UC San Diego
UC San Diego Electronic Theses and Dissertations

Title
Engineering Agrobacterium tumefaciens to Facilitate and Improve Knock-in Efficiency in Plants via Cas9 Editing Strategies

Permalink
https://escholarship.org/uc/item/60x3s3t5

Author
Kirby, Adam John

Publication Date
2020

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA SAN DIEGO

Engineering *Agrobacterium tumefaciens* to Facilitate and Improve Knock-in Efficiency in Plants via Cas9 Editing Strategies

A Thesis submitted in partial satisfaction of the requirements
For the degree Master of Science

in

Biology

by

Adam John Kirby

Committee in Charge:

Professor Martin F. Yanofsky, Chair
Professor Lisa McDonnell
Professor José Pruneda-Paz

2020

Copyright

Adam John Kirby, 2020

All rights reserved.

The Thesis of Adam John Kirby is approved, and it is acceptable in quality and for publication on microfilm and electronically:

Chair

University of California San Diego

2020

DEDICATION

I would like to dedicate this thesis to everyone that has shown me unconditional support and guidance. To my family, and friends thank you for the continual love and encouragement throughout this journey.

TABLE OF CONTENTS

Signature Page iii

Dedication iv

Table of Contents v

List of Figures vi

Acknowledgements vii

Abstract of the Thesis viii

Introduction 1

Materials and Methods 31

Results 34

Discussion 51

References 57

LIST OF FIGURES

Figure 1: DNA Damage and Repair [1] …………………………………………………….5

Figure 2: Sanger Sequencing [10] ……………………………………….……………10

Figure 3: Mutation frequency of a Cas9 system shown in both targeted (CAN1) and non-targeted (LYP1) genes [10] ………………………………………………………….…………13

Figure 4: pTIBO542 is the strain of Agrobacterium used in this study. Above is a schematic showing the different regions of the TI-plasmid within this strain and more specifically the virulence region operons. ……………………………………………...………..……35

Figure 5: Shown above is a schematic illustration the process of generating our mobilizable plasmid that is contained within our donor strain in the triparental mating event. ……..… 39

Figure 6: Illustrated below is the process of triparental mating used in this study to create our VirD2-null Agrobacterium-α. ……………………………………………..…………41

Figure 7: This is a representation of the methods we will use to assemble our mobilizable plasmid within our donor strain. The resulting pUC18T plasmid will be used for the triparental mating event in which we will create our in cis strain of Cas9-VirD2 Agrobacterium. ….. 35

Figure 8: Outlined above is the triparental mating insertion event where our Cas9 fused to VirD2 will replace the existing Kan-SacB genes via homologous recombination. The resulting strain of Agrobacterium will be used for in cis testing of our Cas9-VirD2 system. ………………..44

Figure 9: Generation of pUC18T donor plasmid for triparental mating of Kan-SacB deletion and resulting triparental mating product. ………………………………………………………49

Figure 10: Generation of the pGERARD plasmid for in trans expression of Cas9-VirD2 and transformation into the VirD2-null Agrobacterium-β detailed in figure 6. ……….……… 50

Acknowledgements

Firstly, I would like to thank Dr. Martin Yanofsky for giving me the opportunity to work in the lab three years ago. Since then I have grown tremendously both as an academic and as a person. I will always be grateful for his friendship and support in my academic career.

I would like to thank Dr. Jose Pruneda-Paz for agreeing to be on my thesis committee and taking the time to read my thesis and attend my defense.

I would like to thank Dr. Lisa McDonnell for being a continual mentor of mine and giving me the opportunity to teach with her during my time as a Master's student. The memories and experiences I have gathered during this time will always be a highlight of my time at the University. I appreciate all of the guidance and mentorship she has provided me with, and the confidence she displayed in me.

Finally, I would like to thank Dr. Juan-José Ripoll (JJ) for all he has done for me. JJ has been a mentor, a teacher, and most importantly, a friend. Without his unrelenting support, I would not be where I am today. While there was also frequent moments of unapologetic honesty and brutal critique, I appreciate it all the same. You have been an amazing friend and mentor and I will always remember and be grateful for all you did for me.

ABSTRACT OF THE THESIS

Engineering *Agrobacterium tumefaciens* to Facilitate and Improve Knock-in Efficiency in Plants via Cas9 Editing Strategies

by

Adam John Kirby

Master of Science in Biology

University of California San Diego, 2020

Professor Martin Yanofsky, Chair

The development of efficient tools for plant genome engineering is key to dramatically expedite basic research approaches and, at the same time, facilitate translational and biotechnological strategies in agriculture to improve crops and meet future food demand. For decades, technologies enabling precise Gene Targeting (GT) and efficient DNA knock-in or sequence replacement via Homology-Directed Repair (HDR) in plants has remained challenging, and although some recent reports suggest that progress is being made, no efficient and

reproducible protocols have yet been established. *Agrobacterium tumefaciens*-mediated insertion entails the mobilization of the T-DNA into the host plant genome. However, this process occurs randomly, and the system is very inefficient for targeted knock-ins. For this project, we capitalize on the extensive knowledge of *Agrobacterium tumefaciens* biology and combine it with CRISPR-Cas9 genome editing technologies to develop a novel strategy for high efficiency targeted knock-in. To do so, we decided to engineer Agrobacterium tumefaciens to help us reduce the distance between the T-DNA (harboring the cassette for knock-in) and the Cas9 cut site that marks the location where specifically the targeted insertion will occur. The set of tools we are presenting here will serve not to only accelerate basic research but also assist in engineering crop genomes in order to meet our future food demand in an efficient and sustainable manner.

INTRODUCTION

The Challenge of Feeding the World

7.8 billion, the number of mouths we are challenged to feed today. 9.8 billion, the projected world population in thirty years. Both staggering values and terrifying values. In a mere thirty years the world population is projected to climb 25%. With advances in technology, medicine, and infrastructure it is no mystery as to why the world's population has skyrocketed in past 200 years and why it is projected to continue on this rapid trajectory. While there is a plethora of issues stemming from the increasing world population, there is one issue that will prove to be both immediate and powerful. That issue is food. Feeding a growing population is a constantly adaptive and challenging task that only grows in its importance as time goes on. Factors that are affecting this range all the way from population to limits on natural resources (water, sustainable land, favorable climates, etc.) [24]. In our history growing food demand has been balanced by increasing the amount of agricultural yield largely by expanding cultivated lands [24]. However, this will lead to the inevitable bottleneck of the food demand exceeding the availability of sustainable land for agriculture. This looming moment must be the forefront of our thought as a society as we move forward in our history. Addressing the needs to increases crop yield to meet the needs of our growing world must be done.

The basis of this challenge comes from the inevitable bottleneck between population growth and viable land. So, to combat this we can bypass this bottleneck by attacking the actual yield from the existing agricultural production rather than expanding its footprint. If we can successful increase the efficiency of crop production, we can decrease the amount of space needed to produce it. We also face challenges of crop production with respect to initiated climate

changes and environmental impacts of global expansion. The discovery of increasingly sustainable and low impact crop production will prove to be one of the biggest obstacles to our future existence.

An early example of this can be found in what the coined the Green Revolution during the 1960s and 1970s. During time a man by the name of Norman Borlaug set out to create a strain of wheat that was far more resilient than what was being grown at the time. He was able to accomplish this goal through realizing he could produce more wheat in a smaller field by simply increasing the yield of each individual stock. He was able to increase the yield by growing shorter sturdier wheat stocks that could accommodate for the increases weight of more grains at its head. He originally accomplished this in Mexico and drastically improved their agricultural yield and the rates of malnutrition dropped significantly. He was then asked to expand his creation to other countries experience a food crisis during times of drought and famine. His crosses of different wheat strains known for various specialties netted the creation of an optimized wheat strain that sparked the Green Revolution. This movement lead to improvements in irrigation, and fertilization aimed at increasing yield in major crops. Whereas this movemen set the path to follow to prevent future food crisis, unfortunately they proved to simply delay the crisis we are experiencing today. However, the progress and the "push-forward" that was attained from this revolution is exactly what needs to be built upon for the challenges that our world is currently facing.

A key element to more sustainable agriculture and meeting future demands, resides on further increasing crop yields per unit of area (production footprint). One of the most promising strategies to accomplish this goal is based on genome editing and genome engineering tools to tailor crops. The goal is to create more resistant and efficient (less use of fertilizers, water, …

etc.) crops. There are a variety of avenues that can increase a crops yield, whether it be improved ability to hand biotic/abiotic stress or manipulating growth timelines [25]. We can also address the efficiency of a crop by increasing the nutrients present in it at the time of harvest [25]. All of these goals can be achieved by leveraging the vast knowledge earned from basic and applied agricultural research, together with the implementation of efficient genetic engineering and genome editing. Although this could sound like science-fiction not that long ago, there are already examples that prove this path to be successful. One of the examples is rice in which there exists a genetically modified line that contains traits for increasing the number of grains per plant, the size of these grains, and the amount of plants that can be grown in a given area [25]. These benefits were attained through manipulations of various traits that control for response to different stresses [25]. Therefore, by targeting traits connected to tolerance of various environmental extremes, the overall crop failure rate can be drastically decreased [25]. Expanding on this, genetic elements that confer biotic resistances allows for tolerance of prokaryotic factors such as various bacterial, and viral pathogens which can knock out entire crop fields [25]. This resistance can be engineered in the genome by introduction of genes that manufacture chemicals toxic to various herbivores which allow for an even greater amount of resilience [25]. This makes it evident that genetic engineering of essential crop plants can lead to a significant increase in their efficiency. An increase in a crops efficiency directly correlates to more production from less space which will help move us away from that yield bottleneck.

According to the International Service for the Acquisition of Agri-Biotech Applications, in the year of 2018 around 191.7 million hectares of biotech crops were planted [26]. Of these the most prevalent were soybeans, cotton, maize, and canola [26]. This explosion of biotech crops has directly resulted in roughly 24% yield of the global agricultural market [26]. More

importantly this has increased the crop per unit area output of nearly 17 million farmers globally [26]. This is the progress needed to address the approaching issue of food insecurity on a global scale that will be a reality rather than a fear in the decades to come due to our current population growth and limitations on farmable lands.

As is normal for most advancements and revolutions, there has been push back with regards to genetically modified organisms. GMO’s are criticized by some due to their belief that GMO’s are not natural. However, an important distinction to keep in mind is that none of the crop plants in today’s world are “natural”. All crops plants are the result of extensive informal breeding over the many years we have been growing them. Whether the crop has been chemical modified, breed specifically, or informally selected for they can all be deemed as unnatural. A very key concept that helps clarify the validity of GMO products is the fact that they will not harm you. Manipulating the genome of a crop to improve its yield carries no relation to it being unsafe. Rather a GMO is simply a different version of the crop in the same way two people are different versions of a human.

With this in mind it becomes clear the essential role of genetically engineered crops in our society’s future. In order to effectively continue growth and progress in this field the techniques for gene editing in plant organisms must be a topic of focus. In order to optimize our abilities to produce and engineer favorable crops, we must optimize our strategy for inducing genetic change in plants.

The Birth of Gene Editing

Genome editing can best be described by the act of changing various piece(s) of DNA. Since it is known that DNA is the basis of all cellular life and function, manipulation of DNA becomes not only essential, but powerful in the search for answers. It is searching for answers and solutions to the challenge's scientists face in uncovering the mysteries of nature. Now genome editing is a broad term and encompasses a vast web of DNA modifications. These

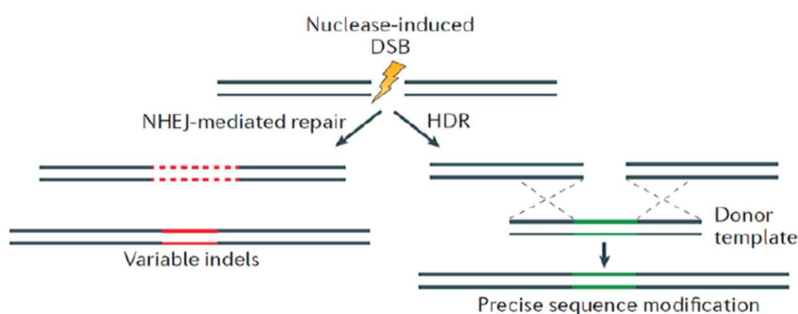

Figure 1: DNA Damage and Repair [1]

modifications of double stranded DNA can be in the form of intentional change, such as breakages through introduction of endonucleases, or accidental cleavage via radiation or chemicals (see Fig. 1) [1]. The accidental changes were actually the birth of modern genome editing as this mechanism takes advantage of known cellular responses to DNA damage. Non-homologous end joining, and Homology directed repair are just two mechanisms by which DNA can repair damage (see Fig. 1) [1]. It is through these processes that mutations and changes of any sort are introduced into the genome. These changes are the first products of genome editing. Unknown and random alterations in a genome sequence that was a direct effect of induced or non-intended damage.

This ability to induce a change in the DNA of cells is essential to our ability to answer questions and on the other hand, it is where the power-train of evolution resides. DNA is the

backbone of all life and any changes to it will result in downstream effects. These effects may not always manufacture a visible phenotype to the eye, however upon close inspection they almost always create a change. The power and importance of genome editing derives from these changes. The quickest way to determine the function of a gene is to remove it or hinder its function. By forcing change of varying degrees, we can begin to understand the functionalities and mechanisms that all originate from nature's code, DNA.

Discoveries of Gene Editing Tools

In order to trace the lineage of the first case of gene editing, it is necessary to understand the steps that had to occur to develop our understanding to a point where gene editing first became a thought let alone a reality. During the late 1960's DNA ligases were first learned of by the Gellert, Lehman, Richardson, and Hurwitz laboratories [2]. The mechanism by which DNA joins the free 3' OH ends with a 5' PO_4 to form a phosphodiester bond became an essential tool in the birth of recombinant DNA technology, the forerunner of genome editing. In later years Hamilton Smith conferred the ideas brought forward by Werner Arber that there were a set of enzymes specifically geared to create double stranded breaks at predesignated spots in the genome [3]. Smith was able to then purify one of these restriction enzymes from a bacterium called *Haemophilius influenzae* (*H. influenzae*) and demonstrated its ability to cut a specific six base sequence segment in DNA [3]. The combination of these discoveries played key roles in the first creation of an engineered DNA plasmid. Paul Berg received the Nobel prize for his groundbreaking work on recombinant DNA methods and was able to combine a cut piece of a lambda bacteriophage, the *E. coli* galactose operon, with SV40 DNA creating the first recombinant DNA plasmid [4]. This first example of DNA recombination served as a basis for

the theories behind genome editing and ushered in a new era of biology. The mystery of how DNA codes for life itself can be distilled to a very simply concept, namely that the order of bases contained in DNA is the key to all life. Thus, DNA is a universal molecule driving functions throughout nature and it is not limited to the organism to which it originates from.

