4939-7003-2_1.

Shaked, H., Melamed-Bessudo, C., and Levy, A. A. (2005). High-frequency gene targeting in Arabidopsis plants expressing the yeast RAD54 gene. *Proc. Natl. Acad. Sci. U. S. A.* 102, 12265–12269. doi:10.1073/pnas.0502601102.

Steinert, J., Schiml, S., and Puchta, H. (2016). Homology-based double-strand break-induced genome engineering in plants. *Plant Cell Rep.* 35, 1429–1438. doi:10.1007/s00299-016-1981-3.

Sun, Y., Zhang, X., Wu, C., He, Y., Ma, Y., Hou, H., et al. (2016). Engineering Herbicide-Resistant Rice Plants through CRISPR/Cas9-Mediated Homologous Recombination of Acetolactate Synthase. *Mol. Plant* 9, 628–631. doi:10.1016/j.molp.2016.01.001.

Timmermans, M. C., Das, O. P., and Messing, J. (1992). Trans replication and high copy numbers of wheat dwarf virus vectors in maize cells. *Nucleic Acids Res.* 20, 4047–4054.

Voytas, D. F. (2013). Plant genome engineering with sequence-specific nucleases. *Annu. Rev. Plant Biol.* 64, 327–350. doi:10.1146/annurev-arplant-042811-105552.

Wang, M., Lu, Y., Botella, J. R., Mao, Y., Hua, K., and Zhu, J.-K. (2017). Gene Targeting by Homology-Directed Repair in Rice Using a Geminivirus-Based CRISPR/Cas9 System. *Mol. Plant* 10, 1007–1010. doi:10.1016/j.molp.2017.03.002.

Wang, Z.-P., Xing, H.-L., Dong, L., Zhang, H.-Y., Han, C.-Y., Wang, X.-C., et al. (2015). Egg cell-specific promoter-controlled CRISPR/Cas9 efficiently generates homozygous mutants for multiple target genes in Arabidopsis in a single generation. *Genome Biol.* 16. doi:10.1186/s13059-015-0715-0.

Weigel, D., and Glazebrook, J. (2009). Quick miniprep for plant DNA isolation. *Cold Spring Harb. Protoc.* 2009, pdb.prot5179. doi:10.1101/pdb.prot5179.

Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., et al. (2015). Cpf1 is a single RNA-guided endonuclease of a Class 2 CRISPR-Cas system. *Cell* 163, 759–771. doi:10.1016/j.cell.2015.09.038.

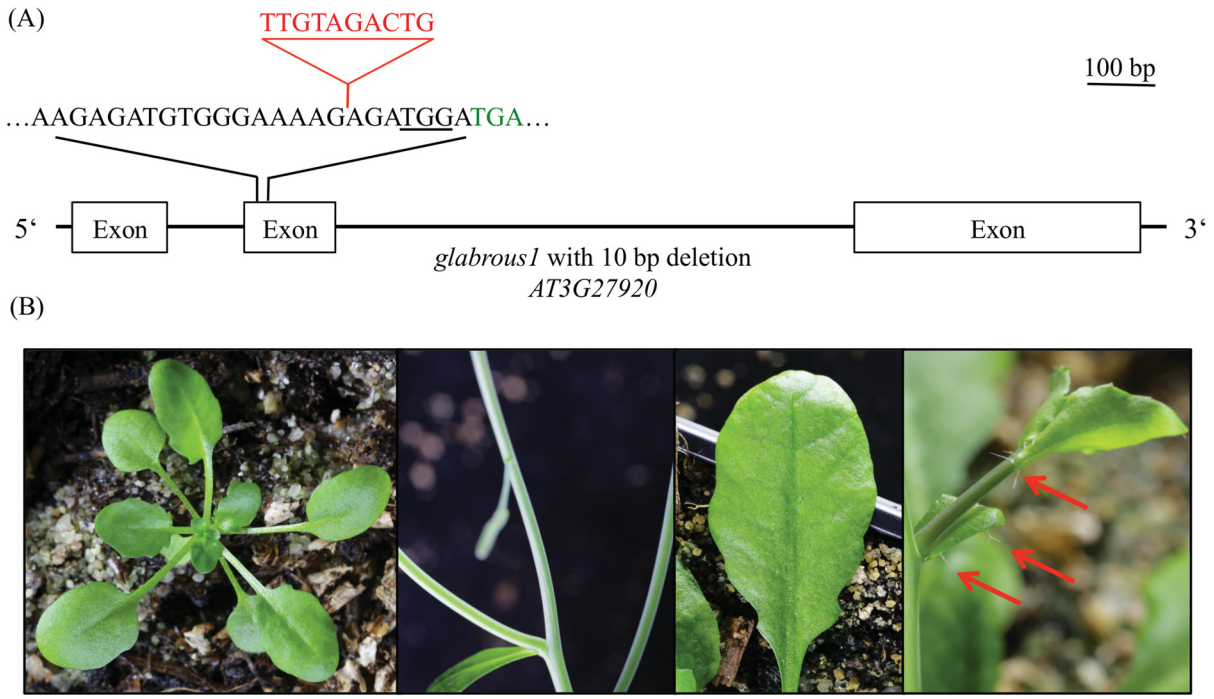
Zhang, K., Raboanatahiry, N., Zhu, B., and Li, M. (2017). Progress in Genome Editing Technology and Its Application in Plants. *Front. Plant Sci.* 8. doi:10.3389/fpls.2017.00177.

Zhao, Y., Zhang, C., Liu, W., Gao, W., Liu, C., Song, G., et al. (2016). An alternative strategy for targeted gene replacement in plants using a dual-sgRNA/Cas9 design. *Sci. Rep.* 6, 23890. doi:10.1038/srep23890.

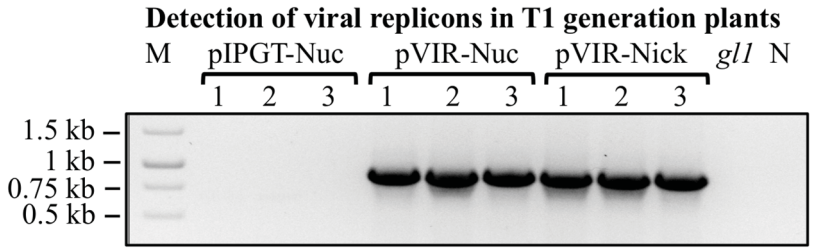
Zhu, L., Mon, H., Xu, J., Lee, J. M., and Kusakabe, T. (2015). CRISPR/Cas9-mediated knockout of factors in non-homologous end joining pathway enhances gene targeting in silkworm cells. *Sci. Rep.* 5. doi:10.1038/srep18103.

Zong, Y., Wang, Y., Li, C., Zhang, R., Chen, K., Ran, Y., et al. (2017). Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nat. Biotechnol.* 35, 438. doi:10.1038/nbt.3811.

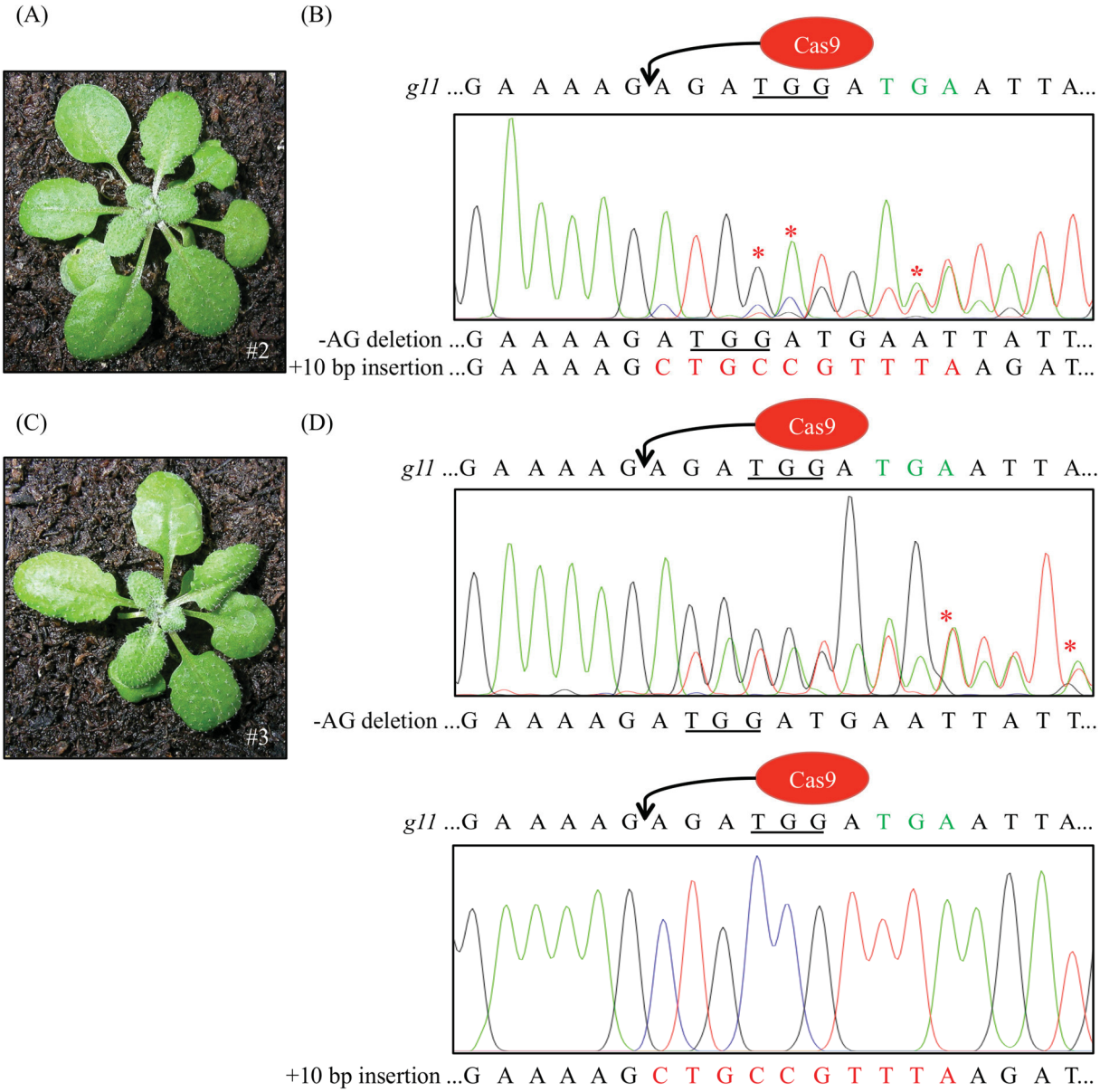
SUPPLEMENTARY MATERIAL



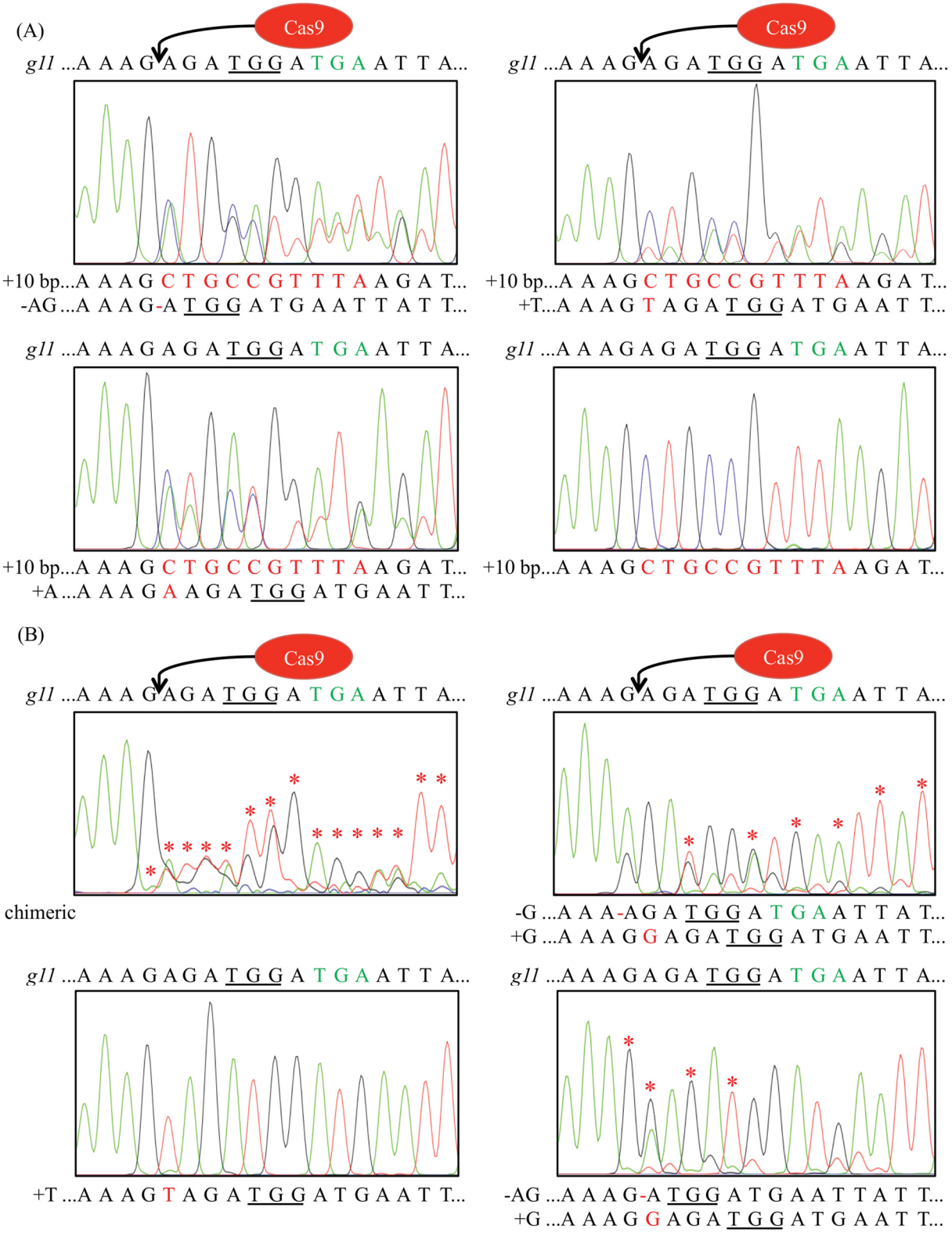
Supplementary Figure S1: A trichomeless *Arabidopsis gl1* mutant is a suitable model for Cas9-mediated repair. In previous experiments, we generated an *Arabidopsis* line with a 10 bp deletion (red) in the second exon of the *GL1* gene (A). The deletion is located three nucleotides in front of a NGG motive (underlined), which functions as PAM motive in *S. pyogenes* Cas9 systems. The deletion leads to a premature STOP codon (green) that disrupts the protein function. Plants carrying the mutations show no trichomes on the leaf surface of rosette leaves and on stems. (B) Some *gl1* mutant plants show single trichomes (red arrows) on leaf margins and on the transition to the petiole, but only on leaves that derive from the stem.



Supplementary Figure S2: Detection of viral replicons in T1 generation plants transformed with pVIR-constructs. We analyzed replicon formation on leaf gDNA of three independent T1 *Arabidopsis* lines transformed with pIPGT-Nuc, pVIR-Nuc, and pVIR-Nick, respectively, using a primer pair that only yields a PCR amplicon on circular replicons. DNA of a *gl1* plant was used as negative control. Replicons could be detected in all T1 plants transformed with a pVIR-construct at the expected size of 919 bp. Image of agarose gel was colour inverted for better visibility. M = Marker, N = water control.



Supplementary Figure S3: Sequence analysis of T3 trichome plants #2 and #3. (A) Image of trichome-bearing plant #2. (B) Sequence analysis of *gl1* gene of plant #2. Multiple peaks that indicate chimerism and unfixed mutations are marked with red asterisks. (C) Image of trichome-bearing plant #3. (D) Top: Sequence analysis of *gl1* gene of plant #3. Bottom: Sequence of a subcloned *gl1* gene of plant #3. Sequence of the dysfunctional *gl1* gene is given on top of each sequencing histogram for comparison. Peak colours: Adenine = green, cytosine = blue, guanine = black, thymine = red; PAM sequence underlined, premature STOP codon marked in green letters.



Supplementary Figure 4: Exemplary sequencing histograms of T4 generation plants with (A) and without (B) trichomes. All plants carrying trichomes contain at least one functional copy of the *GL1* gene with the codon-modified 10 bp integrated in the Cas9 generated DSB. Sequence of the dysfunctional *gl1* gene is given on top of the sequencing histogram for comparison. Multiple peaks that indicate chimerism and unfixed mutations are marked with red asterisks. Peak colours: Adenine = green, cytosine = blue, guanine = black, thymine = red; PAM sequence underlined, premature STOP codon marked in green letters.

Table S3. Primer Sequences.

Primer	Sequence (5' to 3')
FH190	GTTTGCACGACTAATACACTTATG
FH210	ATTGGAAGAGATGTGGGAAAAGAGA
FH211	AAACTCTCTTTTCCCACATCTCTTC
FH214	AAAGACCATCTGAATTTGGGG
FH215	TAAGCACGTGTCACGAAAACC
FH217	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCACCGTTGACCTGCAGGCACG
FH218	GGGGACCACTTTGTACAAGAAAGCTGGGTGTTTAAACTATGAGGGTCGTACG
FH223	AAAGCTGCCGTTTAAGATGG
FH240	GAGCTTTGGGTACGTCACCCACTTTGTTGTATAAATAACTAAAAATAATATG
FH241	TGACTTGAAGTACACTCATATTAAATTCATACCATATAACATAAATTATAAC
FH265	TCTACTTGCACAATGAAATATATTC
FH266	CATACGGTTCAGGTTGTGGA
FH271	GTGACGTACCCAAAGCTCTG
FH278	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCACCTAGGGCGAATTGGCGGAA
FH279	GGGGACCACTTTGTACAAGAAAGCTGGGTGAGCGGGCAGTGAGCGGAA
FH313	CGGAACTAACTCTGTGGGATG
FH314	CCTCAGCGAGATCGAAGTTAG

MANUSCRIPT IV

Towards establishment of a photorespiratory CO_2 pump in *Arabidopsis thaliana* using CRISPR/Cas9-mediated modifications of the *GLYCINE DECARBOXYLASE P-PROTEIN* promoter

Towards establishment of a photorespiratory CO_2 pump in *Arabidopsis thaliana* using CRISPR/Cas9-mediated modifications of the *GLYCINE DECARBOXYLASE P-PROTEIN* promoter

Florian Hahn[1,3], Otho Mantegazza[1,3], Marion Eisenhut[1,3], Urte Schlüter[1,3], Tabea Mettler-Altmann[1,3], Kumari Billakurthi[2,3], Andreas P.M. Weber[1,3]*

[1]Institute of Plant Biochemistry, Heinrich Heine University, Düsseldorf, Germany, Center for Synthetic Life Sciences (CSL), Heinrich Heine University, Düsseldorf, Germany
[2]Institut für Entwicklungs- und Molekularbiologie der Pflanzen, Heinrich Heine University, Düsseldorf, Germany
[3]Cluster of Excellence on Plant Sciences (CEPLAS)

***Correspondence:**
Andreas P.M. Weber
Andreas.Weber@uni-duesseldorf.de

Keywords: CRISPR/Cas9, photorespiration, glycine decarboxylase, *cis*-regulatory elements, C_2 photosynthesis

ABSTRACT

Introducing the highly efficient carbon concentrating mechanisms of C_4 photosynthesis in C_3 plants has been a long-time goal in plant science. Studies in C_3 - C_4 intermediate plants have provided evidence that a photorespiratory glycine shuttle is the initial biochemical step in the evolution of C_4 photosynthesis. This shuttle is based on a bundle sheath cell specific expression of the GLYCINE DECARBOXYLASE system. We aimed at implementing this photorespiratory shuttle in the C_3 plant *Arabidopsis thaliana* by introducing CRISPR/Cas9-induced changes into the promoter region of the *AtGLDP1* gene, an essential component of the GLYCINE

DECARBOXYLASE system. Replacing the globally active *Arabidopsis* promoter with a bundle sheath specific promoter from *Flaveria trinervia* using a gene targeting approach did not yield stable mutant plants. However, we could delete large parts of a *cis*-regulatory element of the *Arabidopsis* promoter, which is responsible for expression of the gene in mesophyll cells. Plants lacking this promoter element accumulated glycine and showed a decrease in *AtGLDP1* expression in leaves. Further analyses are necessary to access the subcellular localization of GLDP1 in these plants.

INTRODUCTION

To meet the rising food demand of the growing world population and to compensate for a decrease of arable land (Ray et al., 2013), one of the major scientific challenges within the next decades is to increase crop productivity. During the last century, crop productivity was often enhanced by increasing the amount of plant biomass investment in harvested parts but partitioning efficiencies have reached a theoretical maximum (Long et al., 2015). Instead, at the present day we could focus on improving how plants convert solar energy into biomass, which is directly connected to improving plant's photosynthetic capacities (Betti et al., 2016; Long et al., 2015; Ort et al., 2015).

RIBULOSE-1,5-BISPHOSPATE CARBOXYLASE/OXYGENASE (RuBisCO), the central enzyme of CO_2 fixation in phototrophic organisms, is also capable of fixing O_2, which results in the formation of 2-phosphoglycolate (2-PG) besides 3-phosphoglycerate (Bowes et al., 1971). 2-PG is a dead-end metabolite and has to be recycled in a complex salvage pathway, called photorespiration (PR) or oxidative C_2 cycle (Peterhansel et al., 2010). Even though photorespiration is essential for phototrophic organisms in an oxygen-containing atmosphere (Eisenhut et al., 2008; Rademacher et al., 2016), it is a wasteful process as it releases CO_2 and NH_3 and consumes ATP and NAD(P)H. Calculations show that PR accounts for yield losses in US wheat and soybean production of 20 % and 36 %, respectively (Walker et al., 2016). This underlines that PR is a major target for crop improvement.

The C_2 cycle (Fig. 1A) starts with the dephosphorylation of 2-PG to glycolate in the chloroplasts, which is further converted to glyoxylate and glycine in the peroxisomes (Somerville and Ogren, 1979; Igarashi et al., 2003; Schwarte and Bauwe, 2007). Glycine is then shuttled into the mitochondria, where the multienzyme system GLYCINE DECARBOXYLASE (GDC) in

cooperation with SERINE HYDROXYMETHYLTRANSFERASE (SHMT) converts two molecules of glycine into one molecule of serine, thereby releasing NH_3 and CO_2 (Voll et al., 2006; Engel et al., 2007, 2011). Serine is then deaminated in the peroxisomes and further reduced to glycerate in peroxisomes or the cytosol (Liepman and Olsen, 2001; Timm et al., 2008). The last step of the C_2 cycle takes place in the chloroplast, where glycerate is phosphorylated to 3-phosphoglycerate, a substrate of the Calvin-Benson cycle (Boldt et al., 2005).

The GDC system is responsible for the high carbon loss of photorespiration and is a bottleneck of the C_2 cycle (Betti et al., 2016). It is found in phototrophic organisms but also in all other eukaryotes and bacteria due to its involvement in C_1 metabolism (Schulze et al., 2016). It consists of four different protein subunits, the H-protein (GLDH), the L-protein (GLDL), the T-protein (GLDT), and the P-protein (GLDP) in a subunit stoichiometry of 27:2:9:4, respectively. This multienzyme system is very fragile and only stable due to the very high concentrations of the subunits as found in the mitochondrial matrix of leaves, where it comprises about 1/3 of the soluble mitochondrial matrix proteins (Oliver et al., 1990). This high amount of GDC allows rapid responses to changes in PR rates and avoids accumulation of glycine, which can be toxic in excess (Eisenhut et al., 2007). Glycine is bound and decarboxylated by the pyridoxal-5-phosphate containing GLDP homodimer, resulting in an aminomethylen residue and CO_2. *Arabidopsis thaliana* (*Arabidopsis*) contains two *GLDP* genes with redundant functions. In contrast to most other photorespiratory genes, a complete knockout of both genes is lethal also under non-photorespiratory conditions due to the role of GDC in other metabolic processes (Engel et al., 2007; Timm and Bauwe, 2013). GLDH carries a lipoyl group that transfers the aminomethylene from GLDP to GLDT, where it is deaminated. The methyl group is transferred to tetrahydrofolate. The methylene-tetrahydrofolate is later used by SHMT to generate serine out of glycine. GLDL reoxidizes the lipoyl group of the GLDH to complete the cycle (Douce et al., 2001; Bauwe, 2010).

Photosynthetic organisms evolved various strategies to avoid energy losses by photorespiration (Sage, 2013; Sage and Khoshravesh, 2016). These strategies mainly rely on CO_2 concentrating mechanisms around RuBisCO to avoid its oxygenation reaction in the first place. The most successful approach in higher plants is C_4 photosynthesis. C_4 plants employ a primary CO_2 fixation (in form of HCO_3^-) via PHOSPHOENOLPYRUVATE CARBOXYLASE (PEPC) in the mesophyll cells (M cells). The resulting C_4 acid is transferred to bundle sheath cells (BS cells) and decarboxylated, which leads to a 10-30-fold enrichment of CO_2 in these cells (Sage, 2013).

As RuBisCO activity is restricted to BS cells in C_4 plants, its oxygenation reaction is repressed and PR highly reduced (Slack and Hatch, 1967; Sage, 2004). However, also in C_4 plants the photorespiratory metabolism is essential to thrive in the current atmosphere (Zelitch et al., 2009). C_4 photosynthesis has evolved manifold independently in over 60 plant lineages (Sage et al., 2011) and international consortia aim at introducing C_4 pathways in important crop plants such as rice (The C_4 rice project - c4rice.irri.org). Therefore, research is focused on understanding the evolution of C_4 photosynthesis, with a special focus on C_3 - C_4 intermediate plants (Schlüter and Weber, 2016). C_3 - C_4 intermediate plants belong to a group of so-called C_2 plants, which are found in the genus *Moricandia*, *Flaveria* and *Cleome* amongst others (Sage et al., 2014). These C_2 plants employ a photorespiratory CO_2 concentrating mechanisms that is thought to be the initial step towards C_4 photosynthesis (Heckmann et al., 2013; Mallmann et al., 2014). In C_2 plants (Fig. 1B), GDC activity is shifted to BS cells by cell-type specific expression of *GLDP* and partly also the other components of the glycine cleavage system (Schulze et al., 2016). Photorespiratory glycine from M cells therefore needs to diffuse to BS cells to complete the C_2 cycle (Hylton et al., 1988; Rawsthorne et al., 1988). Concerted decarboxylation of glycine in the BS cells leads to a 3-fold enhancement of CO_2 concentration in plastids of leaf cells of C_2 plants compared to C_3 plants (Keerberg et al., 2014). This suppresses the oxygenation reaction of RuBisCO in these cells, leading to a more effective photosynthesis under photorespiration-promoting conditions (drought, high temperatures, excessive light) as shown by lowered CO_2 compensation points compared to C_3 plants (Vogan and Sage, 2012). To allow the high fluxes of glycine and enable efficient trapping of released CO_2, C_2 plants show changes in morphology such as higher proportion of BS cells and centripetally arranged chloroplasts. Interestingly, the glycine shuttle might already enforce an evolutionary pressure towards C_4 photosynthesis. Since two molecules of glycine are imported into the BS cells but only one molecule of serine is shuttled back, a nitrogen imbalance occurs between M cells and BS cells that can be rebalanced by metabolic pathways that highly resemble basic C_4 cycles (Heckmann et al., 2013; Mallmann et al., 2014; Schlüter and Weber, 2016). In contrast to implementing a full C_4 cycle into plants, the C_2 photorespiratory pump might be relatively easy to implement into C_3 plants, such as *Arabidopsis* or major crop plants, and is therefore of high biotechnological interest (Bauwe and Kolukisaoglu, 2003; Schulze et al., 2016).

For achieving a BS cell specific expression of GDC genes, it is necessary to understand the regulatory elements that drive cell specific expression in C_2 and C_4 plants. Very diverse control

mechanisms have been identified on various C_4 genes, ranging from control mechanisms in the promoter regions (Gowik et al., 2004; Akyildiz et al., 2007; Engelmann et al., 2008; Wiludda et al., 2012; Adwy et al., 2015; Gowik et al., 2017), the untranslated regions (Kajala et al., 2012; Williams et al., 2016) or the exons (Brown et al., 2011). Extensive studies have been performed on the *GLDPA* promoter of the C_4 plant *Flaveria trinervia* (*F. trinervia*). This promoter shows a complex interplay of transcriptional and post-transcriptional regulation with a dominant BS cell/vein-specific element proximal to the coding region and a general leaf promoter element at the distal region of the promoter. The activity of the unspecific distal promoter is however mostly repressed post-transcriptionally by the proximal, BS cell specific promoter and allows only leaky expression of *FtGLDPA* in M cells to fuel C_1 metabolism (Wiludda et al., 2012). Interestingly, the *FtGLDPA* promoter confines also BS cell/vein specific expression of genes in transgenic *Arabidopsis* plants (Engelmann et al., 2008). Recently, a study comparing the *GLDP* promoter regions of the C_3 plant *Arabidopsis* and the closely related C_2 plant *Moricandia nitens* (*M. nitens*) revealed two conserved motives in the *GLDP* promoters of *Brassicaceae*. The V-Box confers expression in vasculature and BS cells, the M-Box confers expression in M cells and is absent in the C_2 species. This provides a model for the establishment of a C_2 photorespiratory pump by a simple deletion of a promoter element (Adwy et al., 2015).

In this study, we aimed at implementing a photorespiratory CO_2 pump in *Arabidopsis* by restricting the expression of the *AtGLDP1* gene to BS cells. For this, we took advantage of the emerging possibilities of the CRISPR/Cas9 system (Doudna and Charpentier, 2014), which allows subtle changes in the original context of the genome. In this system, the Cas9 nuclease from the bacterium *Streptococcus pyogenes* is guided by a programmable artificial single guide RNA (sgRNA, Jinek et al., 2012) to a DNA sequence of choice where it introduces a double-strand break (DSB). This DSB is repaired by the cell's repair machinery either by error-prone non-homologous end joining (NHEJ) or by homologous recombination (HR) using a homology template (HT) as repair template (Puchta, 2005). HR can be exploited for targeted introduction of foreign DNA, which is called gene targeting (GT, Paszkowski et al., 1988). Additionally, mutations in one of the two cleavage domains of Cas9 allow the introduction of single strand DNA nicks, which further enhance the Cas9 toolbox (Cong et al., 2013; Fauser et al., 2014).

In this manuscript, we explored strategies to restrict *AtGLDP1* expression to the BS cells of *Arabidopsis* through Cas9-based editing. The main strategy was to replace the unspecific *AtGLDP1* promoter with the BS cell specific *FtGLDPA* promoter from *F. trinervia*.

Alternatively, we aimed at deleting the M-Box from the *AtGLDP1* promoter. Both strategies should lead to a BS cell specific expression of *AtGLDP1* and restrict GDC activity to the BS cells, thereby allowing insights into the effects of a photorespiratory CO_2 concentrating mechanism in a C_3 plant without anatomical preconditioning for C_2 photosynthesis.

METHODS

Plant growth conditions

Arabidopsis seeds were surface-sterilized and stratified at 4 °C for three days. Seeds were germinated on 0.8 % (w/v) agar-solidified half-strength MS medium (Duchefa) in growth chambers with a light intensity of 100 µmol photons m^{-2} s^{-1} in a 12-h-light/12-h-dark cycle (22/18 °C). After 14 days, seedlings were transferred to soil for further growth under the same conditions. Plants were either grown under ambient CO_2 conditions (390 ppm CO_2) or at high CO_2 concentrations (3,000 ppm CO_2). For the GT screen, plants were grown in the green house.

Isolation of an *Arabidopsis gldp2-2* mutant

We obtained seeds from a *gldp1-1*$^{\text{heterozygous}}$x*gldp2-2*$^{\text{homozygous}}$ T-DNA line cross (Engel et al., 2007; kindly obtained by Stefan Timm, University of Rostock). T-DNA presence in progeny plants was detected by polymerase chain reaction (PCR) using genomic DNA (Weigel and Glazebrook, 2009) and primers derived from Engel et al. (2007). *AtGLDP2* gene specific primers (P2-2Check-S/P2-A2n) and T-DNA specific primers (P2-A2n/SAIL-LB1) were used to identify plants containing the *gldp2-2* T-DNA insertion. Lack of T-DNA insertion in the *AtGLDP1* gene was verified with *AtGLDP1* gene specific primers (P12-S1/P1-A1n) and T-DNA specific primers (P12-S1/TJLB1). A *gldp1-1* T-DNA line (Campisi et al., 1999; Engel et al., 2007; kindly obtained by Stefan Timm, University of Rostock) was used as control. Seeds homozygous for the *gldp2-2* insertion (SAIL_889_D09, Sessions 2002) and lacking the *gldp1-1* T-DNA insertion were propagated.

Generation of a homology repair template for *AtGLDP1* promoter exchange

We designed a homology repair template (HT) containing the 1,571 bp region upstream of the *F. trinervia GLDPA* gene that includes 5'UTR and promoter (Engelmann et al., 2008). Additionally, we attached the genomic regions up- and downstream of the *AtGLDP1* promoter up to the

predicted cutting site of the Cas9 protein. We added silent nucleotide exchanges in the sgRNA target sites to prevent cleavage after homology-directed repair. The repair template was flanked by homology arms (HA) complementary to the genomic regions adjacent to the cutting site of Cas9 (825 bp and 795 bp, respectively) and contained additionally the sequences of the sgRNA recognition sites including the PAM sequences. The HT was synthesized at GeneART (ThermoFisher Scientific) and integrated into the vector backbone pMA-RQ (GeneART). The resulting vector was named pFH97.

Cloning of a binary T-DNA repair vector for *in planta* gene targeting repair approach

To target the *AtGLDP1* promoter, we used the binary T-DNA vector pDE-Cas9, which contains an *Arabidopsis*-codon-optimized Cas9 under the control of an *Ubiquitin* promoter (PcUBp) from *Petroselinum crispum* (Fauser et al., 2014). For sgRNA subcloning, we used the vector pEN-C.1.1 containing the *Arabidopsis* U6-26 promoter for expression of the sgRNA, the conserved sgRNA scaffold and BbsI-sites for subcloning of the target sequence (Schiml 2014). We digested pEN-C1.1 with BbsI-HF (NewEngland Biolabs) and inserted the protospacer sequences for sgRNA1 (GCTTGTGGATTTGATCGCTG) and sgRNA2 (GCCTCCCGAAAGCCGCCGTG) using two annealed primers FH235/FH236 and FH237/FH238, respectively (Table S1) thereby creating the vectors pFH93 and pFH94. We digested pFH93 with MluI-HF (NewEngland Biolabs) and extracted the sgRNA cassette for ligation into MluI-HF-digested pDE-Cas9 resulting in vector pFH101. We amplified the second sgRNA cassette from pFH94 using primers FH280/FH281 (Table S1) and cloned it into Bsu36I-digested pFH101 via Gibson Cloning (NewEngland Biolabs) to create vector pFH109. We PCR-amplified the HT from pFH97 using primers FH278/FH279 (Table S1) and performed Gateway® BP-reaction with pDONR 201 (Invitrogen) resulting in vector pFH124. By Gateway® LR reaction (Invitrogen) of pFH124 with pFH109, we assembled the IPGT binary T-DNA repair vector pIPGT-Nuc-GLDP1 (Fig. 2B).

Cloning of binary T-DNA repair vectors for viral replicon repair approach

For the viral replicon repair approach, we used the Gateway® compatible vector pFH83 (Hahn et al., 2018), which contains the viral components from *bean yellow dwarf virus* (BeYDV) necessary for replicon formation (Baltes et al., 2014). We amplified the *GLDP1* HT from pFH97 using primers FH244/FH245 (Table S1) and ligated it into PmlI-/BsaXI-digested (NewEngland Biolabs) pFH83 vector, resulting in the vector pFH86. By Gateway® LR reaction (Invitrogen) of

pFH109 with pFH86, we created the viral replicon nuclease repair construct pVIR-Nuc-GLDP1 containing Cas9 nuclease, two sgRNA cassettes and the HT (Fig. 2C).
For the Cas9 nickase repair construct, we ligated the MluI-HF-extracted sgRNA cassette from pFH89 into MluI-HF-digested pDE-Cas9-D10A (Fauser et al., 2014) resulting in vector pFH105. The sgRNA cassette from pFH94 was amplified using primers FH280/FH281 (Table S1) and integrated into Bsu36I-digested pFH105 by Gibson Cloning (NewEngland Biolabs) to create vector pFH113. We performed a Gateway® LR reaction (Invitrogen) of pFH113 with pFH86 to obtain the final viral replicon nickase repair construct pVIR-Nick-GLDP1 (Fig. 2D).

Cloning of a binary T-DNA vector for M-Box deletion

For deletion of the M-box in the *AtGLDP1* promoter, we inserted sgRNA target sequences (GTTATCTTTCTGAGTGATAA and GAGATAATTTACAGGAATTG) into pEN-C.1.1 using annealed primers (FH 298/FH299 and FH300/FH301, respectively; Table S1) as described before. The first sgRNA cassette was inserted into pDE-Cas9-GentR (kindly obtained by Holger Puchta, Karlsruher Institut für Technologie) via MluI-HF, the second sgRNA was inserted afterwards via Gateway® LR Reaction (Invitrogen), resulting in the final vector pCas9-MBOX (Fig. 4B).

Cloning of binary T-DNA vectors for ß -glucuronidase staining

We obtained the constructs P4 and P3Δ1 (Adwy et al., 2015, kindly obtained by Waly Adwy, Leibniz Universität Hannover), which contain the full *AtGLDP1* promoter (including 5'-UTR) or a shortened version of the *AtGLDP1* promoter lacking the 5'-end including the M-Box. We introduced a partial deletion in the M-Box of the P4 construct using the Q5® site-directed mutagenesis kit (NewEngland Biolabs) using primers FH349/FH350 (Table S1) resulting in vector pFH193. We amplified the shortened promoter of the P3Δ1 construct using primers FH356/FH357 (Table S1) and cloned it into SalI-HF/PstI-HF-linearized binary T-DNA ß –glucuronidase (GUS) vector pCambia1381 via Gibson Cloning (NewEngland Biolabs). We amplified the *AtGLDP* promoter sequences of P4 and pFH193 with primers FH364/FH365 (Table S1) and cloned the PCR amplicon into pJET1.2 (ThermoFisher Scientific). We extracted the promoter regions with SalI-HF/PstI-HF and ligated them into pCambia1381 via T4 DNA Ligase (NewEngland Biolabs). The final constructs were named pGUS-GLDPp-FULL (containing the full *AtGLDP1* promoter including 5'-UTR), pGUS-GLDPp-SHORT (containing a shortened

AtGLDP1 promoter including 5'-UTR but lacking the M-Box and the sequences upstream of the M-Box) and pGUS-GLDPp-DELETION (containing the full *AtGLDP1* promoter including 5'-UTR with a partial deletion of the M-Box).

Plant transformation

Agrobacterium tumefaciens, strain GV3101::pMP90 were transformed with T-DNA binary vectors pIPGT-Nuc-GLDP1, pVIR-Nuc-GLDP1, pVIR-Nick-GLDP1, pCas9-MBOX, pGUS-GLDPp-FULL, pGUS-GLDPp-SHORT and pGUS-GLDPp-DELETION. *Arabidopsis gldp2-2* plants were transformed by floral dipping (Clough and Bent, 1998) with pCas9-MBOX while Col-0 plants were transformed with the other constructs. Primary transformants were selected under high CO_2 conditions on 0.8 % (w/v) agar-solidified half-strength MS medium supplemented with 7.5 µg/mL glufosinate (pIPGT-Nuc-GLDP1, pVIR-Nuc-GLDP1), 50 µg/mL kanamycin (pVIR-Nick-GLDP1), 75 µg/mL gentamycin (pCas9-MBOX) or 33 µg/mL hygromycin (pGUS-constructs).

Detection of *Cas9* gene in plants

Presence of the *Cas9* gene in plants was verified by PCR on leaf genomic DNA with primer pairs amplifying specifically the *Cas9* gene (FH313/FH314, Table S1).

Verification of viral replicon formation

Formation of circular viral replicons was verified in T1 generation plants by PCR on leaf genomic DNA using primers FH266/FH271 (Table S1).

Detection of *GLDP* promoter modifications

Leaf genomic DNA was isolated using the PhirePlant Direct PCR Master Mix (Thermo Fisher Scientific). Deletions in the M-Box of the *AtGLDP1* promoter were detected by PCR using primer pair FH302/FH303 (Table S1). Integration of the *FtGLDPA* promoter was verified using primer pair FH293/FH220 (Table S1).

Sequencing of *AtGLDP1* promoter modifications in plants

PCR amplicons amplified by proofreading Q5 Polymerase (NewEngland Biolabs) were analyzed directly by Sanger sequencing (Macrogen) or subcloned into pJET1.2 (ThermoFisher Scientific)

and then sequenced. 4Peaks software and SnapGene Viewer were used to analyze sequencing histograms.

GUS staining

Whole rosettes of *Arabidopsis* plants having 4 to 8 leaves were vacuum infiltrated with GUS staining solution (0.1 M NaP_i, pH 7.0; 10 mM Na_2EDTA; 0.5 mM Potassium-Ferricyanide $K_3[Fe(CN)_6]$; 0.5 mM Potassium-Ferrocyanide $K_4[Fe(CN)_6] \times 3H_2O$, 0.1% Triton X-100, and 2 mM 5-bromo-4-chloro-3-indoyl-beta-D-glucuronide) and incubated overnight at 37 °C. GUS staining was fixed with fixation solution (50 % (v/v) Ethanol, 5 % (v/v) Glacial acetic acid, 3.7 % (v/v) Formaldehyde) for 10 min at 65 °C. Leaves were incubated with 70 % (v/v) Ethanol for removal of chlorophyll.

Immunohistochemical detection of GLDP localization

GLDP localization was visualized by immunogold labeling of embedded leaf tissue followed by transmission electron microscopy (TEM) as described by Khoshravesh et al. (2017). Shortly, leaf tissue stripes of 36 days old plants grown at ambient CO_2 were fixed in a fixation solution (1 % (v/v) glutaraldehyde and 1 % (v/v) paraformaldehyde in 0.1 M sodium cacodylate buffer), dehydrated in a graded ethanol series and embedded in London Resin white. LR white blocks were mounted on Araldite 502 blocks and trimmed. Sectioning was performed on a ultramicrotome using diamond blades. Leaf sections were placed on nickel grids. For immunogold labeling of the GLDP protein, grids were incubated with a rabbit α-GLDP-antibody (dilution 1:5000, kindly obtained by Martha Ludwig, University of Western Australia) followed by incubation with a goat α-rabbit-antibody conjugated with 15 nm gold particles (TED PELLA INC., 1:50 dilution, kindly obtained by Eva Nowack). Immunostained grids were analyzed by TEM (E902, Zeiss).

Measurements of photochemical efficiency of photosystem II

Plants were grown on half-strength MS medium agar plates under ambient or high CO_2 conditions as described above for 28 days. 8 hours after onset of light, plants were dark adapted for 10 minutes and chlorophyll fluorescence was measured with an imaging pulse amplitude fluorometer (PAM M-series, Walz). The F_v / F_m ratio was calculated as indicator for the

maximum quantum yield of photosystem II (PS II).

Expression analysis of *GLDP1* using quantitative real-time PCR analysis

Leaf samples were taken from 24 days old plants grown under controlled conditions in plant growth chambers as described above after 8 hours of light period. Leaf RNA was extracted using EURx GeneMatrix Universal RNA Purification Kit (Roboklon). 2 µg RNA was treated with RQ1 RNase-free DNAse (Promega) to degrade residual DNA and cDNA was synthesized using SuperScript II reverse transcriptase (Invitrogen) with oligo-dT-primers according to the manufacturer's instruction. We used Luna® Universal qPCR Master Mix (NewEngland Biolabs) for quantitative real-time PCR (qRT-PCR) with the StepOne Plus Real-Time system (Applied Biosystems). Primers FH380/FH381 were used as *AtGLDP1* specific primers (efficiency 2.12, Table S1) and black primers P90/P91 binding on *AT4G34270* (*AtBLACK*, efficiency 2.18, Table S1, Czechowski et al., 2005) were used as reference for ΔΔCt analysis. Mean normalized expression was calculated using 3-4 biological replicates with three technical replicates, each, according to Simon (2003).

Metabolite measurements

Plants were grown for 24 days under controlled conditions as described above. Entire rosettes were harvested 7 h after onset of light and snap-frozen in liquid nitrogen. Extraction of plant tissue and metabolite analysis was performed according to Fiehn (2007). Shortly, weighted plant material was transferred in 2 mL storage tubes (Sarstedt) and homogenized using glass beads in a MM 400 ball mill (Retsch) at 30 Hz for 2x30 s. Metabolites were extracted using ice-cold extraction buffer (H_2O, methanol and $CHCl_3$ in volume proportions 1:2.5:1 plus 5 mM Ribitol) and vortexed vigorously, followed by shaking on a rotating shaker for 6 min at 4 °C, 30 rpm. Samples were centrifuged for 2 min at 20,000 x *g*. 10 µl of the supernatant was dried in a speedvac concentrator and derivatized with methoxyamine hydrochloride and N-methyl-N-tri-methylsilyltrifluoracetamid before analysis on a 7200 accurate mass Quadrupole Time-of-Flight GC-MS (Agilent Technologies).

Protein extraction and analysis

We harvested five leaves of 42 days old plants grown at ambient CO_2 7 h after onset of light. Plant material was snap-frozen in liquid nitrogen and homogenized using glass beads in a MM

400 ball mill (Retsch) for 1 min at 30 Hz. 1 mL of extraction buffer (50 mM Tris-HCl, pH 8, 5 mM MgCl2, 1 mM ethylenediaminetetraacetic acid, 1 mM dithiothreitol, 1 mM phenylmethane sulfonyl fluoride, 0.1 mM 4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride, 2 µM Pepstatin A) was added per 250 mg plant fresh weight. Samples were vortexed and centrifuged for 10 min at 10,000 x *g*. 20 µl of supernatant were denatured for 10 min at 98 °C in SDS loading buffer (60 mM Tris-HCl, pH 6.8, 10 % (v/v) glycerol, 0.05 % (w/v) SDS, 0.25 % (w/v) bromophenol blue), separated by SDS-PAGE, and analyzed by Western Blotting according to standard protocols. The primary antibody α-GLDP (kindly obtained by Martha Ludwig) and secondary antibody α-rabbit-IgG-HRP (Merck) were used at 1:5,000 dilutions in 5 % milk powder in TBS-buffer in incubated at room temperature for 1 h.

Promoter sequence analysis

Promoter sequences of various *Brassicaceae* species were derived from Phytozome by BLAST search using *AtGLDP1* as template (https://phytozome.jgi.doe.gov/pz/portal.html). Alignments of the M-Box sequences were performed using MUSCLE alignment tool (Edgar, 2004)

Data analysis

Two-tailed Student's *t*-tests were performed to detect significant differences between samples (Prism 6). In order to perform principal component analysis (PCA), we centered and scaled the intensity of each metabolite separately. The data transformations and the PCA were performed in R (R Core Team, 2017).

RESULTS

We aimed at repressing *AtGLDP1* expression in the M cells of *Arabidopsis* to create a photorespiratory CO_2 pump (Fig. 1), which was suggested to be an initial step of C_4 evolution (Heckmann et al., 2013). The promoter of *AtGLDP1* guides expression in all plant tissues including M cells due to a *cis*-regulatory element, called M-Box (Adwy et al., 2015; Schulze et al., 2013). In contrast, the *FtGLDPA* promoter of the C_4 plant *F. trinervia* shows an expression pattern that is mainly specific to the vasculature and BS cells in *F. trinervia*, but also in *Arabidopsis* (Engelmann et al., 2008; Wiludda et al., 2012). Using two different strategies, we try to change the expression pattern of *AtGLDP1*. (I) We aim at replacing the native *AtGLDP1*

promoter by the *FtGLDPA* promoter using a GT approach. (II) We aim at deleting the M-Box from the native *AtGLDP1* promoter, thereby abolishing expression of the *AtGLDP1* gene in M cells.