From here the ability to manipulate genes became a field of synchronous interest between industry and academia. However, the takeoff of genome editing in recent years was the result of a wide gambit of individual discoveries and milestones of years past. Each one of these moments on their own may have seemed small, however in the grander scope, when put together with the technologies that followed them, they became essential tools in everyday genetics and genomics research with its many applications.

The Growth of Gene Targeting

The concept of gene targeting was first successfully proven in 1979 when Ron Davis and S. Scherer's laboratories inserted a selectable marker into a yeast genome [6]. This was done using the *his3* gene in *Saccharomyces cerevisiae* as a donor region to accept the homology direct recombination of the *ura3* gene used for selectable marking. This proof of concept paper was essential for future gene editing endeavors by proving not only the ability to recombine foreign DNA genes into an existing gene but to use these genes as a phenotypic marker. Due to the inherit inconsistency of homology recombination the presence of a selectable marker would become a vital tool.

This discovery of gene targeting would pave the way for the next milestone in 1986. Mario Capecchi would attempt to replicate this gene replacement in yeast with a model used in mice. The goal was to inject homologous genes of varying mutational patterns directly into the

nucleus of mammalian cells [7]. He would be successful in doing this and pave the first steps in the creation of an Embryonic stem cell bank of varying knockouts in mice. This is a very important step in the history of gene editing as it proves that a functional gene can be targeted, and replacement using a similar gene with a slightly different function. Also it shows the evolution of the model organisms from a simple eukaryotic model in yeast to a more complicated organism in mice. This in ways paved the advancements to using more higher order complex organisms for genetic studies. However, these techniques are still leaning on random chance. Their basis comes from the hope that homology recombination will occur and for Mario and his team this hope was grounded in a 1 in 10,000 cell chance. So, the question then becomes how can this event be triggered?

Early Strategies for Genome Editing in Higher Eukaryotes

That very question was addressed by Maria Jansin in 1994 when she successfully introduced a double stranded break into a mouse genome via an endonuclease, I-SceI[5]. She theorized that in order to maintain the integrity within a genome any form of double stranded break must be repaired. This led to the realization that homology recombination serves as a repair mechanism for DNA damage within cells. Therefore, successful introduction of a double stranded break should lead to a recombinational event. Using I-SceI cleavage sites in neighboring neomycin phosphotransferase genes, she was able to express a rare cutting endonuclease that successful cut at these sites nearly 12 percent of the time, with 70% of these instances being cuts at both sites [5]. This result meant that she could introduce a homology recombination event nearly 120,000 times more efficiently that anyone could prior. But there was still a problem, this endonuclease while highly efficient in the presence of its cut sites, still

required sequence specific sites. So, with the knowledge that double stranded breaks can be induced in a targeted fashion, and these breaks can lead to cell repair mechanisms that will utilize homology directed repair, the question evolved to how can a truly targetable double stranded cut be made?

Introduce, Zinc finger nucleases as a restriction enzyme composed of zinc finger DNA binding domain and DNA cleavage domain. Discovered in 1985 by Aaron Klug through biochemical studies of the Xenopus protein, transcription factor IIIA, Zinc finger nucleases are composed of a three-dimensional domain that allows for distinction and recognition of closely related DNA sequences [8]. Klug revealed that these Zinc finger domains can be linked to recognize specific nucleic acid sequences, and through interaction with its cleavage domain, introduce a double stranded break at these sequence specific locations [8]. This meant that for the first time targeting double stranded breaks within a genome became a reality. Chance had become obsolete and the only obstacle between gene editing/targeting was knowledge of the sequence. The power of this discovery became evident in 1994 when a Zinc finger nuclease was composed to repress a human oncogene that was transformed into a mouse genome [8]. Gene editing is now leaping to the forefront of scientific ingenuity and primed for a major revelation that will change how scientists address questions for the foreseeable future. However, a problem still remains in how can a sequence be targeted if the sequence itself isn't known?

Development of Genome Sequencing

Fast-forward one year to 1995, Craig Ventor and team of his colleagues published the first ever fully sequenced genome [9]. This was accomplished in the organism *Haemophilus Influenzae* Rd by uses Sanger Sequencing to sequence random pieces of the whole genome and

assembling them through analyzing overlapping regions and ultimately creating a map of the entire genome sequence [9]. While groundbreaking this project was not a model for future regular use as sequencing of the entire genome took roughly a year, but this would not last as advances in the field were coming.

The field of Sequencing really took off in 1977 with the introduction of Sanger's chain termination method (see Fig. 2) [10]. This method uses deoxyribonucleotides (dNTP's) chemical

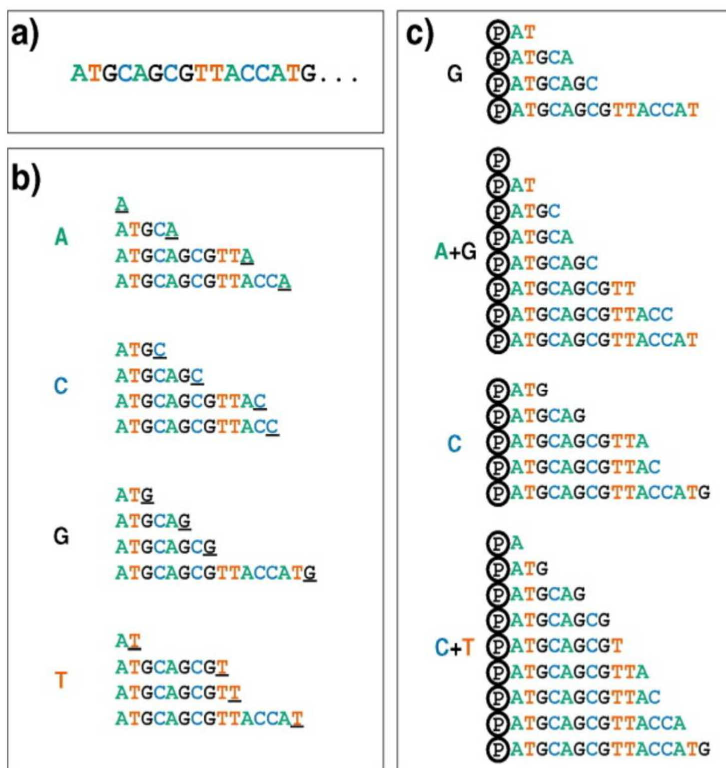

Figure 2: Sanger Sequencing [10]

analogues dideoxyribonucleotides (ddNTP's) as monomers to DNA extension. The dideoxyribonucleotides are composed of a nucleotide that lack the 3' hydroxyl group which is an essential component in the formation of a phosphodiester bond. Thus, in the chain extension during DNA elongation the dideoxyribonucleotides is unable to bind to the 5' phosphate of the sequential nucleotide which in turn terminates the chain [10]. The four ddNTP's are radiolabeled and mixed into four separate reaction pools for DNA extension. Therefore, in the reaction process various chain segments of DNA are formed and these variations cover every possible

combination of the template strand [10]. These mixtures of fragments is then run on four lanes of a polyacrylamide gel which separates the chain fragments by length and using autoradiography the original DNA sequence is revealed [10]. This technique allowed for an explosion of discoveries as knowledge of a genomes sequence allowed for much more specific and focused studies.

Challenges of Genome Editing

Now with the power of targeted nuclease activity the possibilities for gene editing could take off. However, there were several issues with Zinc Finger Nucleases that proved difficult. Firstly, they were extremely difficult to engineer, they were unproven in binding a long stretch of nucleotides with a high affinity, and there were issues when it came to highly specific targets as they were limited to target binding sites every 200bps within the genome [11]. All of these issues could have been addressed and dealt with but the biggest issue was its tendency to cut not only its target region but multiple off-target sites as well [11]. Unfortunately, the coordinated attack and repair of specific gene locus could be completely undermined by an off target cutting by the Zinc Finger Nuclease. Screening for any non-target mutations would have to accompany any successful finding utilizing a Zinc Finger nuclease which inherently handicapped this technique when it came to gene editing.

Discovery of CRISPR-Cas9 System

This major complication opened the door for a system that was primed to alter the landscape of genetics for the foreseeable future. In 2005 a distinct set of repeated DNA sequences were discovered in prokaryotes [12]. This finding was a culmination of screening

across multiple prokaryotic genomes and was found to be present a wide array of them [12]. This highly conserved repeat of intervening sequences led them to the theory of their involvement in an innate biological immune response [12].

Then in May of 2005 a novel Cas protein with suspected nuclease activity was revealed [13]. They discovered this upon investigating *Streptococcus thermophilus* and comparing the novel Cas genes to the known ones at the time [13]. Their findings also revealed a motif called PAM which is a protospacer adjacent motif that was always downstream of the spacer's homologous region in the viral genes [13]. So, with a known Cas protein suspected of nuclease activity (later defined as Cas9), the question became how does it target?

In 2008 findings detailed that these spacers homologous to viral genes are transcribed into small RNA molecules [14]. These RNA molecules, coined CRISPR RNA's, then bind to the endonuclease subunit of the cascade (Cas9) and allow for targeted cleavage of the foreign viral DNA [14]. This finding was revolutionary to the understanding of the system as it proved this mechanism was independent of an RNAi mechanism previously known. The transcribed guide RNA was a sequence specific guide for the attacking Cas9 nuclease.

Application of the CRISPR-Cas9 System

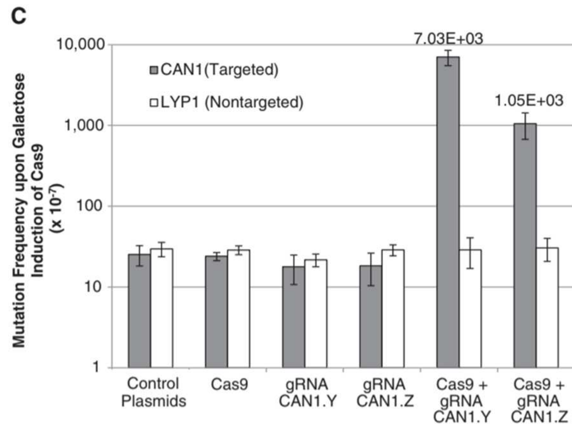

Figure 3: Mutation frequency of a Cas9 system shown in both targeted (CAN1) and non-targeted (LYP1) genes [10]

It was not long before this bacterial defense process was recognized as a tool that could be used for site specific targeting of a genome. In 2013 the viability of the CRISPR-Cas9 system as a tool for targeting gene mutation was published. The data proved the effectiveness and the specificity of the system in *Saccharomyces Cerevisiae* by showing its ability to induce site specific mutations in the *CAN1* gene (see Fig. 3) [15]. Their findings also provided valuable insight that Cas9 alone was not toxic to cells when present in high concentrations [15]. The data proved that engineering and introducing this system into yeast was harmless [15]. However, most importantly these results ruled out the possibility of off target random mutations or cutting [15]. With this knowledge, CRISPR-Cas9 jumped to the forefront of gene editing as the primary way to introduce site specific mutations into virtually any genome of choice.

With this new technology the world of possibilities expanded greatly and now work in higher order eukaryotic cells became a much more feasible option. Prior to the discovery of CRISPR-Cas9 the complexity of eukaryotic cells proved to complicate many genetic studies. Their large genome, gene redundancy, and complex pathways created obstacles that were

otherwise relatively minor in prokaryotic organisms. CRISPR-Cas9 mitigate these challenges by providing a highly specific targeting system for precise and controllable DNA breaks. In 2013, the first genome editing model in eukaryotic cells using CRISPR-Cas9 was published [16]. They were able to target with a guide RNA an endogenous locus in both human and mouse cells and induce a double stranded cut using CRISPR-Cas9 [16]. This break was repaired with a homology directed repair template and revealed little to no mutagenic effects [16]. This was a very important moment for the evolution of CRISPR-Cas9 as it was shown to be applicable and effective in eukaryotic cells. Now the ability to create specific edits in eukaryotic genomes became a viable goal. Further in the paper they illustrate the ability of the CRISPR-Cas9 system to accommodate multiple guide RNA's by combining it into one cassette that can induce simultaneous edits along varying sites within a given gene [16]. This finding proves the power of the system as it can be highly efficient and adaptable.

With the discovery of CRISPR-Cas9 as a highly efficient gene editing system, it has been shown to work in both prokaryotes and eukaryotes. So the question arises, what are some applications for gene editing in plants?

***Arabidopsis thaliana* as a Model Organism**

The field of plant biology has greatly benefited from the identification of *Arabidopsis thaliana* (hereafter, *Arabidopsis*) as its model organism. With other fields already focused on streamlined research in certain model organisms, plant biologists quickly recognized the need to do the same. The field of botany led to studies on a wide array of plants and outlined the many unique characteristics of the various species. In the 1980s *Arabidopsis* became the agreed upon model organism for plant research due to its characteristics making it very research friendly [49].

Arabidopsis is a small dicotyledonous species and very closely related to many important crop plants such as cabbage, canola, and broccoli [18]. It is characterized by a multitude of beneficial traits that make *Arabidopsis* ideal for basic research purposes such as a very small size (allowing for large scale growth in a smaller area), short generation time (typically 5-6 weeks), and a small well-characterized genome [27]. Additionally, through its life cycle it produces many self-progeny which makes for a high yield in each of the different generations when tracking gene inheritance [18]. There is also the ability to do perform cross-pollinations very easily which greatly facilitates genetic studies. These combined traits allow for a very easy and controlled growth in laboratory settings.

Thanks to the research efforts devoted to *Arabidopsis*, the plant research community has facilitated the discovery of widely conserved critical genes involved in primary development programs in plants (flower, fruit or root development; circadian clock, hormone action, etc.) and more recently many tools for gene editing. In the past thirty years publications revolving around *Arabidopsis* has increased significantly. These publications have led to the first complete sequence of a plant genome in 2000, cDNA libraries, new methods of cloning, and one of the most impactful discoveries for an effective transformation of *Arabidopsis* using *Agrobacterium tumefaciens* (hereafter: *Agrobacterium*) [27]. The publication by Feldmann and Marks in 1987 outlined a strategy for *Agrobacterium* transformation of *Arabidopsis* that did not require a tissue culture [28]. This was revolutionary as all other species required a tissue culture, thus making transformation in *Arabidopsis* much easier.