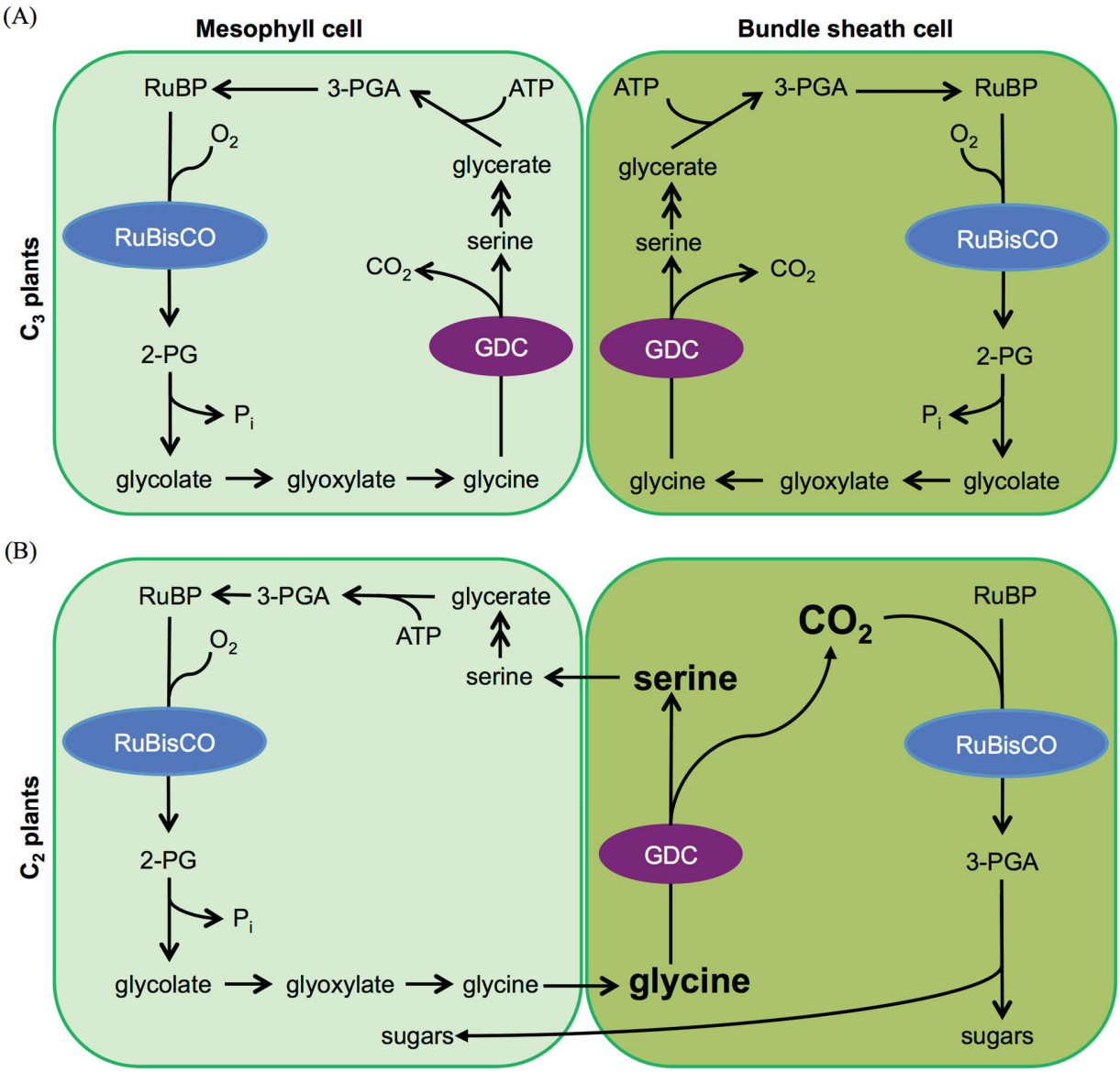


Figure 1: Loss of GDC activity in M cells creates a photorespiratory CO_2 pump. In C_3 plants, photorespiration occurs in M cells and BS cells (A). In C_2 plants, GDC activity is shifted to the BS cells (B). Photorespiratory glycine accumulates in the M cells and diffuses to the BS cells. Due to decarboxylation of high glycine amounts by GDC, CO_2 concentrations are enhanced in the BS cells, which enhances carboxylation reaction of RuBisCO.

Verification of an *Arabidopsis gldp2-2* background line

Arabidopsis contains two redundant *GLDP* genes, *AtGLDP1* (*AT4G33010*) and *AtGLDP2* (*AT2G26080*), which are ubiquitously express in all tissues (Engel et al., 2007; Adwy et al., 2015). For generating a functional photorespiratory pump, it is necessary that none of the two genes keeps its ubiquitous expression pattern. Since *AtGLDP1* plays a more important role in photorespiratory metabolism than *AtGLDP2*, we decided to use a *gldp2* T-DNA knockout line (*gldp2-2*) as background line for genomic modifications of *AtGLDP1* (Engel et al., 2007). We isolated this line by progeny segregation from a *gldp1-1*$^{\text{heterozygous}}$x*gldp2-2*$^{\text{homozygous}}$ parental plant (Engel et al., 2007). Even though we were able to isolate plants, which did not show a T-DNA specific PCR signal for the *gldp1-1* T-DNA insertion (Fig. S1), we noticed, that the isolated *gldp2-2* plants were able to grow on MS-plates supplemented with kanamycin. As this is the selection marker of the plasmid that was used to generate the *gldp1-1* T-DNA insertion line (Campisi et al., 1999), it is likely, that the *gldp1-1* line contains an additional T-DNA insertion on another genomic locus and this insertion was passed on to our isolated *gldp2-2* mutant. In the following physiological experiments, we therefore also used *Arabidopsis* wildtype (Col-0) as additional reference line.

Replacement of the native *AtGLDP1* promoter with a BS cell specific *FtGLDPA* promoter

We aimed at replacing the native *AtGLDP1* promoter in *Arabidopsis* with the *FtGLDPA* promoter using a Cas9-based GT approach. To fully remove the endogenous promoter, we designed two sgRNAs that cut shortly up- and downstream of the promoter (Fig 2A). Simultaneous DSB induction should lead to the removal of the *AtGLDP1* promoter. To integrate the *FtGLDPA* promoter into the DSB by HR, we designed a HT containing the *FtGLDPA* promoter, flanked by the parts of the *AtGLDP1* gene and the end of the upstream gene *AT4G33020* that were also excised by Cas9, and two homology arms of ~800 bp length (Fig 2A).

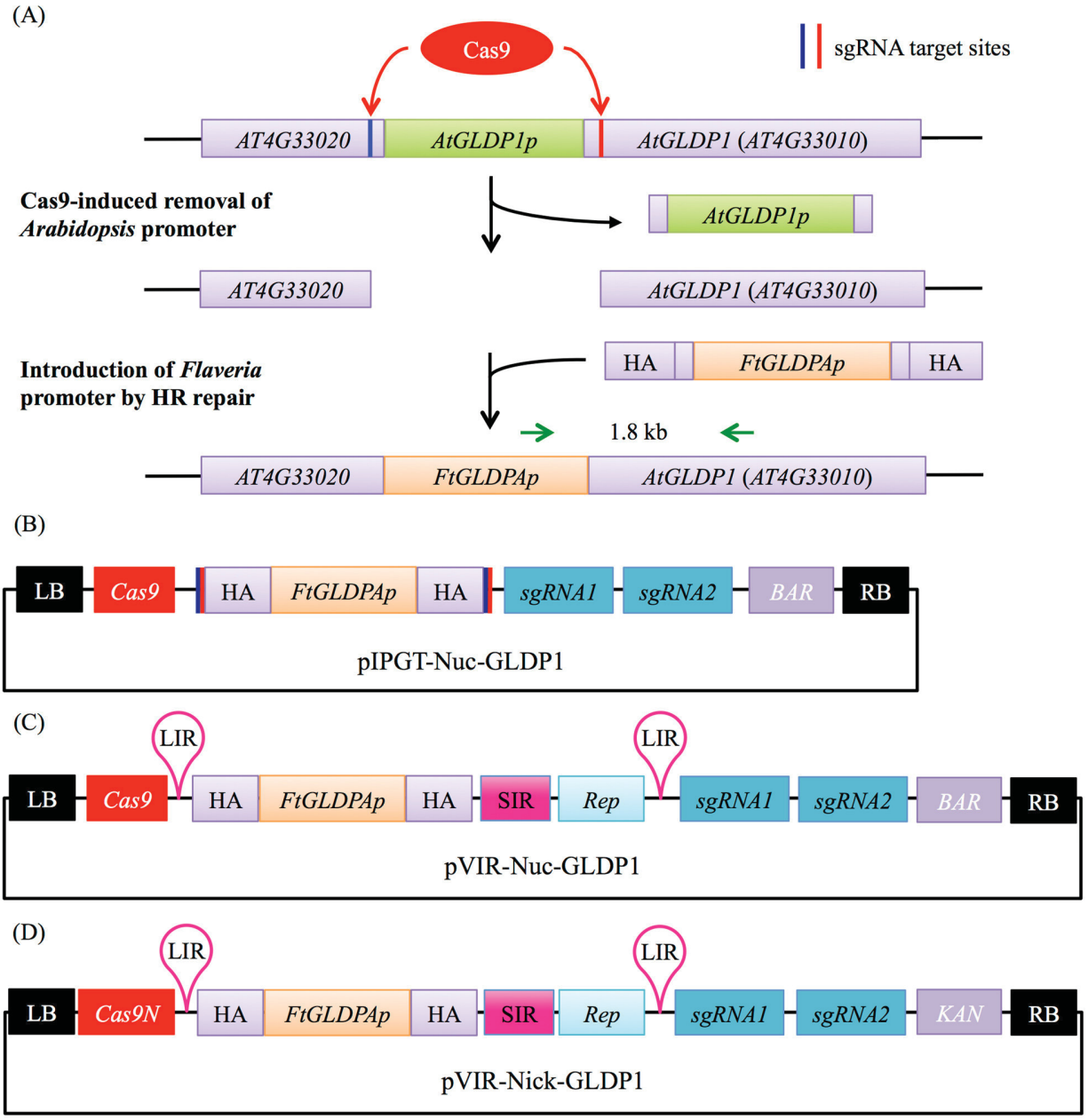


Figure 2: **Strategy for integration of a BS cell specific *Flaveria GLDPA* promoter into *Arabidopsis* by GT.** (A) Targeting of the endogenous *AtGLDP1* promoter with two sgRNAs will lead to loss of the promoter if Cas9-induced cleavage occurs simultaneously at both target sites. If a homology template including the *FtGLDPA* promoter with homology arms (HA) is offered at the time of double-strand break repair, the *FtGLDPA* promoter can be integrated by homologous recombination (HR). HR events can be detected by PCR with a *FtGLDPA* specific promoter and a primer binding on the *GLDP1* gene outside of the HA (green arrows). (B) pIPGT-Nuc-GLDP1 contains a constitutively expressed *Cas9* Nuclease gene (red box) and two sgRNA cassettes (petrol boxes) controlled by the *U6-26* promoter. The repair template consists of the *FtGLDPA* promoter (orange) flanked by HA (purple) and target sites for the Cas9 nuclease (red/blue lines). Selection of transformed plants is possible by integration of a Basta resistance cassette (*Bar*,

violet box). (C) In pVIR-Nuc-GLDP1, the repair template is flanked by the viral long intergenic regions (LIR) and the short intergenic region (SIR, pink box). A viral replicase expressed from the *Rep* gene (turquoise box) can induce rolling circle replication of the repair template. (D) pVIR-Nick-GLDP1 contains the Cas9 nickase (*Cas9N*) instead of the nuclease and a kanamycin resistance (*KAN*) cassette as selection marker. RB = right T-DNA border, LB = left T-DNA border. Size not to scale.

GT frequencies can be enhanced by mobilization of the HT (Fig. S2). This can be achieved by amplification of the HT as viral replicons (Baltes et al., 2014) or by release of the HT from the T-DNA backbone through Cas9 by addition of sgRNA recognition sites (Schiml et al., 2014). We therefore designed the T-DNA binary vector pIPGT-Nuc-GLDP1, which contains the genes for *Cas9*, sgRNAs, selection marker and the HT flanked by sgRNA recognition sites (Fig. 2B). Similarly, we constructed the vector pVIR-Nuc-GLDP1 in which the HT is flanked by viral sequences (long intergenic region, LIR, and short intergenic region, SIR) and a replicase (Fig. 2C), both necessary for amplification of the HT in form of viral replicons (Baltes et al., 2014). Mutation of one cleavage domain of Cas9 turns the Cas9 nuclease into a nickase (Cong et al., 2013). DNA nicks are less prone to DSB repair by NHEJ and might therefore be favorable for HR-mediated repair approaches (Fauser et al., 2014). We therefore also generated a third construct based on the replicon approach in which the Cas9 nuclease was replaced by a Cas9 nickase (pVIR-Nick-GLDP1, Fig. 2D). We decided to transform the three constructs in *Arabidopsis* Col-0 plants (and not in *gldp2-2* background) so that possible lethal effects due to BS cell specific expression of the GDC do not interfere with the generally low GT frequencies in *Arabidopsis* (Fauser et al., 2012; Schiml et al., 2014; Zhao et al., 2016).

We isolated two T1 lines for each construct and could verify replicon formation in both pVIR-constructs (Fig. 3A). We then screened in total 190 progeny plants of each T1 line for successful GT using a primer pair binding to the *FtGLDPA* promoter and the *AtGLDP1* gene outside of the HA region (Fig 2A). We could detect PCR amplicons with a size of 1.8 kb indicating successful promoter replacement at frequencies of up to 19 % in the analyzed plants, depending on the parental line (Fig. 3B and C). Both approaches employing the Cas9 nuclease were similarly efficient, while we obtained less positive PCR amplicons using Cas9 nickase (Fig. 3C). We also observed PCR amplicons of different sizes (Fig. 3B and C), hinting at incorrect repair. We sequenced various correctly sized PCR amplicons derived from all GT approaches and found mostly perfect integration of the HR template in the DSB (Fig. 3D). As our Cas9 system is controlled by a constitutively active promoter, mutations detected on leaf DNA level can be the

result of chimerism. As we aimed at generating stable mutant lines, we screened all plants, which showed a positive PCR amplicon a second time using a different leaf. By this, we narrowed down the number of candidate plants to 12 (pIPGT-Nuc-GLDP1), 8 (pVIR-Nuc-GLDP1) and 2 (pVIR-Nick-GLDP1), respectively.

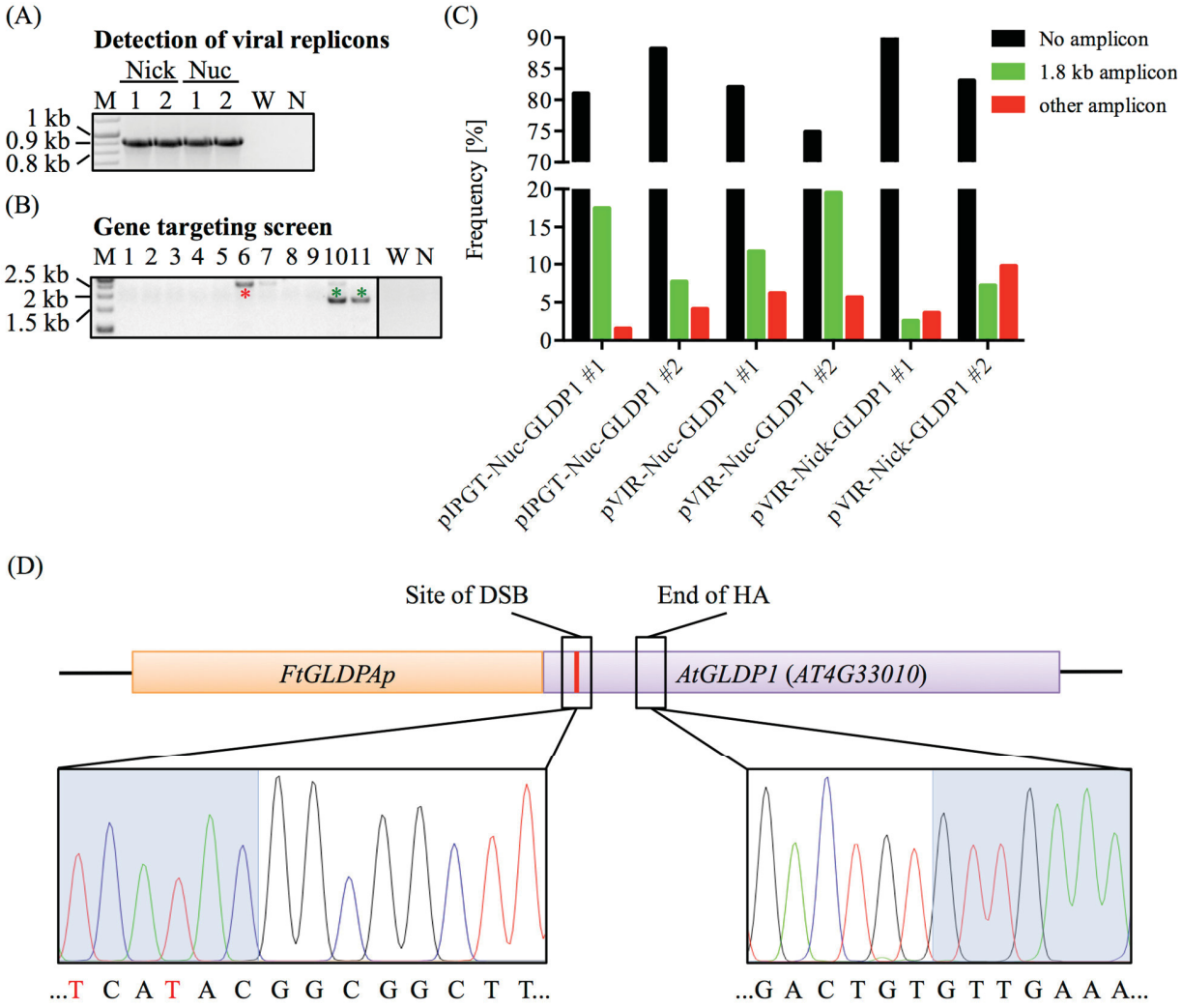


Figure 3: Detection of GT events in stably transformed *Arabidopsis* plants. (A) Detection of viral replicons in T1 generation plants transformed with pVIR-Nuc-GLDP1 and pVIR-Nick-GLDP1 by specific primers that generate only PCR amplicons (919 bp) on circularized T-DNA (W = Col-0 gDNA, N = water control). Image of agarose gel was color inverted for better visibility. (B) Exemplary PCR screening results of 11 plants (T2 generation) transformed with pIPGT-Nuc-GLDP1. PCR was performed with a *FtGLDPA* promoter specific primer and an *AtGLDP1* gene specific primer. PCR amplicons (green asterisks) at 1.8 kb indicate correct GT, PCR amplicons at different sizes (red asterisk) might be the result of incorrect repair pathways. Image of agarose gel was color inverted for better visibility. (C) Quantification of GT events by PCR screening of 190 plants (T2 generation) each, derived from 2 T1 lines (#1/#2) per GT-construct. (D) Exemplary sequencing result of a 1.8 kb PCR amplicon reveals seamless

integration of the *FtGLDPA* promoter. The left box shows the sequencing histogram at the site of the double-strand break, shaded area marks the integrated promoter (red letters indicate silent mutations introduced in the HT to remove sgRNA recognition sites), right box shows the transition between the end of the HA region and the continuation in the *AtGLDP1* gene (shaded). Peak colors: Adenine = green, cytosine = blue, guanine = black, thymine = red.

We screened ~35 progeny T3 plants of each candidate plant by PCR on two independent leaves for presence of a PCR amplicon at the correct size as described before. However, we could not reproduce the positive PCR amplicons from the T2 parental plants on two leaves of the T3 generation plants (not shown), hinting at the presence of somatic mutations only. Eventually, we were not able to generate plants with a stably integrated *FtGLDPA* promoter.

Deletion of an M cell specific element in the *AtGLDP1* promoter

As we could not recover stably mutated plants using a GT approach, we aimed at inducing BS cell specific expression of *AtGLDP1* without relying on homology directed repair mechanisms. Two promoter regions guide expression of *AtGLDP1* in photosynthetic tissues. While the V-Box (-313 to -412) directs expression in BS cells and veins, the M-Box (-700 to -758) guides expression in M cells (Adwy et al., 2015). In an alternative Cas9-approach, we aimed at deleting the M-Box motif from the endogenous promoter by simultaneous usage of two sgRNAs. Residual expression guided by the V-Box should then lead to a BS cell specific expression pattern (Fig. 4A).

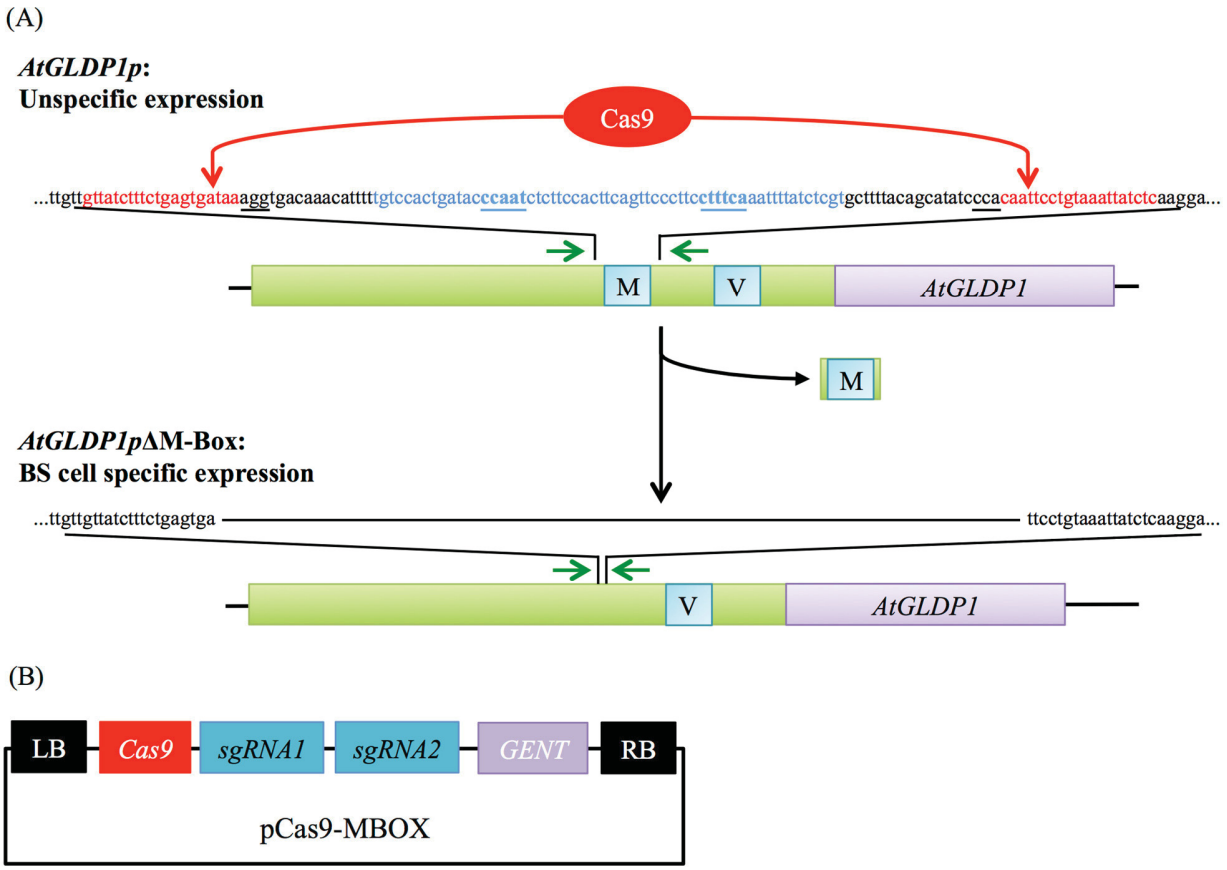


Figure 4: Strategy for deletion of an M cell specific *AtGLDP1* promoter element by CRISPR/Cas9. (A) *Arabidopsis* contains two tissue specific promoter elements in the *AtGLDP1* promoter that guide expression in M cells and in vasculature cells/BS cells, namely the 59 bp long M-Box (M, sequence given in blue, putative transcription factor binding sites (Adwy et al., 2015) marked in bold/underlined) and the V-BOX (V). Simultaneous cutting up- and downstream of the M-Box with a Cas9 nuclease (target sequences marked in red, PAM motives underlined) releases the M-Box and should result in BS cell specific expression of *AtGLDP1*. Deletion events can be monitored by PCR with primers (green arrows) spanning the M-Box. (B) The deletion construct pCas9-MBOX contains the Cas9 nuclease, two sgRNA cassettes and a gentamycin resistance cassette (*GENT*).