Hurdles of Gene Editing in *Arabidopsis*

With *Arabidopsis* as a model organism it is important to identify the biggest hurdles the scientific community faces with regards to modern plant biotechnology. Although targeting insertions into the genomes in a number of model organisms has become routine (for example, *Drosophila*, mice, fungi, bacteria, etc…), the development of highly efficient gene targeting in higher plants has proven difficult. One of the most prominent cases was in 1998 when the Yanofksy lab successfully disrupted the *AGL5*-MADS-box gene in *Arabidopsis* through homologous recombination mediated by T-DNA transformation [37]. However, in the years following there were few papers publishing examples of gene targeting by HDR in plants. More recently there has been progress on the front through the use of CRISPR-related genome editing strategies, but nothing that changes the landscape. Although recent progress using CRISPR-related approaches are beginning to show signs of success, the method most commonly used to introduce DNA constructs into plants is via *Agrobacterium* T-DNA transfer. However, one of the major issues with this strategy is that the *Agrobacterium* T-DNA insertion is a random event. Thus, we need to devise additional tricks in order to allow *Agrobacterium* to transfer DNA into a plant genome where it will integrate into a specific target site by HDR.

History of *Agrobacterium*

Crown gall is the name of the tumor like growth that can be observed as the base of woody plants, grapevines, and roots of trees [28]. The origin of this growth was unknown prior to the year 1897 when Fridiano Cavara published his findings on the cause of crown gall. He noticed these growths in the Botanical gardens of Napoli Italy and performed studies that showed their ability to grow on grapevines after the introduction of a bacterium isolated from an already matured growth [28]. A few years later in 1907, Smith and Townsend were able to isolate the

bacterium from crown gall on daisies and inoculate other plants leading to formation of crown galls [29]. These findings revealed the infectious properties of the bacterium, later defined as *Agrobacterium*.

Tumor Inducing Properties of *Agrobacterium*

With knowledge of its transmittable infectious properties the focus turned to the discovery of its tumor like properties. Armin C. Braun proposed that the plant derived gall tissues were transformed by this bacterium and it was this transformation that allowed the persistent growth [30]. He was able to show that the crown gall tissue was able to grow indefinitely in culture even after there was no longer an *Agrobacterium* present [30]. Studies began illustrating that normal plant tissues when isolated on media were unable to continue growth without additive hormones. However, in the presence of auxin and cytokinin normal plant tissue was able to grow in isolation on the media [30]. When compared to the crown gall tissues that were isolated on media, they found these tissues to be able to grow even without the presence of supplementary auxin and cytokinin [30]. When compared to normal plant tissue, crown gall tissues' ability to grow indefinitely in the absence of these hormones highlighted its tumor-like properties [30].

Now this *Agrobacterium* is responsible for the infectious transmission and tumor-like properties of the crown gall. With this realization the question quickly evolved to how are these tumor-like characteristics induced by the *Agrobacterium*? The goal became figuring out the mechanism by which this bacterium induces tumors on its host.

Theories for Tumor Inducing Properties of *Agrobacterium*

There were three major ideas at the time on the mechanism of its tumor inducing properties. Firstly, it was proposed that *Agrobacterium* caused production of irritating chemicals which led to tumor formation [28]. This idea revolved around the limited understanding of cancer at the time and claimed that the *Agrobacterium* had the ability to interrupt the natural balance and restraint of cellular functions [28]. Research by Erwin F. Smith proposed this irritant was actually ammonia and *Agrobacterium* role in crown gall formation was the facilitation of its release from the plant cells.

Second was that phytohormone auxin was believed to be the major player responsible for tumor formation [28]. This idea originated from the observation that auxin swellings closely resembled that of many galls and pathological growths [28]. Armin C. Braun and Thomas Laskaris showed that an avirulent strain of *Agrobacterium* was able to induce gall like tumors on tomato plants when supplemented with a synthetic growth substance (the phytohormone auxin) [28]. This work highlighted the function of these synthetic growth substances in triggering tumor development. Therefore, it was hypothesized that *Agrobacterium* were able to produce crown gall in part through the introduction of phytohormone auxin. Due to the positive effect of these phytohormones on the avirulent strain, it was theorized that there might be two separate phases in crown gall formation. These two phases were growth of normal cells and then continued stimulation and tumor formation would point to a combination of the two theories (a need for phytohormone auxin and a chemical irritant) [28].

Finally, it was hypothesized that the host organism was conditioned to promote tumor growth through the formation of a tumor-inducing principle [28]. This came from important findings of White and Braun in 1942. Their work illustrated that bacteria-free gall formation from secondary gall tissue pointed to the fact that there was permanent genetic modification

occurring [31]. The lack of bacterium indicated the inheritability of the gall. Moreover, this gall was able to grow in the absence of any phytohormone supplement. Therefore, they postulated there must be a tumor-inducing principle that is transforming the host plant [31]. Additionally, the presence of a guanidine compound found in crown gall tissues pointed to a genetic transformation event [28]. Extensive studies had shown that opine compounds octopine and nopaline are exclusively found in crown gall tissues [28]. Also, the type of guanidine compound is completely dependent on the strain of *Agrobacterium* that was used to transform the plant. This evidence of plants produce *Agrobacterium* dependent versions of opines in their crown galls was more steam in the direction of genetic transformation for the mechanism of *Agrobacterium* [28].

Tumor-Inducing Plasmid of *Agrobacterium*

Numerous studies have reported the presence of a tumor-inducing plasmid that is contained within all virulent strains of *Agrobacterium*. This plasmid is over 200,000 base pairs in length and confers multiple genes involved in the transfer of genetic information to the host organism. Contained in this tumor-inducing plasmid are two major areas, the T-DNA and the virulence region. The T-DNA is the segment that is physically transferred from the infecting bacterium to the plant host. The virulence region encodes a series of operons (A-F) that function specifically to recognize a suitable plant host and promote the transfer of the T-DNA to the suitable hosts genome.

T-DNA of *Agrobacterium*-mediated Gene Transfer

The T-DNA is the region within the TI-plasmid that is excised and transported to the host plant cell for integration. There are many important regions of the T-DNA but there are five to which I will detail. These are the right and left border regions, the genes necessary for enzyme synthesis of important regulators auxin and cytokinin, and the opine synthesis region [32]. The right and left border regions are the repeated sequences that mark the start and end of the T-DNA's region that will be removed via the VirD2/VirD1 complex [37]. These border regions are what the actual VirD2/VirD1 complex recognizes to achieve the targeted removal of the contained T-DNA region. The genes responsible for enzyme synthesis of regulators auxin and cytokinin are carried within the T-DNA to the plant genome [32]. Once in the genome these regions will cause production of these two tumor growth regulators in the plant cell. Production of these hormones leads to stimulation of cell growth and division while simultaneously decreasing the levels of growth control [32]. It is this two-pronged attack on the homoeostasis within the plants growth cycle that leads to the formation of the tumor like growth we know as crown gall [32]. Finally, there is a gene within the T-DNA that encodes for synthesis of octopine [32]. Octopine is a reductive condensation product of pyruvate with arginine and promotes the condensation of pyruvate with other amino acids to create synthesis of octopinic acid [32]. Therefore, the T-DNA's insertion into the plant genome allows for synthesis of octopines that cannot be metabolized by the plant, but can be metabolized by the *Agrobacterium* [32]. A set of uptake and degradation genes on the TI-plasmid allows for transport in and breakdown of opines by the *Agrobacterium* [32]. Thus, we can see the role of all of these regions within the T-DNA. They are needed for successful removal from the TI-plasmid, infection and induction of tumor-like growth, and synthesis of opines for the infecting *Agrobacterium*.

Virulence Genes of Tumor-Inducing Plasmid

Recognition of a suitable plant host by Agrobacterium

The vir region of the Ti-plasmid consists of six operons (VirA, VirB, VirG, VirC, VirD, and VirE). Of these genes there is a portion that is constitutively expressed and a portion that are triggered upon signal reception [33]. The constitutively expressed genes are VirA and VirG which function together to recognize and respond to the presence of a suitable plant host. The remaining operons within the vir region are induced only after the *Agrobacterium* senses a suitable host. VirA encodes for a histidine kinase receptor which recognizes plant-derived phenolic compounds, such as acetosyrinigone [34]. The acetosyrinigone activates the VirA kinase and leads to phosphorylation of the second player in the initial signal pathway, VirG [34]. The now phosphorylated VirG acts as a transcription factor that promotes expression of the vir operons. It is through this activation pathway that VirA and VirG ensure that the transfer of the *Agrobacterium*'s T-DNA only occurs in proximity of a suitable host

The Mechanism of T-DNA Transfer

The first step in T-DNA transfer involves the recognition of cutting at the T-DNA border sequences by VirD1 and VirD2 [38]. VirD1 acts as a topoisomerase and VirD2 as an endonuclease which in conjunction with each other excise the T-DNA segment from the TI-plasmid [34]. These borders become very important as they are the only regions found to be critical for VirD1/VirD2 recognition [34]. Once this excision has occurred the VirD2 remains covalently bound to the 5' end of the free T-DNA fragment [34]. The fact that VirD2 remains covalently attached to the T-DNA will provide the basis for our approach to increase the efficiency of gene targeting in plants. In addition to its role in specifically cutting at the T-DNA

borders, the VirD protein contain Nuclear Localization Signals that help to move the T-DNA complex into the plant cell nucleus, where the T-DNA subsequently integrates into the plant genome.

Next, eleven proteins coded for by the VirB operon in conjunction with VirD4 come together to create a type four secretion system [34]. This secretion system is a molecular complex often used for transport of ssDNA strands, often lead by a 5' bound protein, through the membranes/walls of gram-negative bacteria [34]. VirC genes carry out functions that aid in this process by binding regions outside of the border regions known as overdrive [32]. This binding of VirC1 and VirC2 allows for increased excision and production of free floating VirD2 bound T-DNA fragments [32]. They also act as accessories to aid in the contact between VirD2 and the VirD4 of the secretion system [32]. The remaining vir genes such as VirD5, VirE2, VirE3 and VirF interact with bacterial factors and seem to be required to mediate the transfer of the now free T-DNA segment to the type four secretion system. Specifically, VirE2 has an important role as it binds the T-DNA strand and protects it against nucleases [34]. It is the sum of these virulence genes functions that allow for a successful mobilization of the T-DNA from the TI-plasmid and through the type four secretion system in order to achieve insertion into the hosts genome.

Critical Importance of VirD2 in *Agrobacterium*-mediated Gene Transfer

VirD2 is one of the Vir genes produced by the *Agrobacterium* and used for both excision of the T-DNA from the TI-plasmid and insertion into the new hosts genome. Unlike other DNA mechanisms for insertion such as transposons, the T-DNA doesn't encode any recombinase enzymes when it is in the host cell. Rather the attached VirD2 act as a mediator for insertion

[20]. Studies have shown that VirD2 fulfills a variety of roles once the T-DNA has entered into the host cell such as protection against degradation by native exonucleases and overall maintenance of the integrity for the right border of the T-DNA [20]. It is able to maintain the integrity of this right border of the T-DNA due to its covalent bond to the T-DNA at this spot. So with VirD2 bound covalently at the head of the T-DNA it also encodes an essential factor that allows the T-DNA to enter the cells nuclease. The VirD2 contains nuclear localization site (NLS) on the C terminus region [35]. The NLS is critical to allow for the T-DNA to enter the nucleus of the host cell. Thus, it becomes clear the critical nature of VirD2 in *Agrobacterium*-mediated gene transfer from the first excision of the T-DNA to the end insertion into the host genome with the host cell nuclei. The nature of this conserved/essential protein (VirD2) in the *Agrobacterium*-mediated gene transfer pathway makes it an ideal player in our new system for efficient gene editing in plants.

Overview of *Agrobacterium*-Mediated Gene Transfer Mechanism

VirA is located in the cell envelope of the *Agrobacterium*. This protein acts as a receptor detecting the phenolic compounds emitted from a plant. The binding of VirA propagates an internal pathway that eventually drives the excision and transfer of a specific region in the TI-plasmid known as the T-DNA. When bound, the VirA phosphorylates the VirG protein. VirG when phosphorylated becomes an active transcription factor necessary for expression of the rest of the Vir operon in the TI-plasmid. Amongst these operons is VirD which encodes VirD1 and VirD2. These are endonucleases that cleave the plasmid. They cleave at two twenty-five sequence repeats that flank the region known as the T-DNA. After cleavage the T-DNA is still bound at the end by VirD2 and is guided into the plant cells nucleus by VirD2. This is

accomplished through export via the VirB/VirD4 constructed type four secretion system and entrance into the nuclease via the nuclear localization signal in the VirD2. Once in the nucleus it can be integrated into the plant cells genome. The result of this process is homology recombination of the T-DNA segment into the plants genome. Not only is the gene of interest embedded within the T-DNA inserted directly into the plant's nucleus, but it is incorporated into the plant's genome. *Agrobacterium*-mediated gene transfer has become one of the most used methods for DNA delivery to plant cells due to its ease of use.

Challenges of *Agrobacterium*-mediated Gene Transfer

There are some issues with this strategy for gene editing in plants. Depending on the goal of the experiment one may be interested in knocking out a gene or with regards to *Agrobacterium*-mediated gene transfer doing a knock-in. For a knock-out it is pretty simple to attain using the method of CRISPR-Cas9 above. A targeted double stranded break without a homology repair template will lead to non-homologous end joining and likely a deletion event in the gene knocking out function. However, to perform a knock-in experiment there must be a way to create a targeted insert in the genome and this is where *Agrobacterium*-mediated gene transfer can be used. Due to its ability to insert the T-DNA into the genome of the plant it is possible to have the T-DNA carry your gene of interest and create a knock-in mutant using this method. There are some inherent issues with this strategy and first of which is that the T-DNA inserts randomly into the genome. A major problem with this reality is if a gene knock-in occurs in non-transcribed region this gene will be effectively silent. Stanton B. Gelvin at Purdue University ran into this problem when attempting to utilize *Agrobacterium*-mediated gene transfer to introduce a selectable marker into plant cells [19]. He discovered that the VirD2/T-DNA complex

preferable formed complexes with euchromatic regions of the genome [19]. Thus, this revealed its affinity to insert itself into transcribed regions of the genome. Furthermore, he found that the T-DNA selectively inserted into genes that held a particularly high A/T content which is normally a characteristic of gene promoters [19]. This shows the ability and the likelihood the T-DNA is inserted into a transcribed region but at a random determination. There were no other patterns or selections for certain types of genes or areas within the genome. The issue with this analysis was he realized he was only tested the plants which showed positive for the selectable marker to which was the package of the T-DNA [19]. Upon testing the negative selectable marker plants, it was revealed that the selectable marker had been randomly inserted into non-transcribed regions of the genome [19]. This revealed the true nature of *Agrobacterium*-mediated gene transfer in that is allows for a random knock-in in the genome. So, performing a true targeted knock-in experiment in plants using *Agrobacterium* has an extremely low efficiency which poses a problem.