Thus, we generated the binary T-DNA vector pCas9-MBOX containing a constitutively expressed *Cas9* nuclease gene, two sgRNA cassettes and a selection marker (Fig. 4B). We transformed *Arabidopsis gldp2-2* plants with pCas9-MBOX and recovered three independent T1 lines. We screened 190 to 285 T2 plants per line grown under high CO_2 conditions by PCR using primers that amplify the M-Box. Deletion of the M-Box should be visible as reduction in PCR amplicon size. As a result, we detected two amplicons of different sizes in plant #31 (Fig. 5A), hinting at a deletion on one allele. Sequencing of both amplicons revealed, that the WT-sized

amplicon contained single-bp-insertions at each sgRNA sites resulting from NHEJ mediated error prone repair (Fig. 5B), which should not influence spatial expression pattern. Sequencing of the smaller amplicon demonstrated deletion of about 2/3 of the M-Box (Fig. 5B).

To ensure that the mutation was fixed within the genome of the plant, we analyzed 126 T3 progeny plants of plant #31 by PCR. Indeed, we detected plants homozygous for the deletion, biallelic plants, and plants that showed only the allele with the single-bp-insertions (Fig 5C). The genotype ratio did not deviate significantly from the expected Mendelian segregation ratio of 1:2:1 as calculated by chi-squared test ($P > 0.1$, Table S2).

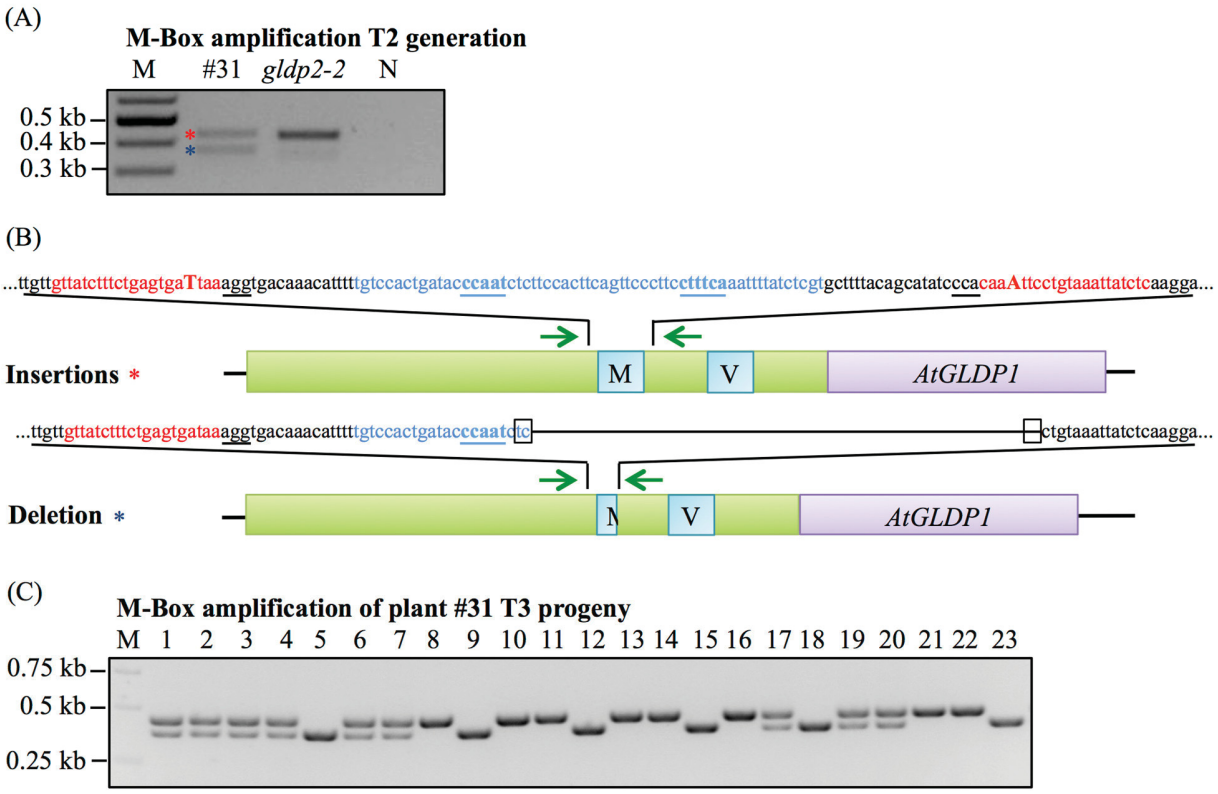


Figure 5: Screening for M-Box deletion in *Arabidopsis* plants transformed with pCas9-MBOX. (A) PCR-based screening of 685 plants (T2 generation) revealed a single plant (#31) showing two PCR amplicons, one (red asterisk) corresponding to the size of an unmodified promoter (441 bp, as shown in *gldp2-2* control plant) and one smaller amplicon (green asterisk, expected size for removal of full M-Box: 341 bp). Image of agarose gel was color inverted for better visibility. N = water control. (B) Sequencing of both PCR amplicons revealed two insertions at both sgRNA target sites on one allele and a deletion of a large part of the M-Box on the other allele. Microhomologies (black boxes) were detected at the borders of the large deletion. (C) Representative image of PCR screening of progeny plants of plant #31. Image of agarose gel was color inverted for better visibility

We also determined if the Cas9-T-DNA was still present in the T3 plants by PCR amplification of the *Cas9* gene. In all tested plants, we could detect a *Cas9* specific PCR amplicon (Fig. S3). This suggests that the *Cas9* T-DNA cassette was already homozygously inserted in the parental plant #31 and is not lost by segregation anymore. Seeds were harvested from both genotypes, in the following called "Insertions" and "Deletion" plants, for further experiments.

Phenotypical analysis of mutant plants

Knockout mutants in the core photorespiratory pathway often show a classical photorespiratory phenotype that is bleached seedlings followed by lethality when grown under ambient CO_2 conditions. Typically the phenotype can be fully or partially rescued by increasing the CO_2 concentration (Timm and Bauwe, 2013). Full knockouts of the GDC components *GLDP* and *GLDT* have been reported to be lethal for the plant even under high CO_2 conditions (Engel et al., 2007; Timm et al., 2017). We therefore compared rosettes of 25 days old "Deletion" and "Insertions" plants grown under ambient CO_2 conditions with *gldp2-2* and Col-0 plants. The mutant plants did not show differences in leaf appearance (Fig. S4A) and we could not detect significant differences in neither rosette diameter (Fig. S4B) nor fresh weight (Fig. S4C). All plants produced normal flowers (Fig. S4D) and viable seeds. However, the "Deletion" and the "Insertions" plants showed a significant delay in development as shown by stem height measurements of 67 days old plants (Fig. 6). As the plants were perfectly viable under ambient CO_2 conditions, all following experiments were conducted with plants grown under ambient CO_2 conditions.

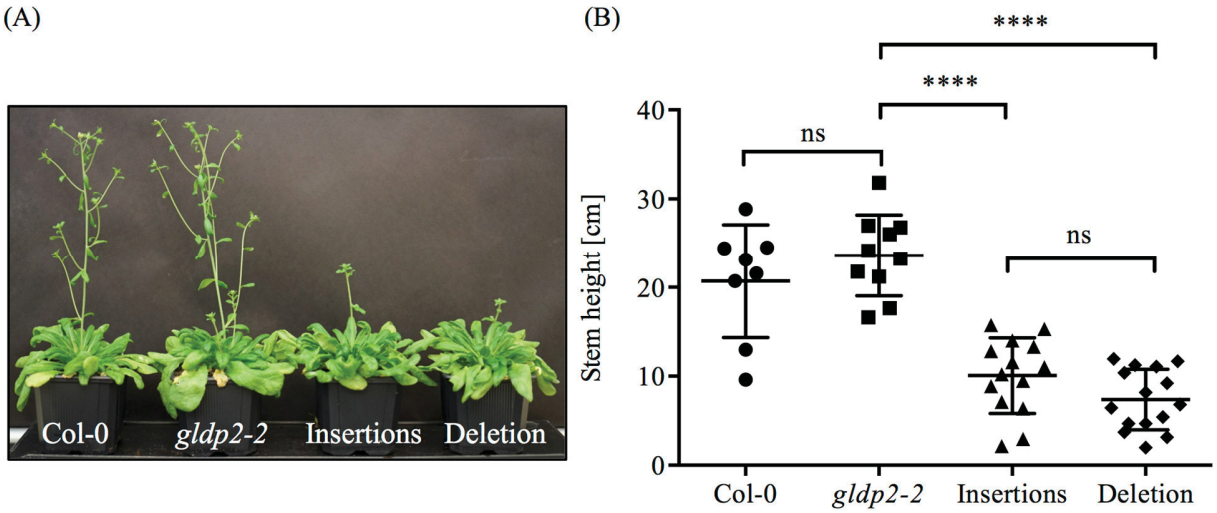


Figure 6: Mutants with modifications in the *AtGLDP1* promoter are delayed in development. (A) Representative picture of a 67 days old WT (Col-0) plant, *gldp2-2* mutant plant, a plant with single-bp-insertions in the sgRNA target sites (Insertions) in the *AtGLDP1* promoter and a plant with large deletion in the M-box (Deletion) grown at ambient CO_2. (B) Determination of stem height of mutant plants in comparison to WT and *gldp2-2* plants. Significant differences of mutant plants compared to the *gldp2-2* background line were determined with a two-tailed Student's *t*-test and are indicated by asterisks (ns: $P > 0.05$; ****: $P < 0.0001$, n = 8-15). No significant (ns) differences were found between Col-0 and *gldp2-2* plants and the Insertions and Deletions line, respectively.

Protein localization was unsuccessful

Full deletion of the 3'-*AtGLDP1* promoter region including the M-Box leads to a loss of expression in M cells as shown by GUS fusion constructs (Adwy et al., 2015). We generated plants with a deletion of large part of the M-Box within the *AtGLDP1* promoter (Fig. 5). To assess the expression pattern of this mutated promoter we generated a GUS fusion construct pGUS-GLDPp-DELETION (Fig. S5A) mimicking the partial deletion of the M-Box in the mutated plants and transformed it into Col-0 plants. Plants transformed with GUS-fusion constructs with the full promoter or with a shortened promoter with a full deletion of the 5'-part of the promoter region including the M-Box (derived from Adwy et al., 2015) were used as controls for expression analysis. Independent T1 plants with 4 to 6 leaves were used for GUS-staining. We could not detect a consistent staining pattern independent of the transformed construct (Fig S5B). To assess AtGLDP1 localization *in planta*, we fixed and embedded leaf tissue of 36 days old plants and stained leaf sections with a GLDP specific antibody followed by immunogold labeling. However, in TEM analysis of immunogold-stained leaf samples, no

immunogold particles could be detected in any sample (data not shown). We could therefore not conclude whether the changes in the *AtGLDP1* promoter structure indeed affected the expression pattern.

Plants with a deletion in the M-Box show slight decrease in *AtGLDP1* transcript and protein amount

We examined if the partial deletion of the M-Box leads to a change of *AtGLDP1* transcript level by measuring *AtGLDP1* transcript in full leaves of 24 days old plant using qRT-PCR. Both promoter modifications led to a decrease between the averages of three replicates of 57-69 % compared to *gldp2-2* (Fig. 7A). However, this difference was statistically not significant (Student's *t*-test, $P > 0.05$), probably due to the unexpectedly high variation between replicates in the transcript levels in *gldp2-2*. Using the GLDP specific antibody, we assessed protein levels in leaf protein extracts of 42 days old plants grown under ambient CO_2 conditions, (Fig. 7B). Strong signals were observed in all genotypes, with a slightly less intense signal in plants with a promoter deletion.

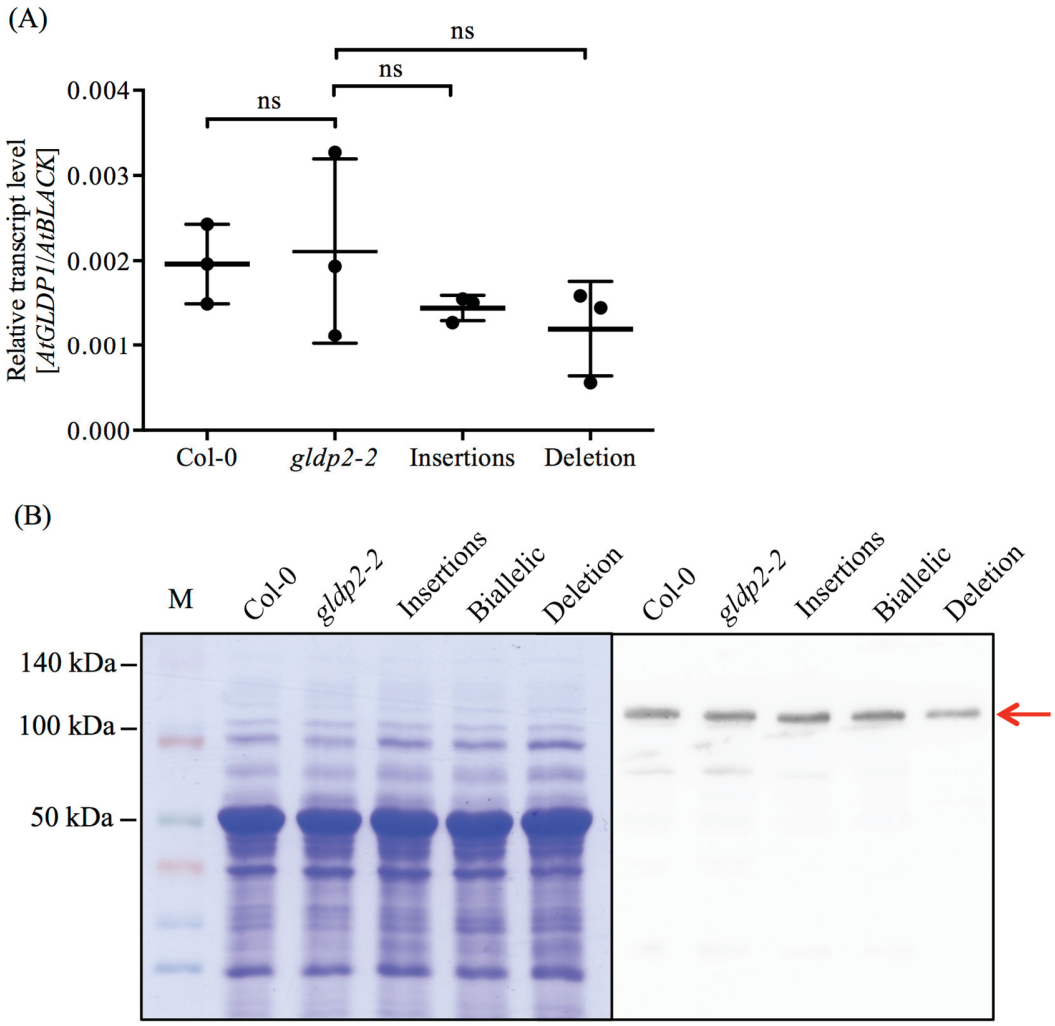


Figure 7: Effects of promoter modifications on *AtGLDP1* transcript levels and protein content. (A) Leaf *AtGLDP1* transcript levels were measured in plants grown under ambient CO_2 conditions (B) for 24 days. Samples were taken 8 hours after onset of light. Shown is the mean normalized expression of three biological replicates with standard deviation. Each value represents the average of three technical replicates. *AtBLACK* was used as reference gene (Czechowski et al., 2005). Significant differences of *AtGLDP1* transcript levels in mutant plants compared to *gldp2-2* plants were analyzed by two-tailed Student's *t*-test (ns: $P > 0.05$). (B) GLDP protein content in leaf extracts grown under ambient CO_2 was assessed by Western Blotting using a GLDP specific antibody. Coomassie stained loading control (left) shows equal loading. Western Blotting reveals a band at the expected size of the AtGLDP1 protein (ca. 106 kDa). A weaker signal can be detected in plants with a homozygous M-Box deletion.

Changes in the *AtGLDP1* promoter do not accelerate photoinhibition of photosystem II

Photoinhibition is often accelerated in photorespiratory mutants (Badger et al., 2009; Timm et al., 2012) due to suppression of PS II repair (Takahashi et al., 2007). We therefore assessed

maximum efficiency of PS II (F_v / F_m) of 18 days old plants. We could not detect differences between Col-0, *gldp2-2*, and plants with *AtGLDP* promoter modifications under the tested conditions (Fig. S8).

Mutant plants accumulate glycine under ambient CO_2 conditions

C_3 - C_4 intermediate plants typically accumulate glycine in the leaf under ambient CO_2 conditions due to missing GDC-activity in M cells, which ultimately allows passive diffusion of glycine from M cells to BS cells. Similarly, serine levels are enhanced compared to C_3 plants (Rawsthorne and Hylton, 1991; Schlüter et al., 2017). Thus, we assessed the glycine and serine levels in leaves of plants grown under ambient CO_2 (Fig. 8A, 8B).

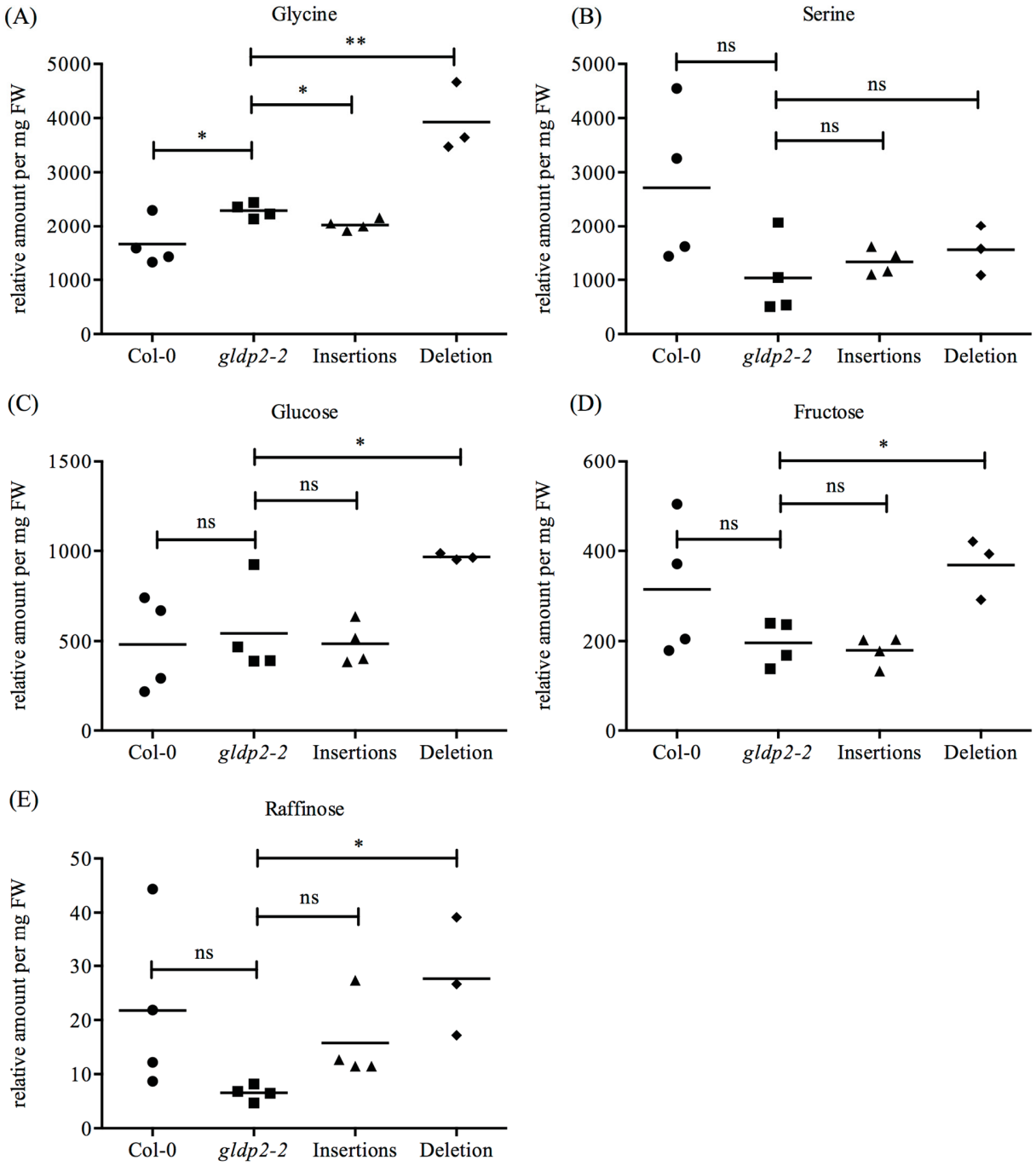


Figure 8: Metabolite levels in plants with a partial M-Box deletion under ambient CO_2 conditions. Metabolite levels of glycine (A), serine (B), glucose (C), fructose (D) and raffinose (E) were measured in rosettes of 24 days old plants grown under ambient CO_2 conditions. Significant differences of metabolite levels of plants compared to *gldp2-2* plants were analyzed by two-tailed Student's *t*-test (ns: $P > 0.05$; *: $P < 0.05$; ** $P < 0.01$).

Indeed, "Deletion" plants accumulated 1.71-fold times more glycine under ambient CO_2 conditions (Fig. 8A) than *gldp2-2* plants ($P = 0.003$), while glycine content in "Insertions" plants

or Col-0 plants was not increased. The serine content (Fig. 8B) was also increased in plants with an M-Box deletion, however, similar or even higher amounts were found in the other genotypes and the changes were not significant. Several metabolites have been proposed to rebalance N content between M cells and BS cells in C_2 plants (Mallmann et al., 2014). However, no significant increases can be detected in the putative nitrogen (N) shuttle metabolites malate, aspartate, alanine, glutamate and α-ketoglutarate under ambient CO_2 conditions (Table S3). Soluble sugars (fructose, glucose, and raffinose) stayed mostly constant in "Insertions" plants but accumulated slightly in the "Deletion" plants (Fig. 8C - E). Otherwise, no consistent changes specific for the plants with the partial M-Box deletion were detected in most other metabolites (Table S3), including no changes in the plant stress indicator γ-aminobutyric acid (GABA). We performed a PCA of all metabolites to verify that the "Deletion" genotype leads to a unique metabolite pattern (Fig. S7). The first principal component (explaining 33% of variation) separated the plants in the *gldp2-2* background from Col-0, the second principal component (explaining 18% of variation) separated the "Deletion" plants from the remaining genotypes.

DISCUSSION

The integration of a C_2 cycle in a C_3 plant is a long-term goal of plant scientists

In this study, we aimed at restricting expression of the *Arabidopsis* At*GLDP1* gene to BS cells, which is an elementary step towards the evolution of C_4 photosynthesis (Heckmann et al., 2013). Three studies already describe similar approaches. In 2010, the group of Hermann Bauwe tried to establish a C_2 shuttle in *Arabidopsis* by integrating the *FaGLDPA* gene from the C_3 - C_4 intermediate plant *F. anomala* under the control of the BS cell specific *FaGLDPA* promoter in the background of a *gldp1*/*gldp2* double-knockout (Engel, 2010). While the lethal phenotype of the double-knockout mutant could be rescued, the mutant "C_2" *Arabidopsis* plants were only able to grow under non-photorespiratory conditions and were sterile. This approach had several drawbacks, caused in part by T-DNA position effects and mostly by use of a foreign *GLDP* gene, as it is unknown if the FaGLDP protein can interact with the other components of the GDC from *Arabidopsis*. Another study in potato aimed to suppress *GLDP* expression in M cells by an antisense construct using a promoter that drives expression preferentially in M cell. These "C_2" potato plants did not show a strong photorespiratory phenotype but accumulated glycine to high levels (Winzer et al., 2001). Similarly, a recent study in rice tried to suppress *GLDH* expression

in M cells by an artificial micro RNA (amiRNA) approach directed against the transcripts of the photorespiratory isoform of the *GLDH* gene (*Os10g37180*). These "C_2" rice plants showed a mild photorespiratory phenotype but could survive under ambient air (Lin et al., 2016). However, amiRNA approaches are prone to varying efficiencies of gene expression knockdown and do not result in a complete loss of transcript. Additionally, the other GLDH isoforms might complement the function of the targeted GLDH photorespiratory isoform, as shown for the two AtGLDP isoforms (Engel et al., 2007). The approaches used in this study employ the advantages of the Cas9 system to introduce targeted changes in the original genomic context.