Role of CRISPR-Cas9 in aiding *Agrobacterium*-mediated Gene Transfer

With respect to *Arabidopsis*, *Agrobacterium*-mediated gene transfer has been consistently used a strategy for gene manipulation. However, for this to be a viable method of gene manipulation it would have to possess the ability to insert at a target location and do so with a relatively high level of efficiency. A connection between the signature associated with a T-DNA insert and that of native DNA repair mechanisms was revealed [22]. It was determined that the sort of the genome scar that is left at the sight of the T-DNA's insertion was similar to that of the Pol θ mediated error-prone double stranded break repair [22]. This led to the realization that the T-DNA likely targets random double stranded breaks in the host genome and integrations at

these locations thus mimicking the scar signature often seen at locations of error prone DNA repair mechanisms [22]. It likely does so through utilization of microhomology areas that allow for targeted insertion into the area of the double stranded break [22]. So, this opened the door of possibilities for a targeted *Agrobacterium*-mediated gene transfer approach in creating knock-in's in plants. If the gene in question could be selected for through creating a double stranded break, the T-DNA could be designed to have homology to this break region and therefore allowing for a specific insertion location within the host genome. Thus, the CRISPR-Cas9 system could provide the perfect match to *Agrobacterium*. With CRISPR's ability to create a targeted double stranded break in the host genome, a site specific insertion of the T-DNA could be promoted.

Creating Heritable Knock-ins in *Arabidopsis*

This raises the question, how efficient can this targeted system of gene manipulation in plants be when utilizing the *Agrobacterium*'s mechanism for insertion. Due to the ability of CRSIPR-Cas9 to generate a site specific double stranded break in the plants genome there can be targetability of the T-DNA's insertion in the plant. The issue with this comes from the inherent inefficiency of homology repair in higher level plant organisms [23]. This inefficiency was highlighted when researchers attempted to create a heritable knock-in in *Arabidopsis* [23]. They utilized an all in one system that had the T-DNA construct containing Cas9, a sgRNA targeting a site near the stop codon of ROS1 and a donor DNA segment for homology repair [23]. In their T1 plants they found 2/30 positive signals for their knock-in while their control not containing a sgRNA yielded 0/30 [23]. Further analysis of the line revealed no T2 progenies of the positive T1's carried the positive signal [23]. This result shows a lack of heritability of this knock-in and

a very low efficiency even in the presence of a targeted double stranded break. With the evidence pointing to an all in one system being largely ineffective they attempted to do a knock-in by a sequential transformation by first providing the Cas9 and then transforming with the T-DNA carrying the sgRNA, donor DNA template, and a selectable marker [23]. They were able to trace the inheritance through the T3 generation, but the overall efficiency of the knock-ins was extremely low which they measured to be only around 5% [23].

However, a major issue with this finding is they provided no comparison to a sample not containing Cas9. Thus, we cannot conclude the findings to be a result of the conjunction with Cas9. Furthermore, there numbers are very consistent with the efficiency already seen even without using Cas9. So, while a technique for using *Agrobacterium*-mediated gene transfer in conjunction with CRISPR-Cas9 proved to be possible it remained an imperfect system for creating efficient heritable knock-in's in plants. In the results section I will detail our use of *Agrobacterium's* mechanism in conjunction with Cas9 by providing the various models we used. These models will include wildtype *Agrobacterium* as a control, and *Agrobacterium* expressing a chimeric Cas9-VirD2 module both in *cis* and in *trans*.

Our Plan

The development of efficient tools for plant genome engineering is key to dramatically expedite basic research approaches and, at the same time, facilitate translational and biotechnological strategies in agriculture to improve crops and meet future food demand.

For decades, technologies enabling precise Gene Targeting (GT) and efficient DNA knock-in or sequence replacement via Homology-Directed Repair (HDR) in plants remain challenging and, so far, no efficient and reproducible protocols have yet been established. This

roadblock to plant research has been often considered one of the “holy grails” of plant biology and represents a major hurdle that plant scientists must overcome in order to dramatically accelerate basic and applied science. The efficiency of creating knock ‘ins in plants, specifically *Arabidopsis*, is limited by the chance that the T-DNA will come in contact with the region that is cut by the CRISPR-Cas9 system. With the novel abilities of *Agrobacterium* to insert a designable T-DNA into a hosts genome and the specificity of the CRISPR-Cas9 system to create a double stranded break in a complex organism we postulated a mechanism by which we could increase the efficiency of knock ‘ins in the model organism *Arabidopsis*.

A region referred to as the T-DNA (Transferred-DNA) in the Gram-negative bacterium *Agrobacterium* has been used as a vector to transfer genes into the plant cell genome and create transgenic strains or insertional mutant alleles among other uses. Many important crop species are routinely transformed using *Agrobacterium*. However, the integration of the T-DNA into the recipient plant genome is random. The molecular events that occur inside the bacterium during its interaction with a plant host have been studied intensively, and a battery of proteins encoded in the bacteria genome have been shown to be required for the successfully transfer of the T-DNA into the pant cell and finally its integration into the genome. VirD2 is one of the key *Agrobacterium* tumefaciens genes involved in T-DNA processing, transferring and integrating the T-DNA once the T-complex is inside the plant cell. It encodes an endonuclease covalently bound to the 5′ end of the T-DNA that contains an NLS (Nuclear localization Domain) domain that guides the T-complex into the plant cell nucleus preserving the integrity of the T-DNA and participating in T-DNA integration. Previous studies have shown that VirD2 interact with plant cell proteins including intracellular nucleoproteins and DNA integration/recombination/repair

apparatus, to facilitate transfer and integration of the T-DNA. A number of studies have tried to increase the specificity of the integration (targeted insertion) but with very modest success.

The recent development of CRISPR-Cas9 technologies has impacted virtually any field in research as it provides a relatively simple way to engineer genomes, including plants. The resulting Cas9-induced double-strand breaks mostly lead to insertions or deletions (In/Dels) due to the prevalence of non-homologous end joining (NHEJ) over HDR. This is in part due to the difficulties in delivering the DNA template to the desired location to favor HDR to achieve targeted knock-in.

In this context, we wanted to capitalize on the extensive knowledge of *Agrobacterium* biology and combine it with CRISPR-Cas9 technologies to develop a novel strategy for high frequency HDR by increasing the proximity between the repair template and the Cas9 cut site, thus facilitating the precise integration of DNA templates into desired locations in the plant genome. This could be achieved by localizing the T-DNA in the same region as our targeted Cas9 cut by fusing to the Cas9 to the VirD2 that borders the 5' end of the T-DNA. Using this hypothesized system, we could effectively bring the T-DNA directly to the cut site by having it attached to the Cas9. Thus, when the Cas9 complex binds the gRNA and cleaves the *Arabidopsis* genome, the T-DNA would be extremely proximal to this cleavage site. The idea is by decreasing the space between the T-DNA and the cut site we can conversely increase the rate of insertion of our T-DNA at our desired location. In the presence of the T-DNA an HDR event would allow for a knock-in of the any intended sequence carried within the T-DNA. By doing this we would be fusing the functionalities of CRISPR-Cas9 and *Agrobacterium* to facilitate more efficient gene targeting in plant models.

Using this model, we would be successful employing the power of the CRISPR-Cas9 systems specificity, with the vast knowledge we have of *Agrobacterium*. This synergistic approach utilizes our understanding of *Agrobacterium's* infectious abilities and attempts to cooperatively mate it with the precision of the CRISPR-Cas9 system for targeting genome sequences. The set of tools and approaches we are presenting here for creating efficient knock-in gene targeting in plants will be vastly important not only for basic research purposes but also for crop engineering and to be able to meet our future food demand in an efficient and sustainable manner.

METHODS AND MATERIALS

Polymerase Chain Reaction, plasmid isolation, DNA digestion, sequencing and electrophoresis:

Polymerase chain reaction was used frequently in the generation of our constructs used in this project. The amplification templates varied based on the target sequence and fragment we were amplifying. Primers were designed to target the specific fragments used in assembling our plasmids.

For regular bacteria screening we used Choice-Taq (Thomas Scientific Ltd.) following manufacture's guidelines but supplementing the reactions with MgCl2 to a final concentration of 2 uM.

For all our cloning related strategies, amplicons were generated using either KOD Taq-polymersae (Millipore Inc.) or Phusion polymerase (New England Biolabs) depending on the length of the fragments. Both polymerases have proof-reading activity which reduces drastically the incorporation of undesired mutations.

Plasmid DNA digestion was performed according to manufactures specifications and using at least 1 ug of DNA. Reactions were incubated at the corresponding temperature.

PCR results and or plasmid digestions was assays by electrophoresis in agarose gels. If required, DNA bands were isolated by gel excision and purification using commercial (Genessee Ltd) and QIAGEN gel isolation solutions.

For plasmid mini-prep isolation DNA, we used commercial columns Spin-Smart from Thomas Scientific and solutions from Promega.

DNA concentrations were measured using a Nanodrop T2000 (ThermoFisher).

For DNA sequencing, samples were processed according to a local company guidelines (Retrogen Inc.). Results were analyzed using SnapGene.

Gibson DNA Assembly:

In order to integrate all of our fragments into a vector backbone in the orientation we needed, we utilized the Gibson DNA method of assembly. In this procedure all of our PCR fragments were designed with homologous base pair overlaps, and in a fashion that would connect our fragments in the desired order. We also utilized restriction enzyme sites within our vector backbones for the overlapping regions on either end of the vector. Then using the combined activities of the DNA Assembly enzyme (exonuclease, polymerase, ligase) we assembled each of the following plasmids; pUC18T-Kan-SacB (see fig. 5D), pUC18T-Cas9-VirD2 (see fig. 7D), pUC18T-LH-RH (see fig. 9C), and pGERARD-Cas9-VirD2 (see fig. 10B).

Triparental Mating:

Recombination of our wildtype *Agrobacterium* strain (AGL-0) with our donor plasmids (pUC18T series) was accomplished via triparental mating. Using this strategy grow separately the three strains needed; the helper *E. coli* strain (pRK2073), our donor *E. coli* with pUC18T donor plasmid, and our *Agrobacterium* strain (AGL-0) with the disarmed pTIBO542 (see fig. 6A). These are each grown in liquid lysogeny broth (LB), and then combined into a singular culture. Once in high concentrations and mixed together the helper strain can mobilize our donor plasmid into the recipient strain. This mixture is then sown on rifamycin containing media plates (additional antibiotic/sucrose for varying triparental mating events depending on resulting *Agrobacterium*) which screens against *E. coli* as they do not contain a chromosomal resistance to

rifamycin while *Agrobacterium* does. This method was used to generate; VirD2-null *Agrobacterium-α* (see fig. 6B), in *cis* Cas9-VirD2 *Agrobacterium* (see fig. 8), and VirD2-null *Agrobacterium-β* (see fig. 9E).

***Agrobacterium* Transformation:**

We transformed our pGERARD plasmids into our different *Agrobacterium* strains using electroporation in conjunction with a helper plasmid to provide replication functions in *trans*, pSOUP [50].

Primers Used for PCR:

RH	**LH**	**SacB**	**Kan**	**Cas9**	**VirD2**
oJJR1727 agacaggaagg accgaataatgg	oJJR1721 tagggcgaatt gggtaccccag catcgatccctg aaag	oJJR1725 cacatatacctg ccgttcactatta tttag	oJJR1729 ctgaaaggtgc cttgtaggcag ctctggcccgt gtctc	oJJR1742 acctgacgggagaa aattggatggataag aaatactcaataggc	oJJR1535 tccggaggtggatcc ggaggtatgcccgat cgcgctcaagtaatc attcg
oJJR1728 ggccgctctaga actagtagcgtttt gccgagcccgc cttgc	oJJR1722 ctacaaggcac ctttcagtaac	oJJR1726 attcggtccttcc tgtctaaacact atcaataagttg gagtc	oJJR1730 gaacggcaggt atatgtgcggg gtctgacgctca gtgg	oJJR1534 acctccggatccacc tccggaacctccgtc acctcctagctgact caaatc	oJJR1743 tattcggtccttcctgt ctctaggtccccccg cgcccatcgtcgcg

RESULTS

I. <u>Vir Region of *pTIBO542* TI-plasmid</u>

In *Agrobacterium* the TI plasmid contains all the required functions for transferring the T-DNA into the plant cell as well as a set of tumor-inducing genes, such as auxin, cytokinin, and opinine synthesis. However, for plant biotechnological purposes, the tumor inducing genes have been eliminated. These genes, after infection, lead to the growth of tissue we know as crown gall. For the purpose of this study we used the *Agrobacterium* strain AGL-0 which contains the disarmed TI plasmid pTIBO542, which lacks the tumor inducing genes and only contains the essential functions for the transfer process to allow transfer to a plant cell and integration into the plant genome (see Fig. 4).

The pTIBO542 plasmid contains a series of genes organized into operons whose role is key for transfer and T-DNA plant genome integration (see Fig. 4). These operons are the virulence operons (vir). We focused our studies on the VirD operon that contains five genes (VirD1-VirD5), and in particular on VirD2. This gene is key for T-DNA transfer and integration when infecting plants. VirD2 remains covalently bound to the 5’ end of the T-DNA during the whole process. We therefore hypothesized we could fuse it to Cas9 to promote induced targeted knock-in insertions by getting the T-DNA containing the HDR template, near the Cas9 site. In order to accomplish this, we needed to create an *Agrobacterium* strain lacking VirD2 (this strain is not virulent and thus not infectious) for then introducing a chimeric Cas9-VirD2 either in *cis* (in the context of the VirD operon) or in *trans* (expressed from a secondary plasmid under the control of the VirD regulatory sequences).