Enhancement of GT frequencies is crucial for regeneration of stable mutant lines

In a first approach, we tried to remove the unspecific *AtGLDP1* promoter by simultaneous cleavage with Cas9 at two target sites. By offering a repair template containing the BS cell specific *GLDPA* promoter from *F. trinervia*, we aimed at introducing the *FtGLDPA* promoter upstream of the *AtGLDP1* gene via HR-mediated repair of the induced DSB. Targeted integration of a transgene in the plant genome via HR is also called GT (Paszkowski et al., 1988) and has been a long-time goal for plant scientists (Puchta, 2002). However, higher plants usually integrate transgenes via NHEJ at random positions in the genome (Puchta, 2005). GT efficiencies are enhanced by priming of the target locus via introduction of a DSB (Puchta et al., 1996) or single strand break (SSB, Fauser et al., 2014), as it is possible by the Cas9-system. Recent reports in various monocotyledons and dicotyledons have shown that GT frequencies can be further enhanced 10-100-fold by mobilization of the HT in form of geminiviral replicons (Baltes et al., 2014; Čermák et al., 2015; Gil-Humanes et al., 2017). For this approach, the HT is flanked with three viral elements derived from geminiviruses as the bean yellow dwarf virus. These elements are the *cis*-acting LIR, the SIR and the *Rep* gene, which encodes for two splice forms of a *trans*-acting viral replicase, namely Rep and RepA (Fig. S2A). Rep initiates rolling circle endoreplication at the LIR site, thereby leading to a replicational release of the HT in form of up to several thousand circular replicons which can be used as repair templates for DSB repair (Chen et al., 2011; Baltes et al., 2014). However, reports of generating stable GT-lines in *Arabidopsis* using viral replicons are lacking. Alternatively, Fauser and colleagues (2012) developed the IPGT approach, in which the template is mobilized by excision of the T-DNA backbone in the plant nucleus by SSN and thus available as repair template for DSB repair (Fig. S2B). In *Arabidopsis*, recovery of stably mutated plants at frequencies of 0.14-0.83 % have been reported using IPGT

(Fauser et al., 2012; Schiml et al., 2014; Zhao et al., 2016). We used a PCR-based screening to identify GT events in leaves of T2 generation plants (Fig. 3B) and could detect GT events in up to 19.5 % of all tested plants (Fig. 3C). The primer combination that we used for screening covers only one of the two induced DSB sites. It cannot be excluded that we therefore counted one-sided GT events as positive events and that the actual number of true GT events is lower. However, IPGT and viral replicons seem to be able to induce GT in *Arabidopsis* on somatic level. While GT frequencies were similar independent of the HT mobilization strategy, clear differences were detected between independent transformant lines (Fig. 3D). Previous reports state a connection between GT efficiencies and the proximity of T-DNA insertion and DSB locus (Fauser et al., 2012). It might therefore be advisable to screen progeny of more independent transformant lines (Schiml et al., 2017). We also used a Cas9 nickase in combination with the viral replicon approach as DNA nicks are less prone to NHEJ-mediated repair and are more likely repaired via HR (Fauser et al., 2014). However, the GT frequencies were lower compared to the nuclease constructs. The nickase approach might not be favorable for our experimental approach as deletion of the *AtGLDP1* promoter (which is the prerequisite for repair by GT) is only efficient if DSB are induced. While we could detect GT events on somatic level in all approaches, we were not able to regenerate fully mutated plants. Previous proof-of-concept studies usually relied on efficient marker systems as GUS or fluorescent proteins to detect GT events (Baltes et al., 2014; Fauser et al., 2012; Schiml et al., 2014; Zhao et al., 2016), which allowed simple selection of positive GT events. The PCR-based screening approach, which we employed due to the lack of a traceable phenotype, is more laborious and its high sensitivity (one cell with successful GT within the leaf from which gDNA was isolated might be enough for a positive PCR amplicon) makes it difficult to distinguish single cell GT events from stably mutated plants. We conclude that even though GT in *Arabidopsis* provides an interesting option for highly specific genomic changes, the available GT systems are yet not efficient enough to generate stable mutants without phenotypes. Optimization of the GT system is possible in several ways. (I) Using of germline specific promoters for Cas9 expression (Wang et al., 2015; Mao et al., 2016) results in loss of somatic mutations. Detection of positive PCR amplicons on leaf level hints therefore at completely mutated plants. (II) Optimization of the PCR screening approach, e.g. by automatization of gDNA isolation and PCR pipetting allows screening of more plants and might enable to quantify PCR amplicon signal strength as a direct readout for the number of mutated cells. (III) Optimization of the HT delivery system e.g. by combination of the IPGT approach and

the viral replicon approach. By this, linear template DNA fragments could be excised by Cas9 from replicated HR template molecules, which might be favored by the HR repair machinery as they resemble natural occurring HR templates.

Only parts of the M-box were deleted by simultaneous cutting with two sgRNAs

Cis-regulatory elements that guide expression in M cells or repress expression in BS cells are an attractive target for genome editing. Mesophyll expression motives have been found in *Flaveria* species but also in *Gynandropsis gynandra* and *Arabidopsis*, amongst others. Distinct changes within these mesophyll expression motives explain differential expressions in C_3 , C_2 and C_4 species (Gowik et al., 2004; Akyildiz et al., 2007; Williams et al., 2016; Gowik et al., 2017). *AtGLDP1* expression in M cells of *Arabidopsis* is controlled by a *cis*-regulatory element in the *AtGLDP1* promoter, called M-Box, while a second *cis*-regulatory element provides expression in BS cells and vasculature (Adwy et al., 2015). Loss of the M-Box should therefore lead to a BS cell specific expression of *AtGLDP1* in leaves. Using a multiplexing Cas9 approach with two sgRNAs, we aimed at deleting the M-Box in the *AtGLDP1* promoter by simultaneous induction of two DSBs in plants with a *gldp2* knockout background (Fig. 4A). Previous studies in *Arabidopsis* reported frequencies of 7.7 to 26 % of deletion events using two sgRNAs (Li et al., 2013; Mao et al., 2013; Zhao et al., 2016). However, screening of more than 600 T2 generation plants from three independent transformant lines yielded only a single plant in which large parts of the M-Box were deleted, albeit not the full region between the two sgRNA sites (Fig. 5B). We cannot exclude that a full deletion of the M-Box is lethal for the plant because they cannot tolerate GDC restriction to the BS cells. However, we also did not detect heterozygous mutations, which should result in viable mutant plants. Alternatively, transformation of pCas9-MBOX in Col-0 plants should allow complete deletion of the M-Box as a functional *AtGLDP2* copy is still available. Crossing of this plant with the *gldp2-2* line would then allow conclusion on a putative lethality effect of the mutation. For a full deletion of the M-Box, both sgRNA target sites must be targeted simultaneously. Timing of the expression of both sgRNA cassettes could improve the deletion efficiencies. This can be achieved by single transcript units, which are then processed by ribozymes or endogenous tRNA machinery (Xie et al., 2015; Tang et al., 2016). Finally, using more independent transformant lines is advisable, as T1 plants can already accumulate mutations at both sgRNA sites. By this, they are resistant against further multiplex cleavage and large deletions cannot occur anymore. However, the mutated plant that we generated contains only a

large 63 bp deletion resulting from cleavage at the downstream sgRNA target site while the upstream target site is unaffected on the same allele (Fig. 5B), hinting that, at least in progeny plants of this transformant line, unmutated target sites were still present in the T2 generation. Classical NHEJ repair usually results in small indel mutations of few bp (Fauser et al., 2014). Larger deletions might result from an alternative NHEJ pathway (microhomology-mediated end joining, MMEJ). MMEJ involves alignment of short microhomologies from resected DSB end leading to large deletions if the ends are rejoined (Sfeir and Symington, 2015; Wang and Xu, 2017). Indeed, a two nucleotides microhomology, namely TC, is present at the borders of the large deletion (Fig. 5B). Similar short microhomologies of 2-4 nucleotides were sufficient to induce large deletions by MMEJ in Cas9 experiments in *Physcomitrella patens* (Collonnier et al., 2017).

Direct evidence for changes in localization patterns are missing

As we could not delete the full M-Box, determination of the expression pattern of our modified promoters is crucial. However, we were not able to access the expression pattern with GUS staining (Fig. S5) or immunogold labeling. The M-Box itself was detected by promoter-deletion studies and we used a promoter::GUS construct with a highly shortened promoter as control, which directed BS cell specificity in previous studies (Adwy et al., 2015). However, we could also not verify the BS cell specificity of this control construct. The most likely reason is that we used a different vector backbone (pCambia 1381) to guide GUS expression. This vector contains a 35S promoter directly upstream of the promoter::GUS fusions for expression of a resistance cassette. This strong promoter might create considerable transcriptional activity also in adjacent open reading frames as noted by the providing company (http://www.cambia.org/daisy/cambia/1209.html). Cloning in another vector backbone is therefore necessary for more reliable GUS staining. Alternatively, the resistance cassette can be removed and a selection marker can be introduced by co-transformation of a second plasmid. The immunogold labeling protocol should be improved as well. As the primary antibody worked well in Western Blots (Fig. 7B), functionality of the secondary immunogold antibody should be verified with suitable positive controls. In theory, separation of M cells and BS cells and subsequent analysis of AtGLDP content via Western Blotting is also an option for localization of the protein. While enrichment of BS cells is possible in C_4 plants (Denton et al., 2017), no efficient protocols have been published to our knowledge for C_3 plants due to their lower amount

of BS cells. Alternatively, fluorescence *in situ* staining or RNA *in situ* hybridization could be used to assess AtGLDP localization, as done in previous studies (Engelmann et al., 2008; Engel, 2010).

Even though we could not verify a different localization, we still think that the large deletion that we created could be sufficient for redirecting the tissue specificity. Adwy et al. (2015) hypothesized that two motives within the M-Box (or the transcription factors binding to them) interact to induce M cell expression. These motives are a NUCLEAR FACTOR Y binding motif (CCAAT) and a CTTTCA motif, which were both found in a screen for M cell specific *cis*-elements in maize (Wang et al., 2014). Indeed, alignment of the *AtGLDP1* M-Box sequence with M-Box sequences from other C_3 *Brassicaceae* shows that both motives are highly conserved (Fig. 9). In our promoter deletion, we could delete the CTTTCA motive. If the interaction between the two motives is therefore disturbed, the remaining CCAAT motif alone should not be functional anymore. Recent unpublished GUS-staining data (Waly Adwy) indeed reveal, that deletion of one of the two motives is sufficient to abolish M cell specific expression. In contrast, the plants with the small insertions at the sgRNA sites are not mutated in any of these motives and should not be affected in their tissue specificity.

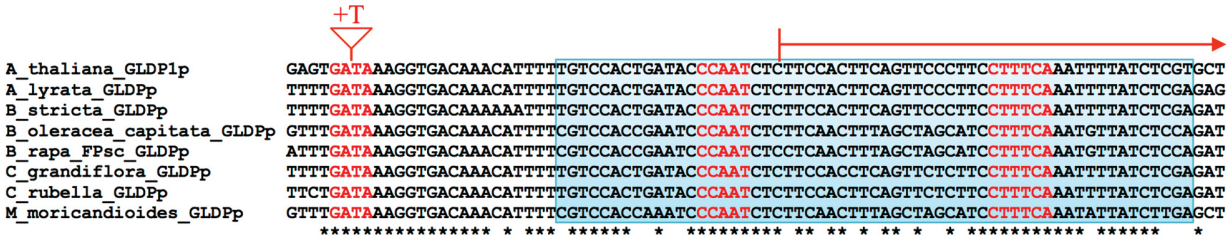


Figure 9: Alignment of the M-Box sequences in *GLDP* promoters from various C_3 *Brassicaceae* species. Annotated M-Box (Adwy et al., 2015) is highlighted by a blue box. Putative binding motives for transcription factors are marked in red. The upstream nucleotide insertion in the "Insertions" plants is marked with a red triangle. The downstream nucleotide insertion, which is not conserved, is not shown here. The part of the M-Box that was deleted by Cas9 in the "Deletion" plants is marked with a red arrow. Asterisks indicate identical nucleotides in all analyzed species (A_thaliana: *Arabidopsis thaliana*, A_lyrata: *Arabidopsis lyrata*, B_stricta: *Boechera stricta*, B_oleracea: *Brassica oleracea capitate*, B_rapa: *Brassica rapa FPsc*, C_grandiflora: *Capsella grandiflora*, C_rubella: *Capsella rubella*, M_moricandioides: *Moricandia moricandioides*).

Decrease in GLDP protein content might explain the developmental phenotype

Previous approaches to introduce a C_2 shuttle in potato (Winzer et al., 2001), *Arabidopsis* (Engel, 2010) or rice (Lin et al., 2016) led to plants with photorespiratory phenotypes in varying degrees. The previously reported "C_2" *Arabidopsis* plants showed drastic phenotypes with bleached leaves and morphologically distinct, sterile flowers, even if grown under high CO_2 conditions (Engel, 2010). In contrast, our mutated plants did not show any chlorotic spots and were able to develop well under ambient CO_2 conditions (Fig. S4). Consequently, the plant stress hormone GABA (Kinnersley and Turano, 2000) content was not increased (Table S3) opposed to other photorespiratory mutants (Eisenhut et al., 2013; Nunes-Nesi et al., 2014). However, we saw a delay in development in older mutant plants, as shown in stem lengths measurements (Fig. 6). Similarly, the "C_2" rice plants were significantly shorter and produced less biomass (Lin et al., 2016). Meristem tissue needs high amounts of folate to fuel the high requirement of cell development, e.g. nucleotide synthesis (Jabrin et al., 2003). A disturbance in C_1 metabolism by varying amounts of GLDP might therefore explain delays in development. However, this disturbance would probably already influence plant development at earlier stages, as reported for the previously generated "C_2" *Arabidopsis* plants, which were delayed in development already at the cotyledon stage (Engel, 2010). Further disruption of developmental or signaling pathways (maybe also caused by the pCas9-MBOX T-DNA insertion) needs to be elucidated.

Analysis of the *AtGLDP1* RNA content in whole leaves revealed decreases in transcript levels in the "Deletion" plants (Fig. 7A). If the *AtGLDP1* gene is indeed only expressed in BS cells, which represent about 15 % of chloroplast containing leaf cells in *Arabidopsis* (Kinsman and Pyke, 1998), a drop in total leaf *AtGLDP1* transcript is expected. A more in-depth study of other transcripts of the "Deletion" plants might allow more conclusions about a successful integration of a photorespiratory pump as the confinement of GDC activity to BS cells has further implications on plant gene expression. As an example, redox balance between M cells and BS cells has to be adjusted as GDC produces high amounts of NADH. In C_2 species from *Moricandia* and *Flaveria*, *ALTERNATIVE OXIDASE* transcripts are therefore increased compared to C_3 species (Schlüter et al., 2017). Interestingly, the GLDP antibody test revealed reduced amounts of AtGLDP protein in the "Deletion" plants (Fig. 7B), which corresponds to the transcript amounts. However, more detailed studies should follow to elucidate AtGLDP protein content in the mutant plants. In the "C_2" rice study, the authors did not detect any GLDH protein in plant leaves using Western Blots after targeting *GLDH* transcripts by amiRNA. Furthermore,

they detected only minimal amounts of GLDP. Similarly, only minimal GLDH amounts were detected by immunolocalization (Lin et al., 2016), which makes it questionable if a real CO_2 pump was created or just a leaf-wide knock down of GDC activity. Similar to the "Deletion" line, we saw a (non-significant) decrease in transcript amounts in the "Insertions" mutants. Analysis of the single-bp insertion sites revealed, that the upstream insertion site destroyed a highly conserved GATA motif (Fig. 9). GATA boxes are associated with light induced gene expression, which is a common feature of photorespiratory genes (Argüello-Astorga and Herrera-Estrella, 1998; Foyer et al., 2009). Mutations of the GATA motif are associated with a decrease in gene expression (Chan et al., 2001; Jeong and Shih, 2003). This might explain the lowered gene expression in the "Insertions" plants.

Glycine accumulation hints at C_2 carbon concentrating mechanism

A common feature of the so-far generated artificial "C_2 " plants is an increase in glycine content, ranging from 5- to 40-fold in potato (Winzer et al., 2001) and 3-to 33-fold in rice (Lin et al., 2016) under ambient CO_2, and 84-fold in *Arabidopsis* even under high CO_2 (Engel, 2010). In contrast, we detected only small increases up to 1.71-fold under ambient CO_2 (Fig. 8A). However, these increases were highly specific for the "Deletion" plants and could not be detected in the other genotypes. A recent comparative study between C_3 and C_2 *Moricandia* species revealed that plants with a functional C_2 shuttle accumulate glycine only 2- to 4-fold compared to the corresponding C_3 species (Schlüter et al., 2017). This shows that extreme glycine accumulations as reported in the "C_2" rice, potato and *Arabidopsis* studies are mostly hinting at a disturbed photorespiratory metabolism and not towards a functional C_2 shuttle. Metabolite measurements of plants that were shifted to non-photorespiratory high CO_2 conditions would be necessary to confirm, that the glycine accumulation in the "Deletion" plants is indeed caused by the photorespiratory flux.

Glycine accumulation is a common feature of knock-down mutants in the glycine cleavage system (Engel et al., 2011; Timm et al., 2017). However, we do not see any glycine accumulation in the "Insertions" mutants (Fig. 8A) even though they showed decreased *AtGLDP1* transcript abundance (Fig. 7A) and might therefore contain less AtGLDP1 protein. The GDC is present in large amounts in mitochondria of photosynthetic tissues (Oliver et al., 1990) to avoid any accumulation of toxic photorespiratory intermediates. This high excess could therefore provide flexibility for a slight decrease in total protein content. In an antisense approach in potato, high

decreases in GLDP content (>60 %) led to high glycine accumulation (Heineke et al., 2001). Similar results have been obtained for SHMT knockdown potato plants (Schjoerring et al., 2006). Here, we did not achieve such high decreases. Increasing photorespiratory conditions by high light and increased temperatures might allow enforcing a more drastic phenotype in the "Insertions" plants. The "C_2 " potato study (Winzer et al., 2001) and the "C_2 " rice study (Lin et al., 2016) reported a decrease in soluble sugars under ambient CO_2. This is a sign for a reduced CO_2 assimilation rate due to inhibition of the Calvin-Benson cycle by disturbed photorespiration. In contrast, we could detect constant or increased sugar contents in the "Deletion" plants (Fig. 8), which hints that sugar production by photosynthesis is not impaired. An increase in soluble sugars was also detected in *GLDL* overexpression lines, which resulted from a partial reprogramming of the plant starch metabolism (Timm et al., 2015). Analysis of starch content in the mutant plants will allow us to verify the partitioning of fixed carbohydrates into starch. PCA analysis of our metabolite data showed a separation of the "Deletion" replicates from the other genotypes (Fig. S7). This result is important as it provides additional evidence, that the "Deletion" plants behave differently and with a slightly stronger effect that the "Insertions" plants.

CO_2 compensation point measurements can verify an efficient C_2 shuttle

Modifications of the photorespiratory pathway are often connected to changes in photosynthetic parameters (Timm et al., 2012). A disrupted photorespiratory pathway leads to interruption of the Calvin-Benson cycle by the inhibitory effects of accumulated photorespiratory intermediates and depletion of Calvin-Benson cycle intermediates. If the Calvin-Benson cycle is impaired, recovery of electron acceptor molecules for the light reaction is hindered. Electrons from the water-splitting reaction of PS II are therefore transferred to oxygen in PS I, which can result in H_2O_2 production. H_2O_2 inhibits synthesis of the D1 protein, an essential part of PS II (Takahashi et al., 2007). The unremarkable F_v / F_m-values of our mutants (Fig. S6) indicate, that the mutations in the *AtGLDP1* promoter do not impair PS II repair, which corresponds to the metabolic and physiological data.

However, a more detailed analysis of photosynthetic parameters of the *AtGLDP1* mutants is indicated. In fact, one of the most striking differences between C_2 and C_3 plants is a lowered CO_2 compensation point in C_2 plants (Krenzer et al., 1975; Ku et al., 1983; Rajendrudu et al., 1986; Sternberg et al., 1986; Schlüter et al., 2017). The "C_2" rice plants showed an increased CO_2

compensation point and reduced carboxylation efficiencies, which further indicates that the aim of introducing a functional C_2 shuttle was not achieved (Lin et al., 2016). The lower CO_2 compensation points of C_2 plants are a result of a more efficient carbon fixation. This is also connected to improved anatomical features in many C_2 plants e.g. increased amounts of mitochondria in BS cells, which are shielded by chloroplasts to avoid leakage of CO_2 (Sage and Khoshravesh, 2016). Interestingly, the "C_2 " rice plants showed a significant reduction of chloroplasts in M cells (Lin et al., 2016), even though this feature is usually associated with a C_4 -like carbon concentrating mechanism (Stata et al., 2016). Also, increase in GDC content has been associated with increase in chloroplast size (Engel, 2010). It will be interesting to use the leaf sections for an in-depth analysis of organelle number, size and location within the different cell types.

CONCLUSION AND OUTLOOK

The Cas9 system allows the generation of highly specific modifications within the original genomic context of the target. We employed this approach to modify the *AtGLDP1* promoter of *Arabidopsis* with the aim of generating a BS cell specific expression pattern of the P-protein, to mimic the CO_2 pump in C_2 plants. Deletion of the M-Box region, which drives expression of the gene in M cells, was the most successful approach to fulfill this aim with feasible screening amount. We were able to create a partial deletion within the M-Box and generated phenotypically normal plants that showed slight accumulations of glycine under ambient CO_2. In the following, further experiments must validate whether we indeed achieved exclusive P-protein localization in the BS cells. Additionally, we need to consider putative effects caused by integration of several T-DNAs within the plants (*gldp2-2* T-DNA, unspecified T-DNA from the *gldp1-1* background, pCas9-MBOX T-DNA). Backcrossing of the mutant lines with a clean *gldp2-2* line without further T-DNA inserts will therefore be necessary and verification of this new *gldp2-2* line is in progress (Master thesis Dominik Daniels). It was striking that we could not generate a full mutation of the M-Box with our approach. Ongoing work (Master thesis Dominik Daniels) aims at generating a full deletion of the M-Box or even larger deletions, as also the upstream areas of the M-Box are conserved in various *Brassicaceae* species (Adwy et al., 2015). For this, we use a different vector system (Hahn et al., 2017), which was already used to generate deletions by using two sgRNAs (Yuichi Yokochi, unpublished data). However, we should not neglect the

possibility, that a full abolishment of *AtGLDP* expression in the M cells is lethal for *Arabidopsis* as indicated by previous experimental approaches (Engel et al., 2007, Engel 2010). During evolution, biochemical adaptions of the photorespiratory pathway occurred probably after adaptions of leaf anatomy (Sage, 2004). This includes the establishment of a proto-Kranz anatomy but also an enrichment of chloroplasts and mitochondria in BS cells as well as increased plasmodesmata frequency for metabolite shuttling (Gowik and Westhoff, 2011; Danila et al., 2016). Even though some anatomical changes seem to be directly induced by a redirected photorespiratory flux (Lin et al., 2016), it might be advisable to decrease the amount of GLDP in M cells gradually, as found in the different evolutionary intermediates between C_3 and C_4 *Flaveria* species (Schulze et al., 2013). Gradual decreases in AtGLDP amount can be achieved by targeting the transcription of the *AtGLDP1* gene. Due to the lack of M cell specific promoters in *Arabidopsis*, we cannot use an amiRNA or antisense approach as done in rice and potato (Winzer et al., 2001; Lin et al., 2016). However, the Cas9 system can be modified to act as a targeted transcriptional repressor (Qi et al., 2013). Therefore, an inactive Cas9 is fused to transcription repressors, such as the transcriptional repressor domain SRDX (Lowder et al., 2015; Tang et al., 2017). By targeting this repressor construct to the M-Box, we might be able to block transcription factors which induce expression in M cells, thereby decreasing the number of transcripts in these cells (Master thesis Dominik Daniels). In summary, the CRISPR/Cas9 system can be exploited for manifold manipulations of genomes as shown in this manuscript, but also for more subtle modifications on transcript level. This makes this technique a promising tool for integration of C_4 cycle components in C_3 plants and will enable us to understand the evolution of C_2 and C_4 photosynthesis in more detail. This understanding can accelerate the integration of efficient carbon concentrating mechanisms in C_3 plants.