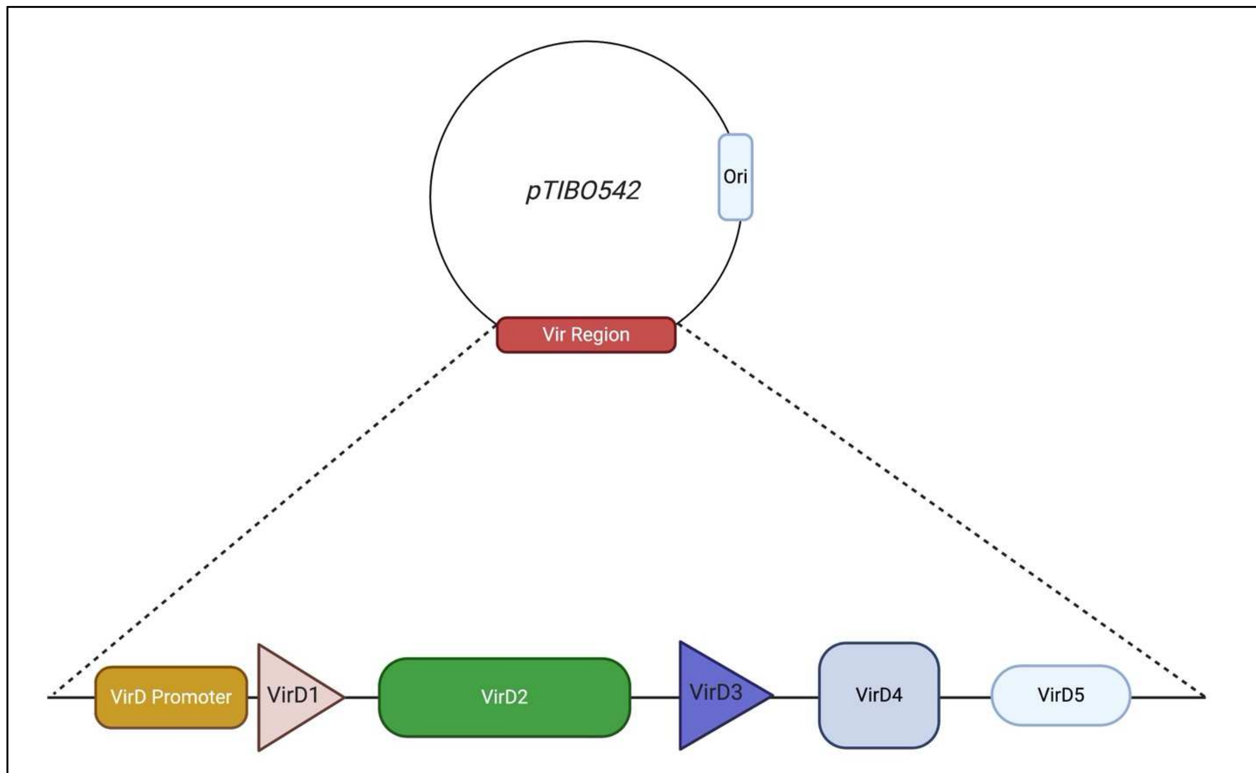

Figure 4: pTIBO542 is the strain of Agrobacterium used in this study. Above is a schematic showing the different regions of the TI-plasmid within this strain and more specifically the virulence region operons.

II. <u>Generation of Donor Plasmid for Triparental Mating</u>

In order to create and screen for a successful knockout of the endogenous VirD2 in wildtype *Agrobacterium*, we followed a triparental mating strategy. To do so, we had to generate a construct compatible with our Agrobacterium strain using the vector pUC18T that can be mobilize into a destination strain (in our case Agrobacterium) with the assistance of a third strain containing the required functions for the mobilization (see III-A for more details).

II-A: Importance of kanamycin in our construct

Since the goal for this triparental mating event is to create a full knockout of the endogenous VirD2 within WT *Agrobacterium*, we needed to replace this region with a form of selection for identification of colonies after successful replacement. In order to do this, we generate a construct with dual selection (positive and negative). This positive selection module contained kanamycin, which will allow us to sow the resulting bacterial conjugate from the triparental mating on kanamycin and rifampicin containing plates (Agrobacterium contrains chromosomal rifampicin resistance). Only the *Agrobacterium* containing a complete deletion will confer the kanamycin resistance, thus allowing for positive selection. This kanamycin resistance gene was amplified from the pENTR plasmid using primers oJJR1729 and oJJR1730 (see Fig. 5A).

II-B: Importance of SacB in our Construct

The second module of our selection process is a negative selection using sucrose. SacB comes from a gram-positive strain of bacteria known as *Bacillus subtilis* and encodes for the enzyme levansucrase. This enzyme converts sucrose into metabolite that is toxic to gram-negative bacteria, such as *Agrobacterium*. Therefore, expression of the SacB gene in *Agrobacterium* will lead to cell death when the *Agrobacterium* is in the presence of sucrose. With this negative selection in place, we can ensure the replacement of VirD2 with our LH-Kan-SacB-RH cassette was successful by ensuring the cells can't grow in the presence of sucrose. This SacB gene was amplified from the pAM5067 plasmid using primers oJJR1725 and oJJR1726 (see Fig. 5A).

II-C: Homologous Border regions of *Agrobacterium* (LH/RH)

Our goal is to achieve a targeted replacement of the VirD2 gene from the VirD operon on the TI-plasmid in our WT *Agrobacterium* with a Kan-SacB cassette. Therefore, we must ensure our mobilizable plasmid is homologous upstream and downstream of the VirD2 gene to promote a targeted replacement (knock-in). and deletion of VirD2. This meant we needed to amplify a 500 base pair region up and down stream of the VirD2 gene, which we labeled as Left Homology (LH) and Right Homology (RH), respectively. Both the LH and RH were amplified from WT *Agrobacterium* using primers oJJR1721/oJJR1722 for the LH and oJJR1727/oJJR1728 for the RH (see Fig. 5A).

II-D: pUC18T with OriT as mobilizable plasmid vector

Finally, we needed a vector to carry our LH-Kan-SacB-RH cassette within the donor *E. coli* strain. We used pUC18T as the vector for our plasmid as it contains an OriT. The OriT is essential for recognition of this plasmid as the mobilizable element by the helper strain. This pUC18T plasmid lacks the mobilization (mob) and transfer (tra) elements needed in bacterial conjugation. Thus, without the presence of a helper strain (pRK2073) providing these genes, it is not able to mobilize. Once in the presence of our helper strain (pRK2073), it will be able to be mobilized into the *Agrobacterium* (AGL-0) and carry our LH-Kan-SacB-RH cassette (see Fig 5D).

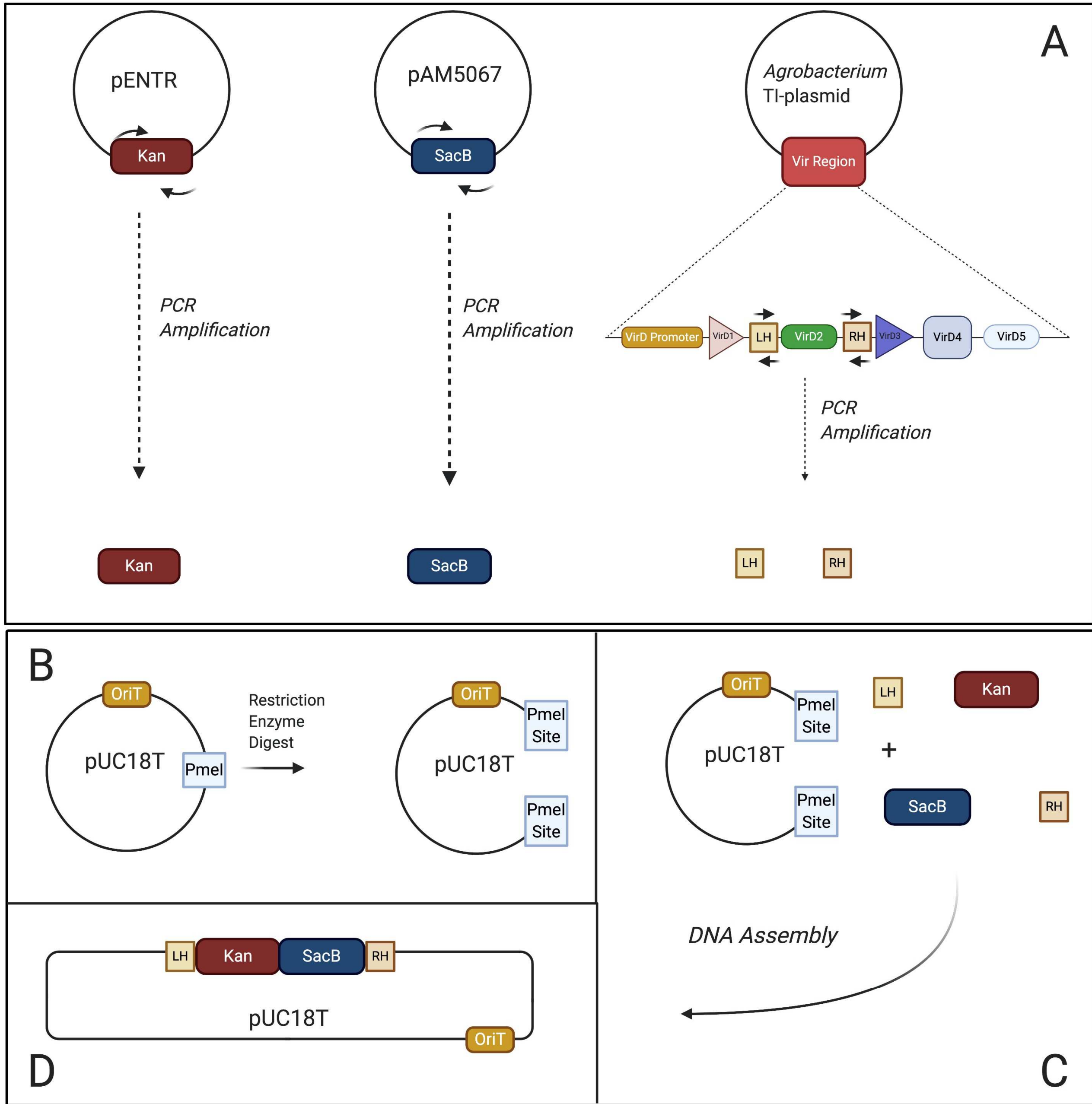

Figure 5: Shown above is a schematic illustration the process of generating our mobilizable plasmid that is contained within our donor strain in the triparental mating event. A: Polymerase chain reaction amplification of various components. B: Restriction enzyme digest of pUCT18T plasmid with PmeI to create sites for assembly. C: Gibson DNA Assembly of fragments to compose completed pUCT18T. D: pUCT18T plasmid to be used in triparental mating.

III. Triparental Mating to create *Agrobacterium* VirD2 Deletion

In order to create an *Agrobacterium* strain containing a Cas9-VirD2 fusion, we must first replace the VirD2 in wildtype *Agrobacterium* with our LH-Kan-SacB-RH cassette. In order to accomplish this, we used a bacterial conjugation method known as triparental mating. In this method of bacterial conjugating we needed three players. There must be a donor strain which contains our mobilized plasmid (pUC18T w/ OriT), a recipient strain by which we will introduce our mobilized plasmid (AGL-0), and finally a helper strain to facilitate this transfer (pRK2073) (see Fig. 6A). In our study we used *E. coli* as our donor strain, pRK2073 as the helper *E. coli* strain, and wildtype *Agrobacterium* as the recipient strain.

III-A: Process of Triparental Mating

Conjugation is initiated by the self-mobilizable helper plasmid within our helper *E. coli* strain DH5a (pRK2073). The helper plasmid (pRK2073) contains tra and mob genes which are essential for bacterial transfer [39]. This helper plasmid is able to transfer itself into our donor strain and is now in the presence of the mobilizable plasmid (pUC18T with OriT). The transfer (tra) and mobilization (mob) genes in our helper plasmid (pRK2073) then recognize the OriT in our donor plasmid (pUC18T) and initiated bacterial transfer to our recipient *Agrobacterium* strain (AGL-0) [40]. The tra genes allow for nicking of the donor plasmid at the OriT site and then the mob genes bind the end of the now single stranded DNA and assist in its transfer to the recipient strain [40]. Once the pUC18T is in the *Agrobacterium* its homologous regions can flip into the VirD operon, replacing the VirD2 with our Kan-SacB gene cassette (see Fig. 6) [41].

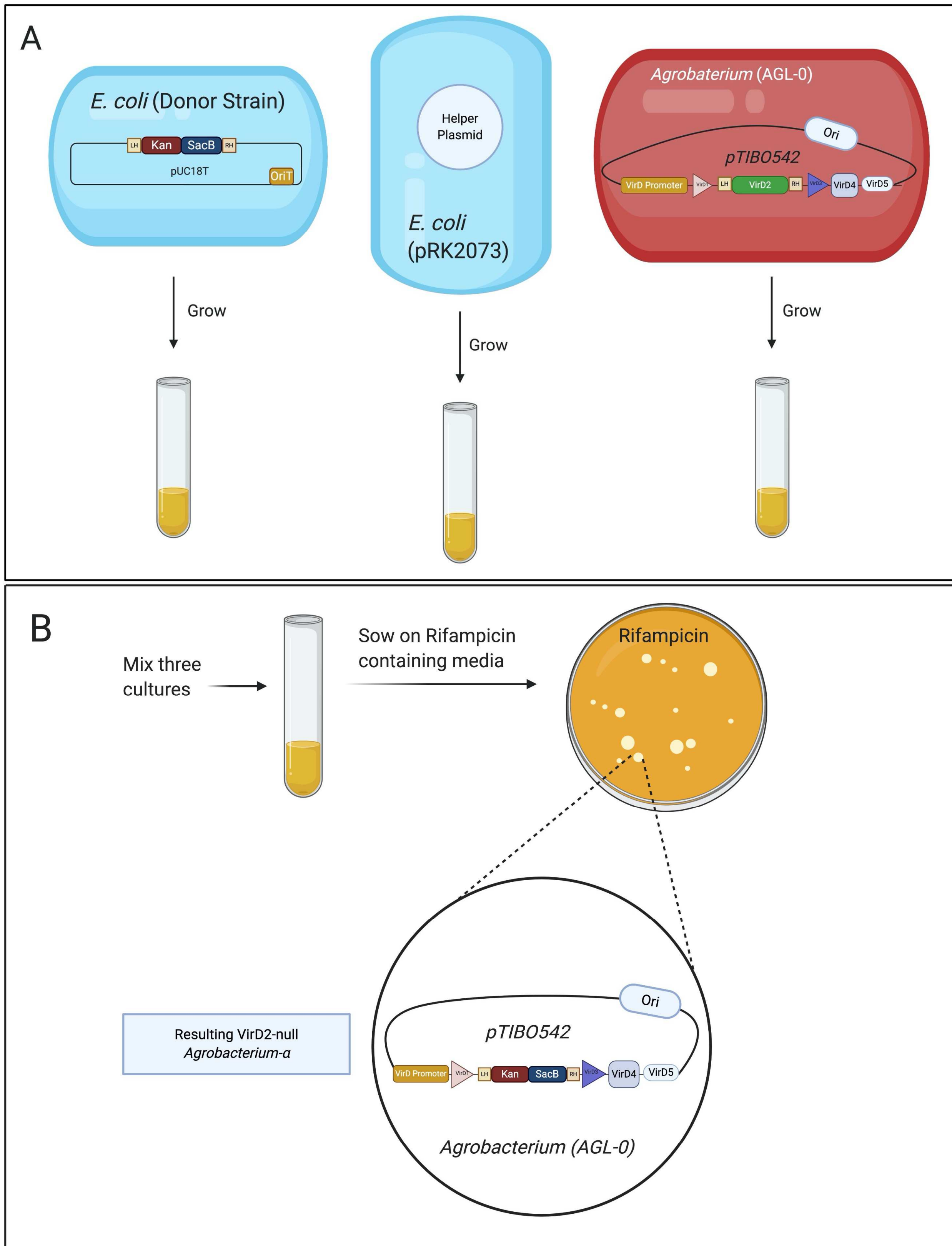

Figure 6: Illustrated below is the process of triparental mating used in this study to create our VirD2-null Agrobacterium-α. A: Growing each of the three strains in liquid media. B: Mixing of the grown cultures and spreading onto a Rifampicin containing plate to select for Agrobacterium.