AUTHOR CONTRIBUTIONS

FH, OM, ME, and AW designed the study. FH carried out all laboratory experiments. OM performed PCA analysis. US contributed promoter sequences. TMA supervised and planned the metabolite measurements. KB assisted in leaf tissue embedding. FH, OM, ME, TMA and AW interpreted the data and wrote the manuscript with critical revisions of US and TMA.

FUNDING

This work was funded by the Cluster of Excellence on Plant Sciences (CEPLAS, EXC 1028) and the HHU Center for Synthetic Life Sciences (CSL).

ACKNOWLEDGMENTS

We thank Marion Nissen for leaf ultracuts preparation and electron microscopy support. We thank Prof. Dr. Holger Puchta, Felix Wolters and Waly Adwy for providing us with vectors and for valuable feedback. The help of Katrin Weber, Maria Graf and Elisabeth Klemm for GC/MS analysis of the metabolites is greatly appreciated. We acknowledge Dr. Eva Nowack and Georg Ehret for providing us with immunogold antibodies and experimental support. Martha Ludwig provided us with a GLDP antibody. We thank Amaya Krusch for provision of *GLDP* promoter sequences and preliminary alignments. We thank the gardeners for plant care and the student helper team for their support.

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

REFERENCES

Adwy, W., Laxa, M., and Peterhansel, C. (2015). A simple mechanism for the establishment of C2-specific gene expression in Brassicaceae. *Plant J.* 84, 1231–1238. doi:10.1111/tpj.13084.

Akyildiz, M., Gowik, U., Engelmann, S., Koczor, M., Streubel, M., and Westhoff, P. (2007). Evolution and Function of a cis-Regulatory Module for Mesophyll-Specific Gene Expression in the C4 Dicot Flaveria trinervia. *Plant Cell* 19, 3391–3402. doi:10.1105/tpc.107.053322.

Argüello-Astorga, G., and Herrera-Estrella, L. (1998). Evolution of Light-Regulated Plant Promoters. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49, 525–555. doi:10.1146/annurev.arplant.49.1.525.

Badger, M. R., Fallahi, H., Kaines, S., and Takahashi, S. (2009). Chlorophyll fluorescence screening of Arabidopsis thaliana for CO2 sensitive photorespiration and photoinhibition mutants. *Funct. Plant Biol.* 36, 867–873. doi:10.1071/FP09199.

Baltes, N. J., Gil-Humanes, J., Cermak, T., Atkins, P. A., and Voytas, D. F. (2014). DNA replicons for plant genome engineering. *Plant Cell* 26, 151–163. doi:10.1105/tpc.113.119792.

Bauwe, H. (2010). "Photorespiration: The Bridge to C4 Photosynthesis," in *C4 Photosynthesis and Related CO2 Concentrating Mechanisms* Advances in Photosynthesis and Respiration. (Springer, Dordrecht), 81–108. doi:10.1007/978-90-481-9407-0_6.

Bauwe, H., and Kolukisaoglu, Ü. (2003). Genetic manipulation of glycine decarboxylation. *J. Exp. Bot.* 54, 1523–1535. doi:10.1093/jxb/erg171.

Betti, M., Bauwe, H., Busch, F. A., Fernie, A. R., Keech, O., Levey, M., et al. (2016). Manipulating photorespiration to increase plant productivity: recent advances and perspectives for crop improvement. *J. Exp. Bot.* 67, 2977–2988. doi:10.1093/jxb/erw076.

Boldt, R., Edner, C., Kolukisaoglu, Ü., Hagemann, M., Weckwerth, W., Wienkoop, S., et al. (2005). d-GLYCERATE 3-KINASE, the Last Unknown Enzyme in the Photorespiratory Cycle in Arabidopsis, Belongs to a Novel Kinase Family. *Plant Cell* 17, 2413–2420. doi:10.1105/tpc.105.033993.

Bowes, G., Ogren, W. L., and Hageman, R. H. (1971). Phosphoglycolate production catalyzed by ribulose diphosphate carboxylase. *Biochem. Biophys. Res. Commun.* 45, 716–722. doi:10.1016/0006-291X(71)90475-X.

Brown, N. J., Newell, C. A., Stanley, S., Chen, J. E., Perrin, A. J., Kajala, K., et al. (2011). Independent and Parallel Recruitment of Preexisting Mechanisms Underlying C4 Photosynthesis. *Science* 331, 1436–1439. doi:10.1126/science.1201248.

Campisi, L., Yang, Y., Yi, Y., Heilig, E., Herman, B., Cassista, A. J., et al. (1999). Generation of enhancer trap lines in Arabidopsis and characterization of expression patterns in the inflorescence. *Plant J.* 17, 699–707. doi:10.1046/j.1365-313X.1999.00409.x.

Čermák, T., Baltes, N. J., Čegan, R., Zhang, Y., and Voytas, D. F. (2015). High-frequency, precise modification of the tomato genome. *Genome Biol.* 16, 232. doi:10.1186/s13059-015-0796-9.

Chan, C.-S., Guo, L., and Shih, M.-C. (2001). Promoter analysis of the nuclear gene encoding the chloroplast glyceraldehyde-3-phosphate dehydrogenase B subunit of Arabidopsis thaliana. *Plant Mol. Biol.* 46, 131–141. doi:10.1023/A:1010602031070.

Chen, Q., He, J., Phoolcharoen, W., and Mason, H. S. (2011). Geminiviral vectors based on bean yellow dwarf virus for production of vaccine antigens and monoclonal antibodies in plants. *Hum. Vaccin.* 7, 331–338.

Clough, S. J., and Bent, A. F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J.* 16, 735–743. doi:10.1046/j.1365-313x.1998.00343.x.

Collonnier, C., Epert, A., Mara, K., Maclot, F., Guyon-Debast, A., Charlot, F., et al. (2017). CRISPR-Cas9-mediated efficient directed mutagenesis and RAD51-dependent and RAD51-independent gene targeting in the moss Physcomitrella patens. *Plant Biotechnol. J.* 15, 122–131. doi:10.1111/pbi.12596.

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

Czechowski, T., Stitt, M., Altmann, T., Udvardi, M. K., and Scheible, W.-R. (2005). Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. *Plant Physiol.* 139, 5–17. doi:10.1104/pp.105.063743.

Danila, F. R., Quick, W. P., White, R. G., Furbank, R. T., and von Caemmerer, S. (2016). The Metabolite Pathway between Bundle Sheath and Mesophyll: Quantification of Plasmodesmata in Leaves of C3 and C4 Monocots. *Plant Cell* 28, 1461–1471. doi:10.1105/tpc.16.00155.

Denton, A. K., Maß, J., Külahoglu, C., Lercher, M. J., Bräutigam, A., and Weber, A. P. M. (2017). Freeze-quenched maize mesophyll and bundle sheath separation uncovers bias in previous tissue-specific RNA-Seq data. *J. Exp. Bot.* 68, 147–160. doi:10.1093/jxb/erw463.

Douce, R., Bourguignon, J., Neuburger, M., and Rébeillé, F. (2001). The glycine decarboxylase system: a fascinating complex. *Trends Plant Sci.* 6, 167–176. doi:10.1016/S1360-1385(01)01892-1.

Doudna, J. A., and Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* 346, 1258096. doi:10.1126/science.1258096.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi:10.1093/nar/gkh340.

Eisenhut, M., Bauwe, H., and Hagemann, M. (2007). Glycine accumulation is toxic for the cyanobacterium Synechocystis sp. strain PCC 6803, but can be compensated by supplementation with magnesium ions. *FEMS Microbiol. Lett.* 277, 232–237. doi:10.1111/j.1574-6968.2007.00960.x.

Eisenhut, M., Planchais, S., Cabassa, C., Guivarc'h, A., Justin, A.-M., Taconnat, L., et al. (2013). Arabidopsis A BOUT DE SOUFFLE is a putative mitochondrial transporter involved in photorespiratory metabolism and is required for meristem growth at ambient CO2 levels. *Plant J.* 73, 836–849. doi:10.1111/tpj.12082.

Eisenhut, M., Ruth, W., Haimovich, M., Bauwe, H., Kaplan, A., and Hagemann, M. (2008). The photorespiratory glycolate metabolism is essential for cyanobacteria and might have been

conveyed endosymbiontically to plants. *Proc. Natl. Acad. Sci.* 105, 17199–17204. doi:10.1073/pnas.0807043105.

Engel, N. (2010). Funktionelle Charakterisierung der mitochondrialen GDC- und SHM-Genfamilien in Arabidopsis thaliana durch Komplementation mit heterologen Genen. Available at: http://rosdok.uni-rostock.de/metadata/rosdok_disshab_0000000531.

Engel, N., Daele, K. van den, Kolukisaoglu, Ü., Morgenthal, K., Weckwerth, W., Pärnik, T., et al. (2007). Deletion of Glycine Decarboxylase in Arabidopsis Is Lethal under Nonphotorespiratory Conditions. *Plant Physiol.* 144, 1328–1335. doi:10.1104/pp.107.099317.

Engel, N., Ewald, R., Gupta, K. J., Zrenner, R., Hagemann, M., and Bauwe, H. (2011). The Presequence of Arabidopsis Serine Hydroxymethyltransferase SHM2 Selectively Prevents Import into Mesophyll Mitochondria. *Plant Physiol.* 157, 1711–1720. doi:10.1104/pp.111.184564.

Engelmann, S., Wiludda, C., Burscheidt, J., Gowik, U., Schlue, U., Koczor, M., et al. (2008). The Gene for the P-Subunit of Glycine Decarboxylase from the C4 Species Flaveria trinervia: Analysis of Transcriptional Control in Transgenic Flaveria bidentis (C4) and Arabidopsis (C3). *Plant Physiol.* 146, 1773–1785. doi:10.1104/pp.107.114462.

Fauser, F., Roth, N., Pacher, M., Ilg, G., Sánchez-Fernández, R., Biesgen, C., et al. (2012). In planta gene targeting. *Proc. Natl. Acad. Sci. U. S. A.* 109, 7535–7540. doi:10.1073/pnas.1202191109.

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

Fiehn, O. (2007). "Validated High Quality Automated Metabolome Analysis of Arabidopsis Thaliana Leaf Disks," in *Concepts in Plant Metabolomics* (Springer, Dordrecht), 1–18. doi:10.1007/978-1-4020-5608-6_1.

Foyer, C. H., Bloom, A. J., Queval, G., and Noctor, G. (2009). Photorespiratory Metabolism: Genes, Mutants, Energetics, and Redox Signaling. *Annu. Rev. Plant Biol.* 60, 455–484. doi:10.1146/annurev.arplant.043008.091948.

Gil-Humanes, J., Wang, Y., Liang, Z., Shan, Q., Ozuna, C. V., Sánchez-León, S., et al. (2017). High-efficiency gene targeting in hexaploid wheat using DNA replicons and CRISPR/Cas9. *Plant J.* 89, 1251–1262. doi:10.1111/tpj.13446.

Gowik, U., Burscheidt, J., Akyildiz, M., Schlue, U., Koczor, M., Streubel, M., et al. (2004). cis-Regulatory Elements for Mesophyll-Specific Gene Expression in the C4 Plant Flaveria trinervia, the Promoter of the C4 Phosphoenolpyruvate Carboxylase Gene. *Plant Cell* 16, 1077–1090. doi:10.1105/tpc.019729.

Gowik, U., Schulze, S., Saladié, M., Rolland, V., Tanz, S. K., Westhoff, P., et al. (2017). A MEM1-like motif directs mesophyll cell-specific expression of the gene encoding the C4 carbonic anhydrase in Flaveria. *J. Exp. Bot.* 68, 311–320. doi:10.1093/jxb/erw475.

Gowik, U., and Westhoff, P. (2011). The Path from C3 to C4 Photosynthesis. *Plant Physiol.* 155, 56–63. doi:10.1104/pp.110.165308.

Hahn, F., Eisenhut, M., Mantegazza, O., and Weber, A. P. M. (2017). Generation of Targeted Knockout Mutants in Arabidopsis thaliana Using CRISPR/Cas9. *BIO-Protoc.* 7. doi:10.21769/BioProtoc.2384.

Hahn, F., Eisenhut, M., Mantegazza, O., and Weber, A. P. M. (2018). Homology-directed repair of a defective glabrous gene in Arabidopsis with Cas9-based gene targeting. *bioRxiv*, 243675. doi:10.1101/243675.

Heckmann, D., Schulze, S., Denton, A., Gowik, U., Westhoff, P., Weber, A. P. M., et al. (2013). Predicting C4 Photosynthesis Evolution: Modular, Individually Adaptive Steps on a Mount Fuji Fitness Landscape. *Cell* 153, 1579–1588. doi:10.1016/j.cell.2013.04.058.

Heineke, D., Bykova, N., Gardeström, P., and Bauwe, H. (2001). Metabolic response of potato plants to an antisense reduction of the P-protein of glycine decarboxylase. *Planta* 212, 880–887. doi:10.1007/s004250000460.

Hylton, C. M., Rawsthorne, S., Smith, A. M., Jones, D. A., and Woolhouse, H. W. (1988). Glycine decarboxylase is confined to the bundle-sheath cells of leaves of C3/C4 intermediate species. *Planta* 175, 452–459. doi:10.1007/BF00393064.

Igarashi, D., Miwa, T., Seki, M., Kobayashi, M., Kato, T., Tabata, S., et al. (2003). Identification of photorespiratory glutamate:glyoxylate aminotransferase (GGAT) gene in Arabidopsis. *Plant J.* 33, 975–987. doi:10.1046/j.1365-313X.2003.01688.x.

Jabrin, S., Ravanel, S., Gambonnet, B., Douce, R., and Rébeillé, F. (2003). One-Carbon Metabolism in Plants. Regulation of Tetrahydrofolate Synthesis during Germination and Seedling Development. *Plant Physiol.* 131, 1431–1439. doi:10.1104/pp.016915.

Jeong, M.-J., and Shih, M.-C. (2003). Interaction of a GATA factor with cis-acting elements involved in light regulation of nuclear genes encoding chloroplast glyceraldehyde-3-phosphate dehydrogenase in Arabidopsis. *Biochem. Biophys. Res. Commun.* 300, 555–562. doi:10.1016/S0006-291X(02)02892-9.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821. doi:10.1126/science.1225829.

Kajala, K., Brown, N. J., Williams, B. P., Borrill, P., Taylor, L. E., and Hibberd, J. M. (2012). Multiple Arabidopsis genes primed for recruitment into C4 photosynthesis. *Plant J.* 69, 47–56. doi:10.1111/j.1365-313X.2011.04769.x.

Keerberg, O., Pärnik, T., Ivanova, H., Bassüner, B., and Bauwe, H. (2014). C2 photosynthesis generates about 3-fold elevated leaf CO2 levels in the C3–C4 intermediate species Flaveria pubescens. *J. Exp. Bot.* 65, 3649–3656. doi:10.1093/jxb/eru239.

Khoshravesh, R., Lundsgaard-Nielsen, V., Sultmanis, S., and Sage, T. L. (2017). Light Microscopy, Transmission Electron Microscopy, and Immunohistochemistry Protocols for Studying Photorespiration. *Methods Mol. Biol. Clifton NJ* 1653, 243–270. doi:10.1007/978-1-4939-7225-8_17.

Kinnersley, A. M., and Turano, F. J. (2000). Gamma Aminobutyric Acid (GABA) and Plant Responses to Stress. *Crit. Rev. Plant Sci.* 19, 479–509. doi:10.1080/07352680091139277.

Kinsman, E. A., and Pyke, K. A. (1998). Bundle sheath cells and cell-specific plastid development in Arabidopsis leaves. *Development* 125, 1815–1822.

Krenzer, E. G., Moss, D. N., and Crookston, R. K. (1975). Carbon Dioxide Compensation Points of Flowering Plants 1. *Plant Physiol.* 56, 194–206.

Ku, M. S. B., Monson, R. K., Littlejohn, R. O., Nakamoto, H., Fisher, D. B., and Edwards, G. E. (1983). Photosynthetic Characteristics of C3-C4 Intermediate Flaveria Species. *Plant Physiol.* 71, 944–948.

Li, J.-F., Aach, J., Norville, J. E., McCormack, M., Zhang, D., Bush, J., et al. (2013). Multiplex and homologous recombination-mediated plant genome editing via guide RNA/Cas9. *Nat. Biotechnol.* 31, 688–691. doi:10.1038/nbt.2654.

Liepman, A. H., and Olsen, L. J. (2001). Peroxisomal alanine : glyoxylate aminotransferase (AGT1) is a photorespiratory enzyme with multiple substrates in Arabidopsis thaliana. *Plant J.* 25, 487–498. doi:10.1046/j.1365-313x.2001.00961.x.

Lin, H., Karki, S., Coe, R. A., Bagha, S., Khoshravesh, R., Balahadia, C. P., et al. (2016). Targeted Knockdown of GDCH in Rice Leads to a Photorespiratory-Deficient Phenotype Useful as a Building Block for C4 Rice. *Plant Cell Physiol.* 57, 919–932. doi:10.1093/pcp/pcw033.

Long, S. P., Marshall-Colon, A., and Zhu, X.-G. (2015). Meeting the Global Food Demand of the Future by Engineering Crop Photosynthesis and Yield Potential. *Cell* 161, 56–66. doi:10.1016/j.cell.2015.03.019.

Lowder, L. G., Zhang, D., Baltes, N. J., Paul, J. W., Tang, X., Zheng, X., et al. (2015). A CRISPR/Cas9 Toolbox for Multiplexed Plant Genome Editing and Transcriptional Regulation. *Plant Physiol.* 169, 971–985. doi:10.1104/pp.15.00636.

Mallmann, J., Heckmann, D., Bräutigam, A., Lercher, M. J., Weber, A. P., Westhoff, P., et al. (2014). The role of photorespiration during the evolution of C4 photosynthesis in the genus Flaveria. *eLife* 3, e02478. doi:10.7554/eLife.02478.

Mao, Y., Zhang, H., Xu, N., Zhang, B., Gou, F., and Zhu, J.-K. (2013). Application of the CRISPR–Cas System for Efficient Genome Engineering in Plants. *Mol. Plant* 6, 2008–2011. doi:10.1093/mp/sst121.

Mao, Y., Zhang, Z., Feng, Z., Wei, P., Zhang, H., Botella, J. R., et al. (2016). Development of germ-line-specific CRISPR-Cas9 systems to improve the production of heritable gene modifications in Arabidopsis. *Plant Biotechnol. J.* 14, 519–532. doi:10.1111/pbi.12468.

Nunes-Nesi, A., Florian, A., Howden, A., Jahnke, K., Timm, S., Bauwe, H., et al. (2014). Is There a Metabolic Requirement for Photorespiratory Enzyme Activities in Heterotrophic Tissues? *Mol. Plant* 7, 248–251. doi:10.1093/mp/sst111.

Oliver, D. J., Neuburger, M., Bourguignon, J., and Douce, R. (1990). Interaction between the Component Enzymes of the Glycine Decarboxylase Multienzyme Complex. *Plant Physiol.* 94, 833–839. doi:10.1104/pp.94.2.833.

Ort, D. R., Merchant, S. S., Alric, J., Barkan, A., Blankenship, R. E., Bock, R., et al. (2015). Redesigning photosynthesis to sustainably meet global food and bioenergy demand. *Proc. Natl. Acad. Sci.* 112, 8529–8536. doi:10.1073/pnas.1424031112.

Paszkowski, J., Baur, M., Bogucki, A., and Potrykus, I. (1988). Gene targeting in plants. *EMBO J.* 7, 4021–4026.

Peterhansel, C., Horst, I., Niessen, M., Blume, C., Kebeish, R., Kürkcüoglu, S., et al. (2010). Photorespiration. *Arab. Book Am. Soc. Plant Biol.* 8. doi:10.1199/tab.0130.

Puchta, H. (2002). Gene replacement by homologous recombination in plants. *Plant Mol. Biol.* 48, 173–182.

Puchta, H. (2005). The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. *J. Exp. Bot.* 56, 1–14. doi:10.1093/jxb/eri025.

Puchta, H., Dujon, B., and Hohn, B. (1996). Two different but related mechanisms are used in plants for the repair of genomic double-strand breaks by homologous recombination. *Proc. Natl. Acad. Sci. U. S. A.* 93, 5055–5060.

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., et al. (2013). Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* 152, 1173–1183. doi:10.1016/j.cell.2013.02.022.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rademacher, N., Kern, R., Fujiwara, T., Mettler-Altmann, T., Miyagishima, S., Hagemann, M., et al. (2016). Photorespiratory glycolate oxidase is essential for the survival of the red alga Cyanidioschyzon merolae under ambient CO2 conditions. *J. Exp. Bot.* 67, 3165–3175. doi:10.1093/jxb/erw118.

Rajendrudu, G., Prasad, J. S. R., and Das, V. S. R. (1986). C3-C4 Intermediate Species in Alternanthera (Amaranthaceae). *Plant Physiol.* 80, 409–414.

Rawsthorne, S., and Hylton, C. M. (1991). The relationship between the post-illumination CO2 burst and glycine metabolism in leaves of C3 and C3-C4 intermediate species of Moricandia. *Planta* 186, 122–126. doi:10.1007/BF00201507.

Rawsthorne, S., Hylton, C. M., Smith, A. M., and Woolhouse, H. W. (1988). Photorespiratory metabolism and immunogold localization of photorespiratory enzymes in leaves of C3 and C3-C4 intermediate species of Moricandia. *Planta* 173, 298–308. doi:10.1007/BF00401016.

Ray, D. K., Mueller, N. D., West, P. C., and Foley, J. A. (2013). Yield Trends Are Insufficient to Double Global Crop Production by 2050. *PLOS ONE* 8, e66428. doi:10.1371/journal.pone.0066428.

Sage, R. F. (2004). The evolution of C4 photosynthesis. *New Phytol.* 161, 341–370. doi:10.1111/j.1469-8137.2004.00974.x.

Sage, R. F. (2013). Photorespiratory compensation: a driver for biological diversity. *Plant Biol.* 15, 624–638. doi:10.1111/plb.12024.

Sage, R. F., Christin, P.-A., and Edwards, E. J. (2011). The C4 plant lineages of planet Earth. *J. Exp. Bot.* 62, 3155–3169. doi:10.1093/jxb/err048.

Sage, R. F., and Khoshravesh, R. (2016). Passive CO2 concentration in higher plants. *Curr. Opin. Plant Biol.* 31, 58–65. doi:10.1016/j.pbi.2016.03.016.

Sage, R. F., Khoshravesh, R., and Sage, T. L. (2014). From proto-Kranz to C4 Kranz: building the bridge to C4 photosynthesis. *J. Exp. Bot.* 65, 3341–3356. doi:10.1093/jxb/eru180.

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

Schiml, S., Fauser, F., and Puchta, H. (2017). CRISPR/Cas-Mediated In Planta Gene Targeting. *Methods Mol. Biol. Clifton NJ* 1610, 3–11. doi:10.1007/978-1-4939-7003-2_1.

Schjoerring, J. K., Mäck, G., Nielsen, K. H., Husted, S., Suzuki, A., Driscoll, S., et al. (2006). Antisense reduction of serine hydroxymethyltransferase results in diurnal displacement of NH4+ assimilation in leaves of Solanum tuberosum. *Plant J.* 45, 71–82. doi:10.1111/j.1365-313X.2005.02598.x.

Schlüter, U., Bräutigam, A., Gowik, U., Melzer, M., Christin, P.-A., Kurz, S., et al. (2017). Photosynthesis in C3–C4 intermediate Moricandia species. *J. Exp. Bot.* 68, 191–206. doi:10.1093/jxb/erw391.

Schlüter, U., and Weber, A. P. M. (2016). The Road to C4 Photosynthesis: Evolution of a Complex Trait via Intermediary States. *Plant Cell Physiol.* 57, 881–889. doi:10.1093/pcp/pcw009.

Schulze, S., Mallmann, J., Burscheidt, J., Koczor, M., Streubel, M., Bauwe, H., et al. (2013). Evolution of C4 Photosynthesis in the Genus Flaveria: Establishment of a Photorespiratory CO2 Pump. *Plant Cell* 25, 2522–2535. doi:10.1105/tpc.113.114520.

Schulze, S., Westhoff, P., and Gowik, U. (2016). Glycine decarboxylase in C3, C4 and C3–C4 intermediate species. *Curr. Opin. Plant Biol.* 31, 29–35. doi:10.1016/j.pbi.2016.03.011.

Schwarte, S., and Bauwe, H. (2007). Identification of the Photorespiratory 2-Phosphoglycolate Phosphatase, PGLP1, in Arabidopsis. *Plant Physiol.* 144, 1580–1586. doi:10.1104/pp.107.099192.