IV. Cas9-VirD2 Fusion in *Cis*

We devised two main strategies for expressing the chimeric Cas9-VirD2 fusion in *Agrobacterium*. The first strategy involved an in *cis* insertion into the VirD operon of the TI-plasmid in our *Agrobacterium* strain. Under this method we needed to create a donor plasmid for a triparental mating event that contained Cas9 fused directly upstream of VirD2 (see Fig. 7). Along with the help of the right and left homology arms, insertion of this fused construct would provide in *cis* expression of the Cas9-VirD2 in our *Agrobacterium* in the context of the VirD operon (see Fig. 8).

IV-A: Amplification of a bacteria optimized Cas9 Coding Sequence

While our system is aimed at creating gene insertion in plants, we needed to use a bacteria codon optimized Cas9 sequence. Codon optimization is a phenomenon noticed in various organisms that can allow for the optimal translational expression of a gene by using a codon optimized sequence [42]. By using codons that are synchronous with organism's codon preferences, we can ensure optimal expressing of our gene [42]. Due to our expression being driven by a bacterial promoter and cell machinery, using a Cas9 coding sequence that is codon optimized for bacteria would help ensure an optimal translation level for the Cas9. By not using this method and inserting a plant codon optimized Cas9 sequence, we could cause a decrease in translation due to the difference in the common repertoire of codons used in bacteria vs. plants. This decrease in Cas9 protein could interfere with the effectiveness of the overall system. For

these reasons, we decided to use a bacterial codon optimized Cas9 version present in the pCas plasmid from the addgene available collection (https://www.addgene.org/42876/).

IV-B: Triparental Mating to Attain Cas9-VirD2 insertion

Once we generate the pUC18T plasmid containing the Cas9-VirD2 cassette, we will utilize once again triparental mating to create an insertion of our Cas9-VirD2 segment replacing the existing Kan-SacB sequence (see Fig. 8). This triparental mating event will again be grown on Rifampicin media and supplemented with sucrose. This will allow for selection against *E. coli*, and, through a negative selection using SacB marker gene, screening for positive colonies. Since SacB presence would lead to cell death, the *Agrobacterium* colonies that contain our Cas9-VirD2 cassette instead of the Kan-SacB segment, will survive.

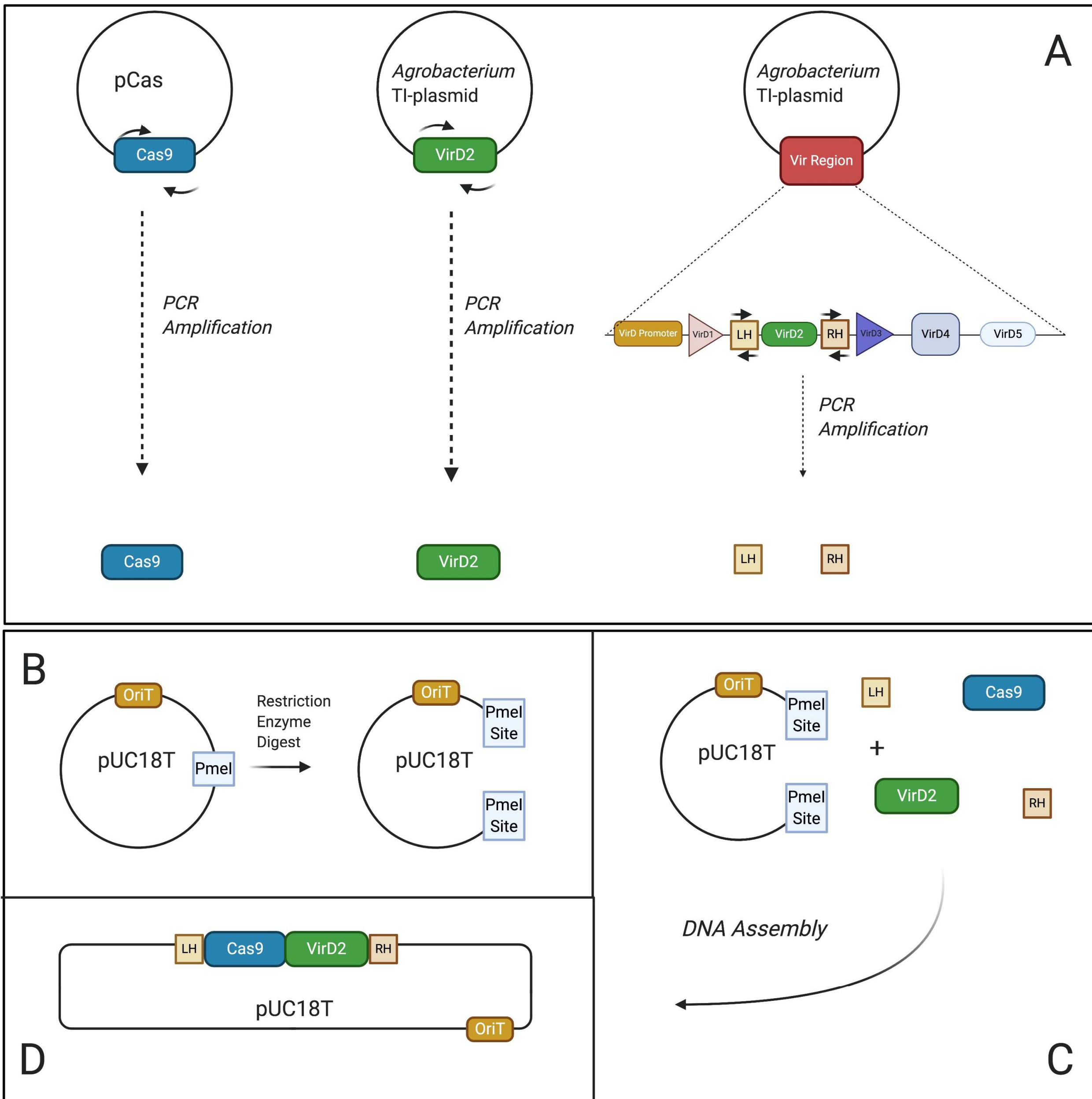

Figure 7: This is a representation of the methods we will use to assemble our mobilizable plasmid within our donor strain. The resulting pUC18T plasmid will be used for the triparental mating event in which we will create our in cis strain of Cas9-VirD2 Agrobacterium. A: Polymerase chain reaction amplification of fragments to assemble pUC18T. B: Restriction enzyme digestion with PmeI to create sites for assembly with PCR products. C: Gibson DNA Assembly of PCR products with digested plasmid. D: Schematic of pUC18T plasmid for triparental mating.

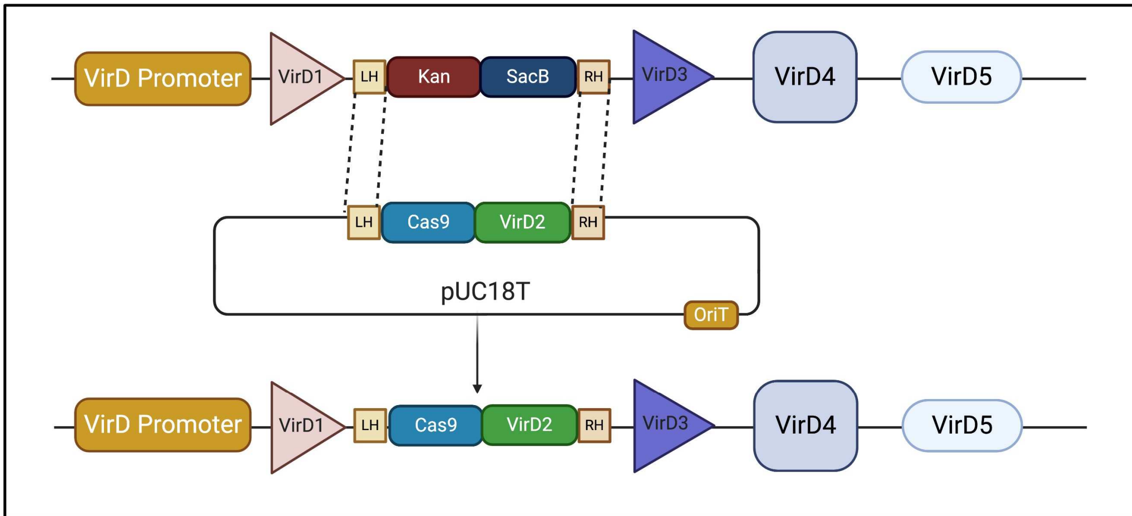

Figure 8: Outlined above is the triparental mating insertion event where our Cas9 fused to VirD2 will replace the existing Kan-SacB genes via homologous recombination. The resulting strain of Agrobacterium will be used for in cis testing of our Cas9-VirD2 system.

V. In *trans* expression of Cas9-VirD2

The second method we devised involves providing the Cas9-VirD2 fusion in *trans*. We hypothesized that perhaps the insertion of the Cas9 cDNA (around 4 Kilobases in length), could interfere with the expression of the entire VirD operon, including VirD2, which, in turn, would make the resulting strain non-infective and useless for our experimental approach. Therefore, we created a new plasmid to provide the Cas9-VirD2 chimeric function in trans, as detailed below.

V-A: pGERARD Plasmid for in *trans* expression

To provide the *Agrobacterium* with Cas9-VirD2 expression in *trans* we designed a plasmid referred to as pGERARD (Gentamicin Resistance Agrobacterium/Rhizobium Derivative) that contains an origin of replication for Agrobacterium (pSa Ori, see Fig. 10) and a gentamicin resistance gene (GmR). The plasmid pGERARD is a derivative of the pGreen0000 plasmid that requires pSOUP for replication in Agrobacterium and was created by a senior member of the Yanofsky lab. At this point we have to alternative approaches for our in trans Cas9-VirD2 strategy.

I) By the insertion of the Kan-SacB cassette into the VirD operon we most likely inactivated the entire operon. So, one alternative might be based on the expression of the entire operon from the pGERARD plasmid but instead of having VirD2, we would replace it by Cas9-Vird2.

II) This second approach involves the elimination of the Kan-SacB cassette within the context of the VirD operon and then the expression of Cas9-VirD2 from the VirD regulatory sequences both cloned into pGERARD (see Fig. 10D). In this strategy the

endogenous VirD operon would remain intact but without VirD2 that would be provided in trans (and fused to Cas9) from the pGERARD plasmid.

For our first strategy we simply have to clone the entire VirD operon into pGERARD but replacing VirD2 by Cas9-VirD2. For this end, a DNA assembly strategy can be used to clone all fragments into pGERARD using the unique SwaI present in its multiple cloning site. The resulting plasmid can then be transformed into our Kan-SacB strain for further experimentation and tests. This new strain should be infective, which would be consistent with the "rescue" of the VirD operon activities that are provided in trans from the pGERARD plasmid. At this point, we could initiate the assays to evaluate whether the Cas9 fused to VirD2 is capable of doing edits by, for example transforming Arabidopsis wild-type plants using this new Agrobacterium strain and a T-DNA containing a gRNA expressing module. If successful, this strain could be used to evaluate whether targeted Knock-Ins can be achieved easily.

V-C: Removal of Kan-SacB Genes

For our second strategy we first need to eliminate the Kan-SacB cassette present in the VirD operon. To do so, a triparental mating strategy will be utilized to effectively remove the existing Kan-SacB genes from the *Agrobacterium* using a pUC18T plasmid version containing LH-RH modules (see Fig. 9E). So upon introducing via the triparental mating, the Kan-SacB sequence will be excised in favor of the empty region between the LH and RH in our pUC18T donor plasmid. These will be sown on rifampicin and sucrose for screening against the presence of SacB containing *Agrobacterium* (VirD2-null *Agrobacterium*-α). In this defective VirD2-null *Agrobacterium*-β, the rest of the VirD operon should be functional with the exception of VirD2.

We then can use this strain to transform into a pGERAD plasmid version containing Cas9-VirD2 chimeric module under the control of VirD regulatory sequences (see Fig. 10D). After transformation on gentamicin this strain can be then tested for Arabidopsis infection and Cas9 activity similarly as described above. If successful, a knock-in strategy can be then elaborated.

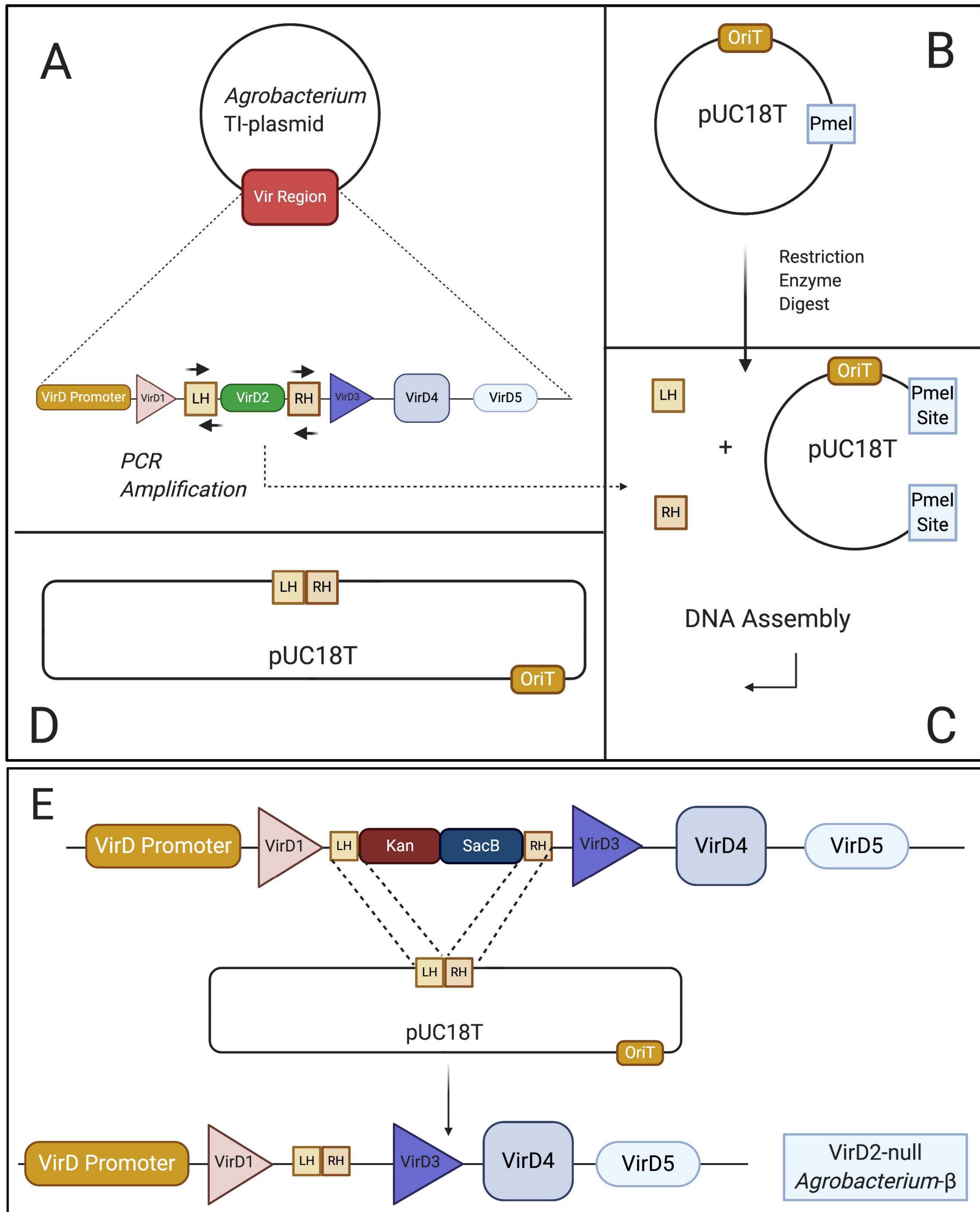

Figure 9: Generation of pUC18T donor plasmid for triparental mating of Kan-SacB deletion and resulting triparental mating product. A: Polymerase chain reaction amplification of right and left homology arms (LH/RH) from WT Agrobacterium. B: Restriction enzyme digestion of pUC18T to create sites for assembly. C: Gibson DNA Assembly of pUC18T with LH/RH to create donor plasmid. D: Resulting pUC18T plasmid that will be used as the donor plasmid in a triparental mating event. E: Resulting VirD2-null Agrobacterium-β strain from triparental mating event with pUC18T.

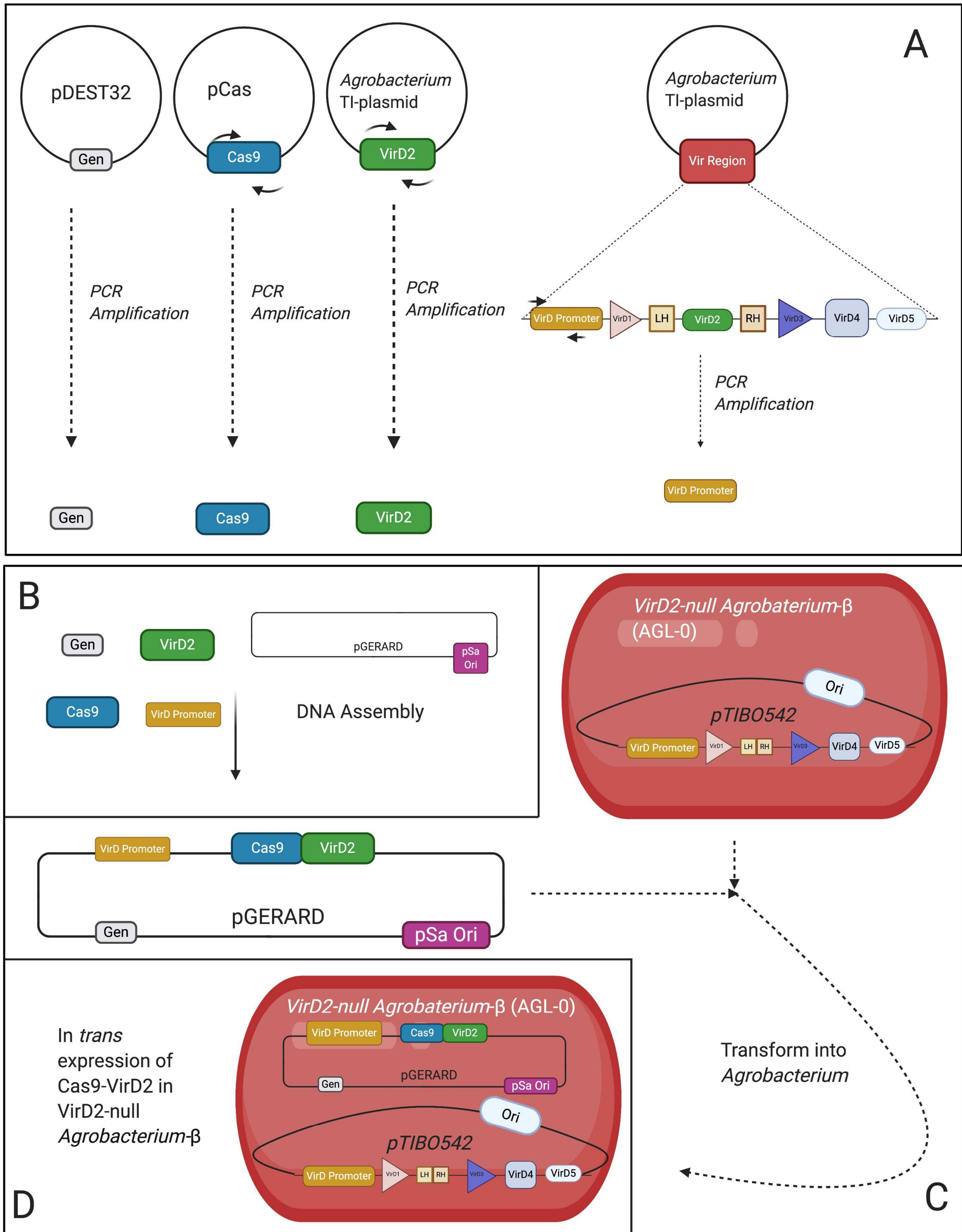

Figure 10: Generation of the pGERARD plasmid for in trans expression of Cas9-VirD2 and transformation into the VirD2-null Agrobacterium-β detailed in figure 6. A: PCR amplification of different modules needed to assembly pGERARD. B: Gibson DNA Assembly to generate complete pGERARD plasmid for in trans expression of Cas9-VirD2. C: Transformation of pGERARD plasmid into VirD2-null Agrobacterium-β. D: Schematic of in trans expression of pGERARD in VirD2-null Agrobacterium-β.

DISCUSSION

Creating a knock-in event in any organism is a difficult task as it requires the insertion of a gene in the right location, and importantly, without disturbing nearby elements. This careful coordination of multiple factors inherently challenges the success and ability to create a knock-in. To knock-in a specific gene/sequence, homology direct repair must be utilized over the much more frequent cellular repair machinery, non-homologous end joining. While this process can be difficult in general, it is more challenging in plants. Plants have been shown to have very low efficiency of HDR when compared to other species [44]. This was highlighted when researchers attempted to induce a homologous recombination of a defective kanamycin gene in *Arabidopsis* using HDR via a T-DNA [47]. Of the beginning 3.6×10^7 protoplasts, only 109 kanamycin resistant samples were detected [47]. PCR screening of these 109 samples revealed a stark 1 out of 109 samples that results in a success recombination of the repaired kanamycin gene in the T-DNA [47]. The result of this low efficiency is a domination of the NHEJ repair mechanism which makes knock-in's extremely rare to successfully induce. The low efficiency of HDR in plant somatic cells could be due to their limitations on the mitotic phase to which HDR can occur. HDR is only active as a cellular repair machinery in the S/G_2 phase and NHEJ dominates the rest of the cycle [43]. Thus, creating an efficient knock-in using HDR in plants has proven to be a major hurdle in plant gene editing.

This issue however is not replicated in some nonvascular plants such as certain algae's and mosses. For example, it has been shown that *Physcomitrella patens* can be a useful model for the efficient integration of various transgenes via HDR [45]. Publications have cited the seamless integration of HDR templates paired with CRISPR-Cas9 resulting in a knock-in of a fluorescent protein [45]. So, what could account for this drastic difference of HDR efficiency in

some mosses and algae compared to other plant organisms? Potentially the nature of the moss's high efficiency of gene knock-ins could be tied to its haploid state. Due to these organisms only having one copy of each chromosome that reduces the complications of heterozygous recessive mutations being masked in diploid cells [46]. Or maybe it could due to the overall complexities of these organisms. However, the fact that creating these knock-ins in higher level plant organisms (in our case, *Arabidopsis*) remains a challenge to progress in the field. This motivated us to leverage our vast knowledge of *Agrobacterium* biology and bring it in conjunction with the specificity of the CRISPR-Cas9 system to create a new approach to accomplish knock-ins efficiently.

While there are some reports of successful knock-in's in *Arabidopsis*, the percentage of success is extremely low. So, to overcome the obstacle of efficient knock-ins we decided to utilize the *Agrobacterium's* insertion machinery which has been proven to be widely successful in plants. The issue with this mechanism is it is random where the T-DNA gets inserted and the efficiency of knock-ins is relatively low. This could be due to a number of reasons such as only one copy of the T-DNA strands actually enters the cell. Therefore, for our HDR knock-in event there is only one copy of our template and the likelihood of an insertion in the desired location is very low. Furthermore, this single copy must insert in a clean fashion without disturbing nearby sequences to constitute a successful knock-in. Another possibility is the genome environment is not favoring the integration due to the plethora of cellular machinery elements such as transcription factors, proteins, and so on that could interfere with the T-DNA's integration. The conformation of the DNA could be considered a limiting factor as well, so condensed or non-condensed chromatin could promote or inhibit insertion at different sites. The third possibility could be the proximity of the T-DNA to the desired insertion site. By bringing the T-DNA very

close to our desired insertion site we could potentially increase the rate at which a successful targeted insertion can occur. In doing it this way we are raising the interaction rate of our T-DNA with our desired location and thus promoting its insertion in this region, rather than relying on random chance. It is therefore on this strategy that we wanted to focus our study on.

Our first thought was to find a way to bring the T-DNA the closest we could to our knock-in site. To achieve this, we decided to utilize CRISPR-Cas9's precise targeting mechanism for creating double stranded breaks in virtually any region of the genome. We used the ability of the CRISPR-Cas9 system to target and break any region of interest and in a highly efficient manner and paired it with our extensive knowledge of *Agrobacterium*. The *Agrobacterium* provides us with a novel and proven way to insert a gene of interest into a plant's genome. As described earlier, the T-DNA of the TI-plasmid serves as the infectious fragment in *Agrobacterium's* gene transfer to plants. Due to this T-DNA being inserted randomly into the plants genome, and CRISPR-Cas9's ability to induce specific double stranded breaks, we had our two systems. The challenge that we address in this project is to successful fuse these two mechanisms together and create a synergistic relationship between them.

We therefore took advantage of the covalent bond that remains between the VirD2 and the T-DNA throughout excision and insertion during the *Agrobacterium*-mediated gene transfer. Knowing that VirD2 remains attached, we theorized that we could fuse the Cas9 protein to VirD2. Thus, when the Cas9 binds our gRNA and cuts our insertion site, it brings along with it the VirD2 and consequently, the T-DNA. Using this strategy of fusing the Cas9 to the VirD2, we can in theory drastically increase the proximity of our T-DNA to our cut site. However, we didn't know what the best way would be to express this chimeric Cas9-VirD2 module. In order to address this, we designed three different strategies for expressing our chimeric module.

The first strategy involved expressing this chimeric module in *cis* with our *Agrobacterium*. This meant replacing the endogenous VirD2 with our chimeric Cas9-VirD2, keeping it within the context of the endogenous VirD operon. To achieve this, we would need to successfully induce two triparental mating events in which we first knock-out VirD2 and replace it with a Kan-SacB cassette for positive and negative selection (see Fig. 6). From this VirD2-null *Agrobacterium*-α strain we would need to replace the Kan-SacB with the chimeric Cas9-VirD2 module via another triparental mating event (see Fig. 8). There are a variety of benefits to this method as we can maintain all of the endogenous cellular regulatory elements. In essential we should be maintaining all of the function in wildtype *Agrobacterium* but expressing a Cas9 protein fused upstream of our VirD2. However, there is the possibility that our Cas9 coding sequence could have unforeseen consequences on the operon's expression. Inserting a 4kb fragment into the middle of the VirD operon could have rebounding effects on not only expression of that chimeric module but the remaining genes in the VirD operon. Additionally, this route requires two successful triparental mating recombination's.

For these reasons we thought best to devise another strategy in which we provide our chimeric module in *trans*. In order to address if this method would be optimal, we needed to test two possibilities. The first would be to express solely the chimeric Cas9-VirD2 module, driven by the endogenous VirD promoter, within the pGERARD plasmid (see Fig. 10D). To achieve this, we would need to do maintain function of the remainder of the VirD operon in our *Agrobacterium*. That would require another triparental mating event to replace our Kan-SacB in our VirD2-null *Agrobacterium-α* with the left and right homology arms (LH and RH respectively), thus restoring function to the VirD operon (see Fig. 9E). This would allow for transforming the VirD2-null *Agrobacterium*-β with our pGERARD plasmid expressing our

chimeric Cas9-VirD2 in *trans*. The possible benefits to this method would be having our chimeric module driven alone in *trans* by the VirD promoter and thus in theory providing close to, or more than endogenous levels of expression. Furthermore, its location in the pGERARD plasmid rather than in the VirD operon would ensure we maintain wildtype expression levels of the remaining VirD operon in our VirD2-null *Agrobacterium*-β. However, this strategy still requires two successful triparental mating events. Therefore, we devised a third option which would be to express the entire VirD operon containing our chimeric Cas9-VirD2 in context with the operon. Using this strategy, we would be providing the entire functionality of the VirD operon in *trans* by the pGERARD plasmid, including our chimeric Cas9-VirD2 module. The advantage of this solution is it requires only one successful triparental mating event, since we can leave the Kan-SacB in the VirD2-null *Agrobacterium-α* to knock-out the operons function.