Sfeir, A., and Symington, L. S. (2015). Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway? *Trends Biochem. Sci.* 40, 701–714. doi:10.1016/j.tibs.2015.08.006.

Simon, P. (2003). Q-Gene: processing quantitative real-time RT-PCR data. *Bioinforma. Oxf. Engl.* 19, 1439–1440.

Slack, C. R., and Hatch, M. D. (1967). Comparative studies on the activity of carboxylases and other enzymes in relation to the new pathway of photosynthetic carbon dioxide fixation in tropical grasses. *Biochem. J.* 103, 660–665.

Somerville, C. R., and Ogren, W. L. (1979). A phosphoglycolate phosphatase-deficient mutant of Arabidopsis. *Nature* 280, 833. doi:10.1038/280833a0.

Stata, M., Sage, T. L., Hoffmann, N., Covshoff, S., Ka-Shu Wong, G., and Sage, R. F. (2016). Mesophyll Chloroplast Investment in C3, C4 and C2 Species of the Genus Flaveria. *Plant Cell Physiol.* 57, 904–918. doi:10.1093/pcp/pcw015.

Sternberg, L. D. S. L., Deniro, M. J., Sloan, M. E., and Black, C. C. (1986). Compensation Point and Isotopic Characteristics of C3/C4 Intermediates and Hybrids in Panicum. *Plant Physiol.* 80, 242–245.

Takahashi, S., Bauwe, H., and Badger, M. (2007). Impairment of the Photorespiratory Pathway Accelerates Photoinhibition of Photosystem II by Suppression of Repair But Not Acceleration of Damage Processes in Arabidopsis. *Plant Physiol.* 144, 487–494. doi:10.1104/pp.107.097253.

Tang, X., Lowder, L. G., Zhang, T., Malzahn, A. A., Zheng, X., Voytas, D. F., et al. (2017). A CRISPR–Cpf1 system for efficient genome editing and transcriptional repression in plants. *Nat. Plants* 3, 17018. doi:10.1038/nplants.2017.18.

Tang, X., Zheng, X., Qi, Y., Zhang, D., Cheng, Y., Tang, A., et al. (2016). A Single Transcript CRISPR-Cas9 System for Efficient Genome Editing in Plants. *Mol. Plant* 9, 1088–1091. doi:10.1016/j.molp.2016.05.001.

Timm, S., and Bauwe, H. (2013). The variety of photorespiratory phenotypes – employing the current status for future research directions on photorespiration. *Plant Biol.* 15, 737–747. doi:10.1111/j.1438-8677.2012.00691.x.

Timm, S., Giese, J., Engel, N., Wittmiß, M., Florian, A., Fernie, A. R., et al. (2017). T-protein is present in large excess over the other proteins of the glycine cleavage system in leaves of Arabidopsis. *Planta*, 1–11. doi:10.1007/s00425-017-2767-8.

Timm, S., Mielewczik, M., Florian, A., Frankenbach, S., Dreissen, A., Hocken, N., et al. (2012). High-to-Low CO2 Acclimation Reveals Plasticity of the Photorespiratory Pathway and Indicates Regulatory Links to Cellular Metabolism of Arabidopsis. *PLOS ONE* 7, e42809. doi:10.1371/journal.pone.0042809.

Timm, S., Nunes-Nesi, A., Pärnik, T., Morgenthal, K., Wienkoop, S., Keerberg, O., et al. (2008). A Cytosolic Pathway for the Conversion of Hydroxypyruvate to Glycerate during Photorespiration in Arabidopsis. *Plant Cell* 20, 2848–2859. doi:10.1105/tpc.108.062265.

Timm, S., Wittmiß, M., Gamlien, S., Ewald, R., Florian, A., Frank, M., et al. (2015). Mitochondrial Dihydrolipoyl Dehydrogenase Activity Shapes Photosynthesis and Photorespiration of Arabidopsis thaliana. *Plant Cell* 27, 1968–1984. doi:10.1105/tpc.15.00105.

Vogan, P. J., and Sage, R. F. (2012). Effects of low atmospheric CO2 and elevated temperature during growth on the gas exchange responses of C3, C3–C4 intermediate, and C4 species from three evolutionary lineages of C4 photosynthesis. *Oecologia* 169, 341–352. doi:10.1007/s00442-011-2201-z.

Voll, L. M., Jamai, A., Renné, P., Voll, H., McClung, C. R., and Weber, A. P. M. (2006). The Photorespiratory Arabidopsis shm1 Mutant Is Deficient in SHM1. *Plant Physiol.* 140, 59–66. doi:10.1104/pp.105.071399.

Walker, B. J., VanLoocke, A., Bernacchi, C. J., and Ort, D. R. (2016). The Costs of Photorespiration to Food Production Now and in the Future. *Annu. Rev. Plant Biol.* 67, 107–129. doi:10.1146/annurev-arplant-043015-111709.

Wang, H., and Xu, X. (2017). Microhomology-mediated end joining: new players join the team. *Cell Biosci.* 7, 6. doi:10.1186/s13578-017-0136-8.

Wang, L., Czedik-Eysenberg, A., Mertz, R. A., Si, Y., Tohge, T., Nunes-Nesi, A., et al. (2014). Comparative analyses of C4 and C3 photosynthesis in developing leaves of maize and rice. *Nat. Biotechnol.* 32, 1158. doi:10.1038/nbt.3019.

Wang, Z.-P., Xing, H.-L., Dong, L., Zhang, H.-Y., Han, C.-Y., Wang, X.-C., et al. (2015). Egg cell-specific promoter-controlled CRISPR/Cas9 efficiently generates homozygous mutants for multiple target genes in Arabidopsis in a single generation. *Genome Biol.* 16. doi:10.1186/s13059-015-0715-0.

Weigel, D., and Glazebrook, J. (2009). Quick miniprep for plant DNA isolation. *Cold Spring Harb. Protoc.* 2009, pdb.prot5179. doi:10.1101/pdb.prot5179.

Williams, B. P., Burgess, S. J., Reyna-Llorens, I., Knerova, J., Aubry, S., Stanley, S., et al. (2016). An Untranslated cis-Element Regulates the Accumulation of Multiple C4 Enzymes in Gynandropsis gynandra Mesophyll Cells. *Plant Cell* 28, 454–465. doi:10.1105/tpc.15.00570.

Wiludda, C., Schulze, S., Gowik, U., Engelmann, S., Koczor, M., Streubel, M., et al. (2012). Regulation of the Photorespiratory GLDPA Gene in C4 Flaveria: An Intricate Interplay of Transcriptional and Posttranscriptional Processes. *Plant Cell* 24, 137–151. doi:10.1105/tpc.111.093872.

Winzer, T., Heineke, D., and Bauwe, H. (2001). Growth and phenotype of potato plants expressing an antisense gene of P-protein of glycine decarboxylase under control of a promoter with preference for the mesophyll. *Ann. Appl. Biol.* 138, 9–15. doi:10.1111/j.1744-7348.2001.tb00080.x.

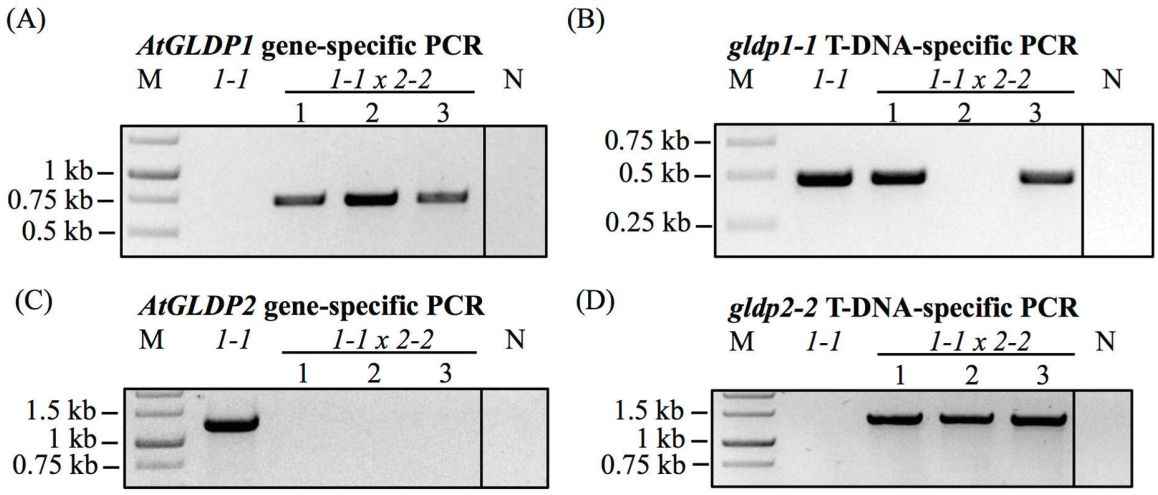
Xie, K., Minkenberg, B., and Yang, Y. (2015). Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc. Natl. Acad. Sci.* 112, 3570–3575. doi:10.1073/pnas.1420294112.

Zelitch, I., Schultes, N. P., Peterson, R. B., Brown, P., and Brutnell, T. P. (2009). High Glycolate Oxidase Activity Is Required for Survival of Maize in Normal Air. *Plant Physiol.* 149, 195–204. doi:10.1104/pp.108.128439.

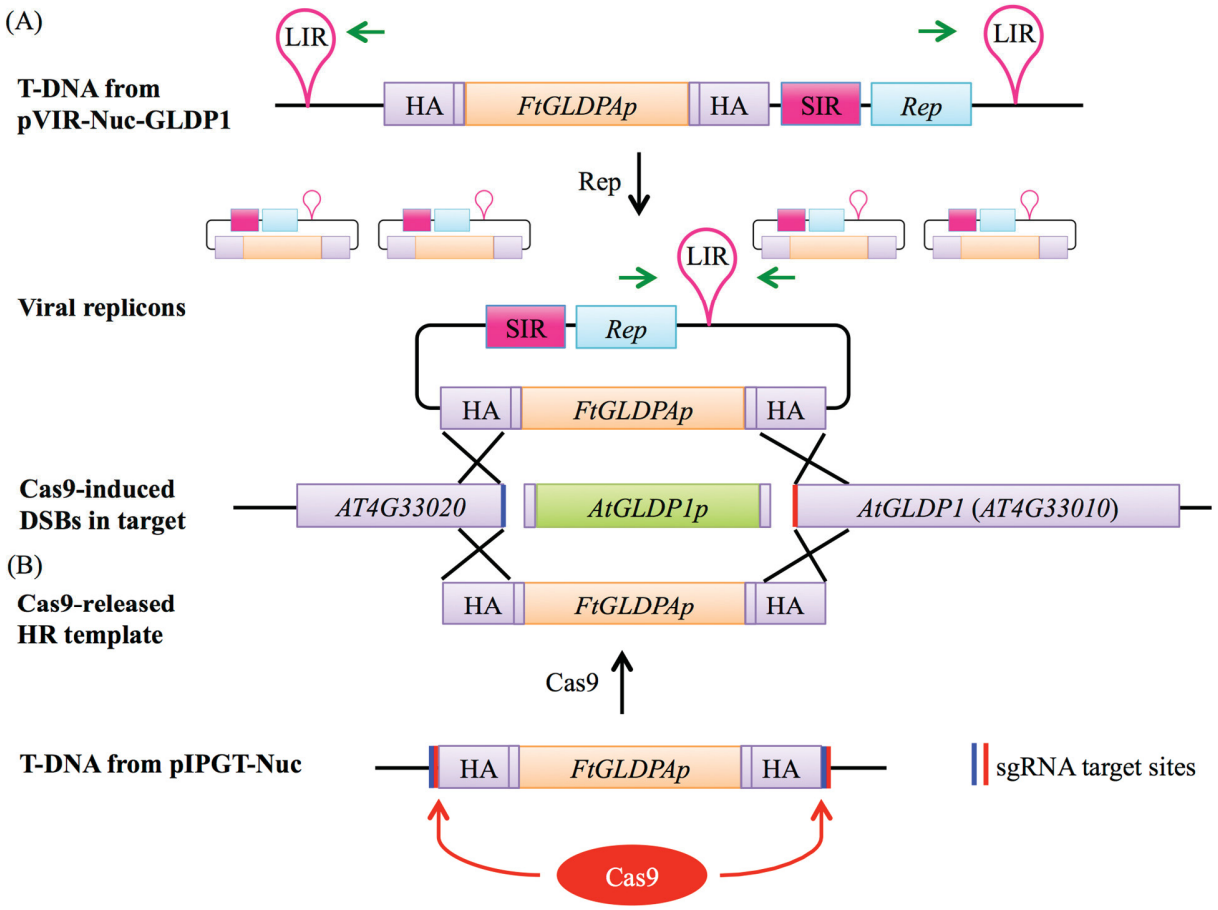
Zhao, Y., Zhang, C., Liu, W., Gao, W., Liu, C., Song, G., et al. (2016). An alternative strategy for targeted gene replacement in plants using a dual-sgRNA/Cas9 design. *Sci. Rep.* 6, 23890. doi:10.1038/srep23890.

.

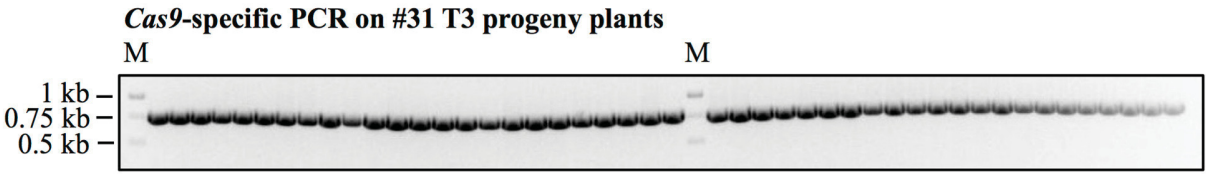
SUPPLEMENTARY MATERIAL



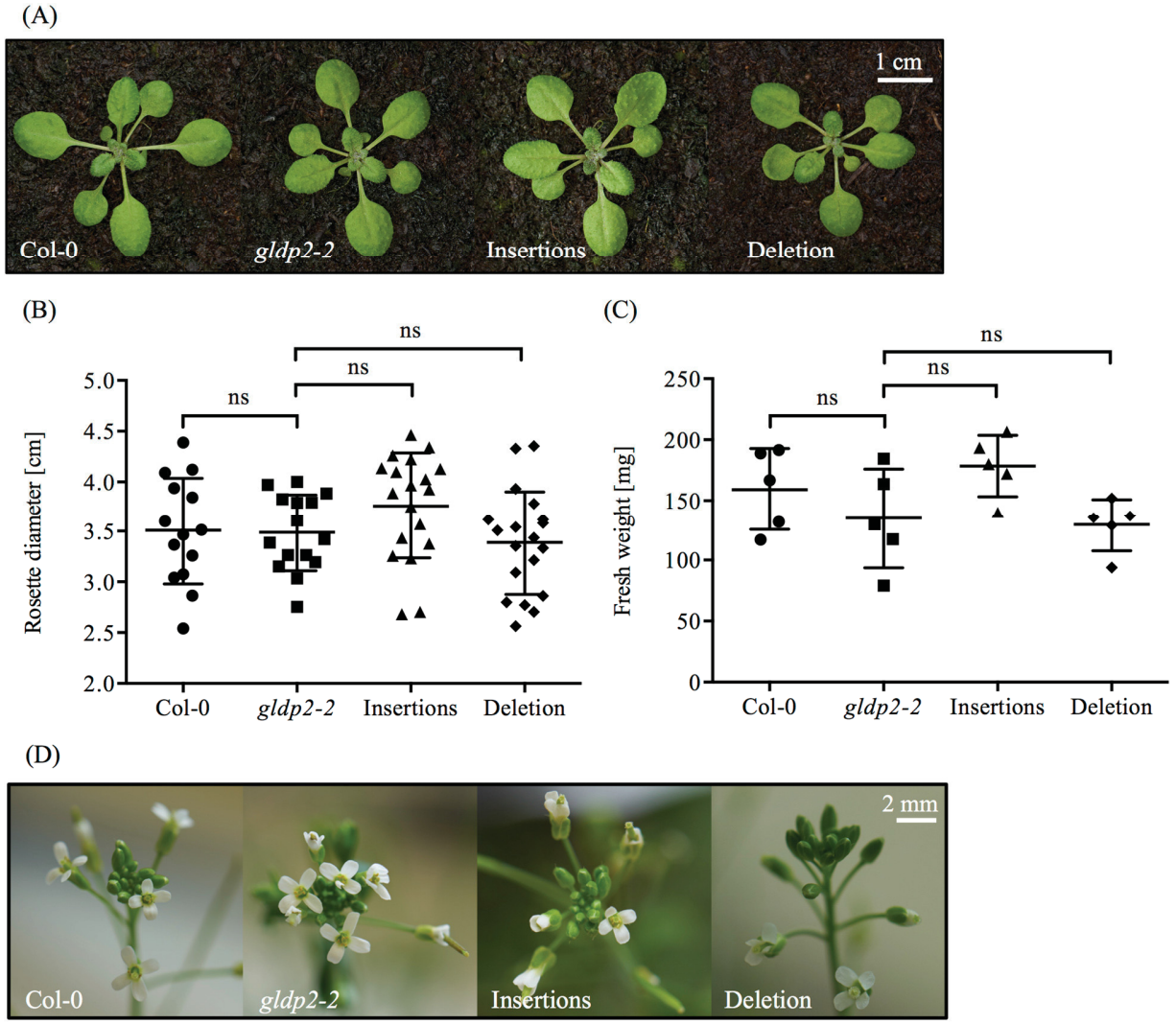
Supplementary Figure S1: Isolation of a *gldp2-2* mutant from a *gldp1-1*^{heterozygous}x*gldp2-2*^{homozygous} T-DNA line cross. Progeny plants of the T-DNA cross (*1-1 x 2-2*) were analyzed for absence of *gldp1-1* T-DNA (A/B) and homozygous insertion of *gldp2-2* T-DNA (C/D). Plant 2 shows an *AtGLDP1* gene specific band (A), no band for *gldp1-1* T-DNA (B), no band for *AtGLDP2* (C) and a T-DNA specific band for *gldp2-2* (D), proving homozygous insertion of the *gldp2-2* T-DNA. DNA from a homozygous *gldp1-1* mutant was used as control (*1-1*). N = water control. Image of agarose gel was color inverted for better visibility.



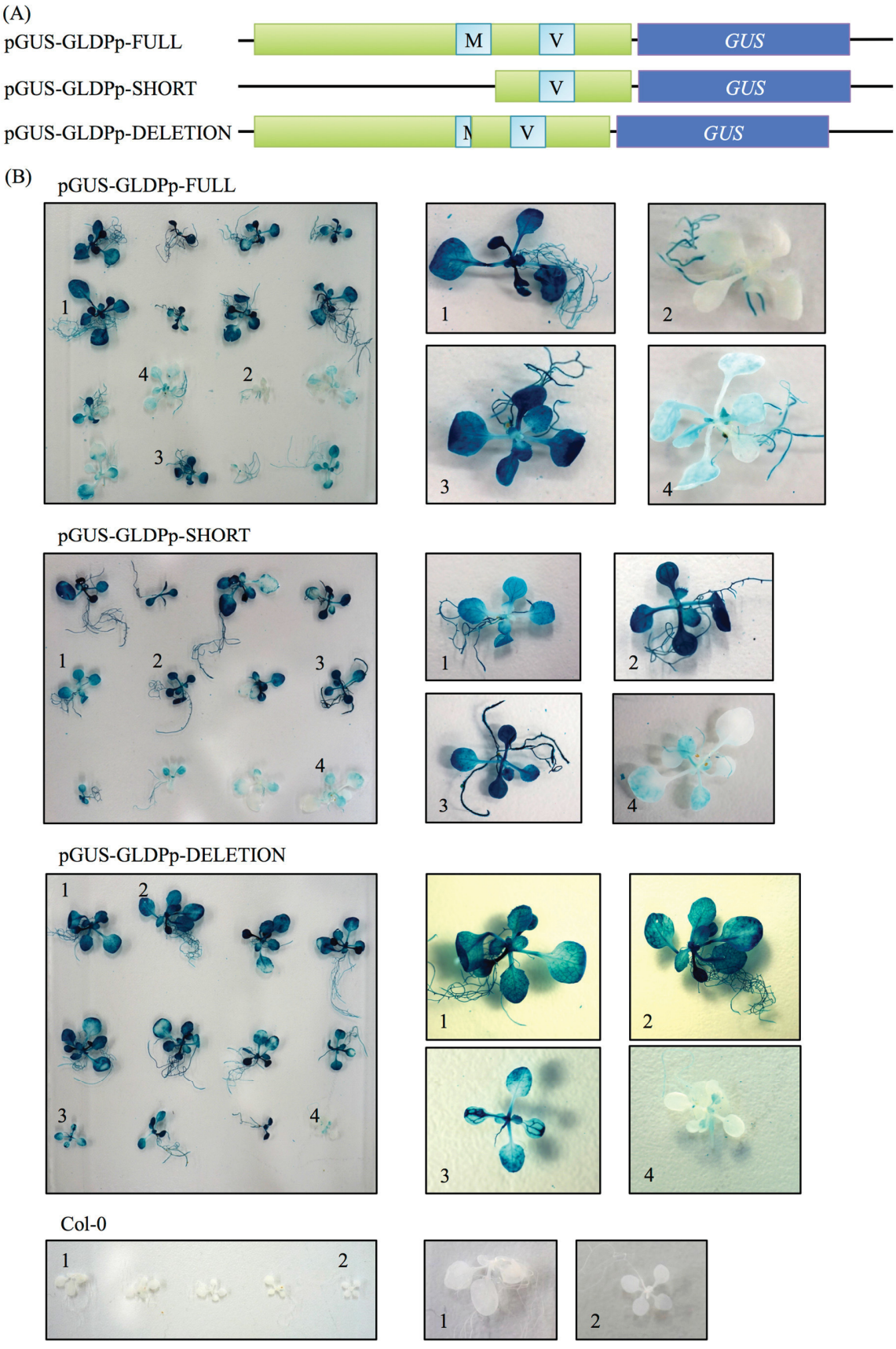
Supplementary Figure S2: Strategies for increasing the availability of the HT. (A) After integration of the T-DNA from pVIR-Nuc-GLDP1 in the *Arabidopsis* nucleus, rolling circle endoreplication of the HT containing the *FtGLDPA* promoter (orange box) is initiated by the replicase Rep. Replication can result in several hundreds of viral replicons, whose presence can be detected by PCR primers (green arrows) that amplify DNA only on circular replicons and not on linear T-DNA. When Cas9 induces two DSBs up- and downstream of the *AtGLDP1* promoter (green box), the *AtGLDP1* promoter is released. The DSBs can be repaired by HR using the viral replicons as repair template due to homologous regions between the HT on the replicon and the genes flanking the DSB sites (purple). This ultimately leads to integration of the *FtGLDPA* promoter. Accordingly, this approach can be combined with a Cas9 nickase, which only induces DNA nicks in the target gene by transformation with pVIR-Nick-GLDP1. (B) Plants transformed with pIPGT-Nuc-GLDP1 contain the HT flanked by the same sgRNA target sites that are used for release of the *Arabidopsis* promoter. Cas9 expression therefore simultaneously targets the *AtGLDP1* promoter and releases the HT, which then functions as repair template.



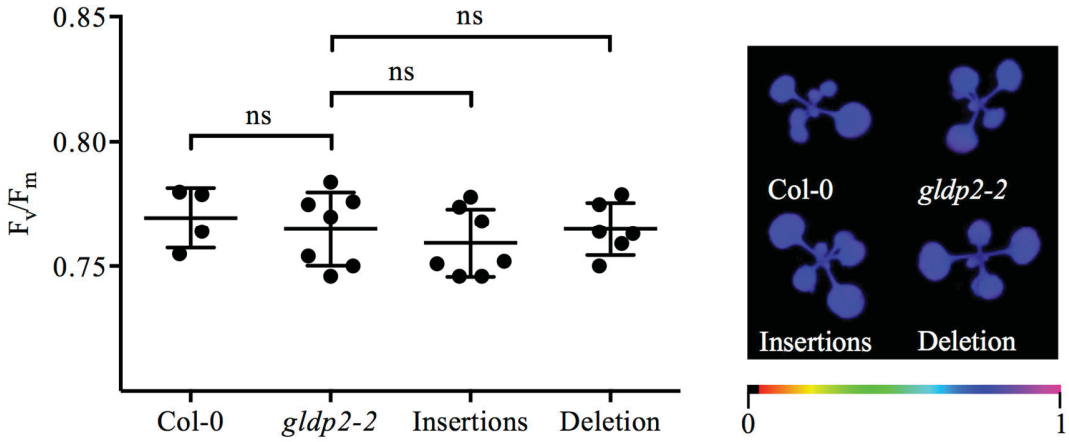
Supplementary Figure S3: Detection of Cas9 in T3 generation progeny of plant #31. Exemplary PCR screening result of 46 T3 progeny plant of plant #31 with *Cas9* primers reveals integration of the *Cas9* gene (expected size: 749 bp) into genomic DNA of all tested plants. M = Marker. Image of agarose gel was color inverted for better visibility.