Because it seems unclear which strategy would provide optimal expression of our chimeric module and the remainder of the VirD operon we are following all of these approaches in parallel. Due to the circumstances of the time, we were unable to complete the experiments that would give insight to the most optimal method. We would have first tested to ensure the VirD2-null *Agrobacterium* strains we generated were not able to create transgenic lines. Any indication of transgenics would be indicative of an incomplete knock-out of the VirD2 in these strains. Once the complete knock-out of VirD2 was confirmed we would need to transform plants with all three of our Cas9-VirD2 *Agrobacterium* strains (in *cis*, in *trans* only Cas9-VirD2, and in *trans* the entire VirD operon with Cas9-VirD2 replacing VirD2). These transformations would be to test that Cas9's being fused to VirD2 is not effecting the infectious ability of the *Agrobacterium*. Assuming we are able to create transgenics, we can conclude that the presence of Cas9 is not effecting VirD2's function. Finally, we need to test the functionality of the Cas9.

To achieve this, we would transform transgenic *Arabidopsis* lines containing a gRNA that targets a gene in *Arabidopsis*. If we see mutations arise in this gene, we known that the Cas9 function is not interrupted by being fused to VirD2, and the VirD2 function is conserved since the *Agrobacterium* was able to transform the plants. Following completion of these experiments we would be able deduce which method of introducing the chimeric Cas9-VirD2 module was most successful and most efficient.

There does remain the possibility that none of these methods would yield success, and in this scenario, we have thought of a few backup strategies. One of these strategies would be to use carbon nanotubes to delivery our constructs. This procedure was recently accomplished in *Arabidopsis* using a CRISPR-Cas9 system to generate edits [48]. It was accomplished by utilizing the positive charge of the carboxylate carbon nanotubes and the negative charge of DNA [48]. This allows the binding of DNA to the carbon nanotube walls and thus delivery into the plant cells upon infiltration with the nanotubes [48]. Thus, we theorized we could provide on the carbon nanotubes copies of the T-DNA and Cas9-VirD2 which will be introduced into the plant via the carbon nanotubes and then associate with each after entry. This possibility mirrors our main strategy of utilizing the T-DNA and VirD2 with Cas9 fused to it, but it changes our delivery method. This is just a possible alternative approach to achieve our primary goal of creating a tool for inducing efficient gene editing in plants.

REFERENCES

1. National Academies of Sciences, Engineering, and Medicine; National Academy of Medicine; National Academy of Sciences; Committee on Human Gene Editing: Scientific, Medical, and Ethical Considerations. Human Genome Editing: Science, Ethics, and Governance. Washington (DC): National Academies Press (US); 2017 Feb 14. A, The Basic Science of Genome Editing. Available from: https://www.ncbi.nlm.nih.gov/books/NBK447276/

2. Shuman, Stewart. "DNA Ligases: Progress and Prospects." Journal of Biological Chemistry, ASBMB Publications, 26 June 2009, www.jbc.org/content/284/26/17365.full.

3. Chial, Heidi. "Restriction Enzymes." Nature News, Nature Publishing Group, 2014, www.nature.com/scitable/spotlight/restriction-enzymes-18458113/.

4. Tirabassi, Rebeccas. "Foundations of Molecular Cloning - Past, Present and Future." NEB, New England Bio Labs Inc, www.neb.com/tools-and-resources/feature-articles/foundations-of-molecular-cloning-past-present-and-future.

5. Rouet, P, Smih, F., and Jasin, M. "Introduction of Double-Strand Breaks into the Genome of Mouse Cells by Expression of a Rare-Cutting Endonuclease." Molecular and Cellular Biology, U.S. National Library of Medicine, 14 Dec. 1994, www.ncbi.nlm.nih.gov/pubmed/7969147.

6. Scherer, S, and R W Davis. "Replacement of Chromosome Segments with Altered DNA Sequences Constructed in Vitro." Proceedings of the National Academy of Sciences of the United States of America, U.S. National Library of Medicine, Oct. 1979, www.ncbi.nlm.nih.gov/pubmed/388424?dopt=AbstractPlus.

7. Capecchi, Mario R, Thomas Kirk R, Folger Kim R. "High Frequency Targeting of Genes to Specific Sites in the Mammalian Genome." Cell, 14 Feb. 1986, doi.org/10.1016/0092-8674(86)90463-0.

8. Klug, A. (2010). The discovery of zinc fingers and their development for practical applications in gene regulation and genome manipulation. Quarterly Reviews of Biophysics, 43(1), 1-21. doi:10.1017/S0033583510000089

9. Fleischmann, R D, Adams M D, White O, Clayton R A, Kirkness E F, Kerlavage A R, Bult C J, Tomb J F, Dougherty B A, Merrick J M. "Whole-Genome Random Sequencing and Assembly of Haemophilus Influenzae Rd." Science (New York, N.Y.), U.S. National Library of Medicine, 28 July 1995, www.ncbi.nlm.nih.gov/pubmed/7542800.

10. Heather, James M, and Benjamin Chain. "The sequence of sequencers: The history of sequencing DNA." Genomics vol. 107,1 (2016): 1-8. doi:10.1016/j.ygeno.2015.11.003

11. Gupta, Rajat M, and Kiran Musunuru. "Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9." The Journal of clinical investigation vol. 124,10 (2014): 4154-61. doi:10.1172/JCI72992

12. Mojica, F.J.M., D ez-Villase or, C.S., Garc a-Mart nez, J.S., and Soria, E. (2005). Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. J Mol Evol 60, 174–182.

13. Bolotin, A., Quinquis, B., Sorokin, A.,and Ehrlich, S.D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology 151, 2551–2561.

14. Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., van der Oost, J. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. Science 321, 960-964.

15. James E. DiCarlo, Julie E. Norville, Prashant Mali, Xavier Rios, John Aach, George M. Church, Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems, Nucleic Acids Research, Volume 41, Issue 7, 1 April 2013, Pages 4336–4343, https://doi.org/10.1093/nar/gkt135

16. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819–823.

17. Li, Yang Grace, and Peter J Christie. "The Agrobacterium VirB/VirD4 T4SS: Mechanism and Architecture Defined Through In Vivo Mutagenesis and Chimeric Systems." Current Topics in Microbiology and Immunology, U.S. National Library of Medicine, 2018, www.ncbi.nlm.nih.gov/pubmed/29808338.

18. "Arabidopsis: The Model Plant." bio0202-Members of the Multinational Arabidopsis Steering Committee, 24 Mar. 2017, www.nsf.gov/pubs/2002/bio0202/model.htm.

19. GelvinDepartment, Stanton B. "Integration of Agrobacterium T-DNA into the Plant Genome." Annual Reviews, Department of Biological Sciences Purdue, Nov. 2017, www.annualreviews.org/doi/10.1146/annurev-genet-120215-035320.

20. Hwang, Hau-Hsuan, and Erh-Min Lai Manda Yu. "Agrobacterium-Mediated Plant Transformation: Biology and Applications." The Arabidopsis Book, The American Society of Plant Biologists, 20 Oct. 2017, bioone.org/journals/the-arabidopsis-book/volume-2017/issue-15/tab.0186/Agrobacterium-Mediated-Plant-Transformation-Biology-and-Applications/10.1199/tab.0186.full.

21. Bakó L, Umeda , Tiburcio AF, Schell J, Koncz C. The VirD2 pilot protein of Agrobacterium-transferred DNA interacts with the TATA box-binding protein and a

nuclear protein kinase in plants. Proc Natl Acad Sci U S A. 2003;100(17):10108□10113. doi:10.1073/pnas.1733208100

22. van Kregten M, de Pater S, Romeijn R, van Schendel R, Hooykaas PJ, Tijsterman M. T-DNA integration in plants results from polymerase-θ-mediated DNA repair. Nat Plants. 2016;2(11):16164. Published 2016 Oct 31. doi:10.1038/nplants.2016.164

23. Miki, Daisuke. Zhang Wenxin, Feng Zhengyan, Zhu Jiang Kang. “CRISPR/Cas9-mediated gene targeting in Arabidopsis using sequential transformation.” Nature communications vol. 9,1 1967. 17 May. 2018, doi:10.1038/s41467-018-04416-0

24. Frona, Daniel, Szenderak Janos, Harangi-Rakos Monika. “The Challenge of Feeding the World.” Sustainabilty , MDPI, 12 Sept. 2019.

25. Sedeek, Khalid E M, Mahas Ahmed, Mahfouz Magdy. “Plant Genome Engineering for Targeted Improvement of Crop Traits.” Frontiers in plant science vol. 10 114. 12 Feb. 2019, doi:10.3389/fpls.2019.00114

26. “Pocket K No. 16: Biotech Crop Highlights in 2018.” Biotech Crop Highlights in 2018 | ISAAA.org, International Service for the Acquisition of Agri-Biotech Applications, Dec. 2019, www.isaaa.org/resources/publications/pocketk/16/.

27. Koornneef, Maarten, and David Meinke. “The Development of Arabidopsis as a Model Plant.” The Plant Journal : for Cell and Molecular Biology, U.S. National Library of Medicine, 10 Mar. 2010, www.ncbi.nlm.nih.gov/pubmed/20409266.

28. Feldmann, K.A., Marks, M.D., Christianson, M.L. and Quatrano, R.S. (1989) A dwarf mutant of Arabidopsis generated by T□DNA insertion mutagenesis. Science, 243, 1351–1354.

29. Kado, Clarence I. “Historical account on gaining insights on the mechanism of crown gall tumorigenesis induced by Agrobacterium tumefaciens.” Frontiers in microbiology vol. 5 340. 7 Aug. 2014, doi:10.3389/fmicb.2014.00340

30. Smith, E.F. and Townsend, C.O. (1907). A plant-tumor of bacterial origin. Science. 25: 671-673.

31. Braun, A C. “A Physiological Basis for Autonomous Growth of the Crown-Gall Tumor Cell.” Proceedings of the National Academy of Sciences of the United States of America vol. 44,4 (1958): 344-9. doi:10.1073/pnas.44.4.344

32. White P. R., Braun A. C. (1942). A cancerous neoplasm of plants autonomous bacteria-free crown-gall tissue. Cancer Res. 2, 597–617

33. Gordon, Jay E, and Peter J Christie. "The Agrobacterium Ti Plasmids." Microbiology spectrum vol. 2,6 (2014): 10.1128/microbiolspec.PLAS-0010-2013. doi:10.1128/microbiolspec.PLAS-0010-2013

34. Stachel, S.E., and Nester, E.W. (1986). The genetic and transcriptional organization of the vir region of the A6 Ti plasmid of Agrobacterium tumefaciens. EMBO J. 5: 1445–1454.
35. Lecroix , Benoit, and Vitoly Citovsky . "Pathways of DNA Transfer to Plants from Agrobacterium Tumefaciens and Related Bacterial Species." Annual Reviews, Annual Review of Phytopathology, 21 June 2019, www.annualreviews.org/doi/pdf/10.1146/annurev-phyto-082718-100101.

36. Gelvin, and Stanton B. "Traversing the Cell: Agrobacterium T-DNA's Journey to the Host Genome." Frontiers in Plant Science: Plant Microbe Interaction, Frontiers, 28 Feb. 2012, www.frontiersin.org/articles/10.3389/fpls.2012.00052/full.

37. Kempin, S., Liljegren, S., Block, L., Rounsley Steven D., Yanofsky Martin F, Lam Eric. Targeted disruption in Arabidopsis. Nature 389, 802–803 (1997). https://doi.org/10.1038/39770

38. Yanofsky, M F, Porter S G, Young C, Albright L M, Gordon M P, Nester E W. "The VirD Operon of Agrobacterium Tumefaciens Encodes a Site-Specific Endonuclease." Cell, U.S. National Library of Medicine, 7 Nov. 1986, www.ncbi.nlm.nih.gov/pubmed/3021341/.

39. Wittmer. "Transfer of the Shuttle Vector pRRI207 between Escherichia Coli and Bacteroides Spp. in Vitro and in Vivo in the Digestive Tract of Axenic Mice and in Gnotoxenic Mice Inoculated with a Human Microflora." *OUP Academic*, Oxford University Press, 1 May 1999, academic.oup.com/femsec/article/29/1/33/766180.

40. Buchanan-Wollaston, V., Passiatore, J. & Cannon, F. The *mob* and *oriT* mobilization functions of a bacterial plasmid promote its transfer to plants. *Nature* **328,** 172–175 (1987). https://doi.org/10.1038/328172a0

41. Griffiths AJF, Gelbart WM, Miller JH. Modern Genetic Analysis. New York: W. H. Freeman; 1999. Bacterial Conjugation. Available from: https://www.ncbi.nlm.nih.gov/books/NBK21351/

42. Camiolo, S, Melito S, Porceddu A. "New insights into the interplay between codon bias determinants in plants." *DNA research : an international journal for rapid publication of reports on genes and genomes* vol. 22,6 (2015): 461-70. doi:10.1093/dnares/dsv027

43. Van Vu, T., Sung, Y.W., Kim, J., Doan Duong Thi Hai, Tran Mil Thi, Kim Jae-Yean. Challenges and Perspectives in Homology-Directed Gene Targeting in Monocot Plants. *Rice* **12,** 95 (2019). https://doi.org/10.1186/s12284-019-0355-1

44. Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW (1983) The double-strand-break repair model for recombination. Cell 33(1):25–35

45. Rensing, Stefan A, Goffinet Bernard, Meyberg Rabea, Wu Shu-Zon, Bezanilla Magdalena. "The Moss Physcomitrium (Physcomitrella) Patens: A Model Organism for Non-Seed Plants." *Plant Cell*, American Society of Plant Biologists, 1 May 2020, www.plantcell.org/content/32/5/1361.

46. Yin, Zixi, and Lingyi Chen. "Simple Meets Single: The Application of CRISPR/Cas9 in Haploid Embryonic Stem Cells." *Stem cells international* vol. 2017 (2017): 2601746. doi:10.1155/2017/2601746

47. Offringa R, Franke-van Dijk ME, De Groot MJ, van den Elzen PJ, Hooykaas PJ (1993) Nonreciprocal homologous recombination between Agrobacterium transferred DNA and a plant chromosomal locus. Proc Natl Acad Sci USA **90**: 7346–7350

48. Demirer, G.S., Zhang, H., Goh, N.S, Gonzalez-Grandio Eduarado, Landry Markita P. Carbon nanotube–mediated DNA delivery without transgene integration in intact plants. *Nat Protoc* **14,** 2954–2971 (2019). https://doi.org/10.1038/s41596-019-0208-9

49. Koornneef M, Meinke D. The development of Arabidopsis as a model plant. *Plant J*. 2010;61(6):909□921. doi:10.1111/j.1365-313X.2009.04086.

50. Hellens, R.P., Edwards, E.A., Leyland, N.R. Bean Samantha, Mullineaux Philip M. pGreen: a versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation. *Plant Mol Biol* **42,** 819–832 (2000). https://doi.org/10.1023/A:1006496308160