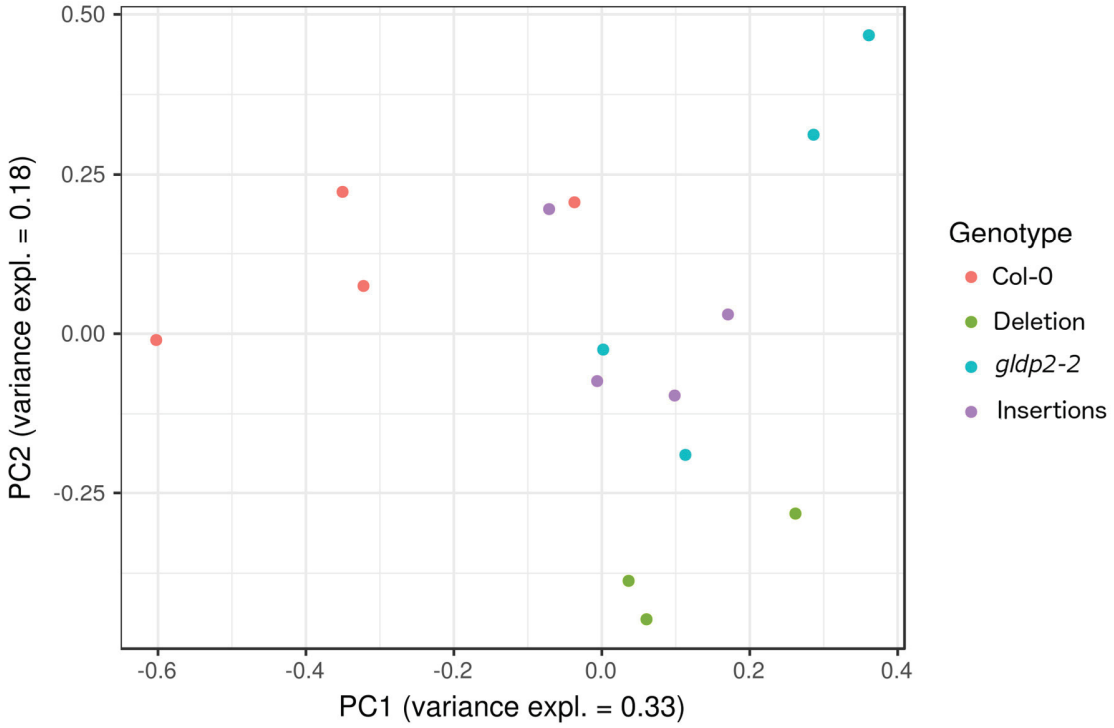
Supplementary Figure S4: Mutants with modifications in the *AtGLDP1* promoter show no significant changes in young rosettes and flowers under ambient CO_2 conditions. (A) Exemplary picture of 25 days old WT (Col-0) plant, *gldp2-2* mutant plant, plant with single-bp-insertions in the sgRNA target sites (Insertions) in the *AtGLDP1* promoter and plant with large deletion in the M-box (Deletion). No significant differences were found between the genotypes in rosette diameter (B, n = 14-19) and fresh weight (C, n = 5) as calculated by two-tailed Student's *t*-test. (D) All genotypes were able to develop fertile flower buds.


(A)
pGUS-GLDPp-FULL
M
V
GUS
pGUS-GLDPp-SHORT
V
GUS
pGUS-GLDPp-DELETION
M
V
GUS
(B)
pGUS-GLDPp-FULL
1
4
2
3
1
2
3
4
pGUS-GLDPp-SHORT
1
2
3
4
1
2
3
4
pGUS-GLDPp-DELETION
1
2
3
4
1
2
3
4
Col-0
1
2
1
2

Supplementary Figure S5: Promoter::GUS-staining reveals no distinct expression pattern. (A) Schematic presentation of the three promoter::GUS constructs that were generated for analysis of the expression pattern. pGUS-GLDPp-FULL (contains the complete *AtGLDP1* promoter) and pGUS-GLDPp-SHORT (contains only the promoter region downstream of the M-Box) were used as controls for unspecific and BS cell specific staining, respectively (Adwy et al., 2015). pGUS-GLDPp-DELETION contains the deletion of a large part of the M-Box as generated in the CRISPR/Cas9 experiments. (B) GUS-staining of independent T1 lines revealed high variances in staining independent of the transformed construct. Col-0 plants were stained as negative control.



Supplementary Figure S6: Mutants with modifications in the *AtGLDP1* promoter show no significant changes in photochemical efficiency of photosystem II. F_v / F_m-values were measured from 28 days old plants grown under ambient CO_2 conditions (left). No significant differences were detected as calculated by two-tailed Student's *t*-test (n = 4-7). Consistent F_v / F_m-values were found across the full rosette, as shown on exemplary pictures from all tested genotypes (right).



Supplementary Figure S7: Principal component analysis of z-score normalized metabolite data.

Table S1: Primer sequences.

Primer	Sequence
FH220	AGCCAACAACAATATCCGC
FH235	ATTGGCTTGTGGATTTGATCGCTG
FH236	AAACCAGCGATCAAATCCACAAGC
FH237	ATTGGCCTCCCGAAAGCCGCCGTG
FH238	AAACCACGGCGGCTTTCGGGAGGC
FH244	GAGCTTTGGGTACGTCACGTCAATTGGATGAATTCTAATAAC
FH245	TGACTTGAAGTACACTCAACAGTCAAAGCCAGCAAATC
FH266	CATACGGTTCAGGTTGTGGA
FH271	GTGACGTACCCAAAGCTCTG
FH278	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCACCTAGGGCGAATTGGCGGAA
FH279	GGGGACCACTTTGTACAAGAAAGCTGGGTGAGCGGGCAGTGAGCGGAA
FH280	TTCCGATCCAGCCTAGGCCCGGGCCCCTGAGGCTTTTTTTCTTC

FH281	TGAACGAAGAAGAAGAAAAAAAGCCGACGCGTTAATGCCAACT
FH293	GCTTTCTTCTTCGCCATCAC
FH298	ATTGGTTATCTTTCTGAGTGATAA
FH299	AAACTTATCACTCAGAAAGATAAC
FH300	ATTGGAGATAATTTACAGGAATTG
FH301	AAACCAATTCCTGTAAATTATCTC
FH302	ATGACGATGGTTCCCATGC
FH303	AACTGTAGTTCCCCTCTAATGGG
FH313	CGGAACTAACTCTGTGGGATG
FH314	CCTCAGCGAGATCGAAGTTAG
FH349	CTGTAAATTATCTCAAGGATTTG
FH350	GAGATTGGGTATCAGTGG
FH356	ACTCCTCTTAAAGCTTGGCTGCAGGTGGGAAAAAAGGTTGCAG
FH357	GGCGCGCCGAATTCCCGGGGATCCGGCTTTTACAGCATATCCC
FH364	CACCGTCGACTCACTTTTCATTATTACTATTTGTTTATG
FH365	CACCCTGCAGTGGGAAAAAAGGTTGCAG

Table S2: Genotype segregation in #31 T3 progeny plants.

Genotype	Number of plants	Ratio (expected)
Insertions	41	0.33 (0.25)
Deletion	28	0.22 (0.25)
Biallelic	57	0.45 (0.5)

Table S3: Summary of metabolite measurements. Given is the mean value of the relative amount per mg FW of 3-4 biological replicates. Significant differences compared to metabolite levels of *gldp2-2* plants were calculated by two-tailed Student's *t*-test (ns: $P > 0.05$; *: $P < 0.05$; **: $P < 0.01$; ***: $P < 0.001$; ****: $P < 0.0001$). FW = fresh weight; rel. = relative; SD = standard deviation; SL = significance level.

Compound	***gldp2-2***			**Col-0**		
	rel. amount / mg FW	SD		rel. amount / mg FW	SD	SL
3-Phosphoglycerate	6.66	0.54		6.68	0.50	ns
α-Alanine	899.11	321.80		713.82	78.15	ns
α-Ketoglutarate	11.44	3.92		17.78	7.02	ns
Asparagine	38.48	10.83		68.83	12.09	**
Aspartic Acid	981.41	179.86		1252.04	149.98	ns
β-Alanine	9.05	1.68		11.25	2.16	ns
Citric Acid	635.28	223.04		1219.61	319.01	*
Cysteine	1.68	0.10		3.07	1.06	*
Fructose	196.01	50.27		314.68	152.01	ns
Fumaric Acid	5075.62	1512.47		4809.46	478.55	ns
GABA	29.47	14.51		19.50	4.02	ns
Gluconic Acid	1.76	2.42		2.14	1.47	ns
Glucose	543.08	256.64		481.35	261.85	ns
Glutamic Acid	1757.06	300.00		2009.04	114.94	ns
Glyceric Acid	167.21	46.22		176.20	29.50	ns
Glycerol	373.10	80.42		354.08	69.06	ns
Glycine	2293.71	134.30		1660.15	439.37	*
Glycolic Acid	16.95	4.24		89.43	83.09	ns
Hydroxyglutarate	3.87	0.86		6.09	1.03	*
Isoleucine	20.03	18.70		8.80	3.11	ns
Lactose	1.24	0.68		2.61	1.41	ns
Leucine	14.40	14.11		6.45	2.58	ns
Lysine	7.01	0.75		11.13	0.38	****

Maleic Acid	14.84	4.36		20.26	4.89	ns
Malic Acid	957.97	267.61		1325.80	186.62	ns
Maltose	50.50	21.12		38.26	3.06	ns
Mannitol	19.43	1.60		25.46	3.25	*
Mannose	7.12	1.79		9.91	1.58	ns
Methionine	17.61	3.02		21.39	3.42	ns
Myoinositol	404.99	19.58		565.96	184.32	ns
Ornithine	4.29	0.28		5.53	0.44	**
Phenylalanine	30.47	4.83		30.71	3.16	ns
Proline	212.80	304.08		108.27	62.13	ns
Putrescine	46.17	9.83		56.75	18.68	ns
Quinic Acid	0.66	0.14		0.87	0.13	ns
Raffinose	6.54	1.44		21.77	16.04	ns
Serine	1037.40	727.92		2714.00	1471.31	ns
Shikimate	140.73	20.51		138.57	15.12	ns
Sorbitol	1.59	0.21		1.98	0.36	ns
Succinic Acid	56.33	21.67		52.06	4.89	ns
Sucrose	4906.29	1054.74		5975.86	882.50	ns
Threonine	230.01	42.33		335.83	66.47	*
Tryptophan	2.79	1.19		4.41	0.40	*
Valine	86.29	59.68		52.51	11.23	ns
Xylose	35.35	4.47		30.59	2.63	ns
Compound	**Insertion**			**Deletion**		
	rel. amount / mg FW	SD	SL	rel. amount / mg FW	SD	SL
3-Phosphoglycerate	7.63	0.03	*	7.61	0.77	ns
α-Alanine	746.50	142.96	ns	1032.50	182.17	ns
α-Ketoglutarate	13.71	2.80	ns	16.63	2.29	ns
Asparagine	40.85	10.78	ns	38.67	8.99	ns
Aspartic Acid	1058.82	136.83	ns	940.74	164.68	ns

β-Alanine	8.29	2.22	ns	12.80	2.33	ns
Citric Acid	859.88	202.08	ns	594.99	177.63	ns
Cysteine	1.79	0.38	ns	1.28	0.19	*
Fructose	179.52	32.76	ns	368.59	67.82	*
Fumaric Acid	6499.72	1344.50	ns	7262.66	322.06	ns
GABA	19.03	3.94	ns	32.08	13.27	ns
Gluconic Acid	1.25	0.63	ns	0.75	0.03	ns
Glucose	485.68	116.29	ns	967.24	17.98	*
Glutamic Acid	1882.55	261.21	ns	1914.35	220.10	ns
Glyceric Acid	174.15	25.58	ns	134.08	24.68	ns
Glycerol	351.00	40.55	ns	276.52	69.72	ns
Glycine	2027.16	104.27	*	3927.60	644.36	**
Glycolic Acid	33.85	8.97	*	17.60	1.91	ns
Hydroxyglutarate	5.38	0.48	*	5.59	0.37	*
Isoleucine	22.66	17.80	ns	36.22	3.91	ns
Lactose	1.44	0.50	ns	0.98	0.20	ns
Leucine	16.91	13.86	ns	23.60	2.34	ns
Lysine	7.08	1.30	ns	8.56	0.99	ns
Maleic Acid	13.38	7.06	ns	13.86	1.32	ns
Malic Acid	1124.65	142.95	ns	1127.75	117.16	ns
Maltose	47.71	9.48	ns	66.02	9.91	ns
Mannitol	17.65	2.98	ns	21.57	1.50	ns
Mannose	11.33	1.62	*	9.22	1.40	ns
Methionine	18.15	3.18	ns	18.05	2.30	ns
Myoinositol	538.36	68.92	**	673.50	209.20	*
Ornithine	4.04	1.32	ns	5.10	1.01	ns
Phenylalanine	30.62	3.56	ns	37.65	1.70	ns
Proline	291.15	278.26	ns	441.28	83.42	ns
Putrescine	37.71	14.96	ns	42.46	1.35	ns
Quinic Acid	0.76	0.18	ns	0.68	0.24	ns

Raffinose	15.76	7.75	ns	27.66	11.02	*
Serine	1331.55	241.49	ns	1555.38	459.87	ns
Shikimate	151.13	3.40	ns	156.72	11.83	ns
Sorbitol	1.69	0.35	ns	1.63	0.11	ns
Succinic Acid	58.07	12.39	ns	83.94	2.38	ns
Sucrose	5874.51	738.63	ns	5995.31	327.33	ns
Threonine	276.07	38.71	ns	276.08	80.60	ns
Tryptophan	4.12	0.80	ns	2.66	1.00	ns
Valine	89.77	50.02	ns	136.59	9.32	ns
Xylose	35.22	4.01	ns	34.12	2.25	ns

CONCLUDING REMARKS

At the beginning of this PhD thesis, the CRISPR/Cas9 system had just started to emerge as the new gold standard of genome editing tools. Only few reports on successful Cas9-based gene editing in plants were published (reviewed in Belhaj et al., 2013) and the inheritance efficiency of generated mutations was not reported. This thesis therefore aimed at establishing a visual marker system to evaluate and track CRISPR/Cas9-mediated genomic modifications in the model plant *Arabidopsis thaliana*.

In the study that is summarized in Manuscript I, two alternative visual markers were compared in *Arabidopsis*. The *BIALAPHOS RESISTANCE* gene, which confers resistance to the herbicide glufosinate, did not provide an efficient marker for regeneration of stable mutants. In contrast, this thesis established the *GLABROUS1* gene as marker for successful genome editing. Using trichome formation as direct read-out for Cas9-induced mutations, T-DNA free knockout mutants were recovered, which transferred the mutations in progeny generations (Hahn et al., 2017b). Following up that work, we published a detailed cloning strategy for a Cas9-knockout vector system as well as a description to detect targeted mutations *in planta* (Manuscript II; Hahn et al., 2017a).

While reports of gene knockouts using CRISPR/Cas9 became more and more frequent in the following years, the introduction of foreign genes into Cas9-induced double-strand breaks remained challenging. This was mainly caused by the intrinsically low frequency of homologous recombination in plant cells (Puchta, 2005). However, two methods were described to enhance frequencies of gene targeting, one employing viral replicons (Baltes et al., 2014), the other one (called *in planta* gene targeting) using template mobilization by Cas9 (Schiml et al., 2014). In this thesis, both approaches were tested in *Arabidopsis* by targeting a defective *glabrous1* gene and using trichome formation as read-out for successful gene targeting. Stable gene targeting events were generated using the *in planta* gene targeting approach, underpinning the high potential of this technique. This work was summarized in Manuscript III that was recently submitted to Frontiers in Plant Science (Hahn et al., 2018).

Last, this thesis applied the CRISPR/Cas9 system to a biological problem. Yields of C_3 plants are highly affected by photorespiration (Walker et al., 2016). This essential, but wasteful salvage pathway detoxifies 2-phophoglycolate, a product of the oxygenation reaction of RuBisCO. In this process, glycine is decarboxylated by the GLYCINE DECARBOXYLASE system, leading to a

net loss of CO_2 (reviewed in Peterhansel et al., 2010). Manuscript IV describes Cas9-induced changes in the promoter region of the *Arabidopsis GLYCINE DECARBOXYLASE P-PROTEIN1* to restrict GLYCINE DECARBOXYLASE activity to the bundle sheath cells. Even though further experiments are necessary to quantify how much the promoter editing has effectively restricted the expression of this protein to the bundle sheath cells, the presented strategies provide promising blueprints for similar approaches in crop plants.

References

Baltes, N. J., Gil-Humanes, J., Cermak, T., Atkins, P. A., and Voytas, D. F. (2014). DNA replicons for plant genome engineering. *Plant Cell* 26, 151–163. doi:10.1105/tpc.113.119792.

Belhaj, K., Chaparro-Garcia, A., Kamoun, S., and Nekrasov, V. (2013). Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. *Plant Methods* 9, 39. doi:10.1186/1746-4811-9-39.

Hahn, F., Eisenhut, M., Mantegazza, O., and Weber, A. P. M. (2017a). Generation of Targeted Knockout Mutants in Arabidopsis thaliana Using CRISPR/Cas9. *BIO-Protoc.* 7. doi:10.21769/BioProtoc.2384.

Hahn, F., Eisenhut, M., Mantegazza, O., and Weber, A. P. M. (2018). Homology-directed repair of a defective glabrous gene in Arabidopsis with Cas9-based gene targeting. *bioRxiv*, 243675. doi:10.1101/243675.

Hahn, F., Mantegazza, O., Greiner, A., Hegemann, P., Eisenhut, M., and Weber, A. P. M. (2017b). An Efficient Visual Screen for CRISPR/Cas9 Activity in Arabidopsis thaliana. *Front. Plant Sci.* 8. doi:10.3389/fpls.2017.00039.

Peterhansel, C., Horst, I., Niessen, M., Blume, C., Kebeish, R., Kürkcüoglu, S., et al. (2010). Photorespiration. *Arab. Book Am. Soc. Plant Biol.* 8. doi:10.1199/tab.0130.

Puchta, H. (2005). The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. *J. Exp. Bot.* 56, 1–14. doi:10.1093/jxb/eri025.

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

Walker, B. J., VanLoocke, A., Bernacchi, C. J., and Ort, D. R. (2016). The Costs of Photorespiration to Food Production Now and in the Future. *Annu. Rev. Plant Biol.* 67, 107–129. doi:10.1146/annurev-arplant-043015-111709.

AUTHOR CONTRIBUTIONS

Manuscript I – An Efficient Visual Screen for CRISPR/Cas9 activity in *Arabidopsis thaliana*

Florian Hahn designed and performed the experiments, except cloning of the vectors pKS Chlamy dual and CasYFP (done by André Greiner and Peter Hegemann). Florian Hahn wrote the manuscript and designed the figures. Otho Mantegazza, Marion Eisenhut and Andreas Weber supervised the experimental design and reviewed the manuscript with comments of Peter Hegemann.

Manuscript II – Generation of Targeted Knockout Mutants in *Arabidopsis thaliana* Using CRISPR/Cas9

Florian Hahn designed the cloning strategy, performed the experiments, wrote the manuscript and designed the figures. Otho Mantegazza, Marion Eisenhut and Andreas Weber helped with the experimental design and drafting the manuscript.

Manuscript III – Homology-directed repair of a defective *glabrous* gene in *Arabidopsis* with Cas9-based gene targeting

Florian Hahn designed and cloned the gene targeting constructs, analyzed the mutants, wrote the manuscript and designed the figures. Otho Mantegazza and Andreas Weber supervised the experimental design and helped interpreting the results. Marion Eisenhut helped with the experimental design. Otho Mantegazza, Marion Eisenhut and Andreas Weber reviewed the manuscript.

Manuscript IV – Towards establishment of a photorespiratory CO_2 pump in *Arabidopsis thaliana* using CRISPR/Cas9-mediated modifications of the *GLYCINE DECARBOXYLASE P-PROTEIN* promoter

Florian Hahn designed the experimental strategy, performed the experiments (if not stated otherwise), wrote the manuscript and designed the figures (except Fig. S7). Otho Mantegazza helped in data interpretation, performed principal component analysis and designed Fig. S7. Tabea Mettler-Altmann advised in metabolite extraction, measured metabolites and helped in data analysis and interpretation. Urte Schlüter provided promoter sequences of various *Brassicacaea* species. Kumari Billakurthi advised and supported leaf tissue embedding. Otho Mantegazza, Marion Eisenhut, Tabea Mettler-Altmann and Andreas Weber supervised the

experimental design. Otho Mantegazza, Marion Eisenhut, Tabea Mettler-Altmann and Urte Schlüter reviewed the manuscript.

Submitted Manuscripts

Hahn, F., Eisenhut, M., Mantegazza, O., and Weber, A. P. M. (2017a). Generation of Targeted Knockout Mutants in Arabidopsis thaliana Using CRISPR/Cas9. *BIO-Protoc.* 7. doi:10.21769/BioProtoc.2384.

Hahn, F., Eisenhut, M., Mantegazza, O., and Weber, A. P. M. (2018). Homology-directed repair of a defective glabrous gene in Arabidopsis with Cas9-based gene targeting. *bioRxiv*, 243675. doi:10.1101/243675.

Hahn, F., Mantegazza, O., Greiner, A., Hegemann, P., Eisenhut, M., and Weber, A. P. M. (2017b). An Efficient Visual Screen for CRISPR/Cas9 Activity in Arabidopsis thaliana. *Front. Plant Sci.* 8. doi:10.3389/fpls.2017.00039.

ACKNOWLEDGMENTS

An dieser Stelle möchte ich den Menschen danken, die diese Arbeit mit möglich gemacht haben, entweder durch ihre fachliche Unterstützung oder durch ihre menschliche Anteilnahme an allen Höhen und Tiefen in den letzten vier Jahren.

Prof. Dr. Andreas Weber – Für die Chance, an einem unglaublich spannenden Thema forschen zu dürfen und die Bereitschaft, mich seit neun Jahren immer wieder in seinem Labor arbeiten zu lassen.

Dr. Otho Mantegazza – Für die alltägliche Unterstützung im Labor, die Diskussionen über Ergebnisse und die nächsten Versuche und den Support beim Zusammenschreiben. Es war eine Freude, mit dir zu arbeiten!

Dr. Marion Eisenhut – Für ihren kritischen Blick auf diese Arbeit und ihr Feedback in der AG Eisenhut und als Teil meines Komitees.

Prof. Dr. Peter Westhoff und PD Dr. Shizue Matsubara – Für die fachliche Unterstützung im Rahmen der Komitee-Meetings.

Das Weber Lab – Für all die schönen Momente und Erfahrungsaustäusche… ob der tägliche Gang zur Mensa mit Angelo und Meng-Ying, die schönen Ausflüge mit unserer Social Activities-Gruppe, der gemeinsame Stiftepool und Spotify-Playlistengeschmack mit Marc, die Therapiesitzungen mit Nadine, die Buzzermaschine von Franzi, Björn und Ana, lange Nachmittage im Frauen-und-Minderheiten-Schlauch und die schönen Karnevalsfeiern und Barbecues mit dem ganzen Rest der Bande.

All die helfenden Hände im Labor und darüber hinaus – Generationen an HiWis, Frau Nöcker, Kirsten, die Gärtner, und noch so viele mehr, die geholfen haben, dass die vielen kleinen Dinge alle reibungslos liefen.

Den CRISPR Kollaborateuren – Für das Hinaustragen von CRISPR/Cas9 in die Welt.

Den iGRAD Plant Teilnehmern – für den Austausch unter Leidensgenossen, das über-den-Tellerrand schauen und die Hilfsbereitschaft untereinander.

Meiner Familie – Für 30 Jahre lang Unterstützung, egal worum es geht.